

Filled Pentagons and Electron Counting Rule for Boron Fullerenes

Kregg D. Quarles, Chernob B. Kah, Rosi N. Gunasinghe, Ryza N. Musin, and Xiao-Qian Wang*

Department of Physics and Center for Functional Nanoscale Materials, Clark Atlanta University, Atlanta, Georgia 30314, United States

S Supporting Information

ABSTRACT: We have revisited the general constructing schemes for a large family of stable hollow boron fullerenes with $80 + 8n$ ($n = 0, 2, 3, \dots$) atoms. In contrast to the hollow pentagon boron fullerenes with 12 hollow pentagons, the stable boron fullerenes constitute 12 filled pentagons and 12 additional hollow hexagons, which are more stable than the empty pentagon boron fullerenes including the “magic” B_{80} buckyball. On the basis of results from first-principles density-functional calculations, an empirical rule for filled pentagons is proposed along with a revised electron counting scheme to account for the improved stability and the associated electronic bonding feature.

Since the discovery of C_{60} buckyball 25 years ago, the fascinating properties and promising applications of the synthetic carbon allotropes—fullerenes, nanotubes, and graphene—overwhelmingly illustrate their unique scientific and technological importance.^{1,2} As boron and carbon share an abundance of bonding similarities,^{3–10} there has been a tremendous amount of interest in the search for nanostructured counterparts of carbon allotropes. Among efforts in exploiting hollow inorganic cage-like structures, the theoretical prediction of a highly stable “magic” B_{80} buckyball³ by the groups of Szwacki, Sadrzadeh, and Yakobson has received a great deal of attention.^{11–20} The boron buckyball B_{80} is structurally analogous to the eminent C_{60} ,¹ with 60 boron atoms placed at the corners of a truncated icosahedron that constitutes 12 pentagons and 20 hexagons, along with an extra 20 boron atoms in the center of each hexagon.³ The 20 capping atoms stabilize the cage of the identical icosahedral (I_h) symmetry as the C_{60} buckyball.

The novel chemical bonding pattern of B_{80} provides crucial insights into the nature of boron nanomaterials and has prompted considerable efforts in designing associated nanostructures such as endohedral complexes,⁹ solids,^{10,11} and hydrogen storage media.¹² Inspired by the B_{80} buckyball configuration, construction rules for a family of stable boron fullerenes were proposed.¹³ Moreover, careful examination of the chemical bonding of B_{80} buckyball with triangular and hexagonal motifs led to re-evaluation of boron sheets and nanotubes composed of purely puckered triangular structures, revealing more stable α -boron sheet (α -BS).⁵

Baruah and co-workers reported that the “magic” I_h - B_{80} buckyball was vibrationally unstable and further showed that a relaxation of the I_h - B_{80} buckyball leads to a vibrationally stable T_h structure.¹⁶ In the present paper, an alternative means for constructing a structurally inequivalent vibrationally stable T_h structure is demonstrated, and a transition state geometry between the two T_h structures is also presented. On the other hand, recent *ab initio* simulation studies have unveiled a few lower energy structures built by an icosahedral B_{12} core along with a shell of pentagonal and hexagonal pyramidal units.^{18,19} These new developments raise questions regarding the relative stability of boron cages.

In accordance with the “Aufbau principle”,²³ stable boron conformations can be constructed from two basic building blocks: the pentagonal pyramid B_6 and the hexagonal pyramid B_7 . The hexagonal B_7 is the precursor for convex and quasi-planar boron clusters and thus closely correlates to the α -BS and the associated boron nanotubes.⁵ On the other hand, the pentagonal B_6 unit is an aromatic component that has attracted revived interest in planar boron clusters such as B_{19}^- .⁴ In this regard, we have studied a volleyball-shaped B_{80} fullerene that is lower in energy than the previously assumed B_{80} buckyball.²⁰ Contrary to the core–shell structured $B_{12}@B_{68}$, it preserves the desired electronic properties as the boron counterpart of C_{60} .²⁰

A natural question arises as to whether there exists a “magic” boron fullerene in lieu of the latest developments.^{18–22} An important consequence of the stability for the B_{80} volleyball is that, in addition to the hollow pentagon, hollow hexagon, and filled hexagon,¹³ it is necessary to add the pentagon pyramid in the building blocks.²⁰ Herein, we present a revised constructing scheme and demonstrate that the migration of capping atoms from hexagonal pyramids to pentagonal rings leads to highly stable boron fullerenes. Consequently, the electron counting rule is revisited by taking into account the effect of enhanced stability related to the B_6 pentagonal pyramids.

We have employed first-principles calculations based on local and semilocal density-functional approaches. For selecting structural conformations and geometry optimizations, gradient-corrected Becke–Lee–Yang–Parr (BLYP) parametrization^{24,25} of the exchange–correlation was used along with a double numerical (DN) basis set as implemented in the DMol3 package.²⁶ The local density-functional calculation results were subsequently rectified through a semilocal approach using Becke–3–Lee–Yang–Parr (B3LYP) for the exchange–correlation functional,^{24,25} with the 6-311G basis set as implemented in the TeraChem package.²⁸ The optimization of atomic positions was performed for local and semilocal calculations, which proceeded until the

Published: June 01, 2011

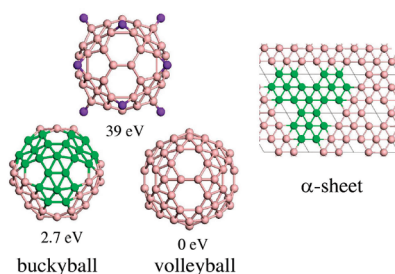


Figure 1. Ball-and-stick representation of the optimized T_h -A B_{80} buckyball and volleyball, along with the transition state between the two conformations and the α -boron sheet. Highlighted with green and purple are the snowdrop pattern and the migrating capping atoms, respectively.

forces were less than 0.01 eV/\AA and the change in energy was less than $5 \times 10^{-4} \text{ eV}$.

It is instructive to examine the buckyball and volleyball structures of B_{80} , as shown in Figure 1. The structure of the vibrational stable T_h - B_{80} (T_h -A)¹⁶ consists of 20 filled hexagons and 12 empty pentagons.³ The filled hexagons are arranged in a snowdrop-like pattern, as highlighted in Figure 1.¹³ The snowdrop constitutes a central hexagonal pyramid surrounded by three hexagonal pyramids and three “hollows” (hollow pentagons or hexagons). By contrast, the B_{80} volleyball can be viewed as the 12 outward capping atoms migrating from hexagonal pyramids to the centers of 12 empty pentagons.²⁰ We illustrate in Figure 1 the transition state between the buckyball and volleyball of B_{80} with an energy barrier of $\sim 39 \text{ eV}$ and the 12 capping atoms highlighted (see the Supporting Information for details of the transition state calculation, along with discussions of models of migrating capping atoms). The exchange between B_6 and B_7 pyramids leads to yet another route of generating a family boron fullerenes.

Carbon fullerenes consist of hexagons and 12 pentagons in conformity with Euler's formula $F - E + V = 2$, where F , E , and V stand for the number of faces, edges, and vertices of the fullerene, respectively. The exactness of 12 pentagons is attributed to the fact that each edge is shared by two faces, each pentagon (hexagon) has five (six) edges, and each vertex is adjacent to three polygons. Let n_p (n_H) denote the number of pentagons (hexagons); one has $F = n_p + n_H$, $E = (5n_p + 6n_H)/2$, and $V = (3n_p + 6n_H)/3$. Therefore, $n_p = 12$.

The isoelectronic requirement for carbon and boron fullerenes implies that the addition of more than 60 carbon atoms needs to be a multiple of 6 for the even number of carbon fullerenes and the trivalent boron, respectively. As a result, the counterpart of C_{60+6n} isolated-pentagon fullerenes is B_{80+8n} . The isoelectronic requirement is important for a closed-shelled electronic structure in that typical high-symmetry boron fullerenes are open-shelled in the absence of such a constraint. For snowdrop B_{80+8n} fullerenes that constitute hexagonal pyramids in addition to pentagonal and hexagonal rings, it is straightforward to conclude that $n_p = 12$, $n_H = n$, and the number of filled hexagons (FHs) $n_{FH} = 20 + 2n$, since each filled hexagon has 12 edges and six triangular faces. For snowdrop-generated boron nanostructures, the total number of valence electrons, $240 + 24n$, is twice that of $120 + 12n$ triangles.¹³ It is worth noting that the snowdrop electron counting rule is applicable to α -BS as well.

Encouraged by the improved stability of the B_{80} volleyball over the B_{80} buckyball,²⁰ herein we propose a filled pentagon model

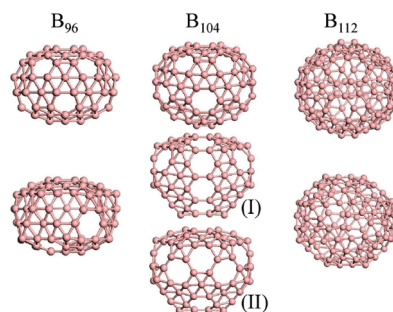


Figure 2. Optimized structures of snowdrop (top panel) and filled-pentagon (bottom panel) fullerenes of B_{96} , B_{104} , and B_{112} .

for boron fullerenes. Specifically, the filled-pentagon (FP) scheme amounts to moving 12 capping atoms from filled hexagons to pentagons, resulting in $n_p = 0$, $n_H = 12 + n$, $n_{FH} = 8 + 2n$, and $n_{FP} = 12$. Consequently, the revised electron counting rule yields more than two electrons per triangle, which is dependent on n but still converged to an α -BS value of 2 as $n \rightarrow \infty$.

Following the nomenclature, hereafter, we refer the filled-hexagon fullerenes as snowdrop fullerenes.^{13,27} In contrast to the snowdrop model where the constructed boron fullerenes are of the same symmetry as the carbon fullerene counterpart, the filled-pentagon fullerenes typically have lower symmetry. The lack of a unique migration path adds another wrinkle to the search for global minimum conformations of boron fullerenes. Owing to the large number of boron atoms involved and the complicated two- and three-center bonding patterns, the associated first-principles density-functional calculations are computationally demanding. Furthermore, owing to the extremely sensitive dependence upon the basis set and the exchange-correlation functional employed, it is necessary to carefully evaluate the local and nonlocal effects in *ab initio* calculations.

Shown in Figure 2 are the optimized structures of B_{96} , B_{104} , and B_{112} , which are the isoelectronic counterparts of C_{72} , C_{78} , and C_{84} , respectively. The snowdrop B_{96} fullerene has a round pillow shape with two hexagonal rings located at the D_{6d} axis. The filled-pentagon B_{96} structure is formed by moving 12 of the 24 capping atoms to fill the 12 pentagons, leaving 14 hollow hexagons. On the equator of D_{6d} - B_{96} , three alternating pairs of hexagons move the capping atoms to the centers of nearest-neighbor pentagons, while the two snowdrop structures near the D_{6d} axis rotate to fill three empty pentagons each. The B_{104} counterpart of C_{78} has five isomers,²⁷ and the snowdrop B_{104} is of D_{3h} symmetry. Among various ways of migrating 12 capping atoms, we show in Figure 2 two low-energy C_{2v} conformations: one has empty hexagons around the equator, while the other has more isolated hexagonal rings. The construction of both conformations involves rotating two snowdrops and hexagonal pair migrations. The B_{112} fullerene consists of 32 hexagons and 12 pentagons.²⁷ Among 24 isomers, we consider the counterpart of C_{84} ground-state structure with D_2 symmetry and an elliptical pillow shape.²⁷ The D_2 snowdrop B_{112} can be transformed to a C_2 filled-pentagon B_{112} through successful rotation of four snowdrops.

An important criterion for structural stability is that vibrational modes are all real. We have performed vibrational analysis for all of the above conformations. While the vibrational stability of filled-pentagon B_{80} , B_{96} , and B_{104} is confirmed, there exist two imaginary modes for snowdrop D_{6d} - B_{96} , analogous to that in

Table 1. Calculated Binding Energies (E_B in eV Relative to Atomic Boron), HOMO–LUMO Gap (E_g in eV), the Energy Difference (ΔE in eV), and Symmetries (S) of Optimization for B_{80+8n} Fullerene Structures Using Local (BLYP) and Semilocal (B3LYP) Approaches, Respectively

n	method	structure	S	E_B (eV)	E_g (eV)	S	E_B (eV)	E_g (eV)	ΔE (eV)
0	BLYP	B_{80}	T_h -A	−406.82	0.94	T_h	−409.53	0.18	2.73
2	BLYP	B_{96}	C_2	−490.48	0.64	D_3	−494.86	0.27	4.38
3	BLYP	B_{104}	D_{3h}	−532.83	0.59	C_{2v} (I)	−534.44	0.27	1.61
3	BLYP					C_{2v} (II)	−536.05	0.20	3.22
0	B3LYP	B_{80}	T_h -A	−453.85	1.87	T_h	−458.13	0.86	4.29
2	B3LYP	B_{96}	C_2	−548.08	1.36	D_3	−553.48	0.93	5.40
3	B3LYP	B_{104}	D_{3h}	−595.81	1.22	C_{2v} (I)	−598.73	0.79	2.93
3	B3LYP					C_{2v} (II)	−598.93	0.91	3.12

I_h - B_{80} .¹⁶ The existence of imaginary frequency modes is attributed to the symmetry constraint of the calculation. A subsequent eigenmode following analysis leads to a lower energy (~ 0.03 eV) and lower symmetry C_2 - B_{96} conformation that has all real vibrational frequencies. We list in Table 1 the calculated binding energies, the gap between the highest-occupied molecular orbital (HOMO) and the lowest-unoccupied molecular orbital (LUMO), the symmetry of the fullerene, and the energy differences between the snowdrop and filled-pentagon models. As is readily observable from Table 1, the filled-pentagon fullerene is systematically lower in energy than the snowdrop counterpart. Furthermore, in contrast to the monotonic decrease of the snowdrop B_{80+8n} with the increase of n , the filled-pentagon B_{80+8n} shows a “magic” number of $n = 2$, corresponding to a filled-pentagon B_{96} . Closer scrutiny of the structure of the filled-pentagon B_{96} reveals that the enhanced stability of B_{96} correlates to the “isolated” empty hexagons or hexagon pairs.

Whereas the snowdrop B_{80+8n} gives rise to a large class of stable boron fullerenes, the filled-pentagon B_{80+8n} leads to structures with improved stability. For $n = 0$, the buckyball and volleyball B_{80} are the first members of the corresponding family. The α -BS is the $n \rightarrow \infty$ analogue for both models. The construction of filled-pentagon cages amounts to transforming a pair of neighboring pentagon rings to filled pentagons, accompanied by the generation of a pair of adjacent hollow hexagons. An optimal arrangement of those building blocks allows for the tuning of the charge transfer, which gives rise to an attractive fluxional behavior and improves the stability. In all of the cases, the improvement of energy is remarkable, ranging from 3.1 eV for B_{104} to 5.4 eV for B_{96} . The semilocal results are in good accordance with those of local results regarding the binding energy, while the correction to the HOMO–LUMO gap is substantial. As seen in Table 1, the rectified gap is about 0.7–0.9 eV for filled-pentagon fullerenes.

In conformity with the donor–acceptor hypothesis on α -BS,⁵ electron transfers in the snowdrop B_{80+8n} can be classified as the capping atoms in the center of hexagonal pyramids as electron donors, while other atoms can be classified as acceptors.¹³ For the filled-pentagon model, the capping atoms of hexagonal and hexagonal pyramids stand for electron donors, while others stand for acceptors.²⁰ As such, it is expected that uniformly distributed hollow hexagons are desired.²⁰ Shown in Figure 3 are the extracted charge-density distributions of HOMO (LUMO) is an evident increase in the π (π^*) bonding in the filled-pentagon B_{96} , as compared to that for snowdrop B_{96} . This

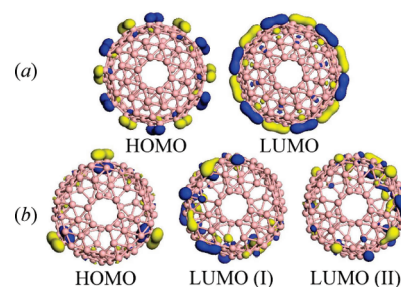


Figure 3. Isodensity surfaces (with an isovalue of 0.02 au) of HOMO and LUMO for (a) the snowdrop and (b) filled-pentagon models of B_{96} , respectively. The positive and negative components are colored with blue and yellow, respectively.

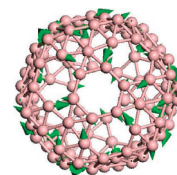


Figure 4. The motions of the infrared-active vibrational mode of 1972 cm^{-1} .

is attributed to the redistribution of the bonding and antibonding patterns related to the capping of pentagon rings.²⁰

A few remarks are immediately in order. (i) Our results demonstrate that the filled pentagon model yields improvement in energy for a family of boron fullerenes not limited to B_{80} . The energy improvement is primarily attributed to the migration of the capping atoms from filled hexagons to pentagons (see Figure S1, Supporting Information) and bonding arrangements associated with changes in aromaticity.^{18,20} (ii) The filled-pentagon fullerenes prefer evenly distributed hollow hexagons in connection to the donor–acceptor charge transfers. (iii) The stability of the filled-pentagon fullerenes is also manifested in the vibrational frequencies. We depict in Figure 4 the motions of the highest vibrational mode for B_{96} . The rest of the vibrational frequencies is in the range of $81\text{--}1411\text{ cm}^{-1}$ (see the Supporting Information). The lowest vibrational modes are significantly softer than the counterparts of carbon fullerenes.^{13,16} (iv) In contrast to the snowdrop building block, for filled-pentagon fullerenes, the basic unit is the adjacent pentagon-hexagon pyramids. (v) The main thrust of the present work is to assert that the filled-pentagon model is energetically preferred. To this end, we have

systematically investigated a large set of snowdrop fullerenes and revealed the existence of a “magic” filled-pentagon B₉₆. Notwithstanding this, there are still a variety of boron fullerenes that remain unexplored. Our results, nevertheless, illustrate that the intriguing bonding pattern of boron keeps bringing surprises.

In summary, we have described a filled-pentagon constructing scheme for a large family of stable boron fullerenes. The improved stability of the filled-pentagon fullerenes over the snowdrop ones is confirmed by intensive first-principles simulations. A revised empirical electron counting rule indicates that a slightly increased electron counting per triangle is energetically preferred, which is useful for seeking for stable boron nanostructures. These results shed important light on the improved stability of boron fullerenes in connection to the aromaticity. We hope that these results can promote revived experimental synthesis of boron fullerenes along with future applications.

■ ASSOCIATED CONTENT

S Supporting Information. The optimized atomic coordinates and vibrational frequencies of B₉₆. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: xwang@cau.edu.

■ ACKNOWLEDGMENT

This work was supported by the National Science Foundation (Grant No. DMR-0934142), Army Research Office (Grant No. W911NF-10-1-0302), and Air Force Office of Scientific Research (Grant No. FA9550-10-1-0254).

■ REFERENCES

- (1) Kroto, H.; Heath, J.; O'Brien, S.; Curl, R.; Smalley, R. C₆₀: Buckminsterfullerene. *Nature* **1985**, *318*, 162.
- (2) Geim, A. K.; Novoselov, K. S. The Rise of Graphene. *Nat. Mater.* **2007**, *6*, 183.
- (3) Szwacki, N. G.; Sadrzadeh, A.; Yakobson, B. I. B₈₀ Fullerene: An *Ab Initio* Prediction of Geometry, Stability, and Electronic Structure. *Phys. Rev. Lett.* **2007**, *98*, 166804.
- (4) Huang, W.; Sergeeva, A. P.; Zhai, H. J.; Averkiev, B. B.; Wang, L. S.; Boldyrev, A. I. A Concentric Planar Doubly π -Aromatic B₁₉⁻ Cluster. *Nature Chem.* **2010**, *2*, 202.
- (5) Tang, H.; Ismail-Beigi, S. Novel Precursors for Boron Nanotubes: The Competition of Two-Center and Three-Center Bonding in Boron Sheets. *Phys. Rev. Lett.* **2007**, *99*, 115501.
- (6) Saxena, S.; Tyson, T. A. Insights on the Atomic and Electronic Structure of Boron Nanoribbons. *Phys. Rev. Lett.* **2010**, *104*, 245502.
- (7) Kiran, B.; Bulusu, S.; Zhai, H.-J.; Yoo, S.; Zeng, X.-C.; Wang, L. S. Planar-to-Tubular Structural Transition in Boron Clusters: B₂₀ as the Embryo of Single-Walled Boron Nanotubes. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 961.
- (8) Zhai, H. J.; Kiran, B.; Li, J.; Wang, L. S. Hydrocarbon Analogues of Boron Clusters-Planarity, Aromaticity and Antiaromaticity. *Nat. Mater.* **2003**, *2*, 827.
- (9) Li, J. L.; Yang, G. W. Ni@B₈₀: A Single Molecular Magnetic Switch. *Appl. Phys. Lett.* **2009**, *95*, 133115.
- (10) Liu, A. Y.; Zope, R. R.; Pederson, M. R. Structural and Bonding Properties of bcc-Based B₈₀ Solids. *Phys. Rev. B* **2008**, *78*, 155422.
- (11) Yan, Q. B.; Zheng, Q. R.; Su, G. Face-Centered-Cubic K₃B₈₀ and Mg₃B₈₀ Metals: Covalent and Ionic Bondings. *Phys. Rev. B* **2009**, *80*, 104111.
- (12) Zhao, Y. F.; Lusk, M. T.; Dillon, A. C.; Heben, M. J.; Zhang, S. B. Boron-Based Organometallic Nanostructures: Hydrogenstorage Properties and Structure Stability. *Nano Lett.* **2008**, *8*, 157.
- (13) Yan, Q. B.; Sheng, X. L.; Zheng, Q. R.; Zhang, L. Z.; Su, G. Family of Boron Fullerenes: General Constructing Schemes, Electron Counting Rule, and *Ab Initio* Calculations. *Phys. Rev. B* **2008**, *78*, 201401(R).
- (14) Szwacki, N. G.; Tymczak, C. J. The Symmetry of the Boron Buckyball and a Related Boron Nanotube. *Chem. Phys. Lett.* **2010**, *494*, 80.
- (15) Sadrzadeh, A.; Pupysheva, O. V.; Singh, A. K.; Yakobson, B. I. The Boron Buckyball and Its Precursors: An Electronic Structure Study. *J. Phys. Chem. A* **2008**, *112*, 13679.
- (16) Baruah, T.; Pederson, M. R.; Zope, R. R. Vibrational Stability and Electronic Structure of a B₈₀ Fullerene. *Phys. Rev. B* **2008**, *78*, 045408.
- (17) Ceulemans, A.; Muya, J. T.; Gopakumar, G.; Nguyen, M. T. Chemical Bonding in the Boron Buckyball. *Chem. Phys. Lett.* **2008**, *461*, 226.
- (18) Li, H.; Shao, N.; Shang, B.; Yuan, L.-F.; Yang, J. L.; Zeng, X. C. Icosahedral B₁₂-Containing Core-Shell Structures of B₈₀. *Chem. Commun.* **2010**, *46*, 3878.
- (19) Zhao, J.; Wang, L.; Li, F.; Chen, Z. B₈₀ and Other Medium-Sized Boron Clusters: Core-Shell Structures, Not Hollow Cages. *J. Phys. Chem. A* **2010**, *114*, 9969.
- (20) Wang, X.-Q. Structural and Electronic Stability of a Volleyball-Shaped B₈₀ Fullerene. *Phys. Rev. B* **2010**, *82*, 153409.
- (21) Ozdogan, C.; Mukhopadhyay, S.; Hayami, W.; Guvenc, Z. B.; Pandey, R.; Boustani, I. The Unusually Stable B₁₀₀ Fullerene, Structural Transitions in Boron Nanostructures, and a Comparative Study of γ - and α -Boron and Sheets. *J. Phys. Chem. C* **2010**, *114*, 4362.
- (22) Zope, R. R.; Baruah, T.; Lau, K. C.; Liu, A. Y.; Pederson, M. R.; Dunlap, B. I. Boron fullerenes: From B₈₀ to Hole Doped Boron Sheets. *Phys. Rev. B* **2009**, *79*, 161403(R).
- (23) Quandt, A.; Boustani, I. Boron Nanotubes. *Chem. Phys. Chem.* **2005**, *6*, 2001.
- (24) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648.
- (25) Lee, C.; Yang, W.; Parr, R. G. Development of the Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785.
- (26) DMol₃; Accelrys Software Inc.: San Diego, CA, 2010.
- (27) Zhang, B. L.; Wang, C. Z.; Ho, K. M.; Xu, C. H.; Chan, C. T. The Geometry of Large Fullerene Cages: C₇₂ to C₁₀₂. *J. Chem. Phys.* **1993**, *98*, 3095.
- (28) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619.

Mechanism of Nitric Oxide Oxidation Reaction ($2\text{NO} + \text{O}_2 \rightarrow 2\text{NO}_2$) Revisited

Oleg B. Gadzhiev,^{*,†,‡} Stanislav K. Ignatov,[†] Shrubha Gangopadhyay,[§] Artëm E. Masunov,[§] and Alexander I. Petrov^{||}

[†]Department of Chemistry, N.I. Lobachevsky State University of Nizhny Novgorod, 23 Gagarin Avenue, Nizhny Novgorod 603950, Russia

[‡]Institute of Applied Physics of the Russian Academy of Sciences, 46 Ul'yanov Street 603950, Nizhny Novgorod, Russia

[§]NanoScience Technology Center, Department of Chemistry and Department of Physics, University of Central Florida, 12424 Research Parkway, Ste 400, Orlando, Florida 32826, United States

^{||}Department of Metallurgy of Noble and Rare Metals, Institute of Non-Ferrous Metals and Materials Science, Siberian Federal University, 81 Svobodny Prospect, Krasnoyarsk 660041, Russia

S Supporting Information

ABSTRACT: The reaction between molecular oxygen and two nitric oxide(II) molecules is studied with high-level *ab initio* wave function methods, including geometry optimizations with coupled cluster (CCSD(T_{full})/cc-pCVTZ) and complete active space with second order perturbation theory levels (CASPT2/cc-pVDZ). The energy at the critical points was refined by calculations at the CCSD(T_{full})/aug-cc-pCVTZ level. The controversies found in the previous theoretical studies are critically discussed and resolved. The best estimate of the activation energy is 6.47 kJ/mol.

This work is focused on the elucidation of the thermal reaction mechanism between dioxygen and two nitrogen monoxide molecules. Very recently, comprehensive experimental^{1,2} and joined quantum chemical and experimental³ investigations were performed for NO/O₂ and related (NO/O₂^{-•}) systems.⁴ The formal kinetics of third order reactions are in the scope of current theoretical developments.⁵ Measurements of the rate constant and activation energy of the reaction $2\text{NO} + \text{O}_2$ were reviewed in ref 6, which covered the years up to the early 1990s. The most recent investigation, which was not included in ref 6, is ref 1. The most reliable value of activation energy determined up to date⁷ is $-4.41 [\pm 3.33]$ kJ/mol in the temperature range 270–600 K (see also ref 8).

The mechanism we report here was found to be different from the mechanisms studied previously by Olson et al.⁹ and other authors.¹⁰ This Letter supports the mechanism proposed by McKee.¹¹ The classical Eyring–Gershinowitz mechanism¹² was studied in more detail by McKee,¹¹ who used *ab initio* and DFT (B3LYP) methods. However, the geometries of intermediates and transition states for the two-step mechanism¹¹ must be refined, and physicochemical parameters of the elementary reactions must be revisited using recent developments in the coupled cluster method.¹³

In ref 11, transition states were not optimized uniformly. In the present report, we will show the two-step mechanism of the reaction $2\text{NO} + \text{O}_2 \rightarrow 2\text{NO}_2$ using the CCSD(T) and CASSCF methods with full geometry optimizations.

Recently, we investigated¹⁰ the potential energy surfaces (PES) of the $2\text{NO}_2 + \text{O}_2$ system using broken symmetry DFT (BS-UDFT), double hybrid DFT, and CCSD(T)//DFT approaches. In the study,¹⁰ we found perfect agreement between

the recommended experimental activation energy^{6,7} and calculated (B3LYP/*aug-pc3*) activation enthalpy (Tables 1, 6, and 7 and Figure 6 in ref 3) when the reaction proceeded via the cyclic intermediate CC–INT (Scheme 1).

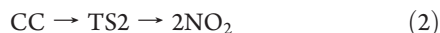
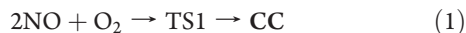
However, the wave function of this system was found to be multiconfigurational, as evidenced by the high values for the T₁ test within the CCSD method. In this Letter, we reinvestigate the PES using CASSCF, CASPT2, and MRMP2 methods that are more appropriate for describing the multiconfigurational character of the CC–INT ground state.

Complete geometry optimization of the stationary points on PES is performed at the RCCSD(T)/cc-pVDZ and UCCSD(T)/cc-pVDZ levels (using the GAMESS-US¹⁴ and Gaussian 2003¹⁵ codes, respectively). Numerical calculations of the force constant matrix were performed to verify the types of stationary points. The intrinsic reaction coordinate (IRC) was identified for all of the transition states. The T₁-test values were below 0.036 for all of the optimized geometries. The structure and relative energies of several stationary points including CC–INT were found to differ from earlier BS-UDFT results.¹⁰ The valley–pitchfork bifurcation point in the main reaction channel, reported previously,¹⁰ was not found here. The elementary steps reported in Figure 1 and in the Supporting Information are as follows (the relative energies at the CCSD(T)/cc-pVDZ theory level are in parentheses): $2\text{NO} + \text{O}_2$ (0 kJ/mol) \rightarrow TS1 (13.98 kJ/mol) \rightarrow CC (–16.41 kJ/mol) \rightarrow TS2 (1.62 kJ/mol) \rightarrow 2NO_2 (–73.95 kJ/mol).

Received: December 29, 2010

Published: May 30, 2011

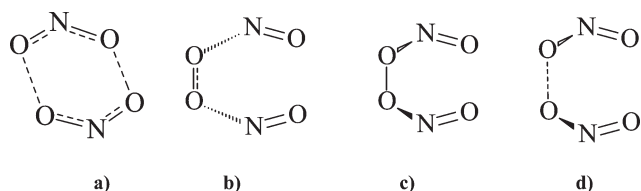
In contrast to the results reported in refs 9 and 10, the mechanism presented here (see Figure 1) takes place entirely on the singlet state potential surface and consists of two steps:



The initial step (1) is the formation of the CC intermediate with the saddle point TS1 and an activation barrier of 13.98 kJ/mol (in refs 9 and 10, this was barrierless elementary reaction $2\text{NO} + \text{O}_2 \rightarrow \text{CC}$). The second step (2) with the saddle point TS2 corresponds to the homolysis of CC (in refs 9 and 10, conformer CC rearranged in CC-INT, which was involved in the main multistep reaction channel, $\text{CC-INT} \rightarrow \text{cis-ONONO}_2 \rightarrow \text{trans-ONONO}_2 \rightarrow \text{O}_2\text{NNO}_2 \rightarrow 2\text{NO}_2$). Elementary reaction 1 (3) is the rate-determining step of the proposed mechanism (in refs 9 and 10, isomerization $\text{CC} \rightarrow \text{CC-INT}$ was the rate-determining step).

To verify the results at the CCSD(T)/cc-pVDZ theory level, the fragments of a potential energy surface corresponding to reactions 1 and 2 have been calculated at the levels CASSCF(10,10)/cc-pVDZ (Gaussian 03) and CASSCF(10,10)/cc-pVTZ (GAMESS-US). These calculations confirm the transition state TS1 (CAS(10,10)/cc-pVDZ) during the formation of the complex CC (CAS(10,10) with cc-pVDZ and cc-pVTZ basis

Scheme 1. Structure of the Planar Cyclic Intermediate CC-INT (a) from Previous Investigations^{9,10} and Structures of TS1 (b), CC (c), and TS2 (d) of the Reaction Path from the Presented Letter (See Text for Details)



sets) in the initial step through the intrinsic reaction coordinate (IRC) method at the CAS(10,10)/cc-pVDZ level starting from TS1 in both directions. For reaction 2, localization of TS2 in the CASSCF method was complicated by the intruder states. The analysis of the configuration interaction (CI) amplitudes for CASSCF(10,10)/cc-pVDZ (obtained in this paper for both intermediate CC and TS1) and CASSCF(10,12)/cc-pVDZ (only for CC) and CAS(26,16)/cc-pVDZ (given earlier for CC-INT and O_2NNO_2)³ shows that the relative weight of the Hartree-Fock configuration (the square of the first coefficient in the CI expansion vector) is greater than 0.8 for all stationary points except CC-INT (CAS(26,16)) and TS1 (CAS(10,10)). This allows us to conclude that the errors of a single-reference CCSD(T) approximation do not influence the results significantly. We consider the CCSD(T) method more reliable than CASSCF since it takes into account the dynamic electron correlation.

To eliminate the contradiction¹⁰ between KS-DFT, BS-UDFT, and CASSCF on the structure of the intermediate CC-INT, the complete optimization of its geometry at the levels CCSD(T)/cc-pVDZ and CAS(2,2)PT2/cc-pVDZ (using MOLPRO software¹⁶) and MR(2,2)MP2/cc-pVDZ (using GAMESS-US software) has been carried out. A good agreement between the optimized geometric parameters has been achieved; the optimized geometries are shown in Figure 2. This allows us to make a conclusion about the reliability of the CCSD(T) results.

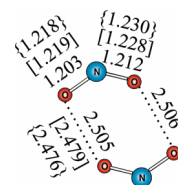


Figure 2. Geometries of intermediate CC-INT optimized at the levels CCSD(T)/cc-pVDZ, CASPT2/cc-pVDZ [square brackets], and MRMP2/cc-pVDZ {curly brackets}. Bond lengths are indicated in Ångstroms.

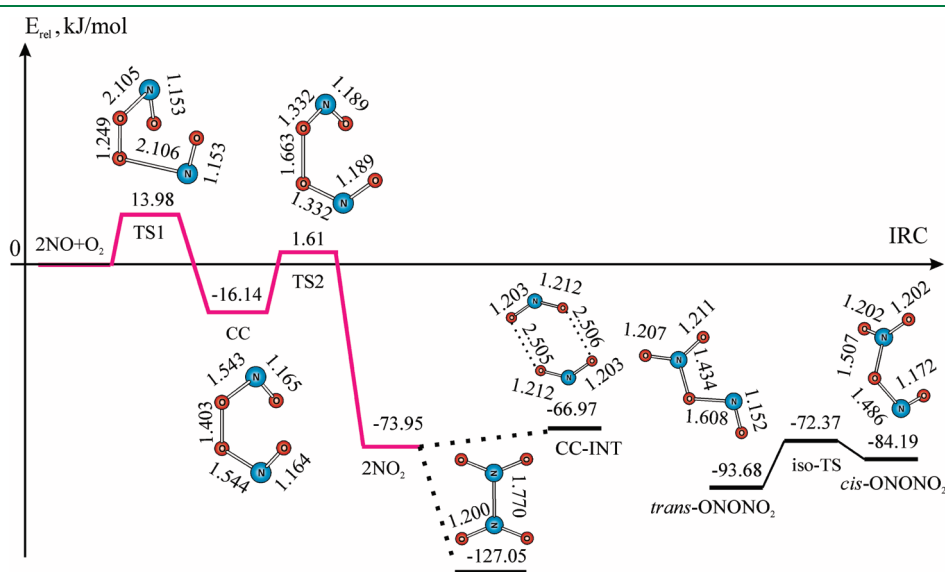


Figure 1. Profile of singlet PES of thermal reaction $2\text{NO}_2 + \text{O}_2 \rightarrow 2\text{NO}_2$, obtained at the CCSD(T)/cc-pVDZ level in the present study. The solid and dotted lines indicate the minimum energy pathways at the CCSD(T)/cc-pVDZ and CAS(26,16)/cc-pVDZ levels, respectively. Bond lengths are indicated in Ångstroms; the relative energies are in kilojoules per mole.

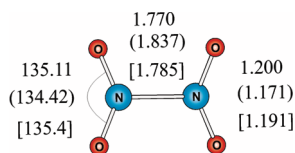


Figure 3. Geometries of dinitrogen tetraoxide O_2NNO_2 optimized at the CCSD(T)/cc-pVDZ and CASSCF/cc-pVDZ (parentheses) levels and experimental data¹⁸ [square brackets]. Bond lengths are indicated in Angstroms; the bond angle is in degrees.

This also proves that the previous¹⁰ KS-DFT results are invalid. The isomerization paths $\text{CC} \rightarrow \text{CC-INT}$ and $\text{CC-INT} \rightarrow \text{cis-ONONO}_2$ proposed previously^{9,10} and $\text{CC-INT} \rightarrow \text{trans-ONONO}_2$ were not found at any levels of CCSD(T) or CASSCF employed in the presented investigation. However, the isomerization reaction $\text{cis-ONONO}_2 \rightarrow \text{iso-TS} \rightarrow \text{trans-ONONO}_2$ was reproduced by calculations at the CAS(26,16)/cc-pVDZ and CCSD(T)/cc-pVDZ levels (see Figure 1). The dissociation reactions $\text{cis-ONONO}_2 \rightarrow 2\text{NO}_2$ ($E_r = -21.62$ kJ/mol) and $\text{trans-ONONO}_2 \rightarrow 2\text{NO}_2$ ($E_r = 1.85$ kJ/mol) with corresponding TSs were established at the CAS(26,16)/cc-pVDZ level and characterized by activation energies of 1.8 and 71.1 kJ/mol, respectively. We were unable to locate both TSs at the CCSD(T)/cc-pVDZ level despite an extra effort (relaxed scans of O–N bonds and Hessian calculations). We suppose that these differences in the calculation results are due to the different amounts of nondynamical electron correlation accounted for by the CCSD(T) and CASSCF methods. Additional calculations may be necessary for unambiguous determination of MEPs for these two reactions.

The existence of the saddle point TS2 with an energy of +1.62 kJ/mol (see Figure 1) detected at the CCSD(T)/cc-pVDZ theory level points to the possibility that reaction $2\text{NO} + \text{O}_2 \rightarrow 2\text{NO}_2$ takes place without the formation of intermediate CC-INT (which is now a local minimum on the potential energy surface with a relatively high energy of -66.97 kJ/mol).

The energy of the dimerization reaction with $2\text{NO}_2 \rightarrow \text{O}_2\text{NNO}_2$ calculated at the CCSD(T)/cc-pVDZ level (-53.01 and -38.48 kJ/mol corrected by ZPVE) agrees with the experimental enthalpy of -53.60 kJ/mol;¹⁷ the CASSCF/cc-pVDZ energy (-36.4 and -6.6 kJ/mol with ZPVE) is noticeably underestimated even for the large active space CAS(26,16). Calculated geometrical parameters (bond distances $r(\text{N-N}) = 1.770$ Å, $r(\text{N=O}) = 1.200$ Å, and bond angle $\angle(\text{ONO}) = 135.1^\circ$) at the CCSD(T)/cc-pVDZ level shown in Figure 3 are closer to the corresponding experimental¹⁸ data (bond distances $r(\text{N-N}) = 1.782$ Å, $r(\text{N=O}) = 1.191$ Å, and bond angle $\angle(\text{ONO}) = 135.4^\circ$) than to the geometrical parameters optimized at the CAS(26,16)/cc-pVDZ level (bond distances $r(\text{N-N}) = 1.837$ Å, $r(\text{N=O}) = 1.171$ Å, and bond angle $\angle(\text{ONO}) = 134.4^\circ$).

The found activation character of the elementary reaction $2\text{NO} + \text{O}_2 \rightarrow \text{TS1} \rightarrow \text{CC}$ agrees with the weak negative dependence of the reaction ($2\text{NO} + \text{O}_2 \rightarrow 2\text{NO}_2$) rate constant on the temperature if we assume that the positive barrier height is a consequence of the basis set superposition error (BSSE), and the basis set expansion will make it possible to lower E_a to be in quantitative agreement with the apparent activation energy⁷ of about -4.41 kJ/mol.

We performed an additional search of stationary points O_2 , NO , and TS1 at the CCSD(T,full)/cc-pCVTZ level and energy

calculations using the compound approach CCSD(T,full)/aug-cc-pCVTZ//CCSD(T,full)/cc-pCVTZ by means of the CFOUR suite of programs.¹⁹ The addition of diffuse functions to the basis set (augmentation) significantly improved the barrier height of reaction 1 from 20.99 to 6.47 kJ/mol, which is close to the 0 kJ/mol of the classical Eyring–Gershinowitz¹² estimation of activation energy of the reaction ($2\text{NO} + \text{O}_2 \rightarrow 2\text{NO}_2$) using the Bodenstein experimental data.²⁰ This is also in reasonably good agreement with the latest value⁷ of -4.41 [± 3.33] kJ/mol recommended for the temperature range 270–600 K.

The presented results show that a balanced treatment of both dynamic and static electron correlation is necessary for the correct energy estimation of PES stationary points for the $2\text{NO} + \text{O}_2$ system, and application of a coupled-cluster method requires the inclusion of at least triple excitations. To achieve chemical accuracy in order to quantitatively compare physicochemical parameters of the reaction ($2\text{NO} + \text{O}_2 \rightarrow 2\text{NO}_2$), focal-point analysis (with extrapolations) is probably required. On the basis of the results obtained here, we conclude that the CCSD(T) approach, in contrast with DFT theory, supports the two-step mechanism of nitric oxide oxidation.

■ ASSOCIATED CONTENT

S Supporting Information. The calculated Cartesian coordinates, total energies, vibrational frequencies, and IRC curves. This material is available free of charge via the Internet at <http://pubs.acs.org>

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: euriscomail@mail.ru.

■ ACKNOWLEDGMENT

This work was supported in part by the U.S. National Science Foundation (CHE-0832622). O.B.G. is thankful for the partial support of this study provided by a grant from the Government of the Russian Federation under agreement no. 11.G34.31.0014. S.K.I. acknowledges the RFBR project 11-03-00085. A.I.P. is thankful to SFU Supercomputer Centre for a generous donation of CPU time. A.E.M. and S.G. are grateful to DOE NERSC (project m513), I2lab, and Institute for Simulation and Training (IST) for computer time provided.

■ REFERENCES

- (1) Galliker, B.; Kissner, R.; Nauser, T.; Koppenol, W. H. *Chem.—Eur. J.* **2009**, *15*, 6161.
- (2) Beckers, H.; Willner, H.; Jacox, M. E. *ChemPhysChem* **2009**, *10*, 706.
- (3) Beckers, H.; Zeng, X.; Willner, H. *Chem.—Eur. J.* **2010**, *16*, 1506.
- (4) Botti, H.; Möller, M. N.; Steinmann, D.; Nauser, T.; Koppenol, W. H.; Denicola, A.; Radi, R. *J. Phys. Chem. B* **2010**, *114*, 16584.
- (5) Williams, B. J. *Math. Chem.* **2011**, *49*, 328.
- (6) Tsukahara, H.; Ishida, T.; Mayumi, M. *Nitric Oxide: Biol. Ch.* **1999**, *3*, 191.
- (7) Atkinson, R.; Baulch, D. L.; Cox, R. A.; Crowley, J. N.; Hampson, R. F.; Hynes, R. G.; Jenkin, M. E.; Rossi, M. J.; Troe, J. *Atmos. Chem. Phys.* **2004**, *4*, 1461.
- (8) Manion, J. A.; Huie, R. E.; Levin, R. D.; Burgess, D. R., Jr.; Orkin, V. L.; Tsang, W.; McGivern, W. S.; Hudgens, J. W.; Knyazev, V. D.; Atkinson, D. B.; Chai, E.; Tereza, A. M.; Lin, C.-Y.; Allison, T. C.;

Mallard, W. G.; Westley, F.; Herron, J. T.; Hampson, R. F.; Frizzell, D. H. NIST Chemical Kinetics Database, NIST Standard Reference Database 17, Version 7.0 (Web Version), Release 1.4.3, Data version 2008.12; National Institute of Standards and Technology, Gaithersburg, MD. Web address: <http://kinetics.nist.gov> (accessed Apr 15, 2011).

(9) Olson, L. P.; Kuwata, K. T.; Bartberger, M. D.; Houk, K. N. *J. Am. Chem. Soc.* **2002**, *124*, 9469.

(10) Gadzhiev, O. B.; Ignatov, S. K.; Razuvaev, A. G.; Masunov, A. E. *J. Phys. Chem. A* **2009**, *113*, 9092.

(11) McKee, M. L. *J. Am. Chem. Soc.* **1995**, *117*, 1629.

(12) Gershinowitz, H.; Eyring, H. *J. Am. Chem. Soc.* **1935**, *57*, 985.

(13) Harding, M. E.; Metzroth, T.; Gauss, J.; Auer, A. A. *J. Chem. Theory Comput.* **2007**, *4*, 64.

(14) Schmidt, M.; Baldrige, K. K.; Boatz, J. A.; Elbert, S.; Gordon, M.; Jensen, J. H.; Koeski, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(15) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Version D.02; Gaussian, Inc.: Wallingford, CT, 2007.

(16) Werner, H.-J.; Knowles, P. J.; Manby, F. R.; Schütz, M.; Celani, P.; Knizia, G.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. *MOLPRO*, version 2009.1; University College Cardiff Consultants: Cardiff, Wales, 2009.

(17) Chase, J. M. W.; Curnutt, J. L.; Downey, J. J. R.; McDonald, R. A.; Syverud, A. N.; Valenzuela, E. A. *J. Phys. Chem. Ref. Data* **1982**, *11*, 695.

(18) McClelland, B. W.; Gundersen, G.; Hedberg, K. *J. Chem. Phys.* **1972**, *56*, 4541.

(19) Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G.; Auer, A. A.; Bartlett, R. J.; Benedikt, U.; Berger, C.; Bernholdt, D. E.; Bomble, Y. J.; Cheng, L.; Christiansen, O.; Heckert, M.; Heun, O.; Huber, C.; Jagau, T.-C.; Jonsson, D.; Jusélius, J.; Klein, K.; Lauderdale, W. J.; Matthews, D. A.; Metzroth, T.; O'Neill, D. P.; Price, D. R.; Prochnow, E.; Ruud, K.; Schiffmann, F.; Schwalbach, W.; Stopkowitz, S.; Tajti, A.; Vázquez, J.; Wang, F.; Watts, J. D. *CFOUR*. Integral packages *MOLE-CULE* (Almlöf, J.; Taylor, P. R.), *PROPS* (Taylor, P. R.), *ABACUS* (Helgaker, T.; Jensen, H. J. Aa.; Jørgensen, P.; Olsen, J.), and ECP routines by Mitin, A. V.; van Wüllen, C. For the current version, see <http://www.cfour.de> (accessed Feb 23, 2011).

(20) Bodenstein, M. *Z. Physik. Chem.* **1922**, *100*, 68.

Replica Temperatures for Uniform Exchange and Efficient Roundtrip Times in Explicit Solvent Parallel Tempering Simulations

Meher K. Prakash,* Alessandro Barducci, and Michele Parrinello

Department of Chemistry and Applied Biosciences, ETH Zurich USI Campus Via Giuseppe Buffi 13, CH 6900 Lugano, Switzerland

ABSTRACT: The efficiency of parallel tempering simulations is greatly influenced by the distribution of replica temperatures. In explicit solvent biomolecular simulations, where the total energy is dominated by the solvent, specific heat is usually assumed to be constant. From this, it follows that a geometric distribution of temperatures is optimal. We observe that for commonly used water models (TIP3P, SPC/E) under constant volume conditions and in the range of temperatures normally used, the specific heat is not a constant, consistent with experimental observations. Using this fact, we derive an improved temperature distribution which substantially reduces the round-trip times, especially when working with a small number of replicas.

1. INTRODUCTION

Parallel tempering (PT) is a popular choice for obtaining enhanced sampling in molecular simulations.^{1–5} In standard PT, multiple NVT simulations are performed in parallel on the same system at different temperatures T_i , $i = 1 - N_T$. At regular intervals, attempts are made to exchange the replica pairs $i \leftrightarrow i + 1$, and the exchanges are accepted with a probability that conserves the detailed balance. One of the advantages of PT is that barriers and bottlenecks can be overcome in the high temperature replicas. In addition, these multiple replica simulations can be carried out in an embarrassingly parallel mode, thus offsetting the added computational costs.

Recently, there have been several studies aimed at estimating^{6–8} and improving the PT efficiencies either by the choice of replica temperature^{8–19} or by enhancing the energy fluctuations of the replicas.²⁰ In a PT simulation, assuming that there are no bottlenecks, the individual replicas perform a random walk in the temperature space. The efficiency of the simulation is measured by the replica round-trip time (τ) across the temperatures, and this optimal value is obtained when the probability of exchange between any two neighboring replicas $P_{i,i+1}^{\text{ex}}$ is constant.¹⁷

$$P_{i,i+1}^{\text{ex, opt}} = \text{const} \quad (1)$$

Thus, for improving the PT efficiency, one usually tries to achieve probabilities of exchange ($P_{i,i+1}^{\text{ex}}$) between neighboring replicas as uniform as possible. One of the assumptions in these theoretical analyses is a constant specific heat (C_V), which leads to the choice of geometrically distributed replica temperatures.^{17,18}

However, when it comes to constant volume biomolecular simulations in explicit solvent, the assumption of a constant C_V is not valid. In fact, the specific heat in these systems is dominated by water, and it has been shown experimentally²¹ and theoretically^{22,23} that for water C_V is far from being a constant in the temperature interval normally used in biomolecular PT simulations. In fact, C_V decreases with temperature. Thus, in order to optimize the distribution of replica temperatures, the actual behavior of C_V under these conditions needs to be considered.

We have here recalculated the dependence of the average potential energy (\bar{E}) and C_V on T for the commonly used TIP3P²⁴

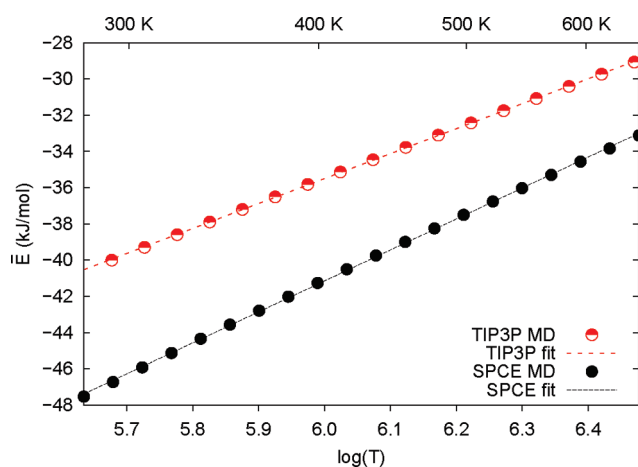


Figure 1. Average potential energy vs $\log(T)$ from NVT simulations of TIP3P and SPC/E waters over the temperature range 280–650 K. Starting with a box of 15 500 water molecules, box size was initially equilibrated at 300 K and 1 bar. NVT simulations at all temperatures were performed using a Nose–Hoover thermostat and this box size.

and SPC/E²⁵ models of water. It can be seen from Figure 1 that for both models the sublinear temperature dependence of \bar{E} in the interval (280–650 K) of relevance in biomolecular PT is well approximated by the expression

$$\bar{E} = a \log(T) + E_0 \quad (2)$$

where the constants (a , E_0) are 14 and -118 kJ/mol and 17 and -143 kJ/mol for TIP3P and SPC/E, respectively. From this dependence, and the relation $C_V = \partial \bar{E} / \partial T$, it follows, as shown in Figure 2, that $C_V = a/T$ in qualitative agreement with the experimental decrease of C_V . The origins of this surprising behavior are not entirely clear, but we take this as our empirical observation. On the basis of this, we reconsider here the issue of optimal choice of temperature distribution for explicit solvent calculation.

Published: June 06, 2011

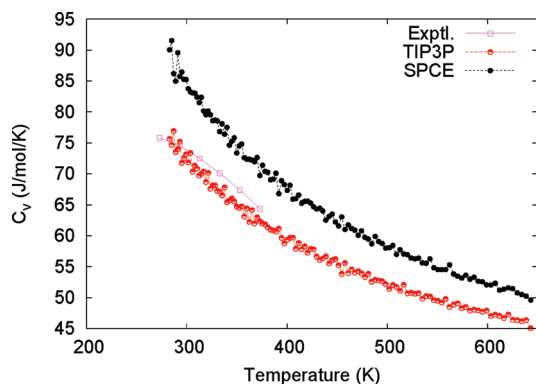


Figure 2. C_V dependence on temperature from experiments²¹ and our MD simulations under NVT conditions. C_V from TIP3P and SPCE/E water simulations was computed as a numerical derivative of the average energies with respect to T .

We use the following arguments. The exchange probability $P_{i,i+1}^{\text{ex}}$ in a replica exchange simulation is $\min\{1, \exp(-\Delta\beta_{i,i+1}\Delta E_{i+1,i})\}$, where $\Delta E_{i+1,i} = E_{i+1} - E_i$, $\Delta\beta_{i,i+1} = \beta_{i+1} - \beta_i$ and $\beta_i = 1/k_B T_i$. To obtain a uniform P^{ex} , as in eq 1, we consider a temperature distribution where the neighboring replicas satisfy the first-order condition with the mean energies^{16,26}

$$\exp(-\Delta\beta_{i-1,i}\Delta\bar{E}_{i-1}) = \exp(-\Delta\beta_{i,i+1}\Delta\bar{E}_{i+1,i}) \quad (3a)$$

$$\text{i.e., } \left(\frac{1}{T_{i-1}} - \frac{1}{T_i}\right)(\bar{E}_i - \bar{E}_{i-1}) = c \quad (3b)$$

where c is a constant related to the exchange probability. Using eq 2, one has

$$\begin{aligned} \frac{c}{a} &= \left(\frac{1}{T_i} - \frac{1}{T_{i-1}}\right)(\log T_{i-1} - \log T_i) \\ &= \left(\frac{1}{T_i} - \frac{1}{T_{i-1}}\right) \left[\log\left(1 + \frac{T_{i-1} - T_i}{T_i}\right)\right] \end{aligned} \quad (4)$$

This equation is satisfied at all i if the temperature distribution follows the relationship

$$\frac{1}{T_i} = \frac{1}{T_{i-1}} - \sqrt{\frac{c/a}{T_{i-1}}} \quad (5)$$

The above equation is a simple analytical expression for generating the replica temperature distribution starting from an initial temperature T_1 and T_2 whose choice is determined by $P_{1,2}^{\text{ex}}$.

As a check of this derivation, we consider a hypothetical system with a constant C_V

$$\bar{E}(T) = bT + E_0 \quad (6)$$

Using the same arguments as before, we find

$$\frac{c}{b} = \left(\frac{1}{T_{i-1}} - \frac{1}{T_i}\right)(T_i - T_{i-1}) = \frac{(T_i/T_{i-1} - 1)^2}{T_i/T_{i-1}} \quad (7)$$

which is satisfied by the usual geometric distribution:

$$\frac{T_i}{T_{i-1}} = \text{const} \quad (8)$$

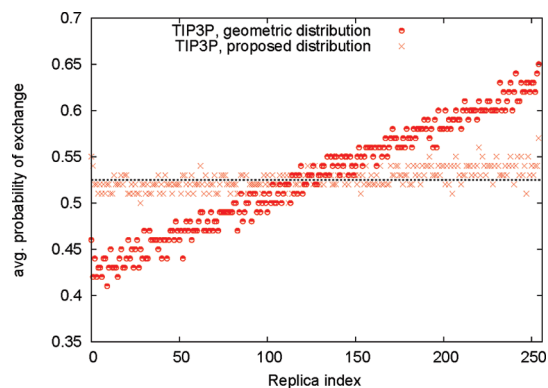


Figure 3. Probability of exchange between replicas i and $i + 1$ ($P_{i,i+1}^{\text{ex}}$) for 2 ns simulations using geometric series distribution and the proposed distribution of temperatures in the range of 280–650 K. $P_{i,i+1}^{\text{ex}}$ are calculated by default by Gromacs. The solid line is a guide for the eye, set at $P^{\text{ex}} = 0.52$, and shows that $P_{i,i+1}^{\text{ex}}$ is uniform for all i 's.

We have numerically checked the behavior of P^{ex} in PT molecular dynamics simulations using eqs 5 and 8 for 256 replicas in the temperature range 280–650 K. Tests were performed using Gromacs²⁷ on a system of 15 500 waters, which is the size required in protein folding and large scale conformational changes.²⁸ Figure 3 shows the uniform distribution of $P_{i,i+1}^{\text{ex}}$ that was obtained using the present T_i distribution; in contrast, a geometric distribution of T_i leads to an increasing $P_{i,i+1}^{\text{ex}}$ with a replica index.

Performing the simulations to obtain a large number of roundtrips required for an estimate of the converged average round-trip time (τ) as a function of N_T is an expensive proposition. Thus, we resort to a simplified model. We assumed in each replica the energy fluctuations are Gaussian distributed with mean \bar{E} and width σ and obtain these parameters using eq 2. Assuming that the exchange attempts between i and $i + 1$ are made at an interval larger than the energy correlation time, the energies E_i and E_{i+1} at every swap attempt will be uncorrelated and Gaussian distributed. Thus, we simulate swapping attempts comparing two random energies extracted from their respective distributions. In such a way, it is a simple exercise to evaluate τ . We checked the validity of this model by comparing the τ evaluated with those obtained in the explicit solvent simulations of a smaller system with 215 water molecules. The results are shown in Figure 4.

Using this model validated in a small system, we computed the τ with the simplified model for a box of 15 500 waters. As a function of N_T , one can identify three regions (Figure 5). In the low N_T region, the geometric distribution τ is dominated by the low $P_{i,i+1}^{\text{ex}}$ in the cold replicas. In the same regime, this shortcoming is avoided by the use of eq 5, which places colder replicas at closer intervals and results in a higher P^{ex} in the colder replicas (Figure 3). The computational gain obtained using eq 5 is particularly advantageous in the simulations of large systems, where one is constrained to work with lesser than optimal number of replicas.

In the mid N_T range, τ is low and N_T is optimal. Assuming a well-defined minimum in τ and for an alternative odd, even replica-pair exchange scheme, theoretical estimates of the optimal number of replicas ($N_T^* = 1 + (0.594(mC_V)^{1/2} - 1/2)\log(T_{N_T}/T_1)$) for simulating m water molecules have been discussed.¹⁹ This estimate for our system with a 15 500 water molecules and using the C_V at room temperature is $N_T^* \approx 185$. The optimal τ in our model however is obtained at 120 replicas. In

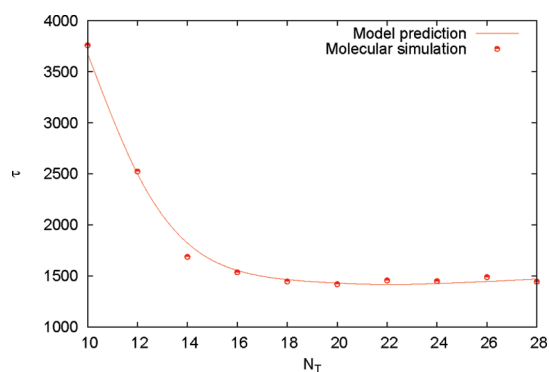


Figure 4. τ obtained from molecular simulations of a box of 215 waters is compared with that from the model. To obtain the τ corresponding to N_T , 100 ns simulations were performed on N_T replicas. Converged round-trip times were obtained from these simulations.

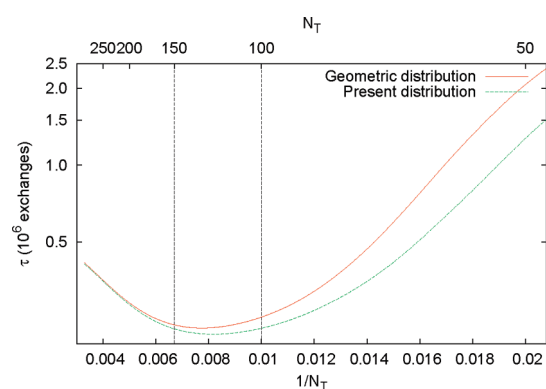


Figure 5. τ in units of millions of exchanges is shown in log scale for a box of 15 500 waters. τ was computed using the model validated in Figure 4. The three zones of decreasing, intermediate, and increasing τ with N_T are marked with dotted lines.

addition, τ has a weak dependence on N_T in this mid- N_T region. Because of this weak dependence, the simulations can be performed with a lower number of replicas, say 100, only marginally compromising efficiency. In the high N_T region, which is suboptimal, τ increases with N_T , and the difference between the two distributions becomes small, as it should.

To conclude, eq 5 offers a simple and practical way of generating replica temperatures for obtaining a uniform exchange and improving the computational efficiency, especially when performing the calculations on a smaller number of replicas. The temperature distribution proposed in the present work is based on the realistic dependence of C_V in the context of explicit solvent NVT, thus becoming directly relevant for improving the efficiency of biomolecular PT simulations.

AUTHOR INFORMATION

Corresponding Author

*Phone: +41-58-666-4811. E-mail: meher@phys.chem.ethz.ch.

ACKNOWLEDGMENT

We thank Max Bonomi for the round-trip simulation code and a critical reading of the manuscript and Hagai Eshet for

helpful discussions. We thank Swiss National Supercomputing Center (CSCS), Manno and High Performance Computing Group of ETH Zurich for computational resources. Financial support from the European Union grant (ERC-2009-AdG-247075) and Canton Ticino (“CSCS in rete”) are gratefully acknowledged.

REFERENCES

- (1) Geyer, C. J.; Thompson, A. *J. Am. Stat. Assoc.* **1995**, *90*, 909–920.
- (2) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.
- (3) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
- (4) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (5) Earl, D. J.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910–3916.
- (6) Rosta, E.; Hummer, G. *J. Chem. Phys.* **2009**, *131*, 165102.
- (7) Lingenheil, M.; Denschlag, R.; Mathias, G.; Tavan, P. *Chem. Phys. Lett.* **2009**, *478*, 80–84.
- (8) Nymeyer, H. *J. Chem. Theory Comput.* **2008**, *4*, 626–636.
- (9) Kofke, D. A. *J. Chem. Phys.* **2002**, *117*, 6911–6914.
- (10) Predescu, C.; Predescu, M.; Ciabanu, C. *J. Chem. Phys.* **2004**, *120*, 4119–4128.
- (11) Kofke, D. A. *J. Chem. Phys.* **2004**, *121*, 1167.
- (12) Kone, A.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 206101.
- (13) Rathore, N.; Chopra, M.; de Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 024111.
- (14) Katzgraber, H. G.; Trebst, S.; Huse, D. A.; Troyer, M. *J. Stat. Mech.* **2006**, *6*, P03018.
- (15) Trebst, S.; Troyer, M.; Hansmann, U. H. E. *J. Chem. Phys.* **2006**, *124*, 174903.
- (16) Garcia, A. E.; Herce, H.; Paschek, D. *Ann. Rep. Comp. Chem.* **2006**, *2*, 83–95.
- (17) Nadler, W.; Hansmann, U. H. E. *Phys. Rev. E* **2007**, *75*, 026109.
- (18) Nadler, W.; Hansmann, U. H. E. *J. Phys. Chem. B* **2008**, *112*, 10386–10387.
- (19) Denschlag, R.; Lingenheil, M.; Tavan, P. *Chem. Phys. Lett.* **2009**, *473*, 193–195.
- (20) Bonomi, M.; Parrinello, M. *Phys. Rev. Lett.* **2010**, *104*, 190601.
- (21) Dass, N.; Varshneya, N. C. *J. Phys. Soc. Jpn.* **1968**, *25*, 1452–1454.
- (22) Efimov, Y. Y.; Naberukhin, Y. I. *J. Struct. Chem.* **2000**, *41*, 433–439.
- (23) Efimov, Y. Y.; Naberukhin, Y. I. *Spectrochim. Acta, Part A* **2005**, *42*, 1789–1794.
- (24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (25) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (26) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins* **2002**, *46*, 225.
- (27) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (28) Pitera, J. W.; Swope, W. C.; Abraham, F. F. *Biophys. J.* **2008**, *94*, 4837–4846.

Generalized Normal Coordinates for the Vibrational Analysis of Molecular Dynamics Simulations

Gerald Mathias^{*,†,§} and Marcel D. Baer^{‡,§}

[†]Lehrstuhl für BioMolekulare Optik, Ludwig–Maximilians Universität München, Oettingenstrasse 67, 80538 München, Germany

[‡]Chemical and Materials Science Division, Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352, United States

ABSTRACT: The computation of vibrational spectra via molecular dynamics (MD) simulations has made lively progress in recent years. In particular, infrared spectra are accessible employing ab initio MD, for which only the total dipole moment has to be computed “on the fly” from the electronic structure along the trajectory. The analysis of such spectra in terms of the normal modes of intramolecular motion, however, still poses a challenge to theory. Here, we present an algorithm to extract such normal modes from MD trajectories by combining several ideas available in the literature. The algorithm allows one to compute both the normal modes and their vibrational bands without having to rely on an equipartition assumption, which hampered previous methods. Our analysis is based on a tensorial definition of the vibrational density of states, which spans both the frequency resolved cross- and auto-correlations of the molecular degrees of freedom. Generalized normal coordinates are introduced as orthonormal transforms of mass-weighted coordinates, which minimize their mutual cross-correlations. The generalized normal coordinates and their associated normal modes are iteratively constructed by a minimization scheme based on the Jacobi diagonalization. Furthermore, the analysis furnishes mode local temperatures, which provide not only a measure for the convergence of the computed intensities but also permits one to correct these intensities a posteriori toward the ensemble limit. As a first non-trivial test application we analyze the infrared spectrum of isoprene based on ab initio MD, which is an important building block of various dye molecules in molecular biology.

1. INTRODUCTION

Vibrational spectra of molecules, such as infrared (IR), Raman, or resonance Raman spectra, encode rich information about chemical bonds, intramolecular forces, and molecular structures. Moreover, in the condensed-phase vibrational spectra can be strongly modified by interactions with polar solvents or more complex environments, such as a protein matrix. Here, vibrational spectra also serve as a probe for intermolecular interactions and, thus, the solvent environment.

The decoding of vibrational spectra, i.e., the assignment of vibrational bands to molecular structures and motions, is often tedious for polyatomic molecules. On the experimental side, it involves, for example, site specific isotope substitutions or mutations of amino acid side chains, when it comes to protein spectra. Therefore, theory has become more and more important to help decipher information encoded in vibrational spectra.^{1–11} With the development of density functional theory (DFT) and reliable gradient corrected functionals, one is nowadays able to compute intramolecular forces to high accuracies at the computational expense of a Hartree–Fock calculation. On the analysis side, the routine tools are the normal-mode analysis (NMA) techniques based on an expansion of the potential energy surface around the equilibrium structure, i.e., the minimum structure at zero temperature. For medium-sized quasi-rigid molecules in the gas phase, such NMA methods based on the harmonic approximation can reach excellent accuracy if method specific scaling factors are employed.¹² In addition to the frequency spectrum, these

methods deliver the normal modes of intramolecular motion and the approximate intensities for IR and Raman absorption.

However, finite temperature effects, such as shifting and broadening of bands, due to the fluctuating structure of the molecule or its solvent environment in condensed phase are beyond the scope of traditional NMA methods. Here, proper sampling of these fluctuations is indispensable, and, thus, molecular dynamics (MD) simulations are the method of choice whenever the system is too large to allow a quantum mechanical treatment of the nuclei.

Within MD simulations the calculation of linear absorption cross-sections is straightforward, as it relates to the Fourier transform of the dipole autocorrelation function, if a proper quantum correction is employed.^{3,5,9–11,13–16} Also peaks in the power spectra of the atomic velocities give a reasonable overview of frequencies that are important for the intramolecular motion.^{16,17}

Assigning these bands to vibrational motifs is by no means straightforward. A simple but often sufficient approach for band assignment is to project the trajectory on some internal coordinates of the molecule and to take the Fourier transforms of their correlation functions.^{5,7,18} This approach is of course neither universal nor complete. Much more involved are instantaneous normal-mode techniques^{19–21} and the instantaneous normal-mode analysis (INMA),^{13,22–25} for which one computes full or partial Hessians of the system along the trajectory and then

Received: February 22, 2011

Published: May 10, 2011

averages over the corresponding frequencies and intensities. To avoid negative frequencies, the INMA technique employs local structural relaxations prior to the calculation of the Hessian. These methods get quite involved, when no analytic second derivatives are at hand, and thus, the number of sampling points is limited for larger molecules.

It is much more appealing to use the information encoded in the MD trajectory directly to construct the normal modes, as such an approach only involves an a posteriori effort for the analysis. Here, advances like the principal mode analysis (PMA)^{22,26–30} and related methods^{16,31} use cross-correlations of Cartesian coordinates or momenta to derive eigenvalue equations for the normal modes. Particularly Martinez et al. provide a general framework of Fourier transformed time correlation functions of coordinates and momenta and relate these to both NMA and PMA.^{16,31}

These methods rely to a lesser or greater extent on the equipartition assumption for the construction of the normal modes. As a result, if equipartition is violated, then the normal coordinates given by Martinez et al. are in general not orthogonal transforms of mass-weighted coordinates, even for a purely harmonic system.^{16,31} As a remedy the authors suggest to orthonormalize the normal coordinate vectors a posteriori. Furthermore, Schmitz and Tavan show that for methods relying on the diagonalization of correlation matrices, one can get artificial mixing of normal coordinates.^{22,28} For PMA this is the case if $T_k/\omega_k^2 \approx T_l/\omega_l^2$, i.e., if the ratios of the mode local temperatures T_k and squared eigenfrequencies ω_k of two modes become comparable.

In this article we combine ideas of these cited works and describe a method to determine generalized normal coordinates directly from (ab initio) MD simulations. The method does not rely on an equipartition assumption and is, therefore, also applicable if even approximate equipartition cannot be reached. We define these generalized normal coordinates as orthogonal linear combinations of mass-weighted Cartesians, which minimize their mutual correlation. To measure this correlation, we define a functional based on a tensorial formulation of the vibrational density of states (VDOS) and provide a corresponding minimization strategy based on the Jacobi diagonalization. Employing the convolution theorem we show how these correlation functions can be efficiently calculated in the Fourier domain.

For the analysis in a molecular frame of reference, the Eckart frame,³² we compare the transformation of coordinates and velocities via a mass-weighted root-mean-square deviation (rmsd) fit and, alternatively, the transform via internal coordinates. In particular, the performance of both methods for large amplitude motions, such as the rotation of a methyl group, are examined.

Furthermore, we derive mode specific IR intensities, which are efficiently and accurately calculated within the framework of correlation functions in Fourier space. These mode local intensities decompose the global IR spectrum obtained from the dipole auto-correlation functions and allow to assign peaks in the IR spectrum to normal-mode vibrations. Furthermore, these local intensities permit us to correct the overall IR spectrum a posteriori toward the ensemble limit.

As a non-trivial application we examine the vibrational modes of isoprene (*trans*-2-methyl-1,3-butadiene) in the gas phase. To mimic the gas phase, we consider the isolated but rotating molecule, whose angular momentum is sampled by 41 initial conditions from a canonical distribution. Thereby we include the rovibrational couplings, albeit classical, in our model

and can observe rotationally induced frequency shifts and line broadenings. These effects are not accessible by standard frequency calculations that ignore molecular rotation.

Isoprene has been chosen because it is an important building block of biologically important dye molecules, such as carotenoids and retinal. In particular, the latter is an important probe for vibrational spectroscopists, as it plays a central role in bacterial photosynthesis and mammalian vision. Still, the decoding of changes in the corresponding spectra due to changes in the environment and in the molecule remains a challenge to theory.^{2,33} For isoprene we are, therefore, particularly interested in how well the C–C and C=C stretching modes, as well as the hydrogen out of plane (HOOP) modes, are described by the ab initio MD simulations, as these modes serve as marker modes for this class of molecules. Furthermore, we demonstrate how our analysis aids to check the convergence of the computed IR spectra with respect to sampling and examine the quality of the suggested temperature correction.

Beyond the comparatively well-behaved isoprene molecule, an extended version of the algorithm, which can deal with multiple reference structures and conformations, has been already employed for the vibrational analysis of floppy molecules, like CH_5^+ and its isotopologues, as well as for microsolvated hydronium and Zundel ions,^{9,11,34} which exhibit large amplitude motion during the dynamics. Here, the algorithm has demonstrated its ability to deal with such challenging problems. The necessary extension to multiple reference structures is, however, beyond the scope of this article and will be described in a subsequent publication.

2. THEORY

We start our derivation of generalized normal coordinates with the basic relations of atomic velocities, kinetic energy, and temperature, which ultimately lead us to a tensorial definition of the VDOS. Close examination of the latter object in the case of purely harmonic motion will result in a principle of minimal cross-correlation between generalized normal coordinates, together with a corresponding minimization procedure. After discussing the transform to the Eckart frame employing an rmsd fit or internal coordinates, we give a computationally convenient recipe to assign individual IR intensities to the vibrational modes.

2.1. Vibrational Density of States (VDOS). MD simulations render the time evolution of a system from time t_i to t_f by a finite trajectory of Cartesian coordinates $\mathbf{x}(t): [t_i, t_f] \rightarrow \mathbb{R}^n$ and velocities $\mathbf{v}(t) = \dot{\mathbf{x}}(t)$ of n DOFs, where the time derivative of a property a is denoted by \dot{a} . A molecular system of N atoms consists of $n = 3N$ Cartesian coordinates $\mathbf{x} = (\mathbf{r}_1^T, \dots, \mathbf{r}_N^T)^T$, where the superscript ‘T’ denotes a matrix transpose. In the course of our vibrational analysis we will make extensive use of the Fourier transform (FT):

$$\hat{a}(\omega) = \int_{t_i}^{t_f} a(t) e^{-i\omega t} dt \quad (1)$$

of dynamical properties, such as velocities and dipole moments, where we mark the FT of a property by the caret accent. In order to write the finite FT integral as an infinite one we introduce a window function $W(t)$ and define the FT integral as

$$\hat{a}(\omega) = \int W(t) a(t) e^{-i\omega t} dt \quad (2)$$

In a rather general fashion, we demand that W is non-negative, bounded and that

$$\text{supp}(W) \subseteq [t_i, t_f] \quad (3)$$

i.e., $W(t) = 0$ outside the interval $[t_i, t_f]$. Within these limits, W can additionally serve to remedy artifacts of the FT, which arise from the finite simulation length and the discrete sampling.^{35,36} The simplest form of W is a rectangular window, which can be trivially realized³⁵ with the Heaviside step function θ as

$$\text{rect}(t; t_i, t_f) = \theta(t - t_i)\theta(t_f - t) \quad (4)$$

Having established the basic notation we start the derivation of our vibrational analysis by looking at the kinetic energy of the system along the trajectory, which is given by

$$E_{\text{kin}}(t) = \frac{1}{2} \mathbf{v}^T(t) \mathbf{M} \mathbf{v}(t) \quad (5)$$

where the mass matrix \mathbf{M} is diagonal with elements $M_{kl} = \delta_{kl} m_k$ containing masses m_k associated with each DOF k , δ being the Kronecker symbol. From the theory of small vibrations we introduce mass-weighted coordinates:

$$\mathbf{c}(t) = \mathbf{M}^{1/2} \mathbf{x}(t) \quad (6)$$

which simplifies the kinetic energy:

$$E_{\text{kin}}(t) = \frac{1}{2} \dot{\mathbf{c}}(t) \cdot \dot{\mathbf{c}}(t) \quad (7)$$

to comprise only a scalar product of the mass-weighted velocities. The average temperature T of the system is computed from the time average of the kinetic energy by

$$T = \frac{2}{k_B(n - n_c)} \frac{\int W^2(t) \dot{\mathbf{c}}(t) \cdot \dot{\mathbf{c}}(t) dt}{\int W^2(t) dt} \quad (8)$$

where n_c denotes the number of constrained DOFs, e.g., the global translation or rotation. Here we used the rather unusual weighting function $W^2(t)$ for which the choice $W(t) = \text{rect}(t)$ leads, however, straight to the common temperature definition, as $\text{rect}^2 = \text{rect}$. Due to the weighting function $W^2(t)$ we can now employ Plancherel's theorem,³⁷ which states that the FT transform preserves scalar products of functions and write the temperature:

$$T = \frac{2}{k_B(n - n_c)} \frac{\int \hat{\mathbf{c}}^*(\omega) \cdot \hat{\mathbf{c}}(\omega) d\omega}{\int \hat{W}^*(\omega) \hat{W}(\omega) d\omega} \quad (9)$$

in the Fourier domain (see eq 2), where the asterisk marks the complex conjugate. From the integral kernel of the numerator, which is an even function in ω , we readily derive the scalar VDOS as

$$\kappa(\omega) = \frac{4}{k_B T} \frac{\hat{\mathbf{c}}^*(\omega) \cdot \hat{\mathbf{c}}(\omega)}{\int \hat{W}^*(\omega) \hat{W}(\omega) d\omega} \quad (10)$$

which is normalized such that

$$\int_0^\infty \kappa(\omega) d\omega = n - n_c \quad (11)$$

Thus, in the ensemble limit, where each and every DOF contributes equally to $\kappa(\omega)$ due to equipartition, the integral:

$$n[\omega_s, \omega_e] = \int_{\omega_s}^{\omega_e} \kappa(\omega) d\omega \quad (12)$$

“counts” the number of vibrational modes contributing with their kinetic energy to the interval $[\omega_s, \omega_e]$. For this property, it is essential to include the mass dependence of the kinetic energy in the definition of the VDOS in eq 10. Note that $\kappa(\omega)$ is the Fourier transform of the time correlation function of the mass-weighted velocities, due to the cross-correlation theorem:

$$\hat{a}^*(\omega) \hat{b}(\omega) = \iint a^*(\tau) b(\tau + t) d\tau e^{-i\omega t} dt \quad (13)$$

which is closely related to the convolution theorem.

For an assignment of vibrational modes to bands in the VDOS we have to analyze the individual contributions of the atomic DOFs. For this purpose, we define the tensorial VDOS Θ : $\rightarrow \mathbb{R}^n \times \mathbb{R}^n$ as

$$\begin{aligned} \Theta(\omega) &= \frac{2}{k_B T} \frac{\hat{\mathbf{c}}^*(\omega) \otimes \hat{\mathbf{c}}(\omega) + \hat{\mathbf{c}}(\omega) \otimes \hat{\mathbf{c}}^*(\omega)}{\int \hat{W}^*(\omega) \hat{W}(\omega) d\omega} \\ &= \frac{4}{k_B T} \frac{\mathcal{R}[\hat{\mathbf{c}}^*(\omega) \otimes \hat{\mathbf{c}}(\omega)]}{\int \hat{W}^*(\omega) \hat{W}(\omega) d\omega} \end{aligned}$$

Here, we have replaced the scalar product in the definition of $\kappa(\omega)$, given in eq 10, by the outer or tensorial product $\hat{\mathbf{c}} \otimes \hat{\mathbf{c}} = \hat{\mathbf{c}} \hat{\mathbf{c}}^T$ and inserted the definition of the temperature from eq 8. Furthermore, we have included the symmetrization $\mathcal{R}(\hat{\mathbf{c}}^* \otimes \hat{\mathbf{c}}) = 1/2(\hat{\mathbf{c}}^* \otimes \hat{\mathbf{c}} + \hat{\mathbf{c}} \otimes \hat{\mathbf{c}}^*)$ to ensure that the off-diagonal elements of $\Theta(\omega)$ are real valued. This symmetrization corresponds to averaging over the time forward and the time reversed trajectories, $\mathbf{x}(t)$ and $\mathbf{x}(-t)$, which are both valid trajectories in time-reversible dynamics.

The diagonal elements $\Theta_{kk}(\omega)$ are related to $\kappa(\omega)$ by the trace:

$$\kappa(\omega) = \sum_k \Theta_{kk}(\omega) \quad (14)$$

and, therefore, $\Theta_{kk}(\omega)$ yields the individual contribution to $\kappa(\omega)$ by DOF k . Since κ is derived from the global temperature eq 8 the integral:

$$\int_0^\infty \Theta_{kk}(\omega) d\omega = \frac{T_k}{T} \quad (15)$$

furnishes the temperature T_k of DOF k .

The off-diagonal elements $\Theta_{k \neq l}(\omega)$ contain the cross-correlations between the k -th and l -th DOF. Note that $\Theta(\omega)$ is an extension of the Fourier transformed time correlation function of the velocities $\mathbf{P}^x(\omega)$ defined by Martinez et al.,³¹ for which we have included the mass matrix \mathbf{M} as a metric tensor in the definition of the correlation function.

2.2. VDOS in the Harmonic Case. In order to establish a procedure of assigning the vibrational bands in $\kappa(\omega)$ to atomic motion we first focus on the properties of $\Theta(\omega)$ in the case of nonergodic sampling of a purely harmonic system with $n_c = 0$. From the harmonic analysis we introduce the orthogonal transform $\mathbf{D}^h \in O_n$ of the mass-weighted coordinates, with O_n

denoting the orthogonal group on \mathbb{R}^n , to the so-called normal coordinates:

$$\mathbf{q} = \mathbf{D}^h \mathbf{c} \quad (16)$$

which factorize the Hamiltonian of the system as

$$H = \frac{1}{2} \sum_{k=1}^n (\dot{q}_k^2 + \omega_k^2 q_k^2) \quad (17)$$

with the eigenfrequencies $\omega_k > 0$.³⁸ Note that \mathbf{D}^h is not unique, because a mere permutation of the coordinate numbering or a linear combination of modes that are degenerate in frequency does not change the physics of the system but leads formally to a different \mathbf{D}^h . The resulting classical equations of motion are solved by

$$q_k(t) = a_k \sin(\omega_k t + \varphi_k) \quad (18)$$

in which the set of amplitudes a_k and phases φ_k depends on the initial conditions.

The finite FT of the normal coordinate velocities are then given as

$$\hat{q}_k = \frac{a_k \omega_k}{2} [e^{i\varphi_k} \hat{W}(\omega - \omega_k) + e^{-i\varphi_k} \hat{W}(\omega + \omega_k)] \quad (19)$$

where \hat{W} is the FT of the window function defined in eq 2. For example, a rectangular window function has the FT:³⁵

$$\int \text{rect}(t; t_i, t_f) e^{-i\omega t} dt = e^{-i\omega \frac{t_i+t_f}{2}} \frac{\sin[(t_f - t_i)\omega/2]}{\omega/2} \quad (20)$$

The elements of $\Theta(\omega)$ evaluate to

$$\Theta_{kl}(\omega) = \frac{2na_k a_l \omega_k \omega_l}{\sum_j a_j^2 \omega_j^2} \times \frac{\cos(\phi_{kl}) \mathcal{R}[\hat{Y}_{kl}(\omega)] - \sin(\phi_{kl}) \mathcal{I}[\hat{Y}_{kl}(\omega)]}{\int \hat{W}^*(\omega) \hat{W}(\omega) d\omega}$$

with the abbreviations $\varphi_{kl} = \varphi_k - \varphi_l$ and

$$\hat{Y}_{kl}(\omega) = \hat{W}(\omega - \omega_k) \hat{W}^*(\omega - \omega_l) + \hat{W}^*(\omega + \omega_k) \hat{W}(\omega + \omega_l) \quad (21)$$

Here, we have assumed that $\hat{W}(\omega + \omega_k) \hat{W}(\omega - \omega_l) \approx 0$, i.e., that bands belonging to positive and negative frequencies do not overlap. The mode temperatures eq 15:

$$T_k = T \frac{na_k^2 \omega_k^2}{\sum_l a_l^2 \omega_l^2} \quad (22)$$

are solely determined by the initial conditions and are independent of the sampling length in the harmonic case.

The corresponding integrals of the off-diagonal elements $\Theta_{k \neq l}$ vanish, if the overlap between $\hat{W}(\omega - \omega_k)$ and $\hat{W}(\omega - \omega_l)$ is negligible. Moreover, also the two-norm:

$$\|\Theta_{kl}\| = \left(\int [\Theta_{kl}(\omega)]^2 d\omega \right)^{1/2} \quad (23)$$

of $\Theta_{k \neq l}$ vanishes in this case, which simply reflects that the q_k and q_l are uncorrelated.

If we employ, instead of \mathbf{D}^h , an arbitrary transform $\mathbf{D} \in O_n$ on the mass-weighted coordinates, the $\|\Theta_{k \neq l}\|$ will be in general nonzero.

In the case of finite overlap between $\hat{W}(\omega - \omega_k)$ and $\hat{W}(\omega - \omega_l)$, e.g., when modes k and l are degenerate, the $\|\Theta_{k \neq l}\|$ of the normal coordinates will, in general, not vanish for a single trajectory. However, for the ensemble average over many initial conditions we find $\|\Theta_{k \neq l}\| = 0$, due to the phase factors $\cos(\varphi_{kl})$ and $\sin(\varphi_{kl})$ in eq 21.

2.3. Generalized Normal Coordinates. Our findings for the purely harmonic case in the previous section now lead us to a generalization of normal coordinates for anharmonic Hamiltonians. We define the measure of the overall intermode coupling as the functional:

$$\text{off}[\Theta(\omega)] = \left[\sum_{k \neq l} \int [\Theta_{kl}(\omega)]^2 d\omega \right]^{1/2} \quad (24)$$

which vanishes for the ensemble average for the normal coordinates of a strictly harmonic system. For the general anharmonic case we call

$$\mathbf{q}^g = \left\{ \mathbf{D} \mathbf{c} \mid \min_{\mathbf{D} \in O_n} \text{off}[\mathbf{D}\Theta(\omega)\mathbf{D}^{-1}] \right\} \quad (25)$$

generalized normal coordinates, which are equivalent to the standard normal coordinates in the harmonic limit. Note that Θ transforms as

$$\Theta'(\omega) = \mathbf{D}\Theta(\omega)\mathbf{D}^{-1} \quad (26)$$

for any $\mathbf{D} \in O_m$ and since the trace is invariant under orthonormal transforms, $\kappa(\omega)$ is preserved, see eq 14.

If Θ would be a regular matrix, then the task of minimizing $\text{off}[\Theta]$ would correspond to diagonalizing Θ . Due to the frequency dependence of Θ this is not achievable by a single transform for all values of ω simultaneously. Hence, we have to find the transform which brings $\Theta(\omega)$ close to diagonal form for all ω in a balanced fashion. For this purpose, we modify the well-known and robust Jacobi diagonalization algorithm.^{39,40} The Jacobi diagonalization relies on the idea that any $\mathbf{D} \in O_n$ can be written as a product:

$$\mathbf{D} = \prod_{k=1}^j \mathbf{J}(u_k, v_k, \vartheta_k) \quad (27)$$

of j Jacobi (or Givens) rotations:

$$\mathbf{J}(u, v, \vartheta) = \begin{pmatrix} & & & u & & v & & & \\ & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \\ & \vdots & \ddots & \vdots & & \vdots & & \vdots & \\ u & 0 & \cdots & c & \cdots & s & \cdots & 0 & \\ \vdots & \vdots & & \vdots & \ddots & \vdots & & \vdots & \\ v & 0 & \cdots & -s & \cdots & c & \cdots & 0 & \\ \vdots & \vdots & & \vdots & & \vdots & \ddots & \vdots & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \quad (28)$$

with $s = \sin(\vartheta)$ and $c = \cos(\vartheta)$. Thus, multiplying \mathbf{x} by $\mathbf{J}(u, v, \vartheta)$ applies a rotation in the u, v plane by the angle ϑ to \mathbf{x} . For the Jacobi diagonalization of a matrix \mathbf{A} one chooses ϑ for a given pair u, v such that A'_{uv} vanishes after the transformation $\mathbf{A}' = \mathbf{J}\mathbf{A}\mathbf{J}^T$ and

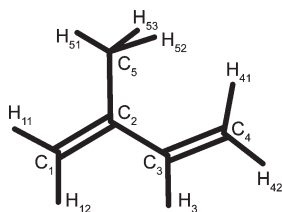


Figure 1. Structure and atomic labels of isoprene.

iterates the procedure over u and v . Considering eq 24 we see that the analogous approach is to minimize $\|\Theta_{u \neq v}\|$, since it is in general not possible to bring it to zero. Thus, the optimal rotation angle ϑ obeys

$$\left. \frac{\partial}{\partial \vartheta'} \int [\mathbf{J}(u, v, \vartheta') \Theta(\omega) \mathbf{J}(u, v, \vartheta')^T]_{uv}^2 d\omega \right|_{\vartheta' = \vartheta} = 0 \quad (29)$$

This leads to an algebraic equation of fourth order, which is given in the Appendix together with its analytic roots $\vartheta_1, \dots, \vartheta_4$. For each iteration complex solutions and values $|\vartheta_j| > \pi/4$ are discarded.⁴¹

For the matrix Jacobi algorithm one often picks the largest off-diagonal element to determine u, v for the next step.⁴¹ Since, for our purpose, we cannot easily judge if such an element can be further reduced, we loop over all possible pairs $u < v$, i.e., perform a Jacobi sweep. When convergence is reached after multiple sweeps, \mathbf{D}^g is reconstructed via eq 27.

2.4. Molecular Frame of Reference. A molecular frame of reference is crucial for any approach that constructs normal modes from atomic motion. Within the frame of reference one, furthermore, assumes that the molecule is quasi-rigid, i.e., that the intramolecular motions describe small oscillations around a single reference structure \mathbf{x}^0 , which corresponds to a local minimum and does not undergo any conformational changes.

During a simulation molecules rotate and translate in general. Therefore, for a vibrational analysis one transforms the molecular coordinates to the Eckart frame of reference connected to an average molecular structure.^{28,31,32} This is achieved by applying a mass-weighted root-mean-square (rms) fit to the Cartesian coordinates of a reference structure \mathbf{x}^0 . Here, one determines new atomic coordinates $\mathbf{r}'_i = \mathbf{R}\mathbf{r}_i + \mathbf{I}$ by a translation vector \mathbf{I} and a rotation matrix $\mathbf{R} \in O_3$ for each time step, which minimize the mass-weighted distance:

$$d_{\text{rms}} = \sqrt{\frac{\sum_{i=1}^N m_i (\mathbf{r}'_i - \mathbf{r}_i^0)^2}{\sum_{i=1}^N m_i}} \quad (30)$$

and transforms the velocities accordingly.

However, if the assumption of quasi-rigidity does not apply globally, then problems with the rms approach can arise, which we discuss on our test molecule isoprene, which is displayed in Figure 1. Here, the methyl group at the C₅ position rotates nearly freely around C₂—C₅ bond at ambient conditions due to the shallow rotational barrier. For the hydrogens this corresponds to a large amplitude motion. In particular, rotations of 120° around this bond preserve the conformation of the molecule but lead to a large local rmsd because of the underlying classical dynamics, for which the methyl hydrogens are discriminable, in contrast to a

quantum mechanical treatment. Correspondingly, averaging over the motion of the hydrogens will render poor results.

An alternative to a rmsd fit which may help to mend these problems is to describe the intramolecular DOF by a complete nonredundant set of internal coordinates.¹⁶ These can comprise primitive coordinates, such as bond distances, bond and dihedral angles, or linear combinations of such primitives.⁴² For example the rotation of a methyl group can be described by a sum of dihedral angles, which is invariant under rotations of 120°. For a molecule we need $n - n_c$ linearly independent internal coordinates $s_i(\mathbf{x})$. Thus $\mathbf{s}(\mathbf{x}) = [s_1(\mathbf{x}), \dots, s_{n-n_c}(\mathbf{x})]^T$ describes a transform to curvilinear coordinates. This transform eliminates the translational and rotational DOFs.

Due to the restriction of our vibrational analysis to orthonormal transforms of mass-weighted Cartesians, we need to map this motion of internal coordinates back to Cartesian displacements from the reference structure, i.e., to the Eckart frame. However, a closed reverse transform to \mathbf{x} is not possible, since \mathbf{s} and \mathbf{x} are of different dimensions.

Because we actually need only a transform to the mass-weighted velocities $\dot{\mathbf{c}}$, we start from the kinetic energy:

$$E_{\text{kin}} = \frac{1}{2} \dot{\mathbf{s}}^T \mathbf{G}^{-1} \dot{\mathbf{s}} \quad (31)$$

of the internal coordinate velocities $\dot{\mathbf{s}}$, where the matrix $\mathbf{G} = \mathbf{B}^T \mathbf{M}^{-1} \mathbf{B}$ is constructed from the derivative of the internal coordinates:

$$B_{ij} = \left. \frac{\partial s_j}{\partial x_i} \right|_{\mathbf{x}^0} \quad (32)$$

at the reference structure \mathbf{x}^0 and the mass matrix \mathbf{M} .³⁸

We can factorize $\mathbf{G} = \mathbf{K}\mathbf{K}^T$ by

$$\mathbf{K} = (\mathbf{M}^{-1/2} \mathbf{B})^T \quad (33)$$

and correspondingly the inverse $\mathbf{G}^{-1} = (\mathbf{K}^+)^T \mathbf{K}^+$ by the pseudoinverse:

$$\mathbf{K}^+ = \mathbf{K}^T (\mathbf{K}\mathbf{K}^T)^{-1} \quad (34)$$

of \mathbf{K} , which satisfies $\mathbf{K}\mathbf{K}^+ = \mathbf{E}_{n-n_c}$. Defining the transformed mass-weighted velocities:

$$\dot{\mathbf{c}}' = \mathbf{K}^+ \dot{\mathbf{s}} \quad (35)$$

yields the correct kinetic energy, which can be checked by inserting these definitions into eqs 7 and 31.

This procedure introduces an error in second order due to the expansion of the internal coordinates around \mathbf{x}^0 . The magnitude of this error mainly depends on the choice of \mathbf{x}^0 and the internal coordinates s .^{42,43} In the Results Section we will assess this error introduced by this forward and backward transform by comparing the resulting $\kappa(\omega)$ for both the rms fit and the internal coordinate transform.

This representation of intramolecular motion allows a characterization of the normal coordinates with respect to the internal coordinates by $\dot{\mathbf{q}} = \mathbf{D}^g \mathbf{K}^+ \dot{\mathbf{s}}$. Correspondingly, the displacement of internal coordinate j associated with normal coordinate k is

$$\left. \frac{\partial q_k}{\partial s_j} \right|_{\mathbf{x}^0} = (\mathbf{D}^g \mathbf{K}^+)_{kj} \quad (36)$$

Normalization of the contributions of s_j to all q_i to account for the different amplitudes of the s_j yields the relative contribution:

$$\lambda_{kj} = \frac{\partial q_k}{\partial s_j} \left[\sum_i \left(\frac{\partial q_i}{\partial s_j} \right)^2 \right]^{-1/2} \quad (37)$$

of s_j to q_k .

2.5. Assignment of IR Intensities. What remains after the construction of the generalized normal coordinates is their assignment to bands in experimental spectra.¹⁶ For the case of IR spectra we now develop a convenient procedure to compute the individual absorption of these modes. From a classical dynamics trajectory we approximate the tensorial IR absorption cross-section of our finite sample as¹⁴

$$\alpha^{\text{QC}}(\omega) = \left[\frac{2\pi\omega}{3V\hbar c_0 \hat{n}(\omega)} \right] \left(1 - \exp\left(-\frac{\hbar\omega}{k_B T}\right) \right) \times Q(\omega) \frac{\mathcal{R}[\hat{\boldsymbol{\mu}}^*(\omega) \otimes \hat{\boldsymbol{\mu}}(\omega)]}{\int \hat{W}^*(\omega) \hat{W}(\omega) d\omega} \quad (38)$$

where V denotes the sample volume, c_0 the speed of light, and $\hat{n}(\omega)$ is the defraction coefficient. Once again, we have used the cross-correlation theorem to express the FT of the time correlation function of the finite dipole trajectory¹⁴ $\boldsymbol{\mu}(t)$ as the (tensorial) product of its FT $\hat{\boldsymbol{\mu}}(\omega)$ and normalized it similar to the definitions of $\kappa(\omega)$ and $\boldsymbol{\Theta}(\omega)$ in eqs 10 and 14. The quantum correction factor $Q(\omega)$ is usually introduced as an ad hoc correction that essentially imposes the detailed balance condition, which proper quantum time correlation functions satisfy, onto the FT of the classical correlation function.¹⁴ For systems without orientational order, such as molecules in the gas phase, the scalar absorption is given as $\alpha^{\text{QC}} = \sum_{k=1}^3 \alpha_{kk}^{\text{QC}}$. For molecular systems the so-called harmonic approximation for the quantum correction factor:

$$Q^{\text{HC}}(\omega) = \frac{1}{k_B T} \frac{\hbar\omega}{1 - \exp\left(-\frac{\hbar\omega}{k_B T}\right)} \quad (39)$$

has proven to render reliable results for the mid-IR range.^{5,13,14,22} Applying the harmonic quantum correction we rewrite eq 38 as

$$\alpha^{\text{HC}}(\omega) = \frac{\gamma}{k_B T \hat{n}(\omega)} \frac{\mathcal{R}[\hat{\mathbf{j}}^*(\omega) \otimes \hat{\mathbf{j}}(\omega)]}{\int \hat{W}^*(\omega) \hat{W}(\omega) d\omega} \quad (40)$$

where the parameter $\gamma = 2\pi/(3Vc_0)$ collects the constants and where we have introduced the current^{15,16} $\mathbf{j}(t) = \dot{\boldsymbol{\mu}}(t)$. For this equality we recall the identity $\hat{a} = i\omega\hat{a}$ for the time derivative, and thus, $\hat{\mathbf{j}} = i\omega\hat{\boldsymbol{\mu}}$.

In order to derive IR intensities for our normal modes, we first have to transform $\boldsymbol{\mu}$ to the frame of reference given by \mathbf{x}^0 . Here, the transformation matrix \mathbf{R} determined for the coordinates along the trajectory is equally applied to $\boldsymbol{\mu}$. Then $\boldsymbol{\mu}$ is expanded to linear order^{28,31} as

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu}_0 + \sum_k \mathbf{a}_k q_k(t) + \mathcal{O}(2) \quad (41)$$

where the yet undetermined expansion coefficients \mathbf{a}_k correspond to the derivatives of the dipole moment along the normal

coordinates q_k in the harmonic case. Defining the coefficient matrix $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n) \in \mathbb{R}^3 \times \mathbb{R}^n$ we write eq 41 as $\boldsymbol{\mu}(t) = \boldsymbol{\mu}_0 + \mathbf{A}\mathbf{q}(t)$ and get, due to linearity,

$$\hat{\mathbf{j}}(\omega) = \mathbf{A}\hat{\mathbf{q}}(\omega) \quad (42)$$

For a numerically stable and tractable way to determine \mathbf{A} within our framework, both sides of eq 42 are multiplied with $\otimes \hat{\mathbf{q}}^*$ to get

$$\hat{\mathbf{j}}(\omega) \otimes \hat{\mathbf{q}}^*(\omega) = \mathbf{A}[\hat{\mathbf{q}}(\omega) \otimes \hat{\mathbf{q}}^*(\omega)] \quad (43)$$

To relate \mathbf{A} and $\boldsymbol{\Theta}$ we define the correlation function:

$$\mathbf{C}(\omega) = (n - n_c) \frac{\mathcal{R}[\hat{\mathbf{j}}(\omega) \otimes \hat{\mathbf{c}}^*(\omega)]}{\int \hat{\mathbf{c}}^*(\omega) \cdot \hat{\mathbf{c}}(\omega) d\omega} \quad (44)$$

Analogous to the definition of $\boldsymbol{\Theta}$ in eq 14, taking the real value here corresponds to a symmetrization in time, which assumes that coordinates and dipoles are in phase. Note that \mathbf{C} can be conveniently calculated along with the initial tensorial VDOS $\boldsymbol{\Theta}$. If we now transform \mathbf{C} to the coordinate system defined by \mathbf{D}^h , then we get with eqs 14 and 43:

$$\mathbf{C}(\omega)(\mathbf{D}^h)^{-1} = \mathbf{A}\boldsymbol{\Theta}(\omega) \quad (45)$$

The effective dipole gradient matrix \mathbf{A} now is determined numerically by a least-squares fit as

$$\frac{\partial}{\partial \mathbf{A}} \int d\omega \left\| \mathbf{A}\boldsymbol{\Theta}(\omega) - \mathbf{C}(\omega)(\mathbf{D}^h)^{-1} \right\|_2 = 0 \quad (46)$$

Note that we can safely restrict this problem to values of ω , where the IR absorption is finite, e.g., $\omega < 4000 \text{ cm}^{-1}$.

The reconstructed IR absorption resulting from \mathbf{A} , eqs 38,40, and 42, is given by

$$\alpha'(\omega) = \frac{\gamma}{\hat{n}(\omega)} \mathbf{A}\boldsymbol{\Theta}(\omega)\mathbf{A}^T \quad (47)$$

where the normalizing factors in the respective definitions of \mathbf{C} and $\boldsymbol{\alpha}$ cancel with $k_B T$ according to eq 8. We can associate the contributions of the diagonal elements of $\boldsymbol{\Theta}$ to $\boldsymbol{\alpha}'$ given by

$$\alpha'_{kk}(\omega) = \frac{\gamma}{\hat{n}(\omega)} (\mathbf{a}_k \otimes \mathbf{a}_k) \boldsymbol{\Theta}_{kk}(\omega) \quad (48)$$

with individual mode absorptions,¹⁶ which describe both the band shapes and the total absorptions of the modes q_k . Note that the reconstructed IR absorption $\boldsymbol{\alpha}'(\omega)$ additionally includes cross-terms $\alpha'_{kl}(\omega)$, which account for the remaining correlations between modes q_k and q_l .

Having decomposed the total IR absorption into individual contributions of the generalized normal coordinates and their intermode couplings, we can approximate the absorption of an ergodic ensemble. If we assume predominantly harmonic modes, then all cross-correlations $\boldsymbol{\Theta}_{k \neq l}(\omega)$ should vanish. Furthermore, in the ergodic limit all modes should contribute according to the mean temperature of the system. Therefore, we estimate the temperature-corrected IR absorption in the ergodic limit as

$$\alpha^{\text{TC}}(\omega) = \sum_{k: T_k > 0} \frac{T}{T_k} \alpha'_{kk}(\omega) \quad (49)$$

Note that we have $T_k = 0$ for the constrained DOFs.

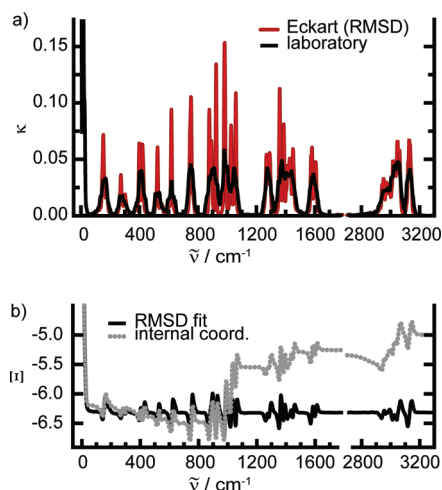


Figure 2. (a) VDOSs κ computed in the laboratory frame of reference (black) and in the Eckart (red) frame of reference obtained by an rmsd fit. (b) Integral difference Ξ (eq 50) computed for the transform to the Eckart frame via either rmsd fit (black line) or internal coordinates (gray dotted line).

3. METHODS

With the primary aim to demonstrate the usefulness of our novel method we have chosen the isoprene molecule, which is not only a basic building block of carotenoids and related dye molecules, like retinal in biological systems, but also serves as a realistic application of our approach.

Ab initio molecular dynamics (MD) simulations⁴⁴ within the framework of Kohn–Sham density functional theory (DFT) have been performed with the CP2k simulation package (see <http://cp2k.berlios.de>). Here, energies, forces, and dipole moments are computed “on the fly”, and the system is propagated on the Born–Oppenheimer surface.

We have employed the BLYP functional^{45,46} together with Goedecker–Teter–Hutter pseudopotentials^{47,48} for core electrons and a TZV2P basis set⁴⁹ for the valence electrons. Within the quickstep algorithm^{50,51} electrostatic interactions have been treated with a density cutoff of 280 Ry. The Poisson equation for the isolated isoprene molecule in a box of 12 Å side length has been solved by the Martyna–Tuckerman solver.⁵² Geometry optimization yielded a reference structure \mathbf{x}^0 , for which harmonic frequencies were calculated for comparison.

For the dynamics a set of 41 initial conditions has been sampled by an NVT simulation employing massive Nosé–Hoover chains at $T = 300$ K.⁵³ From these initial coordinates and velocities subsequent NVE runs of 25 ps each were run. The integration time step was 0.5 fs, and coordinates, velocities, and dipole moments were sampled each step and used for the vibrational analysis.

A set of nonredundant internal coordinates for the reference structure \mathbf{x}^0 has been obtained by the TURBOMOLE v5.9 quantum chemistry package.⁵⁴

The generalized normal coordinate algorithm has been implemented in the C code `normcor`. With standard linear algebra and fast FT libraries the fully converged minimization and normal coordinate analysis for isoprene took less than 10 min on an 3.0 GHz desktop. The optimization problem eq 25 converged to numerical accuracy after 13 Jacobi sweeps.

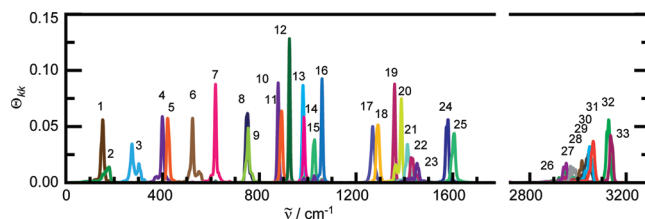


Figure 3. Power spectra $\Theta_{kk}(\omega)$ (colored lines) of the generalized normal coordinates denoted with labels 1–33.

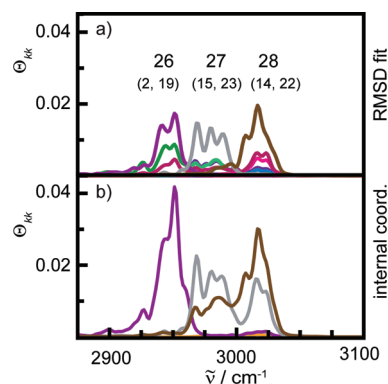


Figure 4. Power spectra $\Theta_{kk}(\omega)$ in the methyl stretching region for: (a) the rmsd fit and (b) the internal coordinate transform.

In order to remedy artifacts from the discrete and finite FT we employed a truncated Gaussian-type windowing function $W(t)$ (see eq 2), which corresponded to a convolution of the FTs by a Gaussian of width 5 cm^{-1} in the frequency domain.

4. RESULTS AND DISCUSSION

To set the stage for the normal coordinate analysis of isoprene, we first compare the VDOSs obtained in the laboratory $\kappa^l(\omega)$ and in the Eckart frame of reference $\kappa^E(\omega)$, which are shown in Figure 2a. After transforming the velocities to the Eckart frame, the intense peak below 50 cm^{-1} vanishes in $\kappa^E(\omega)$, which is present in $\kappa^l(\omega)$. These low-frequency contributions are associated with the rotational and translational DOFs of the molecule. Most other bands sharpen and more distinct features appear.

To check possible errors introduced by the transform to the Eckart frame the integral difference:

$$\Xi(\omega) = \int_0^\omega [\kappa^E(\omega') - \kappa^l(\omega')] d\omega' \quad (50)$$

for which $\kappa^E(\omega)$ was obtained either by an rmsd fit or by the internal coordinate transform, is displayed in Figure 2b. Both integral differences drop to about -6.2 within the first 50 cm^{-1} , accounting for the elimination of the six translational and rotational DOFs. For higher frequencies, however, they behave differently: $\Xi(\omega)$ obtained for the rmsd-fitted VDOS displays oscillations around the level reached after the initial drop. These oscillations correspond to the sharpening of bands already observed in Figure 2a. In contrast, $\Xi(\omega)$ obtained by the internal coordinate transform displays a general drift atop of the oscillations and reaches a limit of -5.0 at 3200 cm^{-1} . Thus, with respect to the rmsd fit the VDOS based on the internal

Table 1. Peak Positions ω_p and Corresponding Harmonic Frequencies ω_h of the Vibrational Bands (in cm^{-1}) as well as the Corresponding Internal Coordinate Coefficients (cf. eq 37) Obtained from the Analysis Based on Internal Coordinates^a

	ω_p	ω_h	internal coordinate contributions
1	152	145	0.99 C ₂ —C ₃ tors. + 0.42 H ₃ wag - 0.36 C ₄ H wag
2	161	171	0.96 Me tors.
3	274	277	0.85 C ₂ —C ₃ =C ₄ bend + 0.80 C ₅ —C ₂ —C ₃ def. + 0.43 C ₁ rock + 0.38 C ₂ —C ₃
4	399	398	0.86 C ₂ outp. - 0.47 C ₃ =C ₄ tors.
5	420	423	0.86 C ₁ rock - 0.57 C ₁ H rock - 0.41 C ₂ —C ₃
6	524	526	0.47 C ₃ =C ₄ + 0.46 C ₁ =C ₂ + 0.46 C ₂ —C ₃ =C ₄ bend - 0.46 C ₄ H rock
7	619	616	0.72 C ₁ =C ₂ tors. + 0.47 C ₃ =C ₄ tors.
8	751	746	0.65 C ₁ =C ₂ tors. - 0.46 C ₃ =C ₄ tors.
9	754	754	0.69 C ₂ —C ₅ + 0.58 C ₂ —C ₃ + 0.45 C ₁ =C ₂
10	878	889	0.92 C ₁ H wag
11	893	899	0.91 C ₄ H wag
12	925	911	0.57 C ₁ H rock + 0.45 C ₂ —C ₅ - 0.36 C ₄ H rock
13	981	984	0.81 H ₃ wag - 0.50 C ₃ =C ₄ tors.
14	985	988	0.62 Me asym. def.' + 0.59 Me rock' - 0.48 Me asym. def. + 0.42 Me rock
15	1029	1024	0.71 Me asym. def. - 0.65 Me rock + 0.51 Me asym. def.' + 0.47 Me rock'
16	1059	1056	0.61 C ₄ H rock - 0.38 H ₃ rock
17	1269	1272	0.39 C ₁ H rock
18	1291	1287	0.79 H ₃ rock - 0.39 C ₃ =C ₄
19	1360	1364	0.90 Me sym. def.
20	1388	1378	0.70 C ₁ H def. + 0.49 C ₄ H def.
21	1414	1414	0.65 C ₄ H def. - 0.47 C ₁ H def.
22	1428	1423	0.54 Me rock + 0.39 Me asym. def.
23	1452	1450	0.53 Me rock' - 0.41 Me asym. def.'
24	1580	1574	0.42 C ₁ =C ₂ - 0.37 C ₃ =C ₄
25	1605	1604	0.40 C ₃ =C ₄ + 0.32 C ₁ =C ₂
26	2951	2908	0.61 H ₅₃ + 0.60 H ₅₁ + 0.52 H ₅₂
27	2968	2956	0.75 H ₅₃ - 0.59 H ₅₁
28	3017	2999	0.82 H ₅₂ - 0.52 H ₅₁
29	3045	3016	0.81 H ₃ - 0.41 H ₁₂
30	3049	3023	0.61 H ₁₂ + 0.55 H ₁₁ + 0.51 H ₃
31	3063	3036	0.72 H ₄₂ + 0.66 H ₄₁
32	3127	3103	0.72 H ₁₁ - 0.64 H ₁₂
33	3135	3118	0.71 H ₄₁ - 0.64 H ₄₂

^a Hydrogen stretches H_{*i*} are marked by their atom labels *i*. The nomenclature, according to Pulay et al.,⁴² for the internal coordinates of the methylenes is: def., $2\alpha - \beta_1 - \beta_2$; rock, $\beta_1 - \beta_2$; and wag, H₁ + H₂ out of plane and of the methyl group (Me) as: Me sym. def., $\alpha_1 + \alpha_2 + \alpha_3 - \beta_1 - \beta_2 - \beta_3$; Me asym. def., $2\alpha_1 - \alpha_2 - \alpha_3$; Me asym. def.', $\alpha_2 - \alpha_3$; Me rock, $2\beta_1 - \beta_2 - \beta_3$; and Me rock', $\beta_2 - \beta_3$, where the H—C—H angles are denoted as α_i and the H—C—C angles as β_i .

coordinate transform gains about one DOF over the full frequency range. Correspondingly, the error introduced by this transform is about 3% if we relate this to the number of internal DOFs of the molecule, 33. This error is not uniformly distributed but has major contributions around 1000, 1300, and 3000 cm^{-1} .

Having inspected the transforms to the Eckart frame, we turn to the contributions of the generalized normal coordinates to the respective VDOS, here based on the rmsd fit. The resulting power spectra $\Theta_{kk}(\omega)$ after minimizing $\text{off}[\Theta]$ are well localized in frequency space, as can be seen in Figure 3. Most band shapes are characterized by single sharp peaks, except for a few, which have small auxiliary bands like modes 3 or 6 or which cover larger frequency intervals, like the overlapping bands of modes 22 and 23 or modes 28–31. Some large peaks appearing in $\kappa(\omega)$ decompose into nearly degenerate and overlapping bands, e.g., modes 8 and 9 at 750 cm^{-1} or 32 and 33 at 3130 cm^{-1} .

However, not all parts of the VDOS spectrum are fully disentangled, which is shown in Figure 4a for modes 26–28. Although these modes dominate $\kappa(\omega)$ in this frequency region, many other modes (given in parentheses), whose main peaks are far away in the spectrum, yield sizable contributions. In contrast, the mode spectra obtained after the internal coordinate transform shown in Figure 4b do not exhibit such admixtures of low-lying modes. Here, however, modes 27 and 28 are still entangled, i.e., they cover the same spectral range and are bimodal.

For an assignment of these modes to atomic motion Table 1 provides the contributions of the internal coordinates *s* to the generalized normal coordinates *q* according to eq 37 based on the data obtained after the internal coordinate transform. Here, modes 26–28 are identified as the symmetric and the two asymmetric stretching modes of the methyl group, respectively. Seemingly the two asymmetric modes cannot be disentangled, whereas the symmetric stretch is clearly identified. The admixtures observed

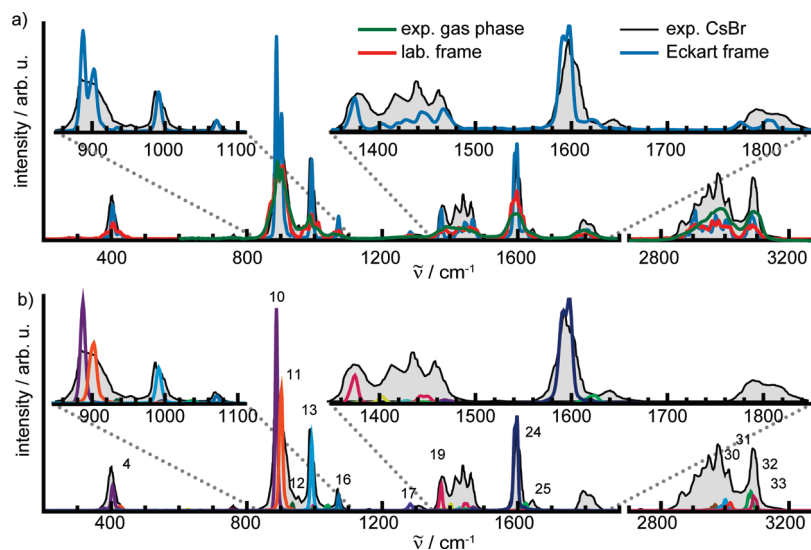


Figure 5. Comparison of the computed IR absorptions to experiment. Experimental data of isoprene in a Cs Br matrix (shaded area) was taken from ref 56, baseline corrected, and intensity scaled to match the dominant methylene wagging band at 900 cm^{-1} . Gas-phase data was obtained from the National Institute of Standards and Technology (NIST) database (green curve).⁵⁵ The frequencies of the computed spectra have been scaled by 1.015 in the lower part of the spectrum and by 0.985 in the C–H stretch region. Panel (a) compares the overall IR absorption cross-section in the laboratory frame (red curve) and in the Eckart frame (blue curve); insets zoom in on interesting spectral regions. Panel (b) compares the individual mode absorptions $\alpha_{kk}(\omega)$ (colored curves) to peaks of the experimental spectrum in CsBr.

for the modes based on the rmsd fitted trajectories stem from other modes of the methyl group: its rotation (mode 2), rocking and deformation (modes 14, 15, 19, 22, and 23). Although the mode classification seems to work better based on the internal coordinate transform, it is still not fully satisfying because the asymmetric modes are not disentangled. The reason is that rotations of the methyl group, which leave the conformation invariant but change the numbering of the hydrogens, are not properly accounted for in both methods. Only modes invariant under such rotations, i.e., the symmetric methyl stretch (mode 26) and the symmetric deformation (mode 19) are correctly identified. A drawback of the internal coordinate transform is that it notably modifies the VDOS in the Eckart frame, in particular near the vibrational contributions of the methyl group at 1000, 1400, and 3000 cm^{-1} , see Figure 2. An extension of our algorithm that can handle such permutations and resolves these problems will be presented in a subsequent publication.

The other modes listed in Table 1 besides those located at the methyl group are nicely characterized by the internal coordinate contributions. For example the modes 24 and 25 represent the asymmetric and symmetric C=C stretches. Also for many other modes in phase and out of phase combination of internal coordinates are found, e.g., modes 7 and 8, 20 and 21, or 31 and 33.

Table 1 furthermore compares the peak positions of the modes ω_p to the frequencies calculated by the harmonic approximation ω_h . The differences are only a few wavenumbers for the first 25 modes and do not exceed 10 cm^{-1} . Thus, anharmonic effects are small for this molecule and induce only subtle shifts. In contrast, for the hydrogen stretches modes 26 to 33 a systematic blue-shift on the order of 25 cm^{-1} is observed. This is somehow counterintuitive if we think of the C–H bond as a Morse-type binding potential, for which we would expect a red-shift due to anharmonicities. However, this blue-shift was no longer observed when we had set the angular momentum to zero within the initial conditions and rerun the MD simulations (data not shown). Thus, these blue-shifts are caused by couplings to rotation.

For comparison to experiment, the interactions of the molecule and its vibrational modes with radiation are required. Figure 5a shows the total IR absorption spectrum α eq 38 computed in the laboratory frame (red curve) and in the Eckart frame of reference (blue curve), for which the dipole moment had been transformed along with the rmsd fit of the coordinates. The frequencies have been slightly scaled by 1.015 in the region below 1900 cm^{-1} and by 0.985 in the region above 2700 cm^{-1} to match experiment. The experimental data are a gas-phase spectrum of isoprene⁵⁵ and a solid-state spectrum obtained in a CsBr matrix.⁵⁶

Comparing the spectra, we find a strikingly close agreement of the computed IR absorption in the laboratory frame (red curve) and the experimental gas-phase data (green curve): band positions, shapes, and relative intensities match almost perfectly. Thus, including molecular rotation within the classical dynamics seems to be sufficient to mimic the rotational broadening of the vibrational bands. Differences are only observed for the C=C stretching modes near 1600 cm^{-1} , whose computed band is slightly sharper and the hydrogen stretch region, where the overall intensity is smaller than in experiment. Note in particular that in the spectral regions of the methyl deformation and stretching modes near 1400 and 2900 cm^{-1} , the computed and the gas-phase spectrum nicely match.

After the dipole moments are transformed to the Eckart frame the computed IR absorption bands sharpen (blue curve), similar to the observations for the VDOS. Here, a comparison to an experimental spectrum that is not rotationally broadened is more appropriate, the spectrum in Cs Br (shaded area). For most peaks in the computed spectrum, we find a close correspondence in the experimental one, see also insets. However, the solid state clearly modifies the experimental spectrum of isoprene, in particular the intensities, e.g., of the hydrogen stretches or in the region between 1400 and 1500 cm^{-1} .

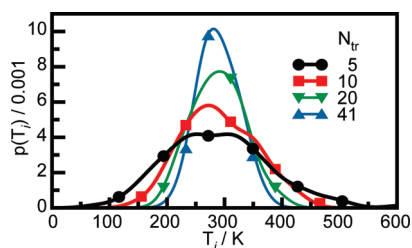


Figure 6. Mode temperature distributions depending on the number of trajectories n .

A mode assignment for these experimental peaks is given in Figure 5b by the IR absorption bands of the individual modes. Most of the peaks in the experimental spectrum are soundly assigned, e.g., the prominent peak of the methylene wagging modes at 900 cm^{-1} , see inset. However, the small band near 1800 cm^{-1} is missing in the reconstructed mode spectrum but present in the computed IR absorption cross-section in Figure 5a. Closer inspection reveals that this band is associated to the overtones of the methylene wagging modes. Because such overtones are not considered in the expansion of the dipole moment eq 41, they are not present in the reconstructed spectrum.

For the methyl modes we find similar deficiencies in the assignment of their IR intensities, as we have observed in the disentangling of these modes. Only the symmetric methyl deformation (mode 19) yields the proper IR absorption, whereas for the other methyl modes, it is grossly underestimated. Here, the sum of mode IR absorptions computed for the Eckart frame (blue line in Figure 5a) does not reach the total IR absorption in the laboratory frame. Again these problems are caused by not properly accounting for the rotation of the methyl group and can be only resolved by considering multiple reference structures.

A major advance of the algorithm presented here is that it provides the mode temperatures along with their generalized normal coordinates. Furthermore, these mode temperatures allow to assess the convergence of the computed IR spectrum. Figure 6 shows the distribution $p(T)$ of these mode temperatures for the full set of 41 trajectories (blue curve) obtained by a kernel density estimate. Here, we have represented each of the 33 mode temperatures by Gaussians of width $\sigma = 24\text{ K}$ and normalized their sum. This resulting distribution has a standard deviation of about 39 K , which is about 13% of the mean temperature 300 K . This seems quite sizable if we recall that this distribution is based on more than 1 ns of ab initio trajectory.

We examined the convergence of this distribution by selecting blocks of $N_{\text{tr}} = 5, 10,$ and 20 trajectories from the whole set and recalculated the mode temperatures and their distribution for each block size. These distributions are compared to the full set in Figure 6. The corresponding standard deviations amount to 91 K for $N_{\text{tr}} = 5$, 71 K for $N_{\text{tr}} = 10$, and 51 K for $N_{\text{tr}} = 20$. Their ratios approximately follow the rule that doubling the number of trajectories scales the standard deviation of the resulting mode temperature distribution by about $1/\sqrt{2}$, as is expected for the standard deviation. Note in particular, that if the number of trajectories is small one can have quite a few modes exhibiting a temperature twice the mean temperature or only half of it. According to eqs 15 and 48, however, this implies that the IR absorption of such modes is grossly over or underestimated.

Finally, we check how well the suggested temperature correction can remedy such deficiencies due to insufficient sampling.

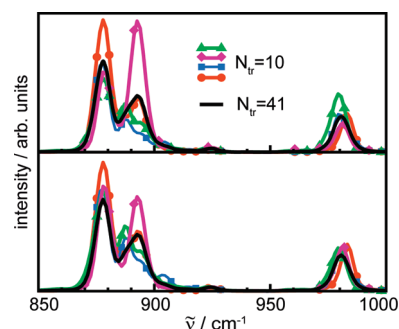


Figure 7. Uncorrected (upper panel) and temperature-corrected (lower panel) IR absorption for sets of 10 trajectories (colored curves) and all 41 trajectories (black curve).

Figure 7 compares the uncorrected and corrected spectrum for sets of 10 trajectories and for all 41 trajectories for the prominent double band of the methylene wagging modes, 11 and 12, and the HOOP mode, 13, at 981 cm^{-1} . For the latter the temperature correction brings the spectra of the small and the unconverged sets close to the one of the full set. In contrast, for the two overlapping methylene wagging bands, the discrepancies of intensities remain larger. However, the temperature corrections are qualitatively correct and shift the intensities toward the relative heights of the full set. Seemingly, to resolve such details extended sampling is inevitable. Nevertheless, the temperature correction can help to identify unconverged parts of the spectrum, in particular, if they are connected to statistical outliers of the mode temperatures.

5. CONCLUSIONS AND OUTLOOK

We have developed a consistent and automated procedure to deduce generalized normal coordinates from finite MD trajectories. It is based on the definition of the VDOS in a tensorial form and does not rely on any equipartition assumption. The generalized normal coordinates are defined as orthonormal transforms of mass-weighted coordinates, which exhibit minimal correlation as measured by the off-diagonal norm of the tensorial VDOS. A minimization scheme based on the Jacobi diagonalization algorithm robustly converges and renders the generalized normal coordinates, their band positions and shapes, their mode temperatures, and their remaining cross-correlations, which result from the finiteness of the sampling and anharmonic intermode couplings. Within this framework we have embedded the assignment of the vibrational modes to internal coordinate motion. Furthermore, mode-specific IR intensities are derived which also take cross-correlations into account.

As a non-trivial application of the algorithm, we presented the analysis of the IR spectrum of isoprene in the gas phase, i.e., an isolated but rotating molecule, which has been sampled by 41 ab initio MD trajectories. The algorithm renders rich and detailed information on the vibrational modes, especially their compositions in terms of internal coordinates and their respective absorption strength. Computed band positions and widths, which are rotationally broadened, nicely agree with an experimental gas-phase spectrum and assign the experimental peaks of a solid-state spectrum.

The distribution of mode temperatures still shows a sizable width if the full set of trajectories is used for the analysis and dramatically broadens for smaller sets. The corresponding temperature correction for the absorption qualitatively rectifies the

relative intensities for overlapping modes and even quantitatively for isolated peaks.

We have extended the algorithm presented here to manage multiple conformations of molecules and to resolve permutational symmetries, which will be described in a subsequent publication. This extension fully resolves the problems observed for the methyl group of isoprene in this article by employing three instead of one reference structure. Furthermore, it allows to treat large amplitude motion more complex than a methyl group rotation,^{9,11,34} for which also multiple conformations of a molecule have to be considered.

APPENDIX A: OPTIMAL JACOBI ANGLE

To determine the optimal rotation angle ϑ as defined by eq 29, we first note that since the overall Frobenius norm of Θ is invariant with respect to orthonormal transforms,⁴¹ it is sufficient to consider just the term $N_{uv} = \int [\Theta_{uv}(\omega)]^2 d\omega$. Thus, according to eqs 24 and 29 we have to solve

$$\left. \frac{\partial N'_{uv}}{\partial \vartheta'} \right|_{\vartheta' = \vartheta} = 0 \quad (51)$$

for the transformed term:

$$N'_{uv} = \int [(c^2 - s^2)\Theta_{uv}(\omega) + sc(\Theta_{uu}(\omega) - \Theta_{vv}(\omega))]^2 d\omega \quad (52)$$

again with $c = \cos \vartheta$ and $s = \sin \vartheta$. With $t = \tan(\vartheta)$, the identities $c = 1/(t^2 + 1)^{1/2}$ and $s = tc$, and the integrals:

$$\begin{aligned} \chi_1 &= \int [\Theta_{uv}(\omega)]^2 d\omega = N_{uv}, \\ \chi_2 &= \int \Theta_{uv}(\omega)[\Theta_{uu}(\omega) - \Theta_{vv}(\omega)] d\omega, \\ \chi_3 &= \int [\Theta_{uu}(\omega) - \Theta_{vv}(\omega)]^2 d\omega \end{aligned}$$

Equation 52 transforms to

$$N'_{uv} = \frac{\chi_1(t^2 - 1)^2 - 2\chi_2 t(t^2 - 1) + \chi_3 t^2}{(1 + t^2)^2} \quad (53)$$

and, with $\eta = (4\chi_1 - \chi_3)/\chi_2$ its derivative is given by

$$\frac{\partial N'_{uv}}{\partial t} = 2\chi_2 \frac{(t^4 + \eta t^3 - 6t^2 - \eta t + 1)}{(1 + t^2)^3} \quad (54)$$

which has the four roots

$$\begin{aligned} t_{kl} &= \frac{1}{4}(-\eta + (-1)^k \sqrt{16 + \eta^2} + (-1)^l \sqrt{2} \\ &\quad \times \sqrt{16 + \eta^2 - (-1)^k \eta \sqrt{16 + \eta^2}}) \quad (55) \end{aligned}$$

$k, l \in \{1, 2\}$. For our transform we choose the root t_{kl} with $|t_{kl}| \leq 1$ and

$$\begin{aligned} \left. \frac{\partial^2 N'_{uv}}{\partial t^2} \right|_{t=t_{kl}} &= -\frac{2\chi_2}{(t_{kl}^2 + 1)^4} [2t_{kl}(t_{kl}^4 - 14t_{kl}^2 + 9) \\ &\quad + \eta(3t_{kl}^4 - 8t_{kl}^2 + 1)] > 0 \end{aligned}$$

to obtain a minimum of N'_{uv} .

AUTHOR INFORMATION

Corresponding Author

*E-mail: gerald.mathias@physik.uni-muenchen.de.

Former Addresses

⁵Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany.

ACKNOWLEDGMENT

We thank Sergei Ivanov and Harald Forbert for helpful discussions and Dominik Marx for encouragement and support. Theodoros Zelleke is acknowledged for implementing the internal coordinate transform. Funding was provided by the Deutsche Forschungsgemeinschaft through grant MA 1547/3 to D.M. within FOR 436.

REFERENCES

- (1) Nonella, M.; Mathias, G.; Eichinger, M.; Tavan, P. *J. Phys. Chem. B* **2003**, *107*, 316–322.
- (2) Vogel, R.; Siebert, F.; Mathias, G.; Tavan, P.; Fan, G.; Sheves, M. *Biochemistry* **2003**, *42*, 9863–9874.
- (3) Rousseau, R.; Kleinschmidt, V.; Schmitt, U. W.; Marx, D. *Angew. Chem., Int. Ed.* **2004**, *43*, 4804–4807.
- (4) Asvany, O.; Kumar, P. P.; Redlich, B.; Hegemann, I.; Schlemmer, S.; Marx, D. *Science* **2005**, *309*, 1219–1222.
- (5) Mathias, G.; Marx, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6980–6985.
- (6) Masia, M.; Forbert, H.; Marx, D. *J. Phys. Chem. A* **2007**, *111*, 12181–12191.
- (7) Baer, M.; Mathias, G.; Kuo, I.-F. W.; Tobias, D. J.; Mundy, C. J.; Marx, D. *ChemPhysChem* **2008**, *9*, 2703–2707.
- (8) Cimas, A.; Maitre, P.; Ohanessian, G.; Gaigeot, M.-P. *J. Chem. Theory Comput.* **2009**, *5*, 2388–2400.
- (9) Ivanov, S. D.; Asvany, O.; Witt, A.; Hugo, E.; Mathias, G.; Redlich, B.; Marx, D.; Schlemmer, S. *Nat. Chem.* **2010**, *2*, 298–302.
- (10) Heyden, M.; Sun, J.; Funkner, S.; Mathias, G.; Forbert, H.; Havenith, M.; Marx, D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 12068–12073.
- (11) Baer, M.; Marx, D.; Mathias, G. *Angew. Chem., Int. Ed.* **2010**, *49*, 7346–7349.
- (12) Neugebauer, J.; Hess, B. A. *J. Chem. Phys.* **2003**, *118*, 7215–7225.
- (13) Nonella, M.; Mathias, G.; Tavan, P. *J. Phys. Chem. A* **2003**, *107*, 8638–8647.
- (14) Ramírez, R.; López-Ciudad, T.; Kumar, P. P.; Marx, D. *J. Chem. Phys.* **2004**, *121*, 3973–3981.
- (15) Gaigeot, M.-P.; Vuilleumier, R.; Sprik, M.; Borgis, D. *J. Chem. Theory Comput.* **2005**, *1*, 772–789.
- (16) Gaigeot, M.-P.; Martinez, M.; Vuilleumier, R. *Mol. Phys.* **2007**, *105*, 2857–2878.
- (17) Pauwels, E.; Verstraelen, T.; De Cooman, H.; Van Speybroeck, V.; Waroquier, M. *J. Phys. Chem. B* **2008**, *112*, 7618–7630.
- (18) Kumar, P. P.; Marx, D. *Phys. Chem. Chem. Phys.* **2006**, *8*, 573–586.
- (19) Stratt, R. M. *Acc. Chem. Res.* **1995**, *28*, 201–207.
- (20) Keyes, T. J. *J. Phys. Chem. A* **1997**, *101*, 2921–2930.
- (21) Kalbfleisch, T.; Keyes, T. J. *J. Chem. Phys.* **1998**, *108*, 7375–7383.
- (22) Schmitz, M.; Tavan, P. *J. Chem. Phys.* **2004**, *121*, 12247–12258.
- (23) Babitzki, G.; Mathias, G.; Tavan, P. *J. Phys. Chem. B* **2009**, *113*, 10496–10508.
- (24) Miani, A.; Helfand, M. S.; Raugei, S. *J. Chem. Theory Comput.* **2009**, *5*, 2158–2172.
- (25) Rieff, B.; Mathias, G.; Bauer, S.; Tavan, P. *Photochem. Photobiol.* **2011**, *87*, 511–523.

- (26) Wheeler, R. A.; Dong, H.; Boesch, S. E. *ChemPhysChem* **2003**, *3*, 382–384.
- (27) Wheeler, R. A.; Dong, H. *ChemPhysChem* **2003**, *4*, 1227–1230.
- (28) Schmitz, M.; Tavan, P. *J. Chem. Phys.* **2004**, *121*, 12233–12246.
- (29) Carbonniere, P.; Dargelos, A.; Ciofini, I.; Adamo, C.; Pouchan, C. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4375–4384.
- (30) Calvo, F.; Parneix, P.; Van-Oanh, N. T. *J. Chem. Phys.* **2010**, *133*, 074303.
- (31) Martinez, M.; Gaigeot, M.-P.; Borgis, D.; Vuilleumier, R. *J. Chem. Phys.* **2006**, *125*, 144106/14.
- (32) Eckart, C. *Phys. Rev.* **1935**, *47*, 552–558.
- (33) Babitzki, G.; Mathias, G.; Tavan, P. *J. Phys. Chem. B* **2009**, *113*, 10496–10508.
- (34) Baer, M.; Marx, D.; Mathias, G. *ChemPhysChem* **2011**, DOI: 10.1002/cphc.201000955.
- (35) Roos, B. W. *Analytic Functions and Distributions in Physics and Engineering*; John Wiley & Sons: New York, 1969.
- (36) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C*; Cambridge University Press: Cambridge, U.K., 1988, pp 553–558.
- (37) Reed, M.; Simon, B. *Methods of Modern Mathematical Physics: II Fourier Analysis, Self-Adjointness*; Academic Press: New York, 1975, p 10.
- (38) Wilson, E.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw Hill: New York, 1955, pp 19–27, 54–63.
- (39) Bunse-Gerstner, A.; Byers, R.; Mehrmann, V. *SIAM J. Matrix Anal. Appl.* **1993**, *14*, 927–949.
- (40) Cardoso, J.-F.; Souloumiac, A. *SIAM J. Mat. Anal. Appl.* **1996**, *17*, 161–164.
- (41) Golub, G.; van Loan, C. *Matrix Computations*, 3rd ed.; Johns Hopkins University Press: Baltimore, MD, 1996, pp 426–438.
- (42) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. *J. Am. Chem. Soc.* **1979**, *101*, 2550–2560.
- (43) Fogarasi, G.; Zhou, X.; Taylor, P. W.; Pulay, P. *J. Am. Chem. Soc.* **1992**, *114*, 8191–8201.
- (44) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press: Cambridge, U.K., 2009, pp 11–75.
- (45) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (46) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (47) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.
- (48) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B* **1998**, *58*, 3641–3662.
- (49) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.
- (50) Lippert, G.; Hutter, J.; Parrinello, M. *Mol. Phys.* **1997**, *92*, 477–487.
- (51) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
- (52) Martyna, G.; Tuckerman, M. *J. Chem. Phys.* **1999**, *110*, 2810–2821.
- (53) Martyna, G. J.; Klein, M. L.; Tuckerman, M. E. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
- (54) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
- (55) Stein, S. E. *NIST/EPA Gas-Phase Infrared Database*; National Institute of Standards and Technology: Gaithersburg, MD, 2004.
- (56) Schrader, B. *Raman/Infrared Atlas of Organic Compounds*; VCH: Weinheim, Germany, 1989, p C2–01.

Reactive Molecular Dynamics Simulation of Fullerene Combustion Synthesis: ReaxFF vs DFTB Potentials

Hu-Jun Qian,[†] Adri C. T. van Duin,^{*,‡} Keiji Morokuma,[§] and Stephan Irle^{*,†}

[†]Institute for Advanced Research and Department of Chemistry, Nagoya University, Nagoya 464-8602, Japan

[‡]Department of Mechanical and Nuclear Engineering, Pennsylvania State University, University Park, Pennsylvania 16802, United States

[§]Fukui Institute for Fundamental Chemistry, Kyoto University, Kyoto, 606-8103, Japan

 Supporting Information

ABSTRACT: The dynamic fullerene self-assembly process during benzene combustion was studied using classical Reactive Force Field (ReaxFF) nonequilibrium molecular dynamics (MD) simulations. In order to drive the combustion process, the hydrogen to carbon (H/C) ratio was gradually reduced during the course of the MD simulations. Target temperatures of 2500 and 3000 K were maintained by using a Berendsen thermostat. Simulation conditions and hydrogen removal strategies were chosen to match closely a previous quantum chemical MD (QM/MD) study based on the density-functional tight-binding (DFTB) potential (Saha et al. *ACS Nano* **2009**, *3*, 2241) to allow a comparison between the two different potentials. Twenty trajectories were computed at each target temperature, and hydrocarbon cluster size, C_xH_y composition, average carbon cluster curvature, carbon hybridization type, and ring count statistics were recorded as a function of time. Similarly as in the QM/MD simulations, only giant fullerene cages in the range from 155 to 212 carbon atoms self-assembled, and no C_{60} cages were observed. The most notable difference concerned the time required for completing cage self-assembly: Depending on temperature, it takes between 50 and 150 ps in DFTB/MD simulations but never less than 100 ps and frequently several 100s ps in ReaxFF/MD simulations. In the present system, the computational cost of ReaxFF/MD is about 1 order of magnitude lower than that of the corresponding DFTB/MD. Overall, the ReaxFF/MD simulations method paints a qualitatively similar picture of fullerene formation in benzene combustion when compared to direct MD simulations based on the DFTB potential.

I. INTRODUCTION

Since the discovery of buckminsterfullerene C_{60} (BF),¹ fullerenes have been the focus of research due their unique structures, chemistry, and potential applications in nanotechnology. Historically, fullerenes were synthesized on the gram scale by evaporating carbon atoms from graphite at high temperatures on the order of several thousand Kelvin.^{1,2} Nowadays, industry-scale production of fullerenes is achieved by continuous combustion synthesis in low-pressure fuel-rich flames of certain hydrocarbons.^{3–7} This technique is sensitive to operating conditions, such as fuel type,⁶ fuel/oxygen ratio, temperature, pressure in the combustion chamber,^{4,7} and even chamber design.⁸ Although considerable advances have been made in optimizing the synthesis conditions on a phenomenological basis,⁹ the elementary reaction mechanisms involved in the self-assembly of fullerene cages are still subject to investigation.¹⁰

Experimentally, combustion of hydrocarbon fuels in oxygen-lean flames was found to produce substantial amounts of polyaromatic hydrocarbons (PAHs), which are presumed to be the precursors of fullerenes.^{7,9–14} However, the molecular structures of the intermediate PAH species with more than ~ 10 carbon atoms remain largely unknown. Experimentally, temperature/pressure profiles, fuel/oxygen ratio as a function of distance from the burner, and *in situ* mass spectra have been recorded. A combination of these data with thermodynamic considerations and kinetic modeling only indirectly sheds light on the PAH H/C ratio as a function of time evolution. On the basis of the recorded

H/C ratio of PAHs in acetylene flames, Homann proposed a picture of “ordered” growth of PAHs along a pathway involving only maximally condensed and fully hydrogenated graphene-like platelets.¹⁰ However, the assumption that such species dominate the aggregation process is only based on thermodynamic stability arguments and neglects entropic effects that are important due to the high environmental temperature. To date, there is no proof that thermodynamically maximally stable species as proposed by Homann and others are true intermediate species in the dynamic fullerene cage self-assembly.

Atomic-scale modeling of complex reaction systems *in silico* has become a useful tool of study, capable of reproducing fullerene self-assembly from benzene, atomic carbon, and C_2 molecules in computer simulations.^{15–22} Over the past six years, Irle, Morokuma, and co-workers have discovered and elaborated on a “shrinking hot giant” (SHG) road of fullerene formation, based on direct quantum chemical molecular dynamics simulations (QM/MD) using a density-functional tight-binding (DFTB) potential.¹⁷ According to the SHG road, the fullerene formation process follows two stages: (1) self-assembly of giant fullerenes (GFs) via a “size-up” process, followed by (2) a shrinking process (termed “size-down”) by irreversibly evaporating C_2 units from vibrationally excited, highly unstable and defective cages. The shrinking process has experimentally been recorded in an *in situ*

Received: March 22, 2011

Published: May 31, 2011

HRTEM movie sequence.²³ Regarding the preceding “size-up” stage, no direct confirmation from experimental results is available. DFTB/MD simulations of pure carbon vapor systems predict three consecutive steps:¹⁶ (1) polyynes chain formation (σ -bond formation between small linear carbon chain fragments) and nucleation of a pentagon/hexagon network into “octopus”-like bowl structures with polyynes “antennas” attached to their openings, (2) ring condensation growth at the bowl openings, and (3) GF cage closure by saturation of dangling bonds at edges, leaving antennas attached to sp^3 cage defects. DFTB/MD simulations of the benzene combustion process by hydrogen removal¹⁵ predicted the following steps: (1) radical creation and ring-opening/fragmentation, (2) growth of polyacetylene-like chain structures, akin to the polyynes chain formation process in pure carbon simulations, followed by a (3) ring condensation process and (4) cage closure. Here, ring condensation was found to ensue after a significant fraction of hydrogen was stripped from the chains, since only then can carbon avoid disruption of favorable conjugation through sp^2 carbons. As a consequence, ring condensation in combustion occurred later than in carbon-only systems, typically after the cluster size was already determined in the hydrocarbon oligomerization process, leaving not a lot of free carbon species to regrow chains at the cluster boundaries. Therefore, GF antennas were formed less frequently in the case of combustion in comparison to carbon-only simulations.

The DFTB method has allowed the direct MD study of the self-assembly process of fullerenes with approximate density functional theory (DFT) accuracy for up to several hundred picoseconds. However, this time scale is still much shorter compared to experimental fullerene synthesis time scales, which occurs on the order of microseconds. Moreover, fullerene synthesis by benzene combustion requires the presence of O_2 molecules for hydrogen abstraction, which occurs only on the order of several hundred picoseconds.^{24,25} Saha et al. had therefore resorted to randomly removing H atoms during the MD, simulating in this way the decreasing H/C ratio during combustion.¹⁵ On the other hand, the semiclassical reactive force field (ReaxFF) method by van Duin and co-workers²⁶ is apparently computationally much more economical and enables the simulation of combustion processes on a nanosecond time scale.^{24,25} The computational efficiency of ReaxFF is paid for by the fact that the force field is local beyond a four-body interaction term and therefore does not explicitly describe resonance structures. Therefore, formally, the famous Hückel aromaticity and Clar²⁷ rules, which are important for the relative stability of PAH isomers, cannot be accurately represented in the ReaxFF potential. We note that in practice, ReaxFF energetics of π -conjugated species are generally in good agreement with quantum chemical DFT benchmark data by virtue of their empirical parametrization.²⁸ However, a systematic comparison of fullerene growth as described by a quantum chemical and a classical potential has never been directly performed. The main goal of this work is therefore to compare ReaxFF/MD with previous DFTB/MD simulations regarding the respective mechanism of fullerene formation during combustion simulations and the computational cost associated with these methods.

II. COMPUTATIONAL METHODOLOGY

ReaxFF is a general bond-order-dependent force field method fitted to potentials derived from first principles DFT calculations,

allowing chemical reactions to take place *via* the cleavage or formation of covalent chemical bonds during a MD run. Atomic bond orders are computed on the fly from the atomic connectivity matrix, with updates at every MD time step. The potential is adjusted as a function of these bond orders, similar as in Tersoff–Brenner type potentials.^{29–31} Nonbonded interactions (van der Waals and Coulomb interactions) are calculated between every pair of atoms, irrespective of connectivity, while a shielding term is introduced for short distances. A full description of the derivation and parametrization in the ReaxFF method is given in refs 24 and 26.

All simulations were performed under constant temperature and constant volume (NVT) conditions, where we maintained target temperatures of 2500 and 3000 K by employing a Berendsen thermostat with a weak coupling constant of $\tau = 100$ ps. This thermostat is closest to, although not identical with, the velocity-scaling thermostat that was employed in our previous DFTB simulations. A relatively short time integration interval of $\Delta t = 0.1$ fs was chosen in the velocity Verlet algorithm to ensure the smoothness of the ReaxFF potential at the employed high temperatures, because in reactive force fields a switching function needs to be frequently updated to handle changes in atomic coordination numbers and corresponding changes in the potential energy. Since the DFTB method does not require a bond order switching function, MD simulations performed on this potential can tolerate a larger time step. For comparison, in the previous DFTB/MD simulations, a time integration interval of $\Delta t = 0.48$ fs had been employed. We adopted the same hydrogen removal strategy as in ref 15, and the initial Cartesian coordinates of benzene molecules are identical as well. In this geometry called g1, 36 benzene molecules are stacked in four layers of nine molecules arranged in a 3×3 quadratic plane within a cubic simulation box with an initial edge length of 21 Å. The interlayer distance was set to 3.4 Å, and the closest intermolecular H contacts are 2.2 Å. Periodic boundary conditions (PBC) were applied on the basis of this cubic unit cell. Using the g1 geometry as a reference, Saha et al. created the three other geometries g2, g3, and g4 by varying interlayer distances. However, the authors mention that the exact initial coordinates of individual benzene molecules have little influence on the trajectories due to rapid randomization of positions at the high temperatures, and thus we only employed the geometry labeled “g1” in this study.

For each temperature, 20 trajectories were simulated starting from the same geometry “g1” with different random H removal sequences. Trajectories are labeled either “2500K_n” or “3000K_n” to distinguish temperature and trajectory number “n”, ranging from 1 to 20. As in the previous DFTB/MD simulations, the combustion process was simulated by a programmed H atom removal process with a random removal rate of 70 hydrogen atoms every 5 ps. This means that for each trajectory, the first H removal process was performed after 5 ps of equilibration, followed by the second and third H removals at 10 and 15 ps. On the occasion of each H removal process, free H atoms/ H_2 molecules (isolated H atoms without any covalent bonds to C atoms) were also removed from the system. At a simulation time of 20 ps, any H remaining in the systems was also removed, and the system was annealed for 500 additional picoseconds. In the previous DFTB/MD simulation, it was found necessary to increase the simulation box edge size from 21 Å to 30 Å after the third H removal at 15 ps because the main cluster size was found to approach the size of the PBC box in

Table 1. ReaxFF/MD Trajectories at 2500 K^a

trajectories	H's removed			t_f/t_{end} (ps)	#C _{cage}	#C _{cluster}	$\langle\langle C^2 \rangle\rangle^{1/2}$ (1/Å)
	@5 ps	@10 ps	@15 ps				
2500K-1	70	70	70	520	212	214	0.232
2500K-2	70	70	70	520	186	216	0.235
2500K-3	70	70	70	674	188	209	0.209
2500K-4	70	70	70	370	158	179	0.260
2500K-5	70	70	70	700.5	186	214	(0.262)
2500K-6	70	70	70	370.5	180	214	(0.237)
2500K-7	70	70	70	630	181	190	0.226
2500K-8	70	70	70	711	172 (open)	203	(0.278)
2500K-9	70	70	70	520	182 (defective)	214	(0.250)
2500K-10	70	70	70	520	183 (defective)	214	(0.237)
2500K-11	70	70	70	268	155	170	0.251
2500K-12	70	70	70	470	176	203	0.240
2500K-13	70	70	70	520	194	216	0.238
2500K-14	70	70	70	520	186 (defective)	214	0.236
2500K-15	70	70	70	520	159 (defective)	202	0.260
2500K-16	70	70	70		unsuccessful	216	
2500K-17	70	70	70		unsuccessful	216	
2500K-18	70	70	70	503.5	191	216	0.239
2500K-19	70	70	70	520	169 (defective)	212	(0.243)
2500K-20	70	70	70	489	184	211	0.263
average ^a				503	183	204	0.241

^a The last row records the average values of t_f , #C_{cage}, #C_{cluster}, and $\langle\langle C^2 \rangle\rangle^{1/2}$ over all successfully formed cages. t_f corresponds to time of closed cage formation in the successful cases, and t_{end} is the point when the simulation is terminated in the unsuccessful cases. #C_{cage} is the number of carbon atoms in the cage only, and #C_{cluster} is the number of total carbon atoms in the largest cluster at the time of cage formation. $\langle\langle C^2 \rangle\rangle^{1/2}$ is the root mean square (RMS) curvature of the cluster, where curvature C is defined as the inverse radii of spheres best fitted to an sp²-carbon atom plus its three bond neighbors. Brackets $\langle\dots\rangle$ denote an average over sp²-carbons. “Defective” denotes a defective cage is formed, and “open” denotes an open cage.

most of the simulated trajectories, preventing fullerene cage formation due to the formation of an extended graphitic sheet across periodic boundaries. Therefore, the carbon density of the model system was initially 0.46 g/cm³ during the first 15 ps of the trajectories and 0.16 g/cm³ afterward. However, for the ReaxFF/MD simulations presented here, we found that we had to employ an even larger box with 50 Å edge length after 15 ps to prevent graphitic sheet formation due to structure formation across periodic images. The previous DFTB/MD trajectories were computed up to 220 ps simulation time. The present ReaxFF/MD simulations were computationally significantly less expensive, allowing us to follow trajectories up to 700 ps.

III. RESULTS AND DISCUSSIONS

III.A. Self-Assembled Giant Fullerenes. Tables 1 and 2 present the size (# of carbon atoms) of the final formed largest clusters and cages at both temperatures of 2500 and 3000 K, respectively. We note the relatively frequent occurrence of partially nested or spiroid-like structures, which are labeled “defective”. In all cases, we recorded the number of hydrogen atoms randomly removed after 5.0, 10.0, and 15.0 ps. Unlike in previous DFTB/MD simulations, free hydrogen atoms were frequently formed and removed, yet only a single free H₂ molecule was found in trajectory 3000K_9 at 10 ps (Table 2). At 2500 K, 12 out of 20 trajectories formed closed cage structures (GFs) during the simulation time, equivalent to a fullerene yield of 60%. Six of them were found to yield defective or partially open

cages (see the corresponding Cartesian coordinates in the Supporting Information, Tables S1 and S2). At 3000 K, the fullerene yield was even higher with 17/20 = 85% successfully formed GFs. In all successful trajectories, we recorded the time when the closed fullerene cages are formed t_f , the number of carbon atoms both in the cage structure #C_{cage} and in the whole cluster #C_{cluster} at time t_f and the root-mean-square (RMS) curvatures calculated from the carbon atoms belonging to the largest cluster. The RMS curvature is an average over all of the sp² carbon atoms in the cluster. Similar to that in previous DFTB/MD simulations, we found that a higher temperature favors the formation of smaller cages: #C_{cage} at 2500 K is distributed between 155 and 212 (DFTB/MD: 174–212) with an average of 183 (DFTB/MD: 191), while the cage size at 3000 K is distributed between 157 and 199 (DFTB/MD: 74–201) with an average of 178 (DFTB/MD: 161). Consequently, RMS curvatures are very similar for ReaxFF/MD and DFTB/MD simulations. The number of carbons contained in exohedrally attached polyene chains (so-called “antennas”) is given as the difference between #C_{cluster} and #C_{cage} in Tables 1 and 2. The average number of antenna carbons is 21 at 2500 K (DFTB/MD: 7) and 16 at 3000 K (DFTB/MD: 25). The successfully formed cages of this study are displayed in Figures 1 and 2 for 2500 and 3000 K, respectively. They occasionally contain endohedrally encapsulated small C_n chains or rings, marked by yellow color, which were not found in our previous DFTB/MD simulations. However, previous tight-binding-based MD³² and reactive force field MD²¹ simulations also predicted similar endohedrally encapsulated carbon clusters.

Table 2. ReaxFF/MD Trajectories at 3000 K^a

trajectory name	H's removed			t_f/t_{end} (ps)	#C _{cage}	#C _{cluster}	$(\langle C^2 \rangle)^{1/2}$ (1/Å)
	@5 ps	@10 ps	@15 ps				
3000K_1	70	70	70	264	173	212	0.241
3000K_2	70	70	70	265.5	168	177	0.215
3000K_3	70	70	70	289.5	199	212	0.240
3000K_4	70	70	70	291.5	170	205	0.232
3000K_5	70	70	70	305.5	181	196	0.239
3000K_6	70	70	70	380	184	195	0.226
3000K_7	70	70	70	221.5	168	182	0.224
3000K_8	70	70	70	234.5	181	189	0.232
3000K_9	70	2/70 ^c	70	484.5	182	192	0.250
3000K_10	70	70	70	353	185	202	0.236
3000K_11	70	70	70	410	186	201	0.227
3000K_12	70	70	70	454.5	183	198	0.230
3000K_13	70	70	70	236.5	175	194	0.250
3000K_14	70	70	70	305	157	174	0.232
3000K_15	70	70	70	254	176	202	0.236
3000K_16	70	70	70	423	175	183	0.219
3000K_17	70	70	70	520	190 (open)	212	0.215
3000K_18	70	70	70	520	177 (PAH)	210	0.214
3000K_19	70	70	70	520	163 (PAH)	201	0.227
3000K_20	70	70	70	431.5	182	186	0.237
average ^b				329.6	178	194	0.233

^a Listed quantities have the same meaning as in Table 1. ^b Refer to Table 1. ^c One hydrogen molecule (2 H atoms) was found and removed at 10 ps in the second H removal step in this trajectory.

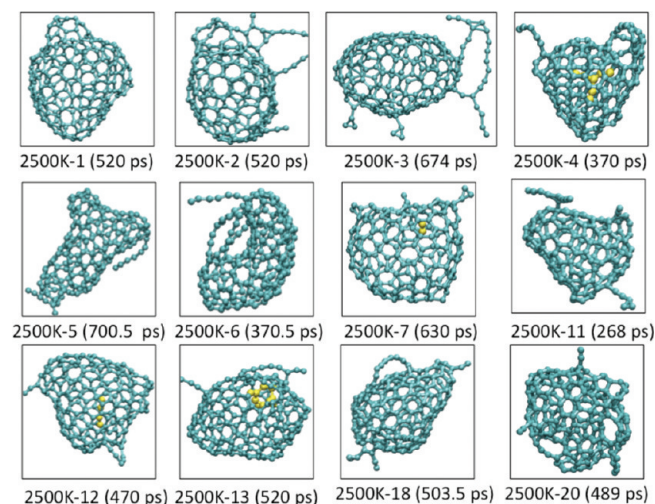


Figure 1. GF cages at 2500 K from the 12 successful trajectories. Yellow spheres indicate encapsulated fragments inside the formed cage.

Similar as in previous DFTB/MD simulations, the cage closure time also depends on the temperature: The average t_f at 2500 K was 503 ps (DFTB/MD: 82), while at 3000 K it was 329.6 ps (DFTB/MD: 46 ps). The most striking difference between DFTB/MD and ReaxFF/MD simulations is the much longer time required for cage formation time in the case of the latter. After 220 ps in ReaxFF/MD (the longest DFTB/MD simulation runs), not a single GF had formed. The reason for this significant discrepancy between ReaxFF and DFTB will be discussed below.

III.B. Dynamics of Giant Fullerene Self-Assembly. In the following sections, we analyze the general properties and processes (key events) of fullerene cage formation for both temperatures. Successful trajectories at the same temperatures followed qualitatively similar patterns, although t_f differed substantially between them. We here discuss only one representative trajectory for each temperature, namely, trajectories 2500K-1 and 3000K-2. Other trajectories follow similar time evolution patterns, although the detailed processes are of course different.

Figure 3 displays the evolution of the potential energy during the course of these trajectories. Discontinuous jumps up in the potential energy curves mark the H removal at 5, 10, and 15 ps, followed by relaxation associated with a potential energy decrease. The potential energy evolution on such a scale was similar to our previous DFTB/MD simulations.

III.B.1. Cluster Growth. Figure 4 displays for both trajectories a histogram of #C_{cluster} sizes of all occurring clusters, where different colors indicate separate molecular clusters. During the initial 5 ps equilibration, no chemical reactions occurred. This is in stark contrast to our previous DFTB/MD study, where pyrolytic H abstraction and radical reactions were observed, particularly at the higher temperature. At both low and high temperatures, the growth of the largest cluster started dramatically after the first H removal at 5 ps but slowed down after all the H atoms were removed from the system after 20 ps. At this time, a large cluster comprised of about half (2500K-1) or less than half (3000K-2) the number of carbon atoms had formed along with a couple of other, smaller clusters. In contrast, the apparently more reactive DFTB/MD simulations exhibited already after 20 ps a cluster consisting of almost all available

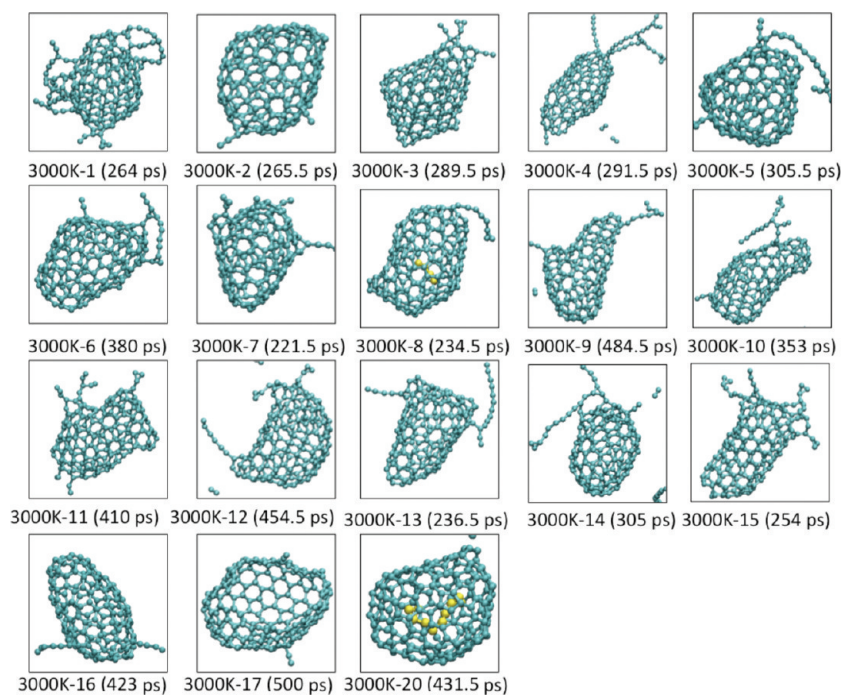


Figure 2. GF cages at 3000 K from the 18 successful trajectories. Yellow spheres indicate encapsulated fragments inside the formed cage.

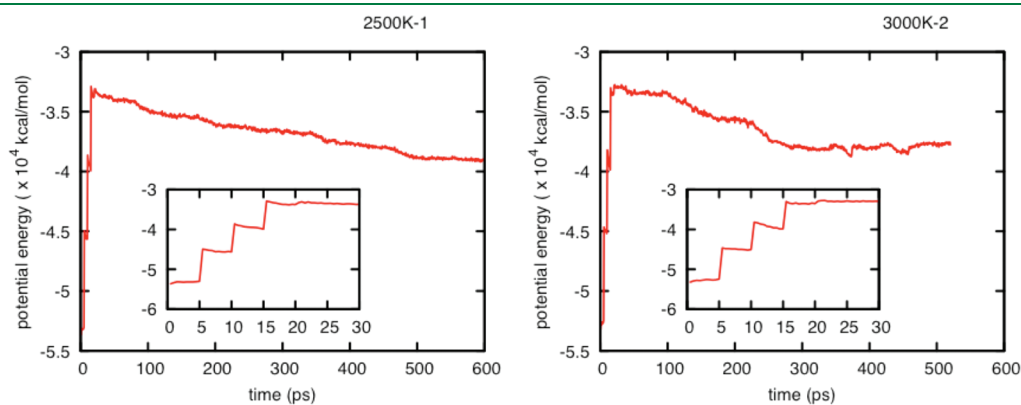


Figure 3. Potential energy variation curve during the course of 2500K-1 (left) and 3000K-2 (right) trajectories. The inset shows the early stage during the initial 30 ps. Jumps in energy correspond to hydrogen removals.

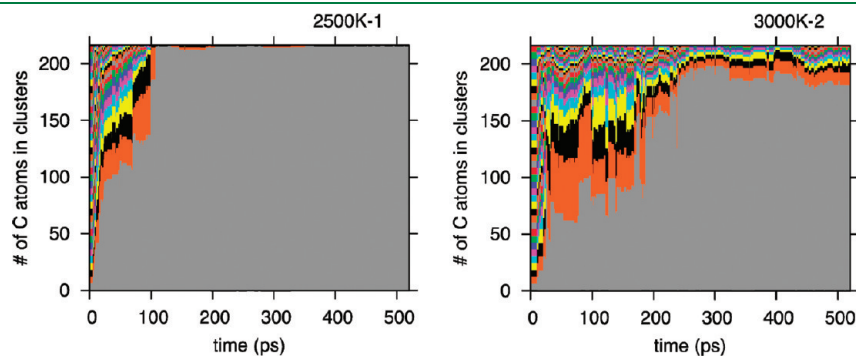


Figure 4. Cluster size ($\#C_{\text{cluster}}$) evolution during the course of 2500K-1 (left) and 3000K-2 (right) trajectories. There are 216 total carbon atoms, and the number of carbon atoms belonging to separate molecular clusters are indicated by vertical bars at a time interval of 1 ps. Initially, the reaction system contains 36 benzene molecules, and gradually small fragments coalesce into one major cluster (shown in gray).

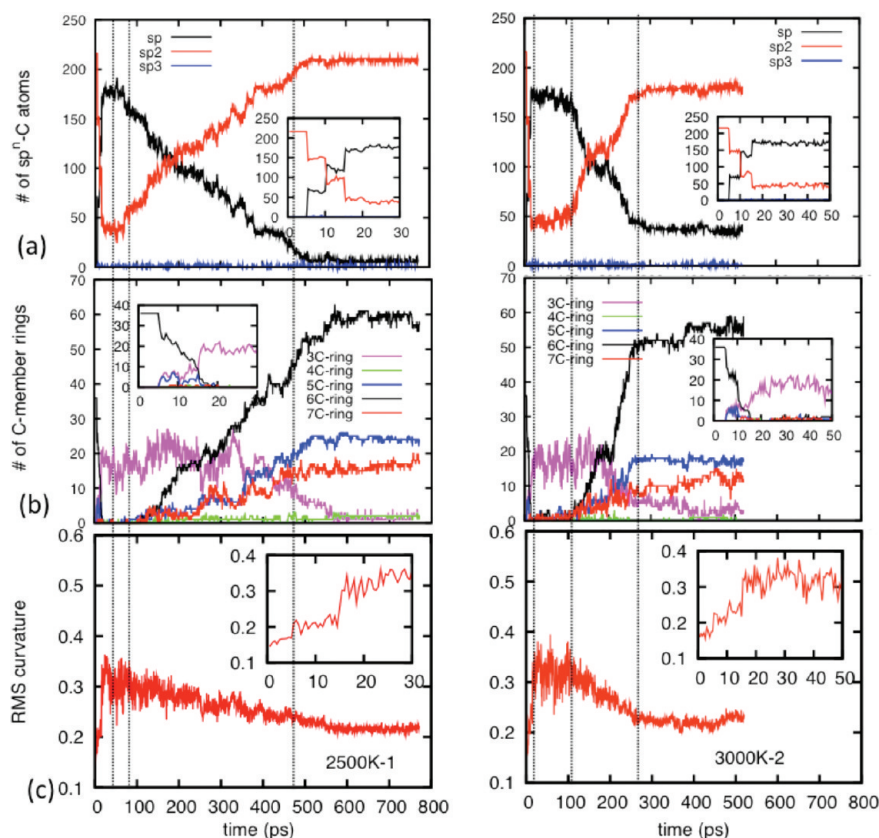


Figure 5. (a) Carbon hybridization count, (b) ring count statistics, and (c) root-mean-square curvature plot as a function of time during the course of trajectory 2500K-1 (left) and 3000K-2 (right).

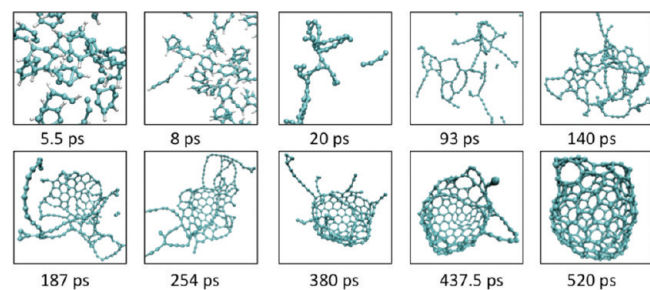


Figure 6. Key event snapshots at different stages in trajectory 2500K-1. Cyan and white spheres represent carbon and hydrogen atoms, respectively.

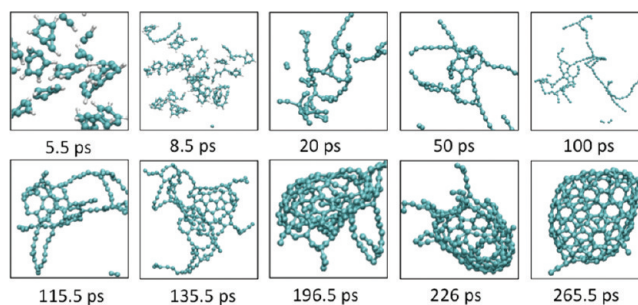


Figure 7. Key event snapshots at different stages in trajectory 3000K-2. Cyan and white spheres represent carbon and hydrogen atoms, respectively.

carbon atoms in the system. In ReaxFF/MD simulations, the largest clusters required a long time to coalesce with the smaller clusters after 20 ps.

As in our previous DFTB/MD study, we monitored the variations in the numbers of mono- or divalent (sp -hybridized), trivalent (sp^2 -hybridized), and tetravalent (sp^3 -hybridized) type carbon atoms as a function of simulation time. The curves are plotted in Figure 5a for trajectories 2500K-1 and 3000K-2. Since the number of tetravalent sp^3 carbons is always negligible under the conditions of the simulations, the plots record mainly the interconversion between sp and sp^2 carbons. Hence, the sp and sp^2 curves are roughly symmetric around 108 (half the number of totally available carbon atoms). In order to better understand and analyze the mechanism of the entire self-assembly process, we

also plot in the same figures the variations in the number of carbon rings from triangles up to heptagons (Figure 5b) and the variation of RMS atomic curvature (Figure 5c) in the system. Three vertical dashed lines in all panels of Figure 5 indicate qualitatively stages in the self-assembly processes. Figures 6 and 7 present the key event snapshots at different stages for trajectories 2500K-1 and 3000K-2, respectively. The data from Figures 5–7 will be discussed in the following paragraphs.

III.B.2. Events during and after H Removal. Each H removal changed the coordination number of the carbon (from which H was abstracted) from three to two. Thus, each of the three initial H removal steps caused a sudden decrease in the total number of sp^2 carbons and the corresponding increase in the number of sp carbons, as can be seen in the insets of Figure 5a. Along with this

change in carbon hybridization, we observed a decrease in the number of hexagons, indicating ring-opening processes. Consequently, linear hydrocarbon chains were present in the system, for instance, shown in the snapshots at 8 ps in trajectory 2500K-1 (Figure 6) and at 8.5 and 20 ps in trajectory 3000K-2 (Figure 7). The ring-opening process producing linear polyacetylenic chains is similar to our observations in DFTB/MD simulations. Interestingly, several pentagons were created after the first H removal at 5 ps, but these pentagons disappeared quickly in the simulations at both temperatures. More significantly, around the same time, three-membered carbon rings (“triangles”) started to appear in large numbers (approximately 20 per trajectory), reaching an abundance plateau at ~ 20 ps and *remaining the dominant ring species until 180–200 ps* (see insets of Figure 5b). Initially, triangles and pentagons were often created in an *m*-benzynes ring by bridge bond formation between the two hydrogen-devoid carbon atoms, forming bicyclo[3.1.0]hexatriene [see for instance the snapshots at 5.5 ps of trajectories 2500K-1 and 3000K-2 (Figures 6 and 7)]. This is an isomer of *m*-benzynes, which was speculated upon in the literature but shown in high-level *ab initio* quantum chemical calculations not to exist.³³ The pentagons of bicyclo[3.1.0]hexatriene were not stable and started to disappear from around 10 ps due to ring-opening processes, while the triangles continued to form abundantly, this time at the tails of polyyne chains and between the chains of Y-junction corners. Examples for such triangles at terminal or Y-junction positions are visible in snapshots of trajectory 2500K-1 at 20 and 93 ps (Figure 6) and 3000K-2 at 20 and 50 ps (Figure 7). In DFTB/MD simulations, triangles never appeared as stable species. Such carbon triangles were also discussed in a systematic ReaxFF benchmark study²⁸ and a ReaxFF/MD study of the aggregation of carbon in an atmosphere of hydrogen molecules.³⁴

As mentioned, the most noticeable difference between present ReaxFF/MD and previous DFTB/MD simulations concerns the time required for cage formation: ReaxFF/MD seems an order of magnitude slower to produce the first fully formed GF cages. The difference in time evolution is clearly visible in the time evolution of sp - and sp^2 -hybridized carbon when comparing ReaxFF/MD (Figure 5a) and DFTB/MD (Figure 3 in ref 15) results: Both simulations predict a minimum number of sp^2 carbon atoms after the third H removal, but it takes vastly longer in ReaxFF/MD simulations (20–30 ps in DFTB/MD, 150–200 ps in ReaxFF/MD) to recover and reach a 50:50 ratio to sp carbon atoms. At the same time, practically no polygonal rings were found at this stage, with the exception of the above-mentioned triangles. In this context, we note another important difference between DFTB/MD and ReaxFF simulation results: DFTB/MD simulations predicted that *ring-condensation processes* would occur rapidly once all hydrogen atoms left the system, easily bringing the number of hexagons back up to around the original number of 36. The same recovery process was found to be very slow in ReaxFF/MD simulations. It took roughly 450 ps (starting from ~ 45 to 500 ps) in the case of 2500K-1 and roughly 150 ps (from ~ 100 ps to ~ 250 ps) in the case of 3000K-2. Similarly as in the DFTB/MD simulations, hexagons were the dominant ring species, followed by pentagons and heptagons. As the ring condensation reactions continued in ReaxFF/MD simulations, the number of triangular carbons became gradually reduced. We conclude that the ring condensation process of pure carbon chains occurs on a different time scale with different intermediate ring species in the two methods. This difference may be explained

at least partially by the fact that the angle strain in linear carbon chains is overestimated by $\sim 20\%$ in ReaxFF relative to the quantum chemical B3LYP/6-31G(d) level of theory,²⁸ while the corresponding DFTB angle strain reproduces the first principles result well.

In Figure 5c, we trace the variation of the root-mean-square (RMS) of the atomic curvatures of all sp^2 carbon atoms in the system. It is clear that the RMS curvature has a small value ($< 0.2 \text{ \AA}^{-1}$) for initial benzene rings at the simulated temperatures. Once the H removal began, the benzene rings started to open, and the molecules were transformed into linear fragments, which have larger flexibility, and thus larger local curvatures can be expected for sp^2 carbon atoms, for instance, at Y-junctions. During this stage, the curvature values were oscillating around a relatively large value > 0.3 . Large oscillations in the curvature curve are attributed to the large flexibilities of Y-junction points during this stage. During the following ring condensation stage, the RMS curvature values slowly converged to values of the final cage structures.

Figure 8 displays the chemical composition of hydrocarbons (C_xH_y) before the second and third H removal step at 10 and 15 ps from all of the 20 trajectories at both temperatures. This figure corresponds to Figure 5 in ref 15. It is clear that at higher temperatures, the trajectories contain larger clusters. After 10 ps simulations, the largest cluster consisted of ~ 43 carbon atoms for $T = 3000$ K, whereas the largest cluster at $T = 2500$ K consisted of only ~ 36 carbon atoms; after 15 ps, the largest cluster was found to have ~ 170 carbon atoms at 3000 K but only ~ 130 at 2500 K. The clusters at both stages are mainly dominated by open-chain polyacetylene-like structures. In the case of DFTB/MD simulations, PAH species and fullerene precursors were already present at ~ 15 ps; in the case of ReaxFF/MD simulations, only polyacetylene-like chain structures were found. Similarly to DFTB/MD predications, larger clusters tend to have slightly lower H/C ratios than the overall system's H/C ratio, while smaller fragments have somewhat higher H/C ratios.¹⁵

III.B.3. Cage Self-Assembly. The mechanism of dynamic fullerene self-assembly is qualitatively similar in both DFTB/MD and ReaxFF/MD simulations, with major differences concerning chemical reaction speed and differences in predicted intermediate structures as discussed above. The **first step** is the *ring-opening and fragmentation* process. In ReaxFF/MD simulations, during H removal, all carbon rings were destroyed, giving way to polyacetylene-like chains. In contrast to DFTB/MD simulations, this step continues beyond the last H removal (until ~ 20 ps). The **second step** in ReaxFF/MD simulations is the *linear chain growth process*. Short, linear carbon chains devoid of hydrogen fuse, creating larger linear and branched chains. This step is characteristic for the ReaxFF/MD simulations and does not occur during DFTB/MD simulations, where rings were formed already during H removal from ring-opened polyacetylene-like chains. The **third step** is dominated by *ring condensation*, starting from small nuclei of condensed rings. Hexagons are most abundant, while a significant number of pentagons provided enough positive curvature to the structure, causing the appearance of basket-shaped structures with linear polyyne chains attached (“octopus on the rock” structures). The number of heptagons is competitive with that of the pentagons but remains smaller. Since pure carbon chains are abundant in ReaxFF/MD simulations after H removal, the “arms of the octopus” are somewhat more abundant when compared to DFTB/MD simulations. Finally, the **fourth step** of fullerene cage self-assembly, *cage-closure*,

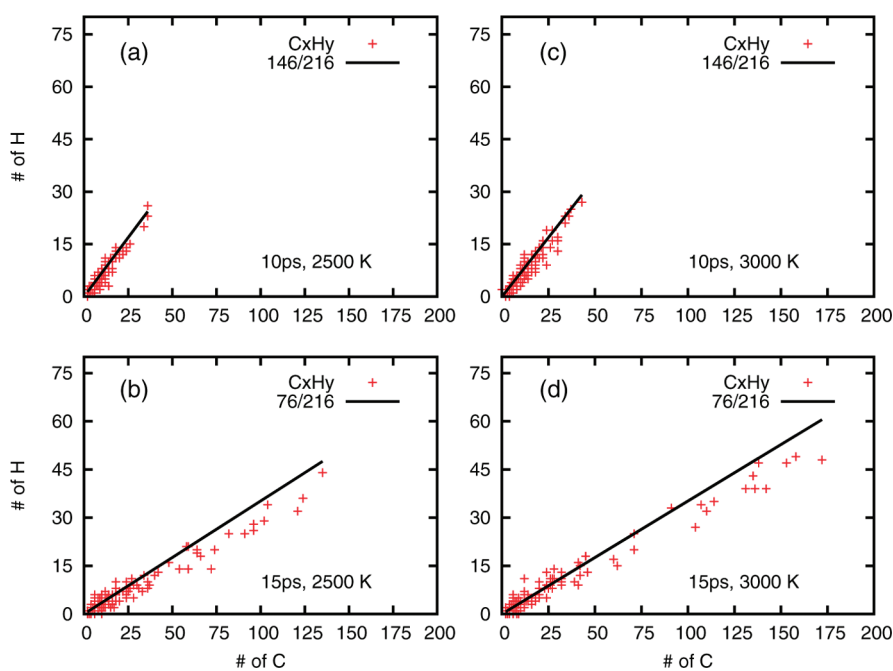


Figure 8. Hydrocarbon cluster compositions at (a) 10 ps for 2500 K, (b) 15 ps for 2500 K, (c) 10 ps for 3000 K, and (d) 15 ps for 3000 K (before the H removal steps). Each point corresponds to a C_xH_y species. The continuous lines represent the overall H/C ratio of the entire system. Data points are plotted for all 2500 K (left) and 3000 K (right) trajectories.

occurred during a drawn-out process and is shown in Figures 6 (520 ps) and 7 (265.5 ps). Similar to the DFTB/MD predictions, in the final closed cages, the numbers of hexagon rings were dominant, followed by around 20 pentagons and a comparable number of heptagons. Clearly, the final self-assembled cages are very large, highly defective, and far from obeying the isolated pentagon rule, similar as in DFTB/MD simulations.

III.C. Computational Efficiency. To accurately evaluate the differences in CPU times required for corresponding trajectories, we timed DFTB/MD and ReaxFF/MD simulations for a system with the following characteristics: 36 benzene molecules were placed in a 21 \AA^3 PBC box, and a Berendsen thermostat with 2500 K target temperature was employed in both simulations. For this benchmark, the time integration interval in DFTB was lowered to $\Delta t = 0.1$ fs so that we could estimate the performance difference of the two methods using the same number of time intervals for the same simulated time. A total of 1000 time integration steps were performed, yielding two trajectories of 100 fs length in both ReaxFF- and DFTB-based MD simulations. The CPU time required was 130.73 and 2396.31 s, respectively, on an Intel Xeon E5460@2.66 GHz CPU core, indicating that ReaxFF is roughly 20 times faster than DFTB in such a system. When taking into account that DFTB/MD simulations can be performed with a larger time interval of $\Delta t = 0.48$ fs (while still conserving total energy in NVE simulations with an error of only a few kcal/mol), the speedup of ReaxFF reduces effectively to a factor of about 4.

IV. SUMMARY AND CONCLUSIONS

ReaxFF/MD simulations predict similar patterns of fullerene formation during benzene combustion to those of previous DFTB/MD simulations,¹⁵ with major differences concerning chemical reaction speed and some differences in predicted

intermediate structures. Common to both methods is the prediction that, under the given carbon concentration in the absence of a carrier gas, giant fullerene cages with sizes between 155 and 212 (DFTB/MD: between 174 and 212) carbon atoms are formed in relatively high yields of 60% (2500 K, DFTB/MD: 50%) and 85% (3000 K, DFTB/MD: 42%). The cage self-assembly in both types of MD simulations involves the benzene ring-opening, carbon chain, and ring growth *via* ring-condensation reactions of sp -hybridized carbon chains attached to sp^2 carbon ring networks. In both simulations, hexagons finally outweigh by far pentagons and heptagons, and the final giant fullerene (GF) cages possess linear carbon chains attached to the cages at sp^3 carbon defects. The major difference between classical ReaxFF and quantum chemical DFTB potentials concerns the reactivity of the partially hydrogenated and pure carbon species: DFTB predicts a much faster succession of events, leading to fullerene cage closure being up to a full order of magnitude faster (~ 50 ps vs ~ 500 ps) than ReaxFF. It appears that the bending of sp^1 chains is probably the deciding factor in the reaction: A comparison of the potential energy profiles for the bending of the C_3 molecule (see Figure S1 in the Supporting Information) shows that this mode is stiffer in ReaxFF than in the DFTB and DFT methods, even though ReaxFF outperforms DFTB in the prediction of isomer energies of the C_{28} species (Table S1 in Supporting Information). Further differences were found as follows:

- 1 Final cages often appear to be “nested” or “spiroid”-like in ReaxFF/MD simulations, exhibiting major discontinuities in the sp^2 carbon cage walls.
- 2 GFs occasionally exhibit endohedrally encapsulated small C_n chains or rings, different from previous DFTB/MD simulations.
- 3 In ReaxFF simulations, carbon triangles are common occurrences as termini of linear carbon chains and during hydrogen removal as part of the hypothetical bicyclo[3.1.0]hexatriene

isomer of *m*-benzyne (see also Figure S1 in the Supporting Information).

- 4 In ReaxFF simulations, carbon rings open up to form linear chains during hydrogen removal.

It is difficult to pinpoint the exact contribution of these fundamental differences between both methods to each phenomenological difference. Since DFTB is an approximate method of density functional theory, which itself is an approximation of higher *ab initio* quantum chemical methods, we conclude that ReaxFF/MD is well suited to describe fullerene formation processes on nanosecond time scales. These simulations for the presented systems are between 4 and 20 times faster than corresponding quantum chemical DFTB/MD simulations, depending on the employed time integration intervals.

■ ASSOCIATED CONTENT

S Supporting Information. Frozen potential energy scan of the \angle CCC angle of the C_3 molecule in its singlet ground state using self-consistent-charge (SCC) and nonself-consistent-charge (NCC) DFTB levels of theory, ReaxFF, and B3LYP/6-311G(d); relative energies of C_{28} isomers in electronvolts and corresponding linear regression coefficient R^2 for correlation between B3LYP/6-31G(d) and SCC- and NCC-DFTB, B3LYP/6-31G, AM1, PM3, and ReaxFF methods. Cartesian coordinates in Ångströms of last snapshot structures from trajectories 1–20 at 2500 and 3000 K corresponding to either t_f or t_{end} , as indicated in Tables 1 and 2. In these structures, label C' denotes endohedrally encapsulated carbon atoms. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: acv13@engr.psu.edu (A.C.T.v.D.); sirle@iar.nagoya-u.ac.jp (S.I.).

■ ACKNOWLEDGMENT

We thank Biswajit Saha for providing the initial g1 coordinates. S.I. acknowledges support by the Program for Improvement of Research Environment for Young Researchers from Special Coordination Funds for Promoting Science and Technology (SCF) commissioned by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. This work was supported in part by a CREST (Core Research for Evolutional Science and Technology) grant in the Area of High Performance Computing for Multiscale and Multiphysics Phenomena from the Japan Science and Technology Agency (JST).

■ REFERENCES

- (1) Kroto, H. W.; Heath, J. R.; O'Brien, S. C.; Curl, R. F.; Smalley, R. E. *Nature* **1985**, *318*, 162.
- (2) Krätschmer, W.; Fostiropoulos, K.; Huffman, D. R. *Chem. Phys. Lett.* **1990**, *170*, 167.
- (3) Howard, J. B.; McKinnon, J. T.; Makarovskiy, Y.; Lafleur, A. L.; Johnson, M. E. *Nature* **1991**, *352*, 139.
- (4) Hebgren, P.; Goel, A.; Howard, J. B.; Rainey, L. C.; Vander Sande, J. B. *Proc. Combust. Inst.* **2000**, *28*, 1397.
- (5) Frenklach, M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2028.
- (6) Alford, J. M.; Bernal, C.; Cates, M.; Diener, M. D. *Carbon* **2008**, *46*, 1623.
- (7) Goel, A.; Hebgren, P.; Vander Sande, J. B.; Howard, J. B. *Carbon* **2002**, *40*, 177.

(8) *Masters of the Flame: Industrial Production of Fullerenes Becomes a Reality*; Nano-C: Westwood, MA, 2004.

(9) Takehara, H.; Fujiwara, M.; Arikawa, M.; Diener, M. D.; Alford, J. M. *Carbon* **2005**, *43*, 311.

(10) Homann, K.-H. *Angew. Chem., Int. Ed.* **1998**, *37*, 2434.

(11) Pope, C. J.; Marr, J. A.; Howard, J. B. *J. Phys. Chem.* **1993**, *97*, 11001.

(12) Ahrens, J.; Bachmann, M.; Baum, T.; Griesheimer, J.; Kovacs, R.; Weilmünster, P.; Homann, K. H. *Int. J. Mass. Spectrosc. Ion Proc.* **1994**, *138*, 133.

(13) Grieco, W. J.; Howard, J. B.; Rainey, L. C.; Vander Sande, J. B. *Carbon* **2000**, *38*, 597.

(14) Richter, H.; Labrocca, A. J.; Grieco, W. J.; Taghizadeh, K.; Lafleur, A. L.; Howard, J. B. *J. Phys. Chem. B* **1997**, *101*, 1556.

(15) Saha, B.; Shindo, S.; Irle, S.; Morokuma, K. *ACS Nano* **2009**, *3*, 2241.

(16) Irle, S.; Zheng, G.; Elstner, M.; Morokuma, K. *Nano Lett.* **2003**, *3*, 1657.

(17) Irle, S.; Zheng, G.; Wang, Z.; Morokuma, K. *J. Phys. Chem. B* **2006**, *110*, 14531.

(18) Zheng, G.; Irle, S.; Morokuma, K. *J. Chem. Phys.* **2005**, *122*, 014708.

(19) Yamaguchi, Y.; Maruyama, S. *Chem. Phys. Lett.* **1998**, *286*, 336.

(20) Maruyama, S.; Yamaguchi, Y. *Chem. Phys. Lett.* **1998**, *286*, 343.

(21) Powles, R. C.; Marks, N. A.; Lau, D. W. M. *Phys. Rev. B* **2009**, *79*, 075430.

(22) Irle, S.; Zheng, G.; Elstner, M.; Morokuma, K. *Nano Lett.* **2003**, *3*, 465.

(23) Huang, J. Y.; Ding, F.; Jiao, K.; Yakobson, B. I. *Phys. Rev. Lett.* **2007**, *99*, 175503.

(24) Chenoweth, K.; van Duin, A. C. T.; Goddard, W. A. *J. Phys. Chem. A* **2008**, *112*, 1040.

(25) Page, A. J.; Moghtaderi, B. *J. Phys. Chem. A* **2009**, *113*, 1539.

(26) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. *J. Phys. Chem. A* **2001**, *105*, 9396.

(27) Clar, E. *Polycyclic Hydrocarbons*; AP: London, 1964.

(28) Nielson, K. D.; van Duin, A. C. T.; Oxgaard, J.; Deng, W.-Q.; Goddard, W. A., III. *J. Phys. Chem. A* **2005**, *109*, 493.

(29) Brenner, D. W. *Phys. Rev. B* **1990**, *42*, 9458.

(30) Brenner, D. W. *Phys. Rev. B* **1992**, *46*, 1948.

(31) Tersoff, J. *Phys. Rev. B* **1988**, *37*, 6991.

(32) Bogana, M.; Ravagnan, L.; Casari, C. S.; Zivelonghi, A.; Baserga, A.; Bassi, A. L.; Bottani, C. E.; Vinati, S.; Salis, E.; Piseri, P.; Barborini, E.; Colombo, L.; Milani, P. *New J. Phys.* **2005**, *7*, 1.

(33) Kraka, E.; Anglada, J.; Hjerpe, A.; Filatov, M.; Cremer, D. *Chem. Phys. Lett.* **2001**, *348*, 115.

(34) Lümmen, N. *Comput. Mater. Sci.* **2010**, *49*, 243.

A Hirshfeld Partitioning of the MP2 Correlation Energy: Method and Its Application to the Benzene Dimers

Alisa Krishtal,^{*,†} Sergei F. Vyboishchikov,^{*,‡} and Christian Van Alsenoy^{*,†}

[†]Department of Chemistry, University of Antwerp, Universiteitsplein 1, B2610 Antwerp, Belgium

[‡]Institut de Química Computacional, Campus de Montilivi, Universitat de Girona, 17071 Girona, Catalonia, Spain

ABSTRACT: In this work a method is presented for the partitioning of MP2 correlation energies through a grid-space partitioning using the iterative Hirshfeld weight function. The correlation energies are partitioned into mono- and diatomic contributions using two alternative schemes, which allow different levels of parallelization. The method is tested on a set of 24 molecules containing various atoms, leading to the conclusion that, while the numerical results of the two schemes slightly differ, the chemical information contained in them is similar. The method is subsequently applied to the analysis of the interaction energy of three benzene dimers.

1. INTRODUCTION

The appeal in studying the energy of a system lies in its hold on information, such as bonding, stability, and reactivity. The information provided by an energy calculation can be greatly enriched if the total energy can be partitioning into chemically meaningful components. An obvious way of partitioning the energy within a correlated ab initio method is to do it for the Hartree–Fock and correlation energy separately. Since a number of energy partitioning schemes are available for the Hartree–Fock energy,^{1–10} the present paper focuses on the correlation energy partitioning.

Since the energy expression depends on the level of theory, the number and the nature of various contributions differ accordingly. In particular, within the range of the correlated ab initio methods, such as Møller–Plesset¹¹ perturbation theory and coupled clusters,^{12,13} the quality of the energy obtained is determined by the sophistication level of the correlation energy expression. For a closed-shell system, the correlation energy can be expressed through the amplitude τ_{ij}^{ab} in eq 1:

$$E_c = \sum_{ij} \sum_{ab}^{\text{occ virt}} \tau_{ij}^{ab} [2(ia|jb) - (ib|ja)] \quad (1)$$

where i and j denote occupied molecular orbitals, a and b are virtual molecular orbitals, and $(ia|jb)$ are two-electron repulsion integrals. Although this expression is, strictly speaking, of a nonlocalized nature, it is tempting for a chemist to rationalize it in terms of atomic and bond (diatomic) contributions, in order to extract chemical information about the system.

Therefore, several methods have been developed in the past decade for the partitioning of correlation energies.^{14–22} These methods can be divided into two major categories. The first one includes methods based on a population analysis partitioning,^{14–17} where one makes use of the linear combination of atomic orbitals (LCAO) expansion of the molecular orbitals and rewrites eq 1 in terms of a sum over atomic orbitals with common centers, thus partitioning the total correlation energy into a sum of mono- and diatomic contributions.^{14–16} On the other hand, for variational methods such as configuration interaction (CI), the two-electron

energy can be partitioned directly using the second-order density matrix Γ :¹⁷

$$E_{cc} = \sum_{ijkl} \Gamma_{ijkl} (ij|kl) \quad (2)$$

where E_{cc} contains both the correlation energy as well as the two-electron Hartree–Fock energy contributions. The methods of the second category are based on a partitioning in the three-dimensional physical space,^{3,18–21} where the two-electron integrals $(ia|jb)$ are partitioned into mono- or diatomic contributions by inserting the atomic weight functions into the integral. In this case, too, one can make use of either eqs 1¹⁸ or 2.^{19–21}

This broad collection of methods reflects the nonuniqueness of partitioning, which is due to the lack of a quantum mechanical “atomic Hamiltonian” within a molecule. Therefore, none of the wabove-mentioned methods can be a priori declared superior to the others. The usefulness of each method is to be judged by practical criteria, such as the consistency of the results with the chemical knowledge and intuition. For instance, when partitioning the total energies into diatomic contributions, one would desire the values to be comparable with dissociation energies for bonding interactions, to increase with the bond multiplicity, and to provide clearly different values for bonding and nonbonding interactions.

In this work, we propose to partition the correlation energy of the second-order Møller–Plesset method (MP2) through the use of the Hirshfeld method,^{23–25} which has proved to be useful for the analysis of a broad gamma of properties, including the energy at Hartree–Fock⁹ and density functional theory (DFT) levels.²⁶ A short description of the Hirshfeld method, together with the outline of our methodology for the partitioning of the MP2 correlation energy into mono- and diatomic contributions is presented in Section 2, followed by some details on the implementation in Section 3 and by a discussion of the results on a limited set of 24 molecules in Section 4. An application of the method to the interaction energy of three benzene dimers is

Received: December 3, 2010

Published: June 09, 2011

presented in Section 5. Finally, Section 6 contains the summary and the conclusions of this work.

The correlation energy decomposition can be used in two manners. First, for the systems in which the electron correlation essentially determines the molecular structure (such as van der Waals complexes), the decomposition can be applied as such to understand which pairwise interactions are more important than the others. On the other hand, the correlation energy decomposition can be used for any molecule on the top of a suitable Hartree–Fock energy decomposition (for instance, according to ref 9). This gives rise to a total correlated energy decomposition.

2. METHOD

The Hirshfeld method allows to partition molecular properties into atomic contributions through insertion of a normalized weight function into the integral that determines the property in question. For instance, the population of atom A is determined by the following expression:

$$N_A = \int w_A(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} = \int \rho_A(\mathbf{r})d\mathbf{r} \quad (3)$$

where $\rho(\mathbf{r})$ is the molecular density and $w_A(\mathbf{r})$ is the atomic weight function. The total number of electrons is recovered by summation over all the atoms, since the atomic weight functions at each point \mathbf{r} sum to one:

$$N = \sum_A \int w_A(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} = \int \left(\sum_A w_A(\mathbf{r})\right)\rho(\mathbf{r})d\mathbf{r} \quad (4)$$

In the classic version of the Hirshfeld method (H-C),²³ the weight function of each atom is constructed from the free-atom densities $\rho_A^{[0]}(\mathbf{r})$ normalized by a superposition of the free-atom densities of all the atoms in the molecule, referred to as “promolecule”:

$$w_A^{\text{H-C}}(\mathbf{r}) = \frac{\rho_A^{[0]}(\mathbf{r})}{\sum_B \rho_B^{[0]}(\mathbf{r})} \quad (5)$$

These free-atom densities are obtained from a self-consistent field (SCF) calculation on an isolated atom in its spectroscopic ground state with the same basis set as used in the calculation of the electron density of the molecule. This form of a weight function assumes that the atoms within the molecules resemble the free spherically symmetric neutral atoms. Since this choice is somewhat arbitrary, an improved iterative version of the Hirshfeld method (H-I) was developed by Bultinck et al.,²⁵ in which the promolecule is constructed from densities of atoms which resemble actual atoms in the molecule rather than free atoms. This is achieved through an iterative procedure, where the weight function at the n -th iteration is constructed from the atomic densities obtained in the previous iteration:

$$w_A^{[n]}(\mathbf{r}) = \frac{\rho_A^{[n-1]}(\mathbf{r})}{\sum_B \rho_B^{[n-1]}(\mathbf{r})} \quad (6)$$

where $\rho_A^{[n-1]}(\mathbf{r}) = w_A^{[n-1]}(\mathbf{r})\rho(\mathbf{r})$ and $w_A^{[1]}(\mathbf{r}) = w_A^{\text{H-C}}(\mathbf{r})$ is the classic Hirshfeld weight function. Here, $\rho_A^{[n-1]}(\mathbf{r})$ is an atomic

density which integrates to the number of electrons $N_A^{[n-1]}$ calculated in the previous iteration as

$$N_A^{[n-1]} = \int w_A^{[n-1]}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \quad (7)$$

Since $N_A^{[n-1]}$ is generally a noninteger number, this is achieved by interpolating between the densities of atoms with the upper bound integer of $N_A^{[n-1]}$ and the lower bound integer of $N_A^{[n-1]}$ number of electrons. This procedure is repeated until convergence, at which point the converged atomic weight functions become the H-I weight functions. The practical details of the iterative procedure can be found in ref 25.

The partitioning of a two-electron integral can be obtained by inserting the Hirshfeld atomic weight function into the two-electron integral as follows:

$$\begin{aligned} (ia|jb) &= \sum_A (ia|jb)^A \\ &= \sum_A \int w_A(\mathbf{r}_1)\varphi_i(\mathbf{r}_1)\varphi_a(\mathbf{r}_1) \left[\int \frac{\varphi_j(\mathbf{r}_2)\varphi_b(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 \right] d\mathbf{r}_1 \end{aligned} \quad (8)$$

The integral in square brackets can be easily evaluated at each gridpoint \mathbf{r}_1 using analytical formulas for nuclear attraction integrals, which allows avoiding double numerical integration. A similar partitioning of the two-electron integral was employed by Imamura et al.¹⁸ However, those authors used the Becke weight function²⁷ rather than the Hirshfeld partitioning. More importantly, Imamura et al. only decompose the amplitude in the correlation energy expression (eq 1), which leads to a correlation energy expressed by a sum of monatomic contributions only. In contrast, we prefer to partition also the two-electron integrals in the second multiplier of eq 1 in order to obtain both mono- and diatomic terms. Eventually, this yields the following MP2 correlation energy partitioning into diatomic contributions $E_{\text{MP2}}^{\text{AB}}$ and monatomic ones $E_{\text{MP2}}^{\text{AA}}$:

$$E_{\text{MP2}} = \sum_A E_{\text{MP2}}^{\text{AA}} + \sum_{A < B} E_{\text{MP2}}^{\text{AB}} \quad (9)$$

where

$$E_{\text{MP2}}^{\text{AB}} = 2 \sum_i^{\text{occ}} \sum_a^{\text{virt}} \sum_j^{\text{occ}} \sum_b^{\text{virt}} \frac{(ia|jb)^A [2(ia|jb)^B - (ib|ja)^B]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (10)$$

and

$$E_{\text{MP2}}^{\text{AA}} = \sum_i^{\text{occ}} \sum_a^{\text{virt}} \sum_j^{\text{occ}} \sum_b^{\text{virt}} \frac{(ia|jb)^A [2(ia|jb)^A - (ib|ja)^A]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (11)$$

Here ε_i is the orbital energy of the i -th molecular orbital. The factor 2 in eq 10 is due to symmetric nature of the expression: $E_{\text{MP2}}^{\text{AB}} = E_{\text{MP2}}^{\text{BA}}$.

A closer look at eq 10 reveals an alternative expression for the $E_{\text{MP2}}^{\text{AB}}$ energy contribution. Indeed, since $(ia|jb) = (ai|bj)^*$, a similar but not equivalent diatomic correlation energy expression $E'_{\text{MP2}}^{\text{AB}}$ can be defined:

$$E'_{\text{MP2}}^{\text{AB}} = 2 \sum_a^{\text{virt}} \sum_i^{\text{occ}} \sum_b^{\text{virt}} \sum_j^{\text{occ}} \frac{(ai|bj)^A [2(ai|bj)^B - (aj|bi)^B]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (12)$$

The difference between the two definitions lies in the terms $(ib|ja)^B$ and $(aj|bi)^B$, which are not equivalent since the atomic weight function of atom B is inserted in the left-hand side of the integral (eq 8). This involves integration over \mathbf{r}_1 in $\int_{w_B(\mathbf{r}_1)} \varphi_i(\mathbf{r}_1) \varphi_b(\mathbf{r}_1) \int \varphi_j(\mathbf{r}_2) \varphi_a(\mathbf{r}_2) r_{12}^{-1} d\mathbf{r}_2 d\mathbf{r}_1$ in the former and in $\int_{w_B(\mathbf{r}_1)} \varphi_a(\mathbf{r}_1) \varphi_j(\mathbf{r}_1) \int \varphi_i(\mathbf{r}_2) \varphi_b(\mathbf{r}_2) r_{12}^{-1} d\mathbf{r}_2 d\mathbf{r}_1$ in the latter.

Partitioning the MP2 correlation energy according to eq 10 can be interpreted in terms of pair correlation energies e_{ij} :

$$E_{MP2} = \sum_{ij}^{occ} e_{ij} \quad (13)$$

where

$$e_{ij} = \sum_{ab}^{virt} \frac{(ia|jb)[2(ia|jb)] - (ib|ja)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (14)$$

$$= \sum_{ab}^{virt} \frac{(ai|bj)[2(ai|bj)] - (aj|bi)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (15)$$

Partitioning e_{ij} into diatomic contributions according to eq 10 can be regarded as decomposing the i -th occupied orbital to atoms A and B while summing over the occupied j . The alternative expression in eq 12, on the other hand, decomposes i to atom A on the left-hand side of the product in the e_{ij} expression, but both i and j are decomposed to atom B on the right-hand side of the product. From this perspective, eq 12 is less theoretically sound than eq 10 but, as will be shown later, is better suited for parallelization while producing similar results.

Yet another option would be to consistently decompose orbital i to atom A and orbital j to atom B, which leads to the expression:

$$\tilde{E}_{MP2}^{AB} = 2 \sum_i^{occ} \sum_a^{virt} \sum_j^{occ} \sum_b^{virt} \frac{(ia|jb)^A [2(jb|ia)^B - (ja|ib)^B]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (16)$$

However, since expression 16 is considerably more computationally demanding, we will focus on eqs 10 and 12 for the rest of this article.

3. COMPUTATIONAL DETAILS

The calculation of all diatomic contributions in eqs 10 and 12 is computationally demanding, since each of the integrals $(ia|jb)^A$ has to be evaluated numerically. As we would like to be able to apply this method in systems of interest without being overencumbered with size limitations, we implemented an MPI-parallelized code with the following structure (Chart 1):

Chart 1

do i=proc_id,num_occ_orb,numprocs	do a=proc_id,num_virt_orb,numprocs
loop over gridpoints	loop over gridpoints
integrate and store in memory	integrate and store in memory
(ia jb)^A for all a, j, b and A	(ai bj)^A for all i, b, j and A
end loop	end loop
calculate E(i,AB)	calculate E(a,AB)
enddo	enddo
reduce all E(i,AB)'s	reduce all E(a,AB)'s

where the left-hand scheme represents parallelization over the occupied molecular orbitals i , according to eq 10, while the right-hand scheme represents parallelization over the virtual molecular orbitals a according to eq 12. Here $E(i,AB)$ and $E(a,AB)$ are the contributions of the i -th molecular orbital into E_{MP2}^{AB} and of the a -th molecular orbital into E_{MP2}^{AB} , respectively.

$$E(i,AB) = (2 - \delta_{AB}) \sum_a^{virt} \sum_j^{occ} \sum_b^{virt} \frac{(ia|jb)^A [2(ia|jb)^B - (ib|ja)^B]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (17)$$

$$E(a,AB) = (2 - \delta_{AB}) \sum_i^{occ} \sum_b^{virt} \sum_j^{occ} \frac{(ai|bj)^A [2(ai|bj)^B - (aj|bi)^B]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (18)$$

The possibility to calculate $E(i,AB)$ and $E(a,AB)$ is the key feature of the parallelization scheme; this cannot be achieved for expressions, such as eq 16. Although the right-hand scheme will always result in a larger total CPU time, the number of numerical integrations to be performed being larger (since there are always more virtuals than occupied), it allows a much higher degree of parallelization. For instance, an ethylene molecule calculated with the aug-cc-pVDZ basis set has 8 occupied and 72 virtual molecular orbitals. In the first version of the method the calculation can be parallelized over a maximum of 8 processes, while in the second version the calculation can be run over 72 processes. In the latter case, the wall-clock time of the calculation will be effectively reduced by a factor of 72 compared to a single-process run of the same code, since the communication needed between the processors is minimal, requiring only the transmission of a symmetric matrix holding the $E(a,AB)$ energies.

The geometries of the 24 small molecules examined in Section 4 were optimized with the Gaussian03²⁸ program at the MP2/aug-cc-pVDZ level. The full MP2 formalism was used throughout the work. The geometries of the three benzene dimers examined in Section 5 are obtained from ref 30, and the interaction energies were calculated using the Gaussian03²⁸ program at the MP2/6-31++G** level. The partitioning of the correlation energies was performed using the STOCK program.²⁴ For the sake of comparison, the Ayala–Scuseria energies¹⁴ were calculated using another program written by the authors.

4. TEST CALCULATIONS

The MP2 correlation energies of a set of 24 molecules were calculated and partitioned using the two alternative methods described above. Table 1 gives the monatomic contributions E_{MP2}^{AA} (eq 10) and E_{MP2}^{AA} (eq 12) to the correlation energies, while Table 2 contains the diatomic contributions E_{MP2}^{AB} and E_{MP2}^{AB} of bonded atoms. The contributions of nonbonded atoms will be discussed at a later stage. For both partitioning schemes, the total correlation energies are reconstructed from eq 9 with an accuracy of 0.01 kcal/mol.

As seen from Table 1, the monatomic values depend strongly on the atom type, getting more negative in the order $H < B < C < N < F < O < Cl$. For a given type of atom, the energies are influenced by the bonding with the neighboring atoms. For instance, the correlation energy of a hydrogen atom becomes less negative for $H(B) > H(C) > H(Cl) > H(N) > H(O) > H(F)$, suggesting that the absolute value of the correlation energy decreases with increasing polarity of the bond. The hybridization

Table 1. Monoatomic Energies E_{MP2}^{AA} and E'_{MP2}^{AA} (eqs 10 and 12, respectively) Calculated Using the Aug-cc-pVDZ Basis Set^a

	H		B		N		C		O		X	
	E_{MP2}^{AA}	E'_{MP2}^{AA}	E_{MP2}^{AA}	E'_{MP2}^{AA}	E_{MP2}^{AA}	E'_{MP2}^{AA}	E_{MP2}^{AA}	E'_{MP2}^{AA}	E_{MP2}^{AA}	E'_{MP2}^{AA}	E_{MP2}^{AA}	E'_{MP2}^{AA}
BH ₃	-7.90	-7.46	-20.75	-19.72								
B ₂ H ₆	-7.44 (2B)	-7.06 (2B)	-26.46	-26.76								
	-7.95 (1B)	-7.49 (1B)										
CH ₄	-5.79	-5.34					-58.94	-57.58				
C ₂ H ₆	-6.30	-5.84					-59.10	-57.58				
C ₂ H ₄	-6.10	-5.63					-60.69	-59.15				
C ₂ H ₂	-5.23	-4.77					-63.44	-67.81				
C ₆ H ₆	-6.71	-6.22					-65.24	-63.51				
H ₂ O	-2.63	-2.36							-124.13	-123.62		
CH ₃ OH	-6.55(C)	-6.08(C)					-59.79	-52.27	-125.75	-124.97		
	-6.34(C...O)	-5.87(C...O)										
	-3.17(O)	-2.87(O)										
CH ₃ OCH ₃	-6.76(C)	-6.29(C)					-55.63	-54.09	-128.51	-127.51		
	-6.43(C...O)	-5.96(C...O)										
H ₂ CO	-6.68	-6.21					-48.11	-46.62	-127.54	-126.67		
HCO ₂ H	-7.01(C)	-6.52(C)					-43.01	-41.58	-129.78	-129.01		
	-2.75(O)	-2.49(O)							-127.39(H)	-126.67(H)		
CO ₂							-37.63	-36.23	-129.61	-128.88		
CO							-45.79	-44.84	-125.48	-124.54		
NH ₃	-3.74	-3.40			-97.85	-96.95						
CH ₃ NH ₂	-6.44(C...N)	-5.96(C...N)			-98.68	-97.57	-55.89	-54.33				
	-6.70(C)	-6.22(C)										
	-4.62(N)	-3.89(N)										
N ₂ H ₄	-4.40	-4.03			-96.01	-94.82						
<i>cis</i> -N ₂ H ₂	-5.10	-4.68			-92.80	-91.50						
<i>trans</i> -N ₂ H ₂	-4.87	-4.46			-92.88	-91.59						
N ₂					-87.88	-86.53						
HF	-1.90	-1.69									-134.95	-134.74
HCl	-5.11	-4.69									-92.67	-92.24
CH ₃ F	-6.21	-5.75					-51.92	-50.49			-137.71	-137.29
CH ₃ Cl	-6.07	-5.62					-59.75	-58.27			-94.19	-93.64

^a All values are in kcal/mol. H(X) denotes a hydrogen atom bonded with atom X. H(X...Y) denotes a hydrogen atom bonded with atom X and sterical interaction with atom Y.

state of the atoms, the hybridization of the neighboring atoms, and the weak interactions also influence the correlation energy, although to a smaller extent. For example, the correlation energy of H increases on going from ethane to acetylene by 1.07 kcal/mol but decreases on going from NH₃ to *cis*-N₂H₂ by 1.44 kcal/mol, when calculated using eq 10. Similar trends are observed for the heavier atoms. For example, the absolute value of the monoatomic correlation energy of carbon increases from ethane to acetylene but decreases when going from CH₃OCH₃ to CO₂. These variations are found to correlate strongly for the H atoms with the in the atomic charges calculated with the iterative Hirshfeld method from the MP2 density, as shown in Figure 1, where all hydrogen atoms present the molecules in Table 1 where included. For heavier atoms, the general trend persists but is less pronounced: A value of $R = 0.87$ is found for C atoms and an even lower value for O and N. This is hardly surprising since one cannot expect a straightforward connection between a nonlocal property as correlation and a local property of condensed density. The correlation found for the H atoms can be explained by the fact that the E_{MP2}^H property can be seen as a special case

among the monoatomic correlation energies, since it describes the correlation energy of an atom containing only one electron. This “self-correlating” property can be rationalized as follows: Consider the symmetric H₂ molecule described by a minimal basis set, such that there is only one doubly occupied molecular orbital i and one virtual orbital a . The monoatomic correlation energy of atom H₁ is then given by:

$$E_{MP2}^H = \frac{(ia|ia)^{H_1}(ia|ia)^{H_1}}{2(\epsilon_i - \epsilon_a)} \quad (19)$$

Since both of the atoms are symmetric, the weight function of both atoms is identical, and the orbitals are equally delocalized over both atoms, so $E_{MP2}^{H_1} = E_{MP2}^{H_2}$ and, due to the factor 2 in the definition in eq 10, $E_{MP2}^{H_1} = 1/2E_{MP2}^{H_1 H_2}$. Now consider a hypothetical molecule HX, still in the framework of the minimal basis set, where X still contains only one electron but is more electronegative. Then the orbital i is more localized on X. Evidently, E_{MP2}^H will decrease and so will the atomic population of H. Thus, an increase in electronegativity of neighboring atoms causes a decrease of the

Table 2. Diatomic Energies E_{MP2}^{AB} and E'_{MP2}^{AB} (eqs 10 and 12, respectively) Calculated Using the Aug-cc-pVDZ Basis Set^a

	BH		CH		CC		NH		NN	
	E_{MP2}^{AB}	E'_{MP2}^{AB}	E_{MP2}^{AB}	E'_{MP2}^{AB}	E_{MP2}^{AB}	E'_{MP2}^{AB}	E_{MP2}^{AB}	E'_{MP2}^{AB}	E_{MP2}^{AB}	E'_{MP2}^{AB}
BH ₃	-5.27	-5.96								
B ₂ H ₆	-5.34(1B)	-5.97(1B)								
	-3.90(2B)	-4.15(1B)								
CH ₄			-6.05	-6.73						
C ₂ H ₆			-5.62	-6.30	-7.74	-8.49				
C ₂ H ₄			-5.46	-6.15	-13.65	-15.03				
C ₂ H ₂			-5.01	-5.67	-22.87	-25.21				
C ₆ H ₆			-4.84	-5.53	-9.42	-10.42				
CH ₃ OH			-5.21(C)	-5.94(C)						
			-5.06(C...O)	-5.78(C...O)						
CH ₃ OCH ₃			-5.12(C)	-5.85						
			-5.02(C...O)	-5.73(C...O)						
H ₂ CO			-4.81	-5.57						
HCO ₂ H			-3.97	-4.74						
NH ₃							-5.84	-6.44		
CH ₃ NH ₂			-5.28(C)	-6.01(C)			-5.29	-5.91		
			-5.15(C...N)	-5.85(C...N)						
N ₂ H ₄							-4.96	-5.61	-6.98	-7.97
<i>cis</i> -N ₂ H ₂							-4.71	-5.40	-15.95	-17.71
<i>trans</i> -N ₂ H ₂							-4.55	-5.22	-16.25	-18.04
N ₂									-32.79	-35.48
CH ₃ F			-5.19	-5.92						
CH ₃ Cl			-5.16	-5.86						

	CN		OH		OC		XH		XC	
	E_{MP2}^{AA}	E'_{MP2}^{AA}	E_{MP2}^{AA}	E'_{MP2}^{AA}	E_{MP2}^{AA}	E'_{MP2}^{AA}	E_{MP2}^{AA}	E'_{MP2}^{AA}	E_{MP2}^{AA}	E'_{MP2}^{AA}
H ₂ O			-5.04	-5.55						
CH ₃ OH			-4.30	-4.84	-6.15	-6.95				
CH ₃ OCH ₃					-4.74	-5.54				
H ₂ CO					-13.69	-15.14				
HCO ₂ H			-3.90	-4.35	-6.44	-7.22				
					-11.48	-12.76				
CO ₂					-12.70	-14.09				
CO					-22.00	-23.89				
CH ₃ NH ₂	-7.60	-8.46								
HF							-4.34	-4.76		
HCl							-6.73	-7.57		
CH ₃ F									-4.58	-5.24
CH ₃ Cl									-7.44	-8.32

^a All values are in kcal/mol.

monatomic correlation energy. Since this picture is too simplistic for heavier atoms, the correlation factor is lower in this case. For carbon atoms, a reasonably strong correlation of $R = 0.96$ (Figure 2) is found between the monatomic correlation energies and the change in atomic population due to correlation, defined as

$$N_A^{corr} = \int w_A(\mathbf{r})(\rho^{MP2}(\mathbf{r}) - \rho^{SCF}(\mathbf{r}))d\mathbf{r} \quad (20)$$

The diatomic correlation energy contributions depend strongly on the bond multiplicity. The correlation energy varies between 4 and 8 kcal/mol for a single bond, 13 to 18 kcal/mol for

a double bond, and 22 to 25 kcal/mol for a triple bond (or even more for the N₂ molecule) for all the atoms in the examined set of molecules. The value obtained for the C–C bond in benzene is quite encouraging, since it is situated between the value for ethane and ethylene.

The dependence on bond multiplicity can be explained by the energy denominator in eqs 10 and 12. When passing from a single to a double to a triple bond, the diatomic component increases strongly due to a much smaller gap between the π -orbitals compared to that between σ -orbitals. Within the same bond type and multiplicity, the values are influenced by the

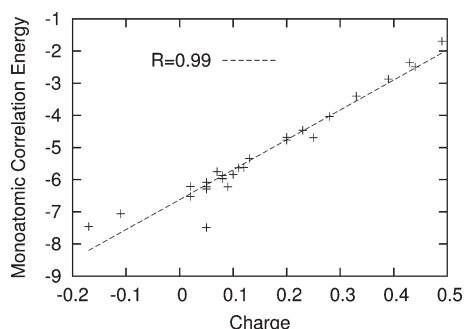


Figure 1. The correlation between the Hirshfeld atomic charge (in au) and the monoatomic MP2 correlation energy (in kcal/mol) $E_{\text{MP2}}^{\text{AA}}$ (eq 11) of the H atoms.

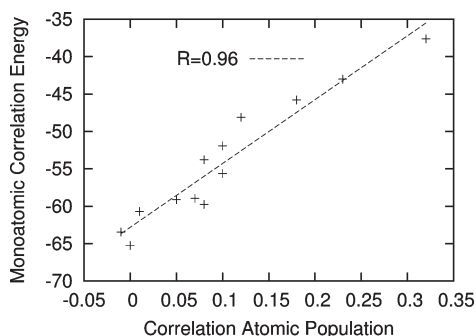


Figure 2. The correlation between the change in the atomic population due to correlation $N_{\text{A}}^{\text{corr}}$ (eq 20) (in au) and the monoatomic MP2 correlation energy (in kcal/mol) $E_{\text{MP2}}^{\text{AA}}$ (eq 11) of the C atoms.

chemical surrounding. For instance, the absolute values for the C–H bond decrease on going from ethane to acetylene, with the absolute values for methane being larger than for ethane and for benzene lower than for acetylene.

The values calculated using eqs 10 and 12 differ by about 0.5 kcal/mol for the mono- and diatomic energies involving the H atom and up to 3 kcal/mol for correlation energies of the heavier atoms. In general, the monoatomic energies $E_{\text{MP2}}^{\text{AA}}$ are slightly less negative than $E_{\text{MP2}}^{\text{AA}}$. Correspondingly, the diatomic energies $E_{\text{MP2}}^{\text{AB}}$ are larger in absolute values than $E_{\text{MP2}}^{\text{AB}}$. Despite these small numerical differences, the trends discussed above are equivalent in both partitioning schemes.

The basis set dependence of the method is summarized in Table 3. Only the mono- and diatomic values calculated using eq 12 are shown, as the basis set dependence of both partitioning schemes is similar. Table 3 also reports the total correlation energy for each basis set. One can see that the MP2 correlation energy is significantly undermined when polarization functions are excluded, while the addition of diffuse functions has only a minor effect. The difference between the Pople double- and triple- ζ basis sets amounts to 20% of the total correlation energy calculated with the triple- ζ basis set, while the values obtained with the aug-cc-pVDZ basis set are situated in between. Comparing the aug-cc-pVDZ values with the values obtained with the largest Pople basis set shows that while the monoatomic energies of the former are slightly higher, the diatomic energies are lower. This effect is especially pronounced for the energies of the C, N, and O atoms. Overall, the addition of polarization functions appears to result in the most significant improvement.

In order to compare our scheme with other partitioning techniques, in Table 3 we give the values obtained using the Ayala–Scuseria¹⁴ method, calculated with the 6-31G* basis set. These results indicate that not only the numerical values of the mono- and diatomic energies differ but also, in some cases, the trends as well. For example, the relation between the diatomic correlation energy and the bond multiplicity, as discussed above, is not present in the Ayala–Scuseria method.

In most molecules examined here, the sum of the monoatomic energies and diatomic energies of the bonding interactions comprises more than 95% of the total correlation energy. The diatomic contributions between two nonbonded atoms are limited, in absolute values, to less than 0.05 kcal/mol between two H atoms, 0.5 kcal/mol for an H atom interaction with a heavier atom, and 1.0 kcal/mol between two heavier atoms. Three exceptions to this rule are present in the 24 molecules studied here, which concern 2 nonbonding interactions between O atoms in HCO₂H and CO₂ and a nonbonding interaction between B atoms in diborane. The diatomic correlation energies for these three pairs are summarized in Table 4, calculated using various basis sets. While $E_{\text{MP2}}^{\text{OO}}$ for the O···O interaction in both molecules decreases in absolute value with increasing the basis set and may therefore be expected to become less significant at the complete basis set limit, the situation is reversed for the B···B interaction. This is probably related to an exchange interaction, which also emerges indirectly in bond orders between nonbonded X···Z atoms in the case of three-center two-electron X–Y–Z bonds.³¹

5. INTERACTION ENERGY OF BENZENE DIMERS

In the previous section we have demonstrated that both schemes (eqs 10 and 12) provide chemically meaningful mono- and diatomic correlation energies. In this section we will focus on applying the method to benzene dimers. The energetics of this system is largely determined by dispersion interaction, which cannot be correctly described neither by Hartree–Fock nor by most density functionals. As mentioned in Section 2, the scheme based on eq 10 is theoretically more sound. However, since the scheme based on eq 12 is much more easily parallelizable, and in light of the quality of the results obtained in the previous section, the CPU-intensive calculations on the benzene dimers were carried out within the second scheme.

The structures of three benzene dimers in the sandwich, parallel displaced, and T-shaped configurations are shown in Figure 3. The geometries optimized at QCISD(T)/aug-cc-pVTZ level were taken from ref 30. The calculations of the MP2 correlation energy were performed using the 6-31++G** basis set. Since the MP2 method overestimates dispersion interaction in the benzene dimer³² and the 6-31++G** basis set is not large enough to fully reproduce the correlation energy, the results only represent a qualitative picture of the interaction energy of these dimers. This qualitative picture is, however, sufficient for our goal, since the chemical information that can be obtained from a partitioning method is confined in the trends, rather than the exact values. For each of the geometries, the total MP2 energy was calculated for the dimer and for each of the monomers using the BSSE correction. Since Tables 1 and 2 already report values of the total mono- and diatomic correlation energies of benzene, we will discuss only the contribution of correlation energy to the interaction energy of the dimers. Table 5 lists the interaction energy at Hartree–Fock and MP2 levels as well as the contributions of the monoatomic correlation energies (E^{C} and E^{H}), the intramolecular diatomic correlation

Table 3. Mono- and Diatomic Correlation Energies Calculated Using eq 12 with Various Basis Sets^a

	6-31G	6-31G**	6-31++G**	6-311	6-311G**	6-311++G**	aug-cc-pVDZ	Ayala–Scuseria
CH ₄								
C	-38.59	-60.14	-59.86	-50.69	-70.42	-69.65	-57.58	-55.42
H	-3.25	-4.93	-5.15	-3.45	-5.2	-5.45	-5.34	-5.15
CH	-2.74	-6.13	-6.12	-2.83	-6.56	-6.53	-6.73	-6.84
total	-63.09	-105.56	-106.27	-76.40	-118.84	-119.04	-107.56	-105.56
C ₂ H ₆								
C	-39.46	-59.59	-60.01	-57.78	-69.93	-69.97	-57.58	-55.32
H	-3.72	-5.62	-5.63	-3.93	-5.87	-5.93	-5.84	-6.93
CC	-2.37	-7.19	-7.32	-2.65	-8.38	-8.25	-8.49	-9.99
CH	-2.55	-5.87	-5.87	-2.65	-6.20	-6.19	-6.30	-6.93
total	-119.87	-198.03	-199.50	-147.00	-224.16	-224.58	-200.73	-198.04
C ₂ H ₄								
C	-41.00	-60.56	-61.27	-53.73	-71.15	-71.50	-59.15	-61.23
H	-3.67	-5.43	-5.43	-3.86	-5.65	-5.69	-5.63	-4.95
CC	-7.89	-14.14	-13.88	-7.95	-14.90	-14.65	-15.03	-8.22
CH	-2.38	-5.69	-5.66	-2.48	-6.00	-5.97	-6.15	-7.04
total	-115.59	-181.85	-183.05	-142.43	-206.19	-206.74	-183.15	-181.85
NH ₃								
N	-60.17	-93.43	-95.33	-74.97	-108.85	-109.71	-96.95	-85.55
H	-2.01	-3.06	-3.25	-2.10	-3.11	-3.40	-3.40	-4.38
NH	-2.46	-5.79	-5.78	-2.65	-6.06	-6.01	-6.44	-6.89
total	-73.86	-120.55	-123.03	-89.51	-136.96	-138.57	-127.24	-120.55
N ₂ H ₄								
N	-61.71	-91.74	-93.73	-76.12	-106.49	-107.92	-94.82	-87.39
H	-2.53	-3.87	-3.83	-2.68	-3.95	-3.98	-4.03	-4.46
NN	-1.08	-6.81	-7.00	-1.34	-7.37	-7.21	-7.97	-7.78
NH	-1.99	-5.22	-5.12	-2.17	-5.40	-5.29	-5.61	-6.34
total	-143.21	-228.48	-232.46	-173.71	-259.84	-262.46	-238.97	-228.47
HCO ₂ H								
H(C)	-4.28	-6.56	-6.37	-4.52	-6.77	-6.60	-6.52	-5.12
H(O)	-1.72	-2.49	-2.36	-1.77	-2.48	-2.43	-2.49	-3.69
C	-29.41	-43.65	-44.28	-41.00	-52.04	-52.67	-41.58	-49.64
O	-85.57	-117.58	-120.42	-102.95	-140.48	-142.39	-129.06	-118.69
O(H)	-81.76	-115.84	-118.29	-98.66	-139.36	-140.97	-126.67	-111.07
CO(1)	-1.61	-6.64	-6.81	-1.64	-6.65	-6.68	-7.22	-6.07
CO(2)	-6.96	-12.45	-12.10	-6.81	-12.44	-12.10	-12.76	-8.43
CH	-1.96	-4.82	-4.54	-1.92	-4.83	-4.60	-4.74	-5.98
OH	-1.30	-4.05	-3.99	-1.47	-4.03	-3.95	-4.35	-5.82
total	-217.05	-316.77	-322.15	-263.35	-371.85	-375.38	-338.79	-316.77

^aThe rightmost column lists the values obtained with the Ayala–Scuseria (ref 14) method with the 6-31G** basis set.

energies ($E_{\text{intra}}^{\text{CC}}$, $E_{\text{intra}}^{\text{HH}}$ and $E_{\text{intra}}^{\text{CH}}$) and the intermolecular diatomic correlation energies ($E_{\text{inter}}^{\text{CC}}$, $E_{\text{inter}}^{\text{HH}}$ and $E_{\text{inter}}^{\text{CH}}$) to the interaction energies. These quantities are defined as follows

$$E^{\text{A}} = E_{\text{MP2}}^{\text{AA}}(\text{dimer}) - E_{\text{MP2}}^{\text{AA}}(\text{monomer}) \quad (21)$$

$$E_{\text{intra}}^{\text{AB}} = E_{\text{MP2}}^{\text{AB}}(\text{dimer}) - E_{\text{MP2}}^{\text{AB}}(\text{monomer}) \quad (22)$$

$$E_{\text{inter}}^{\text{AB}} = E_{\text{MP2}}^{\text{AB}}(\text{dimer}) \quad (23)$$

The calculations were performed according to eq 12.

As seen from Table 5, the correlation energy is essential for correct description of the bonding in the benzene dimers. The interaction energies of all three configurations are positive when calculated using the Hartree–Fock method. The MP2 method predicts the parallel-displaced configuration to be the most stable with an interaction energy of -2.24 kcal/mol, followed by the sandwich configuration with an interaction energy of -2.02 kcal/mol, and the T-shaped configuration with an interaction energy of -1.43 kcal/mol. This order is in agreement with the QCISD-(T)/aug-cc-pVTZ interaction energy reported by Janowski et al.,³⁰ although our MP2/6-31++G** values are somewhat

Table 4. Significant Diatomic Correlation Energies (in kcal/mol) between Nonbonding Atoms Calculated Using eqs 10 and 12

	6-31G	6-31G**	6-31++G**	6-311	6-311G**	6-311++G**	aug-cc-pVDZ
OO(HCO ₂ H)	-1.69(-1.79)	-1.33(-1.41)	-1.44(-1.55)	-1.79(-1.70)	-1.36(-1.43)	-1.45(-1.55)	-1.61(-1.73)
OO(CO ₂)	-6.66(-6.34)	-4.03(-4.05)	-3.94(-3.99)	-6.59(-6.27)	-4.02(-4.05)	-3.93(-3.97)	-3.83(-3.90)
BB(B ₂ H ₆)	-1.93(-2.19)	-3.70(-4.15)	-3.90(-4.37)	-1.97(-2.28)	-4.02(-4.16)	-4.18(-4.68)	-4.11(-4.59)

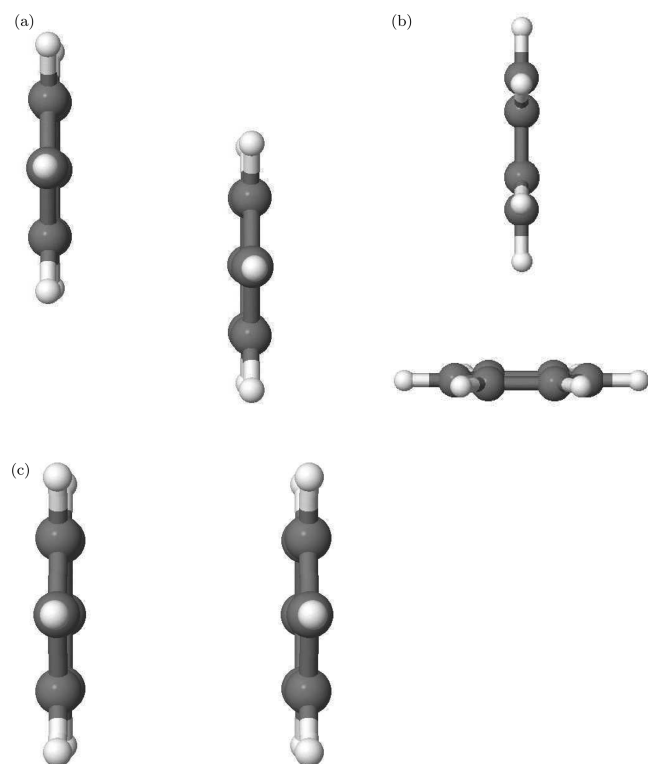


Figure 3. Three benzene dimers used in this work.

underestimated due to a limited basis set size. The MP2 correlation energy is found to be the highest in the parallel displaced dimer and the lowest in the T-shaped dimer, despite the fact that the largest overlap between the π -orbitals is present in the sandwich dimer.

The partitioned correlation energy reveals that due to dispersion interaction in the dimer, the intramolecular diatomic contributions are partially shifted to the monatomic contributions. For example, in the parallel-displaced dimer, the monatomic correlation energies increase by up to 11.4 kcal/mol per benzene moiety, while the diatomic correlation energies within a benzene moiety decrease by up to 6.7 kcal/mol. The net change in correlation energy within a moiety is negative, amounting to an average 80% of the total correlation energy contribution to the interaction energy. The rest of the correlation energy is assigned to the intermolecular diatomic correlation energies. The largest values are found for the C–C interactions in the parallel displaced and sandwich dimers and the C–H interactions in the T-shaped dimer.

Since the physical phenomenon responsible for the negative interaction energy of the benzene dimers is the dispersion interaction between the electronic densities of the two benzene moieties, we are especially interested in the distribution of the pairwise interactions from which these intermolecular correlation

Table 5. Interaction Energies of Three Benzene Dimers Calculated at the Hartree–Fock (E^{HF}) and MP2 (E^{MP2}) Levels of Theory with the 6-31++G** Basis Set and the Contributions of the Monoatomic Correlation Energies (E^{C} and E^{H}), the Intra-molecular Diatomic Correlation Energies ($E_{\text{intra}}^{\text{CC}}$, $E_{\text{intra}}^{\text{HH}}$, and $E_{\text{intra}}^{\text{CH}}$), and the Intermolecular Diatomic Correlation Energies ($E_{\text{inter}}^{\text{CC}}$, $E_{\text{inter}}^{\text{HH}}$, and $E_{\text{inter}}^{\text{CH}}$) to the Interaction Energies^a

	parallel displaced	sandwich	T-shaped
E^{HF}	3.69	3.51	1.02
E^{MP2}	-2.24	-1.43	-2.02
E^{corr}	-5.94	-4.95	-3.04
E^{C}	-9.59	-8.08	-5.55
E^{H}	-1.84	-1.58	-1.56
sum	-11.44	-9.67	-7.12
$E_{\text{intra}}^{\text{CC}}$	4.44	3.69	3.10
$E_{\text{intra}}^{\text{HH}}$	0.03	0.02	0.02
$E_{\text{intra}}^{\text{CH}}$	2.20	1.77	1.69
sum	6.67	5.48	4.81
$E_{\text{inter}}^{\text{CC}}$	-0.74	-0.53	0.05
$E_{\text{inter}}^{\text{HH}}$	-0.06	-0.03	-0.07
$E_{\text{inter}}^{\text{CH}}$	-0.39	-0.22	-0.73
sum	-1.19	-0.78	-0.74

^a Calculated with eqs 21–23. All values are in kcal/mol.

energy contributions consist. These values contain the most important information on the nature of the intermolecular dispersion interaction. The small values are in agreement with the recent findings of Contreras-García et al.³³ that the region of the noncovalent dispersion interaction was found to lie between the two monomers, where the electron density ρ is smaller than 0.02 au and the reduced gradient $|\nabla\rho| \cdot \rho^{-4/3}$ is greater than 0.7 au. As discussed in Section 4, small values of electron density inevitably lead to relatively small intermolecular diatomic correlation energy contributions.

The sum of the intermolecular correlation energy contributions is the largest in the parallel-displaced dimer, which is the most stable dimer of the three configurations examined here. Figure 4 illustrates the major pairwise interactions between the three symmetrically nonequivalent carbon and hydrogen atoms. The strongest intermolecular interactions (solid lines on Figure 4) total to -0.06 kcal/mol and occur between carbon atoms separated by 3.6 Å. Interactions of -0.03 kcal/mol (dotted lines on Figure 4) which occur between carbon atoms separated by an average of 3.9 Å and between carbon and hydrogen atoms separated by an average 3.5 Å. The weakest interactions shown here are -0.02 kcal/mol and occur between carbon atoms separated by 4.2 Å. In the sandwich dimer, where one has only one symmetrically nonequivalent carbon and hydrogen atoms, the strongest intermolecular interactions of -0.03 kcal/mol are

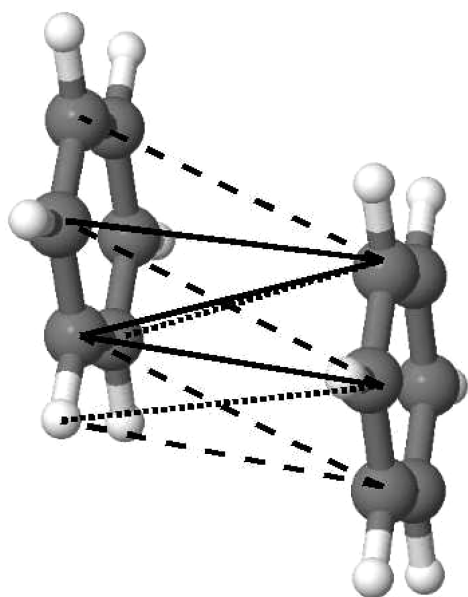


Figure 4. Main intermolecular pairwise interactions $E_{\text{inter}}^{\text{AB}}$ (eq 23) in the parallel-displaced benzene dimer. Strongest interactions (≥ -0.06 kcal/mol) are represented by solid lines, intermediate interactions (-0.03 kcal/mol) are dotted lines, and the weakest interactions (< -0.02 kcal/mol) are shown by dashed lines.

found between two nearest carbon–carbon atoms, which are separated by 3.9 \AA . The second strongest interaction of -0.02 kcal/mol is found between a carbon atom and its two neighbors, separated by 4.2 \AA . The intermolecular interactions between the carbon and hydrogen atoms are all smaller than -0.01 kcal/mol, and as in the case of parallel-displaced dimer, no significant interactions are found between the hydrogen atoms. In the T-shaped, most of the intermolecular interaction energy can be attributed to the six C–H interactions between the H atom of the upper benzene monomer in Figure 2, which is closest to the lower benzene monomer, and the six C atoms in the lower monomer. Each of these six interactions amounts to -0.1 kcal/mol, almost twice as strong as the strongest C–C interaction in the parallel-displaced monomer. The strong interaction is in accordance with the short C–H distance, which amounts to 2.8 \AA . The clear connection between the distance and the intermolecular diatomic interactions is in correspondence with the known R^{-6} dependence of the dispersion interaction.

6. SUMMARY AND CONCLUSIONS

In conclusion, we have proposed a method for partitioning of MP2 correlation energy using the Hirshfeld method. It can be used either on the top of an existing Hartree–Fock energy decomposition or on its own right. Two alternative partitioning schemes were presented, with one of the schemes having a better theoretical foundation but the other allowing a better degree of parallelization. In spite of different formulations, both schemes produce only slightly different numerical results but show the same trends for diatomic and monatomic energies. The correlation energy partitioning can be either performed on its own right or done as a complement to an existing Hartree–Fock energy partitioning, which results in a partitioning of the total correlated energy.

The basis set dependence of the partitioning method is similar to that of the MP2 method itself: A limited basis set was found to undermine the energies, while including polarization functions leads to a major improvement. The monatomic correlation energies were found to depend strongly on the type of the atom. For hydrogen, the monatomic correlation energy is in line with the atomic charge. This does not hold for heavier atoms. The diatomic energies for bonded atoms depend mostly on the bond multiplicity and reflect the strength of the bond between the two atoms. The diatomic contributions of non-bonding interatomic interactions are found to be negligible in most cases, with exception of systems where exchange interactions between nonbonding atoms are present.

Finally, the method was applied for the analysis of the interaction energies of benzene dimers. The intermolecular interactions were found to be determined mainly by the distance between the atoms, with C–C interactions being stronger than C–H interactions for equal distance and H–H interactions being always negligible.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: alisa.krishtal@ua.ac.be; sergey.vy@catalonia.net; kris.vanalsenoy@ua.ac.be.

ACKNOWLEDGMENT

The work was performed under the HPC-EUROPA2 project (project number: 228398) with the support of the European Commission (Capacities Area – Research Infrastructures) using the computer facilities of the Barcelona Supercomputer Center. We also thank the Spanish Ministerio de Ciencia e Innovación (grant CTQ2008-03077/BQU) for financial support. A.K. acknowledges The Research Foundation – Flanders (FWO) for a postdoctoral position.

REFERENCES

- (1) Mayer, I. *Chem. Phys. Lett.* **2000**, *332*, 381.
- (2) Hamza, A.; Mayer, I. *Theor. Chim. Acta* **2003**, *109*, 91.
- (3) Salvador, P.; Mayer, I. *J. Chem. Phys.* **2004**, *120*, 5046.
- (4) Mayer, I. *Chem. Phys. Lett.* **2003**, *382*, 265.
- (5) Ichikawa, H.; Yoshida, A. *Int. J. Quantum Chem.* **1999**, *71*, 35.
- (6) Kollmar, H. *Theoret. Chim. Acta* **1978**, *50*, 235.
- (7) Nakai, H.; Kikuchi, Y. *J. Theor. Comput. Chem.* **2005**, *4*, 317.
- (8) Vyboishchikov, S. F. *Int. J. Quantum Chem.* **2008**, *108*, 708.
- (9) Mandado, M.; Van Alsenoy, C.; Geerlings, P.; De Proft, F.; Mosquera, R. A. *Chem. Phys. Chem.* **2006**, *7*, 1294.
- (10) Imamura, Y.; Takahashi, A.; Nakai, H. *J. Chem. Phys.* **2007**, *126*, 034103.
- (11) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 0618.
- (12) Bartlett, R. J.; Purvis, G. D., III. *Int. J. Quantum Chem.* **1978**, *14*, 561.
- (13) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. *Int. J. Quantum Chem.* **1978**, *14*, 545.
- (14) Ayala, P. Y.; Scuseria, G. E. *Chem. Phys. Lett.* **2000**, *322*, 213.
- (15) Imamura, Y.; Baba, T.; Nakai, H. *Int. J. Quantum Chem.* **2008**, *108*, 1316.
- (16) Mayer, I. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4630.
- (17) Vyboishchikov, S. F.; Salvador, P. *Chem. Phys. Lett.* **2006**, *430*, 204.
- (18) Imamura, Y.; Nakai, H. *J. Comput. Chem.* **2008**, *29*, 1555.
- (19) Alcoba, D. R.; Torre, A.; Lain, L.; Bochicchio, R. C. *J. Chem. Phys.* **2005**, *122*, 074102.

- (20) Alcoba, D. R.; Torre, A.; Lain, L.; Bochicchio, R. C. *J. Chem. Phys.* **2007**, *127*, 104110.
- (21) Alcoba, D. R.; Oña, O.; Torre, A.; Lain, L.; Bochicchio, R. C. *J. Phys. Chem. A* **2008**, *112*, 10023.
- (22) Kikuchi, Y.; Imamura, Y.; Nakai, H. *Int. J. Quantum Chem.* **2009**, *109*, 2464.
- (23) Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129.
- (24) Rousseau, B.; Peeters, A.; Van Alsenoy, C. *Chem. Phys. Lett.* **2000**, *324*, 189.
- (25) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. *J. Chem. Phys.* **2007**, *126*, 144111.
- (26) Krishtal, A.; Senet, P.; Van Alsenoy, C. *Synth. React. Inorg., Met.-Org., Nano-Met. Chem.* **2008**, *38*, 84.
- (27) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 2547.
- (28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.05; Gaussian, Inc.: Wallingford, CT, 2004.
- (29) Molawi, K.; Cohen, A. J.; Handy, N. C. *Int. J. Quantum Chem.* **2002**, *89*, 86.
- (30) Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27.
- (31) Mayer, I. *Simple theorems, proofs, and derivations in quantum chemistry*; Springer: New York, 2003; pp 245–247.
- (32) Pitoňák, M.; Neogrady, P.; Rezáč, J.; Jurečka, P.; Urban, M.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 1829.
- (33) Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W. *J. Chem. Theory Comput.* **2011**, *7*, 625.

The Accuracy of Density Functional Theory in the Description of Cation– π and π –Hydrogen Bond Interactions

Ana Rute Neves, Pedro Alexandrino Fernandes, and Maria João Ramos*

REQUIMTE, Faculdade de Ciências do Porto, Rua do Campo Alegre S/N, 4169-007, Porto, Portugal

ABSTRACT: Cation– π and π –hydrogen bond interactions are ubiquitous in protein folding, molecular recognition, and ligand–receptor associations. As such systems are routinely studied at the DFT level, it becomes essential to understand the underlying accuracy of the plethora of density functionals currently available for the description of these interactions. For that purpose, we carried out theoretical calculations on two small model systems (benzene–Na⁺ and benzene–H₂O) that represent a paradigm for those intermolecular interactions and systematically tested 46 density functionals against the results of high-level post-HF methods, ranging from MP2 to extrapolated CCSD(T)/CBS. A total of 13 basis sets were also tested to examine the convergence of the interaction energy with basis set size. The convergence was surprisingly fast, with deviations below 0.2 kcal/mol for double- ζ polarized basis sets with diffuse functions. Concerning functional benchmarking, the Truhlar group functionals were particularly well suited for the description of the π –hydrogen bond interactions. In the case of cation– π interactions, there was not a clear correlation between accuracy and functional sophistication. Despite the large number of functionals predicting interaction energies within chemical accuracy (five for π –hydrogen bond and 20 for cation– π interactions), not a single functional has shown chemical accuracy in both cases. Moreover, if we calculate the average error for these two interactions, only two density functionals resulted in an average error below 1.0 kcal/mol (M06 and HCTH, with average errors of 0.6 and 0.8 kcal/mol). The obtained results serve as a guide for future computer simulations on this kind of system.

INTRODUCTION

Cation– π and π –hydrogen bond (π -Hbond) interactions are noncovalent molecular interactions between an electron- π -rich system and an adjacent cation or a hydrogen bond donor.

Cation– π interactions between amino acids contribute significantly to protein folding, molecular recognition, and drug-receptor interactions.¹ They are frequently involved in key interactions at protein–protein interfaces, and they might participate in molecular recognition patterns at the active sites of enzymes or receptors rich in aromatic residues (Phe, Tyr, or Trp). In proteins, cation– π interactions can arise between aromatic residues (Phe/Tyr/Trp) as the π component and positively charged amino acids (Lys, Arg, His) as the cation. Although Phe, Tyr, and Trp comprise 9% of the natural amino acids, they are substantially overrepresented at binding sites. One of the reasons for this is their capacity for establishing cation– π interactions.² In fact, on average there is one cation– π interaction for every 77 amino acids in the protein data bank. As a result, essentially all proteins of significant size have at least one cation– π interaction. Over 25% of all Trp residues are involved in cation– π interactions with Lys or Arg,^{2,3} Arg being the most frequent cation.⁴

A large number of examples of protein–ligand associations in which these interactions take place have been described. For instance, the cation– π interaction is exploited for neurotransmitter recognition throughout the nervous system, since receptors are rich in aromatic amino acids and make use of cation– π interactions to bind their ligands, such as acetylcholine, GABA (γ -aminobutyric acid), and serotonin.^{2,4} Cation– π interactions are also responsible for the functioning and selectivity in ion channels.³ The crystal structure of the K⁺ channel shows that the mouth of the extracellular entrance is composed of the aromatic

rings of four conserved tyrosines, which enable the entrance of K⁺ ions into the pocket.⁴ Another example is the nucleosome remodeling factor BPTF, whose aromatic residues surround the Lys amino acids of histones, leaving no doubt that cation– π interactions are important in this binding.²

These interactions are also present in DNA and RNA, as both purine and pyrimidine bases are electron- π -rich systems commonly involved in the binding of cationic species (not only metal ions but also charged amino acid side chains) and also participate in hydrogen bonds with H-donor groups. Therefore, these interactions provide important contributions to the overall stability of enzyme and nucleotide molecules.

π -Hbond interactions are weaker than the conventional hydrogen bonds but also play important roles. This type of hydrogen bond is crucial in solvation, hydrophobic interactions, molecular recognition, protein folding, neurotransmitter conformations, crystal packing, and cluster and micelle formation.⁵ They seem to be particularly prevalent in molecules containing an indole or porphyrin moiety. Frequent π -Hbond donors include Tyr, Ser, Thr, Gln, or Asn residues.

In summary, cation– π and π -Hbond interactions are ubiquitous in biological systems, and their importance and frequent inclusion in theoretical models led us to study how well theoretical calculations at the DFT level describe such important interactions.

Density Functional Theory (DFT) has become one of the most widespread methods for calculating a variety of molecular properties.^{6–8} The main reason for the popularity of DFT is the inclusion of electron correlation without being as computationally

Received: March 10, 2011

Published: May 24, 2011

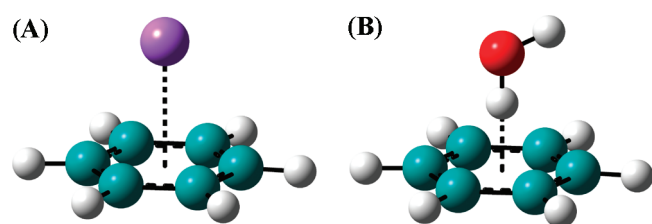


Figure 1. Models used in this study. (A) Benzene–Na⁺, which prototypically describes cation– π interactions, and (B) benzene–H₂O, which prototypically describes π –Hbond interactions. The dashed line represents the interatomic 6-fold symmetry axis that we have explored on the multidimensional potential energy surfaces.

demanding as other computational methods, such as post-Hartree–Fock methods. For this reason, DFT enables one to do calculations on molecules of over 100 atoms, which would be very difficult with other methods of comparable accuracy.^{6,9,10} Today, there is a great variety of density functionals with different levels of accuracy and computational cost, and they seem to increase day after day by the desire to improve the accuracy of the DFT methodology.¹⁰ For these reasons, it becomes very difficult to rate DFT functionals or to assess which one is better for a particular system or property. For readers who are interested in using DFT on systems in which cation– π and π –Hbond interactions play an important role, the key question is probably which functional should be used. In the case of a lack of information, B3LYP is usually used as a default (even though this procedure is quite questionable). In fact, it is difficult to recommend any DFT functional because the number is overwhelming and each method has its own strengths and weaknesses. Therefore, it is useful to identify a small set of functionals that perform well, taking into account the properties and type of system under study, as well as the availability and computational cost associated.^{6,8} The best way of choosing the most suitable functional for the system under study is being aware of new benchmarking studies.

Due to their importance, several studies have been devoted to evaluating the ability of DFT methods to describe nonbonded interactions.^{9,11–18} These studies usually concentrate mostly on dispersion interactions in general terms and not individually on π –Hbonds and cation– π interactions (with the exception of refs 17 and 18, which concentrate on π –Hbond interactions). Moreover, they analyze just a small set of density functionals and do not always address the more modern ones.

The purpose of the present article is to overcome this problem, by investigating systematically how well the plethora of current density functionals represent, in particular, cation– π and π –Hbond interactions. Their correct description is essential for any reliable computer simulation in which those interactions play an important role. The obtained results serve as a guide for future computer simulations on this kind of system.

METHODS

I. Model Systems. Figure 1 illustrates the two small models, benzene–Na⁺ and benzene–H₂O, on which we have carried out theoretical calculations to study cation– π and π –Hbond interactions, respectively.

The model systems were specifically chosen to ensure that (i) they represented well the target biological interactions, (ii) they had a minimum number of atoms, to allow a great number of

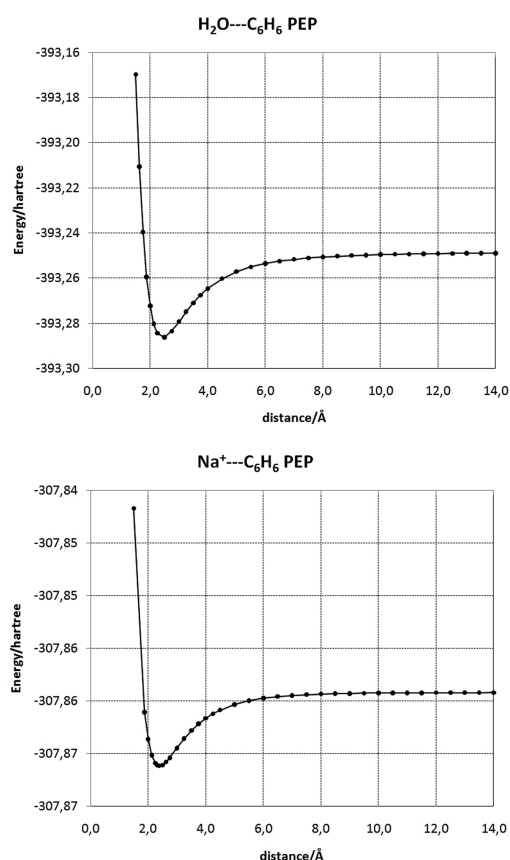


Figure 2. Potential energy profile along the 6-fold axis of benzene for the benzene–water system (top) and benzene–sodium ion system (bottom), calculated at the MP2/6-311++G(d,p)//MP2/6-311++G(d,p) level.

calculations and the use of computational demanding methods, such as CCSD(T) with large basis sets, (iii) they minimize the existence of other intermolecular interactions, which would otherwise complicate the interpretation of the results, and (iv) are as general as possible and not biased toward a specific system. The benzene molecule represents the π system. The reason for the choice of Na⁺ as the metal cation and H₂O molecule as the hydrogen bond donor group was that they represent well the most abundant biologically relevant Hbond/charge donors (Tyr, Ser, Thr/Arg, Lys, His, respectively), and they obey all of the exposed requisites.

II. Potential Energy Profile. In order to study the intermolecular interactions, we have built up a potential energy profile (PEP), by scanning different distances between the benzene molecule and the other molecular entity (Na⁺ or H₂O), as shown in Figure 2 above.

These initial calculations were performed at the MP2/6-311++G(d,p) level and were carried out using the Gaussian 03 suite of programs.¹⁹ The scanned coordinate was the benzene 6-fold symmetry axis. The sodium ion and one of the HO bonds of water were constrained to be along the axis. The purpose was to give a direction to the interaction that broadly mimics the one found in proteins. The PES above the benzene ring is very flat, and the water Hbond donor/cation can move very freely on the face of the ring. In biological systems, the position relative to the ring centroid is mostly determined by the protein structure and interactions. Thus, the minimum of this PEP is not rigorously a

Table 1. The DFT Functionals Tested in This Work

type	year	functional	ref.	type	year	functional	ref.		
GGA	1988	BLYP	31, 32	GGE	2004	TPSSVWN5	33, 34		
	1988	BP86	32, 35		M-GGA	1996	BB95	32, 37	
	1991	PW91PW91	36			1998	VSXC	38	
	1992	BPW91	32, 36			2003	TPSSTPSS	33	
	1996	BPBE	32, 39			2004	PBEKCIS	39, 41	
	1996	G96LYP	31, 40			2004	TPSSKCIS	33, 41	
	1996	PBEPBE	39			2006	M06-L	9, 43	
	1998	HCTH	42			HM-GGA	1996	B1B95	32, 37
	2001	OLYP	31, 44				2003	TPSSh	33
	2004	XLYP	31, 32, 45, 46				2004	BB1K	31, 32, 48
H-GGA	1993	B3P86	35, 47	2004			MPWB1K	36, 37, 48, 49	
	1993	B3PW91	36, 47	2004	MPW1B95		36, 37, 48, 49		
	1993	BHandH	31, 32	2004	MPW1KCIS		36, 41, 49, 50		
	1993	BHandHLYP	31, 32	2004	MPW1K1K		36, 41, 49, 50		
	1994	B3LYP	31, 47	2004	mPWKCIS		36, 41, 49		
	1996	PBE1PBE	39	2004	PBE1KCIS		12, 39, 41		
	1997	B1LYP	29, 35	2005	TPSS1KCIS		33, 41, 51		
	1998	B97-1	42	2005	M05	54			
	1998	B98	52, 53	2006	M05-2X	56			
	2000	MPW1K	36, 49, 55	2006	M06-HF	9, 58			
2001	B97-2	57	2008	M06-2X	9, 60				
2001	O3LYP	31, 44, 59	2008	M06	9, 60				
2004	X3LYP	31, 32, 45, 46							
2004	MPW3LYP	31, 36, 49							

stationary point on the free PES but is close to the structure commonly found in proteins where the residues' translational and rotational degrees of freedom are constrained by the connections to the backbone and by interactions with the remaining protein. We have also considered freely optimizing the water molecule (in which case the results would be less general and more specific for the model), but the results were almost equivalent, and the difference in energy (about 0.06 kcal/mol) was absolutely irrelevant for rating density functionals. The specific points along the 6-fold axis were chosen to give a smooth PEP with appropriate resolution near the minimum. MP2 was chosen because it provides very accurate geometries for this kind of system and because its use will not introduce any bias when comparing the density functionals, eventually favoring the functional with which the geometries were obtained, even though we must emphasize that the energy is not very sensitive to the quality of the geometry, provided that the geometry has good accuracy. At last, the two monomers were independently optimized, to provide energies at infinite separation. The different geometries along both PEPs (benzene–Na⁺ and benzene–H₂O) were then used on the following benchmarking studies.

III. Basis Set Truncation Error. In order to examine the convergence of the interaction energy with the completeness of the basis set, a total of 13 different basis sets were assessed, using the functional B3LYP. We have calculated single point energies for all of the points of both PEPs (benzene–Na⁺ and benzene–H₂O) and for the monomers separately. We have used the Pople basis sets 6-31G(d), 6-31+G(d), 6-31G(d,p), 6-31+G(d,p), 6-311G(d,p), 6-311+G(d,p), 6-311++G(d,p), 6-311++G(2d,2p), and 6-311++G(3df,3pd) and also the

correlation consistent basis sets cc-pVDZ, aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ. The last one was the most complete basis set we could afford in terms of computational demand, so this was the one considered as the reference to calculate the errors on the interaction energies obtained with the other basis sets. Even so, this was not a limitation to this study, because the truncation error of the interaction energy using the aug-cc-pVQZ basis set is very low, as we will see later in the Results section. In principle, it could be possible to extrapolate the energies of the systems (and hence the interaction energy) to the complete basis set limit (CBS) using the Dunning DZ–QZ basis set series. However, the procedure to extrapolate is not well-defined within DFT, and the convergence is not necessarily consistent, contrary to what happens with wave function methods. Moreover, the results are already so close to the CBS limit that the extrapolation is not necessary here. Counterpoise (cp) corrections for basis set superposition error (BSSE) were included in the interaction energy calculations, using the Gaussian 03 software.¹⁹

IV. Density Functional Benchmarking. In this study, we were particularly interested in assessing the performance of DFT functionals in the description of cation– π and π –Hbond intermolecular interactions. This assessment does not represent the global quality of the functionals, which must be measured through the calculation of diverse sets of properties in a representative set of molecular systems. Instead, this study just evaluates specifically the accuracy of the calculation of cation– π and π –Hbond intermolecular interactions. For this purpose, we tested the performance of 46 functionals (Table 1). The extrapolated CCSD(T)/CBS energy was also obtained, with two different extrapolation methods. For the water–benzene system,

the MP2/CBS energy was calculated by the method of Truhlar using the MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ results.²⁰ Upon addition of the CCSD(T) correction (the energy difference between CCSD(T) and MP2, both with the 6-31+G(d) basis set), we obtained the extrapolated CCSD(T)/CBS energy. This procedure takes advantage of significant cancellation of the basis set truncation error in the two methods. The result was very close to other very accurate CCSD(T)/CBS extrapolations.²¹ For Na⁺–benzene, we did not find other very high level calculations in the literature. This prompted us to increase even further the accuracy of our calculations. We have used a set of extrapolation techniques to obtain the extrapolated CCSD(T)/CBS energy. First, the extrapolated MP2/CBS energy was estimated using three methods: the method of Truhlar and the cc-pVXZ ($X = 2-3$) basis sets,²² the method of Truhlar and Zhao and the aug-cc-pVXZ ($X = 2-3$) basis sets,²⁰ and the method of Halkier et al.²³ and the aug-cc-pVXZ ($X = 3-4$) basis sets (which constituted the highest-level extrapolation of the MP2 energy in this work). Note that in this last calculation the extrapolated HF/CBS energy was obtained with the method of Truhlar and Zhao and the aug-cc-pVXZ ($X = 2-3$) pair, and only the correlation energy was extrapolated with the triple and quadruple- ζ basis sets. This procedure is appropriate, as the extrapolation error is completely dominated by the correlation energy. Afterward, the energy was calculated at the CCSD(T)/aug-cc-pVTZ level, and the final extrapolated energy was obtained adding the CCSD(T) correction to the MP2 energy (i.e., the energy difference between CCSD(T) and MP2, both with the aug-cc-pVTZ basis set).

For comparison, we have also calculated the CCSD(T)/cc-pVXZ ($X = 2-3$) energies and used the extrapolation technique of Truhlar²² with these values. All extrapolation schemes resulted in very similar interaction energies for the Na⁺–benzene interaction energy. The basis set superposition error (BSSE) was accounted for in the extrapolation of the MP2 energy with the method of Halkier et al. and the aug-cc-pVTZ and aug-cc-pVQZ basis sets but not with the extrapolations used with the method of Truhlar, because the last was parametrized without BSSE corrections.^{20,22}

The choice of the density functionals was based on recent benchmarking studies,^{8,12,24-30} which presented an overview of the current status of the field. Particular care was taken to ensure a diverse and representative choice of density functionals, including generalized gradient approximation, GGA; generalized gradient exchange, GGE; meta generalized gradient approximation, M-GGA; hybrid generalized gradient approximation, H-GGA; and hybrid meta generalized gradient approximation, HM-GGA.

The results for each density functional were obtained using single point energy calculations for selected points of the MP2/6-311++G(d,p) PEPs (benzene–Na⁺ and benzene–H₂O) and for the isolated monomers, using the 6-311++G(2d,2p) basis set. The specific distances were 1.50 Å, 1.75 Å, 2.00 Å, 2.125 Å, 2.25 Å, 2.50 Å, 2.75 Å, 3.00 Å, 3.50 Å, 4.50 Å, 5.50 Å, 6.50 Å, 8.50 Å, 10.50 Å, 12.50 Å and 14.50 Å for water (hydrogen–benzene centroid distances) and 2.00 Å, 2.25 Å, 2.3125 Å, 2.375 Å, 2.50 Å, 2.625 Å, 2.75 Å, 3.00 Å, 3.25 Å, 3.50 Å, 4.00 Å, 5.00 Å, 6.00 Å, 8.00 Å, 10.00 Å, 12.00 Å and 14.00 Å for sodium ions (sodium–benzene centroid distances). The distances were chosen in order to obtain a smooth PEP with appropriate resolution near the minimum. At a few points, convergence was not achieved with some hybrid-meta functionals (which are known to be numerically unstable), but these cases did not compromise at any rate

Table 2. Electronic Interaction Energies (in kcal/mol) with ($E_{\text{int-cp}}$) and without (E_{int}) Counterpoise Correction for the Benzene–Na⁺ Interaction and the Respective BSSEs Calculated with Several Basis Sets^a

basis set	E_{int}	$E_{\text{int-cp}}$	BSSE	ΔE_{trunc}
6-31G(d)	–28.16	–26.11	–2.05	2.74
6-31+G(d)	–24.67	–23.58	–1.09	0.21
6-31G(d,p)	–28.06	–25.97	–2.09	2.60
6-31+G(d,p)	–24.34	–23.54	–0.80	0.17
6-311G(d,p)	–25.75	–24.53	–1.22	1.16
6-311+G(d,p)	–24.25	–23.51	–0.74	0.14
6-311++G(d,p)	–24.24	–23.51	–0.73	0.14
6-311++G(2d,2p)	–24.16	–23.45	–0.71	0.08
6-311++G(3df,3pd)	–23.90	–23.39	–0.51	0.02
cc-pVDZ	–25.06	–23.84	–1.22	0.47
aug-cc-pVDZ	–24.08	–23.55	–0.53	0.18
aug-cc-pVTZ	–23.57	–23.38	–0.19	0.01
aug-cc-pVQZ	–23.45	–23.37	–0.08	0.00

^a Deviations (ΔE_{trunc}) between the cp-corrected interaction energies and the larger basis set (aug-cc-pVQZ) are also shown.

the correct location of the minimum and the calculation of the binding energy. The choice of the 6-311++G(2d,2p) basis set was the best compromise between the completeness of the basis set and the computational time, according to the basis set benchmarking exposed below. We have found that this basis set presented good results in describing these interaction energies, once the DFT truncation errors were probably below 0.1 kcal/mol in both systems (0.08 kcal/mol in the benzene–Na⁺ model and 0.09 kcal/mol in the benzene–H₂O model), when measured with the B3LYP functional. Moreover, this basis set is the more adequate for the benchmarking because it will be the larger that will be routinely used when dealing with large biological systems (and not model systems as the ones used here), which commonly include over 100 atoms.

The calculations with the functionals M05, M05-2X, M06, M06-2X, M06-L, M06-HF, and X3LYP were performed using the Gaussian 09 suite of programs.⁶¹ All other calculations were carried out using the Gaussian 03 software package.¹⁹

The influence of the grid size used was found to be very small on the systems and properties studied here. The difference in the interaction energy calculated with the default grid (a pruned (75 302) grid) and that with the ultrafine grid (a pruned (99 590) grid) was always below 0.01 kcal/mol for the water–benzene system and below 0.1 kcal/mol for sodium–benzene system. Therefore, we have used the default grid size of both programs.

RESULTS AND DISCUSSION

I. Basis Set Truncation Error. In this section, we analyze the results for the convergence of the interaction energy with basis set size. This is a quite important issue, as it directly affects the obtained accuracy and the feasibility of the calculations. DFT is known for its usual fast convergence with basis set size. In the present study, the appropriate choice of the basis set is crucial in allowing for the benchmarking of a very large number of density functionals. Tables 2 and 3 present the deviations of the interaction energies calculated for each basis set against aug-cc-pVQZ, which was the most complete one.

Table 3. Electronic Interaction Energies (in kcal/mol) with ($E_{\text{int-cp}}$) and without (E_{int}) Counterpoise Correction for the Benzene–H₂O Interaction and the Respective BSSEs Calculated with Several Basis Sets^a

basis set	E_{int}	$E_{\text{int-cp}}$	BSSE	ΔE_{trunc}
6-31G(d)	−2.37	−1.42	−0.95	0.24
6-31+G(d)	−1.88	−1.40	−0.48	0.22
6-31G(d,p)	−2.25	−1.34	−0.91	0.16
6-31+G(d,p)	−1.71	−1.31	−0.40	0.13
6-311G(d,p)	−2.19	−1.33	−0.86	0.15
6-311+G(d,p)	−1.64	−1.30	−0.34	0.12
6-311++G(d,p)	−1.63	−1.29	−0.34	0.11
6-311++G(2d,2p)	−1.49	−1.27	−0.22	0.09
6-311++G(3df,3pd)	−1.36	−1.17	−0.19	0.01
cc-pVDZ	−2.02	−1.32	−0.70	0.14
aug-cc-pVDZ	−1.43	−1.20	−0.23	0.02
aug-cc-pVTZ	−1.26	−1.18	−0.08	0.00
aug-cc-pVQZ	−1.21	−1.18	−0.03	0.00

^a Deviations (ΔE_{trunc}) between the cp-corrected interaction energies and the larger basis set (aug-cc-pVQZ) are also shown.

A more complete basis set is not affordable in terms of computational demand. Anyway, the result will not change significantly beyond the quadruple- ζ basis set. This can be seen in Tables 2 and 3, as increasing the complexity of the basis sets from triple- ζ (aug-cc-pVTZ) to quadruple- ζ (aug-cc-pVQZ) results in a meaningless lowering of $E_{\text{int-cp}}$ by 0.01 kcal/mol on the benzene–Na⁺ system and has no effect at all on the benzene–H₂O system within the accuracy considered in this work. This also renders unnecessary the extrapolation to the CBS limit within DFT/B3LYP, which converges much faster than post-HF methods, for which the extrapolation is mandatory.

Considering now the remaining basis sets, one can easily isolate the three main factors that influence their accuracy: the degree of contraction of the sp shell, polarization functions, and diffuse functions.

Starting with the Pople basis sets, in the benzene–Na⁺ system, when we move from double- ζ 6-31G(d,p) to triple- ζ 6-311G(d,p), the ΔE_{trunc} decreases from 2.60 to 1.16 kcal/mol. The same happens in the benzene–H₂O system, but in this case, the values are much smaller than in the benzene–Na⁺ system (from 0.16 to 0.15 kcal/mol).

Concerning the polarization functions, in the benzene–H₂O system, for example, there is a decrease in ΔE_{trunc} from 0.11 to 0.09 kcal/mol, and then to 0.01 kcal/mol, when we move consecutively from 6-311++G(d,p) to 6-311++G(2d,2p), and to 6-311++G(3df,3pd). The polarization space seems to be saturated at this level. The same is true for the benzene–Na⁺ system, when ΔE_{trunc} decreases from 0.14 to 0.08 kcal/mol, and then to 0.02 kcal/mol with the same basis sets. The increase beyond the first set of polarization functions gives surprisingly small contributions to E_{int} .

We can also see that the influence of diffuse functions on heavy atoms is significant and much larger than in hydrogen atoms, as there is a decrease in ΔE_{trunc} from 6-311G(d,p) to 6-311+G(d,p) of about 1 kcal/mol, in the benzene–Na⁺ system, but when we add diffuse functions to hydrogen atoms (6-311++G(d,p)), the ΔE_{trunc} remains constant. In the benzene–H₂O case, we can see almost the same, but with smaller values, as has been previously said.

Table 4. Electronic Interaction Energies (E_{int}), Including Counterpoise Corrections, and Their Deviation from the Reference Value (ΔE_{int}) in the Benzene–Na⁺ System^a

rank	functional	type	%HF	E_{int}	$ \Delta E_{\text{int}} $
1	mPWKCIS	M-GGA	0	−22.6	0.1
2	B3PW91	H-GGA	20	−22.8	0.1
3	BB95	M-GGA	0	−22.6	0.1
4	BLYP	GGA	0	−22.5	0.2
5	M06	HM-GGA	27	−22.9	0.2
6	HCTH	GGA	0	−22.9	0.2
7	O3LYP	H-GGA	12	−22.9	0.3
8	OLYP	GGA	0	−22.4	0.3
9	B97-2	H-GGA	21	−23.0	0.3
10	MPW1KCIS	HM-GGA	15	−23.1	0.4
11	B3P86	H-GGA	20	−23.1	0.4
12	BP86	GGA	0	−22.0	0.6
13	TPSSKCIS	M-GGA	0	−23.4	0.7
14	B3LYP	H-GGA	20	−23.4	0.8
15	TPSSTPSS	M-GGA	0	−23.6	0.9
16	BPBE	GGA	0	−21.8	0.9
17	BPW91	GGA	0	−21.7	0.9
18	B1LYP	H-GGA	25	−23.6	1.0
19	PBEKCIS	M-GGA	0	−23.7	1.0
20	TPSS1KCIS	HM-GGA	13	−23.7	1.0
21	M06-L	M-GGA	0	−21.6	1.1
22	TPSSh	HM-GGA	10	−23.8	1.1
23	TPSSVWN5	GGE	0	−21.5	1.1
24	MPWKCIS1K	HM-GGA	41	−23.9	1.2
25	B98	H-GGA	22	−24.0	1.3
26	PBE1KCIS	HM-GGA	22	−24.1	1.4
27	B1B95	HM-GGA	25	−24.2	1.5
28	PBEPBE	GGA	0	−24.3	1.6
29	B97-1	H-GGA	21	−24.4	1.7
30	MPW1K	H-GGA	43	−24.5	1.8
31	MPW3LYP	H-GGA	22	−24.5	1.8
32	X3LYP	H-GGA	21.8	−24.7	2.0
33	PW91PW91	M-GGA	0	−24.7	2.0
34	PBE1PBE	H-GGA	25	−24.8	2.1
35	BHandHLYP	H-GGA	50	−24.8	2.1
36	BB1K	HM-GGA	42	−24.9	2.2
37	G96LYP	GGA	0	−19.9	2.8
38	MPW1B95	HM-GGA	31	−25.5	2.9
39	M05	HM-GGA	28	−25.8	3.2
40	MPWB1K	HM-GGA	44	−26.0	3.3
41	M06-2X	HM-GGA	54	−26.6	4.0
42	M05-2X	HM-GGA	56	−27.2	4.5
43	BHandH	H-GGA	50	−29.8	7.1
44	VSXC	M-GGA	0	−30.9	8.2
45	M06-HF	HM-GGA	100	−31.7	9.1
46	XLYP	GGA	0	−33.1	10.4

post-HF methods	E_{int}
MP2/cc-pVDZ	−21.79
MP2/cc-pVTZ	−22.48
MP2/CBS ^b	−23.39
MP2/aug-cc-pVDZ	−21.91

Table 4. Continued

post-HF methods	E_{int}
MP2/aug-cc-pVTZ	−22.46
MP2/aug-cc-pVQZ	−22.66
MP2/CBS ^c	−23.18
MP2/CBS ^d	−22.74
CCSD(T)/cc-pVDZ	−21.20
CCSD(T)/cc-pVTZ	−22.34
CCSD(T)/CBS ^c	−23.53
CCSD(T)/aug-cc-pVDZ	−21.70
CCSD(T)/aug-cc-pVTZ	−22.40
CCSD(T)/CBS ^f	−23.11
CCSD(T)/CBS ^g	−22.67
exptl ^h	−20.70 ± 1.03

^aAll interaction energies in the table include counterpoise corrections, even though the extrapolations based on the Truhlar method were done *without* counterpoise corrections. ^bExtrapolated from MP2/cc-pVDZ and MP2/cc-pVTZ with the method of Truhlar. ^cExtrapolated from MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ with the method of Truhlar et al. ^{20,22} ^dExtrapolated from MP2/aug-cc-pVTZ and MP2/aug-cc-pVQZ with the method of Halkier et al. ²³ ^eExtrapolated from CCSD(T)/cc-pVDZ and CCSD(T)/cc-pVTZ with the method of Truhlar. ²² ^fObtained adding to extrapolation 2 the CCSD(T) correction (the difference between the CCSD(T) and MP2 energies) calculated with the aug-cc-pVTZ basis set. ^gObtained adding to extrapolation 3 the CCSD(T) correction (the difference between the CCSD(T) and MP2 energies) calculated with the aug-cc-pVTZ basis set. This was the most accurate calculation and was taken as a reference to rank the functionals. ^hExperimental value. ⁶² (Note that the ZPE energy was calculated at the MP2/6-311++G(d,p) level and subtracted from the experimental value.)

Moving on to the correlation consistent basis sets, one can say that the results are generally better than those with the Pople basis sets, as was expected. Regarding the diffuse functions, they decrease the ΔE_{trunc} by about 0.3 kcal/mol for the benzene–Na⁺ system and about 0.1 kcal/mol for the benzene–H₂O system, when we move from the cc-pVDZ to aug-cc-pVDZ basis sets. With respect to ζ level and polarization functions, we need to conjecture two in one, because they are inseparably within these basis sets. When we move from double- ζ (aug-cc-pVDZ) to triple- ζ (aug-cc-pVTZ), the ΔE_{trunc} decreases considerably in benzene–Na⁺ (approximately 0.2 kcal/mol), but the value stabilizes here, and there is no change from that point on (i.e., on moving to aug-cc-pVQZ), because we are close to the convergence limit for this property. However, in the benzene–H₂O system, the increase of ζ level and polarization functions does not influence ΔE_{trunc} beyond the double- ζ basis. In fact, the addition of diffuse functions for the double- ζ basis set (aug-cc-pVDZ) has been enough to reduce the ΔE_{trunc} to almost zero, giving no way to improve it.

We opted to choose the 6-311++G(2d,2p) basis set for the functional benchmarking that follows, due to the good compromise between the complexity of the basis set and the computational demand. The truncation errors associated with the use of this basis set are below 0.1 kcal/mol in both systems (0.08 kcal/mol in the benzene–Na⁺ model and 0.09 kcal/mol in the benzene–H₂O model).

It is also interesting to assess the influence of counterpoise corrections (cp) for the basis set superposition error (BSSE) in the interaction energies of both systems. The values are presented in Tables 2 and 3. The influence of the cp correction is significantly greater for the less complete basis sets, where the

Table 5. Electronic Interaction Energies (E_{int}), Including Counterpoise Corrections, and Their Deviation from the Reference Value (ΔE_{int}) for the Benzene–H₂O System^a

rank	funcional	type	%HF	E_{int}	$ \Delta E_{\text{int}} $
1	M06-2X	HM-GGA	54	−3.40	0.04
2	M05-2X	HM-GGA	56	−3.33	0.11
3	M06-HF	HM-GGA	100	−3.65	0.21
4	M05	HM-GGA	28	−2.58	0.86
5	BHandH	H-GGA	50	−4.37	0.93
6	M06	HM-GGA	27	−2.36	1.08
7	MPWB1K	HM-GGA	44	−2.33	1.11
8	M06-L	M-GGA	0	−2.23	1.21
9	MPW1B95	HM-GGA	31	−2.21	1.23
10	PW91PW91	M-GGA	0	−2.12	1.32
11	B97-1	H-GGA	21	−2.10	1.34
12	HCTH	GGA	0	−2.09	1.35
13	PBE1PBE	H-GGA	25	−1.95	1.49
14	PBE1KCIS	HM-GGA	22	−1.92	1.52
15	PBEPBE	GGA	0	−1.92	1.52
16	B98	H-GGA	22	−1.90	1.54
17	PBEKCIS	M-GGA	0	−1.87	1.57
18	MPW3LYP	H-GGA	22	−1.82	1.62
19	BB1K	HM-GGA	42	−1.80	1.64
20	BHandHLYP	H-GGA	50	−1.66	1.78
21	MPWIK	H-GGA	43	−1.64	1.80
22	X3LYP	H-GGA	21.8	−1.63	1.81
23	MPWKICIS1K	HM-GGA	41	−1.61	1.83
24	TPSS1KCIS	HM-GGA	13	−1.54	1.90
25	B1B95	HM-GGA	25	−1.53	1.91
26	TPSSKCIS	M-GGA	0	−1.50	1.94
27	TPSSh	HM-GGA	10	−1.46	1.98
28	MPW1KCIS	HM-GGA	15	−1.40	2.04
29	TPSSTPSS	M-GGA	0	−1.39	2.05
30	B97-2	H-GGA	21	−1.36	2.08
31	B3P86	H-GGA	20	−1.31	2.13
32	mPWKCIS	M-GGA	0	−1.28	2.16
33	B3LYP	H-GGA	20	−1.27	2.17
34	B1LYP	H-GGA	25	−1.19	2.25
35	TPSSVWN5	GGE	0	−1.12	2.32
36	BB95	M-GGA	0	−1.00	2.44
37	O3LYP	H-GGA	12	−0.98	2.46
38	B3PW91	H-GGA	20	−0.95	2.49
39	OLYP	GGA	0	−0.88	2.56
40	BP86	GGA	0	−0.82	2.62
41	BLYP	GGA	0	−0.76	2.68
42	BPW91	GGA	0	−0.59	2.85
43	BPBE	GGA	0	−0.58	2.86
44	G96LYP	GGA	0	−0.16	3.28
45	XLYP	GGA	0	−7.77	4.33
46	VSXC	M-GGA	0	−8.12	4.68
	MP2/aug-cc-pVDZ			−4.52	
	MP2/aug-cc-pVTZ			−3.88	
	MP2/CBS ^b			−3.72	
	MP2/6-31+G(d)			−2.16	
	CCSD(T)/6-31+G(d)			−2.21	
	Δ (CCSD(T)-MP2)			−0.05	

Table 5. Continued

MP2/aug-cc-pVDZ	−4.52
MP2/aug-cc-pVTZ	−3.88
MP2/CBS ^b	−3.72
MP2/6-31+G(d)	−2.16
CCSD(T)/6-31+G(d)	−2.21
$\Delta(\text{CCSD(T)}-\text{MP2})$	−0.05
CCSD(T)/CBS ^c	−3.77
exptl ^d	−3.44 ± 0.09

^a All interaction energies in the table include counterpoise corrections, even though the extrapolations based on the Truhlar method were done *without* counterpoise corrections. ^b Extrapolated from MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ with the method of Truhlar et al.^{20,22}

^c Obtained adding to extrapolation 1 the CCSD(T) correction (the difference between the CCSD(T) and MP2 energies) calculated with the 6-31+G(d) basis set. ^d Experimental value (see ref 18 and references therein).

BSSE reaches −2.05 kcal/mol for the benzene–Na⁺ complex and −0.95 kcal/mol for the benzene–H₂O, and it is very small for the most complete ones, where the BSSE tends to vanish on both systems. This reflects the inherently better description of the latter monomers, needing the basis sets of the interacting partner less for the description of their own electronic density. Therefore, it is strongly advisable to use counterpoise corrections when using less complete basis sets in systems with these types of interactions.

II. Benchmarking of Density Functionals. In this section, we assess the performance of the 46 DFT functionals (Table 1) used to study the description of cation– π and π –Hbond interactions. Tables 4 and 5 include the electronic interaction energies for both systems, the high-level extrapolated MP2/CBS and CCSD(T)/CBS interaction energies, and the experimental values. These last were available for both systems studied here: −3.44 ± 0.09 kcal/mol for benzene–water (see ref 18 and references therein) and −20.70 ± 1.03 kcal/mol for the Na⁺–benzene system.⁶² For the Na⁺–benzene system, we have calculated the ZPE energy at the MP2/6-311++G(d,p) level and subtracted it from the experimental value to get the electronic binding energy. A similar procedure was done by others for the benzene–water system.¹⁸ The difference between the experimental and the extrapolated CCSD(T)/CBS interaction energies was 0.33 and 1.97 kcal/mol for the benzene–water and the Na⁺–benzene systems, respectively.

We have taken the experimental value for the water–benzene system as the reference value for this system, as the accuracy of this value is better than the one we can achieve with the present protocol. In the case of benzene–Na⁺, we have preferred to use the CCSD(T) value as the reference for this study. The experimental value is not very accurate (error bar above 1 kcal/mol) and is shadowed by controversy, with measurements differing by over 7 kcal/mol in recent years.^{62,63} We have calculated the whole PEPs for each of the 46 functionals. Globally, the results show a tendency of DFT functionals to overestimate the interaction energies.

In Table 4, we have ranked the functionals according to their absolute difference from the CCSD(T)/CBS value (in the bottom of the table) for the benzene–Na⁺ system.

The post-HF methods give very satisfactory results and are very close to the experimental value (−20.70 kcal/mol). The difference between the extrapolation with the Truhlar method

and the aug-cc-pVXZ (X = 2–3) basis sets and the Helgaker method and the aug-cc-pVXZ (X = 3–4) basis sets is small (0.4 kcal/mol). The difference between the CCSD(T) and MP2 energies with the aug-cc-pVTZ is only 0.07 kcal/mol. The difference in the extrapolated values using the Truhlar method with and without diffuse functions is only 0.2 kcal/mol at the MP2 level and 0.4 kcal/mol at the CCSD(T) level, suggesting that the extrapolation without diffuse functions can be seen as a viable alternative for larger systems, at least at the MP2 level.

In general, the DFT results are very good (perhaps excellent), as 20 of the 46 functionals calculate this interaction within chemical accuracy and 33 out of 46 within 2 kcal/mol. This result is particularly positive if we consider the magnitude of the interaction that is being calculated. There is no clear correlation between the amount of HF exchange and the binding energy, contrary to what happens with other properties, like activation energies (e.g., see ref 9), even though there is a tendency for functionals with large fractions of HF exchange to give poorer results (e.g., compare the results of M06-L and M06 with the ones of M06-2X and M06-HF). This may be due to the presence of a metal in the system, as it is well documented that large fractions of HF exchange are detrimental for the description of metals (in particular for transition metals). There is no correlation between functional families and accuracy, even though the hybrid meta functionals give slightly poorer average results. On average, HM functionals give the largest binding energies and GGA functionals give smaller. We also noted that all functionals overestimated the interaction energy.

mPWKICIS, B3PW91, BB95, BLYP, M06, and HCTH results are the most accurate. The very popular B3LYP successfully calculates this interaction within chemical accuracy (error of 0.8 kcal/mol).

Moving on to π –Hbond interactions, we have also ranked the functionals according to their $|\Delta E_{\text{int}}|$ values (Table 5). We have used the experimental value for the ranking even though the geometry of the computational complex (with an HO bond along the 6-fold axis) is not fully coincident with the absolute minimum (but is relevant in terms of the typical orientation found in biological systems, where the strain of the protein backbone overcomes the very shallow minimum in this flat region of the PES). Comparing the results obtained at the extrapolated MP2/CBS level with the system fully relaxed (−3.66 kcal/mol²⁰) and our results at the same level, with the HO bond along the 6-fold symmetry axis (−3.72 kcal/mol), we can see that the difference coming from the different geometries is irrelevant.

The two experimental interaction energies reported (using the ZPE corrections of Feller⁶⁴ to obtain D_e from the experimental D_0) are slightly less negative than the extrapolated CCSD(T)/CBS results (with values of −3.24 ± 0.28¹⁸ and −3.44 ± 0.09¹⁸). We have adopted the second because it is the more precise one.

Here, we can find out that the HM-GGA functionals give the best results in describing π –Hbond interactions, in particular the ones from the Truhlar group, mostly with high fractions of HF exchange. A similar result was found in an earlier study on the same system but included a much smaller number of density functionals.¹⁸ The most accurate functionals are M06-2X, M05-2X, M06-HF, and M05. Five density functionals have calculated this π –Hbond interaction within chemical accuracy, all with high fractions of HF exchange. This observation may reinforce the hypotheses that the failure of the functionals with high fractions of HF exchange in the description of the Na⁺–benzene

interaction should not be due to an incapacity to account accurately for the interaction itself but instead to a less accurate description of one of the binding partners (Na^+). Twenty-seven density functionals resulted in deviations below 2 kcal/mol. The very popular B3LYP functional ranks 33rd, with a ΔE_{int} of 2.17 kcal/mol. The failure in accounting for dispersive interactions may be one of the reasons for this. Even though the main source for attraction is electrostatic (in the sense of a polarized average electronic density), dispersion forces also contribute to binding, and these are partially unaccounted for by B3LYP. Note that these interactions at equilibrium distances occur in a medium-overlapping density regime, in which the density functional can still account (at least in part) for dispersion, contrary to what happens in longer-range interactions, where the density overlap is much less existent or essentially absent. The results are also in line with other benchmark studies, which show that the HM-GGA functionals with large fractions of HF exchange perform well in the calculation of polar noncovalent interactions.

CONCLUDING REMARKS

We have performed a basis set benchmarking study, using a total of 13 different basis sets, where we also tested the importance of counterpoise corrections for BSSE. The convergence of the DFT (B3LYP) interaction energy with basis set size was surprisingly fast, with truncation errors (relative to aug-cc-pVQZ) below 0.22 kcal/mol for double- ζ basis sets with (at least) polarization and diffuse functions on heavy atoms. Even smaller truncation errors were found for larger basis sets, as expected. The 6-311++G(2d,2p) basis set represented the best compromise between accuracy and computational time. The basis set superposition error was particularly large for less complete basis sets (up to 2 kcal/mol in the smaller basis sets) and steadily decreased as the basis set increased, which probably reflects the inherently better description of the latter monomers, needing the basis sets of the interacting partner less for the description of their own electronic density.

Cation- π and π -Hbond interactions occur at short/medium range, between a highly polarizable center and a strong dipole/charge, in a region where there is still significant density overlap between the interacting molecules, far from the overlap-free region where dispersion interactions are inherently unaccounted for by DFT. The description of the interactions in this region poses a challenge for DFT, as the density functionals must have a good balance between dispersion, repulsion, and dipolar/electrostatic attraction. In general, the density functionals gave very satisfactory results, with a significant number of predictions within chemical accuracy. In the case of π -Hbond interactions, these mostly belong to the hybrid-meta family and have large fractions of Hartree-Fock exchange. The functionals of the Truhlar group were particularly well suited for this purpose. Beyond their adequacy for the description of intermolecular interactions, they present other advantages such as their good performance in thermochemistry and barrier heights,^{9,60} which makes them a very good choice for other studies on biological systems (e.g., chemical/enzymatic reactivity), where many factors must be consistently addressed to have an accurate result.

In the case of cation- π interactions, there was not a clear correlation between accuracy and functional sophistication (in terms of its dependence on the density gradients) or percentage of HF exchange. One of the most interesting observations is that

despite the large number of functionals predicting interaction energies within chemical accuracy (five for π -Hbond and 20 for cation- π interactions), not a single functional has shown chemical accuracy in both. Moreover, if we calculate the average error for these two interactions, only two density functionals resulted in an average error below 1.0 kcal/mol (M06 and HCTH, with average errors of 0.6 and 0.8 kcal/mol). This epitomizes the limitations that density functionals still present nowadays in terms of the generality of their performance and emphasizes the necessity of a careful choice of the density functional (based on benchmarking studies) before embarking on long and complex electronic structure calculations.

AUTHOR INFORMATION

Corresponding Author

*E-mail: mjramos@fc.up.pt.

ACKNOWLEDGMENT

This work has been financed by the program FEDER/COMPETE and by the Fundação para a Ciência e a Tecnologia (Project PTDC/QUI-QUI/102760/2008).

REFERENCES

- (1) Jiang, Y.; Wu, J.; Zu, J. W.; Lu, Y. X.; Hu, G. X.; Yu, Q. S. *Int. J. Quantum Chem.* **2008**, *108* (7), 1294–1303.
- (2) Dougherty, D. A. *J. Nutr.* **2007**, *137* (6), 1504s–1508s.
- (3) Ruan, C. H.; Rodgers, M. T. *J. Am. Chem. Soc.* **2004**, *126* (44), 14600–14610.
- (4) Liu, T.; Zhu, W. L.; Gu, J. D.; Shen, J. H.; Luo, X. M.; Chen, G.; Puah, C. M.; Silman, I.; Chen, K. X.; Sussman, J. L.; Jiang, H. L. *J. Phys. Chem. A* **2004**, *108* (43), 9400–9405.
- (5) Tarakeshwar, P.; Choi, H. S.; Lee, S. J.; Lee, J. Y.; Kim, K. S.; Ha, T. K.; Jang, J. H.; Lee, J. G.; Lee, H. J. *Chem. Phys.* **1999**, *111* (13), 5838–5850.
- (6) Atkins, P. W.; Friedman, R. S. *Molecular quantum mechanics*, 4th ed.; Oxford University Press: Oxford, 2005; pp xiv, 573.
- (7) Ghosh, A. J. *Biol. Inorg. Chem.* **2006**, *11* (6), 671–673.
- (8) Schultz, N. E.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109* (49), 11127–11143.
- (9) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41* (2), 157–167.
- (10) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111* (42), 10439–10452.
- (11) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. *Chem. Rev.* **2000**, *100* (11), 4145–4185.
- (12) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1* (3), 415–432.
- (13) van Mourik, T. *J. Chem. Theory Comput.* **2008**, *4* (10), 1610–1619.
- (14) Peverati, R.; Baldrige, K. K. *J. Chem. Theory Comput.* **2008**, *4* (12), 2030–2048.
- (15) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8* (45), 5287–5293.
- (16) Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28* (2), 555–569.
- (17) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3* (1), 289–300.
- (18) Zhao, Y.; Tishchenko, O.; Truhlar, D. G. *J. Phys. Chem. B* **2005**, *109* (41), 19046–19051.
- (19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; et al. *Gaussian 03*, Revision C.02.; Gaussian, Inc.: Wallingford CT, 2004.
- (20) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109* (30), 6624–6627.

- (21) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2000**, *122* (46), 11450–11458.
- (22) Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *294* (1–3), 45–48.
- (23) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286* (3–4), 243–252.
- (24) Karton, A.; Tarnopolsky, A.; Lamere, J. F.; Schatz, G. C.; Martin, J. M. L. *J. Phys. Chem. A* **2008**, *112* (50), 12868–12886.
- (25) Silva, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2008**, *129* (10).
- (26) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. C* **2008**, *112* (17), 6860–6868.
- (27) Settergren, N. M.; Buhlmann, P.; Amin, E. A. *THEOCHEM* **2008**, *861* (1–3), 68–73.
- (28) Santra, B.; Michaelides, A.; Scheffler, M. *J. Chem. Phys.* **2007**, *127* (18), -.
- (29) Amin, E. A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4* (1), 75–85.
- (30) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. B* **2007**, *111* (30), 9146–9152.
- (31) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37* (2), 785–789.
- (32) Becke, A. D. *Phys. Rev. A* **1988**, *38* (6), 3098–3100.
- (33) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91* (14).
- (34) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58* (8), 1200–1211.
- (35) Perdew, J. P. *Phys. Rev. B* **1986**, *33* (12), 8822–8824.
- (36) Perdew, J. P. Unified Theory of Exchange and Correlation Beyond the Local Density Approximation. In *Electronic Structure of Solids*; Zieche, P., Eschrig, H., Eds.; Akademie: Berlin, Germany, **1991**.
- (37) Becke, A. D. *J. Chem. Phys.* **1996**, *104* (3), 1040–1046.
- (38) Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109* (2), 400–410.
- (39) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868.
- (40) Gill, P. M. W. *Mol. Phys.* **1996**, *89* (2), 433–445.
- (41) Krieger, J. B.; Chen, J. Q.; Iafrate, G. J.; Savin, A. *Electron. Correl. Mater. Prop.* **1999**, 463–477.
- (42) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109* (15), 6264–6271.
- (43) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125* (19), -.
- (44) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99* (5), 403–412.
- (45) Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (9), 2673–2677.
- (46) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45* (23), 13244–13249.
- (47) Becke, A. D. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.
- (48) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108* (33), 6908–6918.
- (49) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108* (2), 664–675.
- (50) Zhao, Y.; Gonzalez-Garcia, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109* (9), 2012–2018.
- (51) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7* (1), 43–52.
- (52) Schmider, H. L.; Becke, A. D. *J. Chem. Phys.* **1998**, *108* (23), 9624–9631.
- (53) Becke, A. D. *J. Chem. Phys.* **1997**, *107* (20), 8554–8560.
- (54) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123* (16), -.
- (55) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104* (21), 4811–4815.
- (56) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2* (2), 364–382.
- (57) Wilson, P. J.; Bradley, T. J.; Tozer, D. J. *J. Chem. Phys.* **2001**, *115* (20), 9233–9242.
- (58) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110* (49), 13126–13130.
- (59) Hoe, W. M.; Cohen, A. J.; Handy, N. C. *Chem. Phys. Lett.* **2001**, *341* (3–4), 319–328.
- (60) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120* (1–3), 215–241.
- (61) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; et al. *Gaussian 09*, Revision A.02.; Gaussian, Inc.: Wallingford, CT, 2009.
- (62) Armentrout, P. B.; Rodgers, M. T. *J. Phys. Chem. A* **2000**, *104* (11), 2238–2247.
- (63) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97* (5), 1303–1324.
- (64) Feller, D. *J. Phys. Chem. A* **1999**, *103* (38), 7558–7561.

Ab Initio Modeling of Donor–Acceptor Interactions and Charge-Transfer Excitations in Molecular Complexes: The Case of Terthiophene–Tetracyanoquinodimethane

Juan Aragó,^{†,‡} Juan C. Sancho-García,[§] Enrique Ortí,^{*,†} and David Beljonne[‡]

[†]Instituto de Ciencia Molecular, Universidad de Valencia, E-46980 Valencia, Spain

[‡]Laboratory for Chemistry of Novel Materials, Université de Mons, Place du Parc 20, B-7000 Mons, Belgium

[§]Departamento de Química Física, Universidad de Alicante, E-03080 Alicante, Spain

S Supporting Information

ABSTRACT: This work presents a thorough quantum chemical study of the terthiophene–tetracyanoquinodimethane complex as a model for π – π donor–acceptor systems. Dispersion-corrected hybrid (B3LYP-D) and double hybrid (B2PLYP-D), hybrid meta (M06-2X and M06-HF), and recently proposed long-range corrected (LC-wPBE, CAM-B3LYP, and wB97X-D) functionals have been chosen to deal with π – π intermolecular interactions and charge-transfer excitations in a balanced way. These properties are exhaustively compared to those computed with high-level ab initio SCS-MP2 and CASPT2 methods. The wB97X-D functional exhibits the best performance. It provides reliable intermolecular distances and interaction energies and predicts a small charge transfer from the donor to the acceptor in the ground state. In addition, wB97X-D is also able to yield an accurate description of the charge-transfer excitations in comparison to benchmark CASPT2 calculations.

INTRODUCTION

Donor–acceptor (D–A) molecular complexes, formed by π -conjugated materials, have recently attracted a large interest in the field of organic solar cells since these complexes can undergo very efficient photoinduced charge transfer (CT).^{1,2} In addition to the charge generation process, optimum charge transport through donor and acceptor layers is essential to achieve high performances in solar energy conversion.³ In this respect, donor and acceptor materials with fine-tuned semiconducting properties are also needed. Oligothiophenes (*n*Ts) constitute one of the most widely studied classes of semiconducting materials acting as electron-donor compounds. They show suitable electronic and solid-state properties that result into high charge carriers mobilities.^{4–7} On the other hand, 7,7,8,8-tetracyano-*p*-quinodimethane (TCNQ), acting as a strong electron-acceptor, has attracted special attention in the field of molecular materials since the discovery of the first true organic metal, namely, the CT complex formed by this molecule and tetrathiafulvalene.^{8,9} Although both *n*Ts and TCNQ derivatives have been well studied individually, D–A complexes formed by these systems have received much less attention.^{10,11}

Recently, Panda et al. investigated a series of donor π -conjugated oligomers mixed with electron acceptors in chloroform by means of UV–vis absorption spectroscopy and cyclic voltammetry.¹² Compelling spectroscopic evidence for the formation of *n*Ts–TCNQ D–A complexes in solution was provided. More specifically, new absorption bands were identified for mixtures of *n*Ts and TCNQ in chloroform whose intensities showed a marked dependence on the concentration of the donor and the acceptor and did not match those reported for [*n*Ts]⁺ and [TCNQ][–]. These new bands were described as electronic transitions from the ground state (no charge separation) to the

charge-separated excited state of the *n*Ts–TCNQ complexes. Although the complexes were experimentally detected, fundamental information about the geometry of the supramolecular dimer, the charge distribution in the electronic ground state, and the nature of the CT excitations is scarce. This information can be conveniently assessed from state-of-the-art quantum chemical calculations that are able to simultaneously describe van der Waals (dispersion) interactions and CT excitations in a balanced way.

Theoretically, the accurate treatment of both van der Waals interactions and CT excitations is a difficult and challenging task for most quantum chemical methods and is thus the subject of intense ongoing research. Dispersion interactions purely arise from electron correlation effects¹³ and thus need highly correlated wave function methods to treat them adequately. Coupled-cluster theory with singles, doubles, and perturbatively connected triple excitations [CCSD(T)], in conjunction with large basis sets, can accurately describe these effects.^{14,15} Unfortunately, CCSD(T) calculations are computationally very demanding and therefore unfeasible for systems such as D–A complexes of medium-size. On the other hand, second-order Møller–Plesset perturbation theory (MP2)¹⁶ has a lower computational cost (N^5 vs N^7 , where N is related to the size of the system) and can be accordingly seen as a suitable method to partly account for dispersive interactions in large systems. However, the MP2 method tends to generally overestimate binding energies for π -stacked systems.^{17–19} Hence, an alternative to MP2 recently appeared, denoted as spin-component scaled MP2 (SCS-MP2), which significantly improves the MP2 results, namely, the

Received: March 24, 2011

Published: June 06, 2011

overestimation of the binding energies in π -stacks.^{20,21} Therefore, the SCS-MP2 method can be considered as a reliable quantum chemical wave function method to treat noncovalent interactions in large π -systems where the cost of CCSD(T) is prohibitive.

We now turn our attention to Kohn–Sham (KS) density functional theory (DFT), which is by far the most widely used method for electronic structure calculations in condensed matter physics and quantum chemistry. It seems that a general drawback of all the common functionals is their inability to describe long-range electron correlations responsible for noncovalent interactions.^{22–24} The long-range correlation effects can be captured by nonempirical approaches specially devised to account for dispersion interactions explicitly.^{25–31} Another approach is to combine symmetry-adapted intermolecular perturbation theory (SAPT) with a DFT representation of the monomers.³² Since this representation of dispersion interactions is nonempirical in nature, the computational cost still represents a serious bottleneck. In recent years, there has been considerable interest in a less costly yet qualitatively correct DFT-based description of noncovalent interactions, leading to several dispersion-corrected methods.^{32–41} Among them, the new series of M06-class functionals developed by Truhlar and co-workers^{33,34} has extensively been parametrized to take dispersive effects into account. The simplest approach, normally designated as DFT-D, introduces dispersion interactions using an empirical potential of the form C_6R^{-6} .^{35–41} The dispersion energy is calculated separately from DFT calculations and simply added to the converged DFT energy. DFT-D has been applied to calculate the intermolecular interactions energies for large benchmark sets of noncovalent molecules with very satisfactory results.^{42–44} Therefore, this approach can be an excellent alternative to deal with large systems such as those tackled here.

Another well-known shortcoming of DFT methodologies is the poor description of CT excitations for local and hybrid functionals traditionally used within the time-dependent DFT (TD-DFT) approach.⁴⁵ This shortcoming can be overcome using multiconfigurational ab initio methodologies, such as the complete active space method combined with a second-order perturbation approach (CASCF/CASPT2). The CASCF/CASPT2 method is the most appropriate for the study of CT excitations where TD-DFT fails totally.⁴⁶ However, the application of the CASCF/CASPT2 protocol to large systems requires enormous computational resources. A recent promising alternative within the TD-DFT framework, which improves the accuracy of both Rydberg and CT excitations while maintaining good quality local excitations, is the long-range corrected approach (LC).⁴⁷ This approach consists in the splitting of the $1/r_{12}$ two-particle operator into short- and long-range exchange components, with the help of the standard error function (erf):

$$\frac{1}{r_{12}} = \frac{1 - \text{erf}(\omega r_{12})}{r_{12}} + \frac{\text{erf}(\omega r_{12})}{r_{12}} \quad (1)$$

where r_{12} is the interelectronic distance and ω is a parameter defining the range separation. Short-range exchange is treated mainly using a local or hybrid functional, whereas long-range exchange is usually treated using exact orbital exchange. Within this LC scheme, several functionals, such as LC-wPBE,⁴⁸ Coulomb-attenuated CAM-B3LYP,^{49,50} and the wB97^{51–53} family have been recently proposed to improve CT excitation predictions. Double hybrid functionals, like B2PLYP, have also

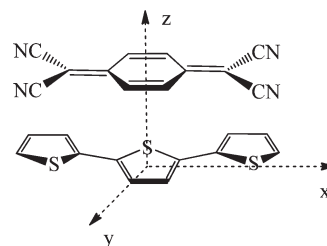


Figure 1. Terthiophene–tetracyanoquinodimethane complex in the cofacial geometry. The two molecular planes are parallel to one another.

shown very good performance in benchmark calculations for vertical excitation energies of a wide variety of organic molecules.^{54,55} Furthermore, the combination of B2PLYP with the dispersion approach (B2PLYP-D) has been repeatedly tested to yield accurate noncovalent interactions.^{43,44,56} Hence, this functional can also be seen as a robust and reliable alternative method.

Here we perform a thorough quantum chemical study of the terthiophene–tetracyanoquinodimethane (3T–TCNQ) complex (Figure 1) focusing mainly on: (i) the performance of several density functionals to account for the energetics of π – π interactions between 3T and TCNQ, (ii) the charge distribution provided by the different functionals for the 3T–TCNQ ground-state, (iii) the identification and the characterization of the most stable structures, and (iv) the description of the nature and the energy of the CT excitations. In a broader perspective, our aim is thus to identify a suitable methodology that can both properly account for π – π interactions and predict reliable CT excitation energies, which can be mostly useful to investigate larger π – π D–A complexes.

THEORETICAL AND COMPUTATIONAL DETAILS

Calculations were performed with the Gaussian 09⁵⁷ and ORCA⁵⁸ program packages. Calculations make use of the cc-pVDZ and cc-pVTZ basis sets.⁵⁹ The former was chosen as a compromise between accuracy and applicability to large molecules. The isolated molecules, 3T and TCNQ, were first optimized using the Becke’s three-parameter B3LYP exchange functional^{60,61} and the cc-pVDZ basis set. Potential energy curves were constructed by plotting the intermolecular interaction energy between 3T and TCNQ monomers, in a purely cofacial geometry, and by varying the intermolecular distance between the molecular centers along the z axis (see Figure 1). The intermolecular interaction energy (ΔE) calculated by the different methods was computed as follows:

$$\Delta E = E_{3T-TCNQ} - E_{3T} - E_{TCNQ} \quad (2)$$

where $E_{3T-TCNQ}$ denotes the energy of the complex and E_{3T} and E_{TCNQ} correspond to the energy of the monomers. The SCS-MP2 method in conjunction with the cc-pVTZ basis set will be used as reference for this part of the study due to its close agreement with CCSD(T) results.²¹ Note that the correlation energy at second order is scaled according to the following equation:

$$E^{\text{corr}}(\text{SCS-MP2}) = p_S E_{\uparrow\downarrow} + p_T E_{(\uparrow\uparrow + \downarrow\downarrow)} \quad (2)$$

where $E_{(\uparrow\uparrow + \downarrow\downarrow)}$ and $E_{\uparrow\downarrow}$ are the second-order contributions from double excitations of electron pairs with parallel- and antiparallel-spin, respectively, with $p_S = 6/5$ and $p_T = 1/3$ being the default scaling parameters.²⁰

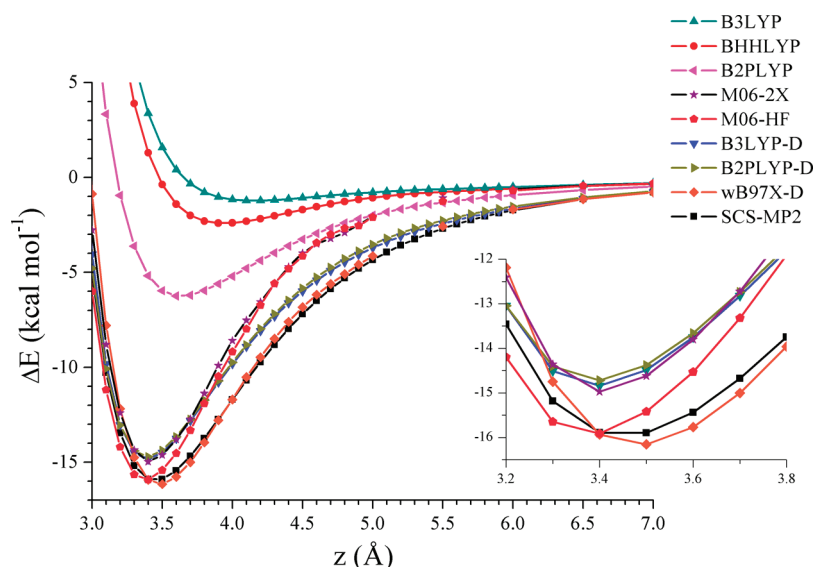


Figure 2. Potential energy curves for the cofacial 3T–TCNQ complex computed at various DFT/cc-pVDZ levels. The SCS-MP2/cc-pVTZ curve is shown as a reference. Intermolecular interaction energies (ΔE) are counterpoise uncorrected.

The interaction energy was then calculated at the B3LYP(-D),^{60,61} BHHLYP,^{60,61} B2PLYP(-D),⁶² M06-2X,³⁴ M06-HF,⁶³ and wB97X-D⁵² levels in combination with the cc-pVDZ and cc-pVTZ basis sets. The -D term denotes the approach originally developed by Grimme to calculate the dispersion energy (E_D) between two weakly overlapping systems, which is calculated separately (in a post-self consistent field fashion) by resorting to a van der Waals-type function explicitly depending on the well-known R_{AB}^{-6} decay of such interactions:³⁸

$$E_D = -s_6 \sum_{AB} f_d(R_{AB}) \frac{C_6^{AB}}{R_{AB}^6} \quad (3)$$

where $f_d(R_{AB})$ is a damping function of the interatomic distance (R), C_6^{AB} is the dispersion coefficient for the atomic pair AB, and s_6 is a scaling factor that only depends on the functional used. The parameters implemented for B3LYP-D and B2PLYP-D in ORCA and for wB97X-D in Gaussian 09 were used as defaults.

Basis set superposition errors (BSSEs) for the interaction energies in the 3T–TCNQ complex were computed by applying the counterpoise (CP) method.⁶⁴

The resolution of the identity (RI) technique^{65,66} was employed for SCS-MP2 and B2PLYP calculations. The error introduced by the RI treatment is completely negligible compared to other effects. The terms SCS-MP2 and B2PLYP will be used instead of RI-SCS-MP2 and RI-B2PLYP, respectively.

The vertical CT excitation energies were first computed by means of the CASPT2 approach.^{67–69} In this method, the first-order wave function and the second-order energy are calculated using the CASSCF wave function as reference. The CASSCF/CASPT2 protocol has been shown to be remarkably accurate for CT excitation energies of organic molecules which are not well described by TD-DFT methodologies.⁴⁶ Therefore, CASSCF/CASPT2 excitation energies will be used as a reference. The active space is constituted by 4π and $4\pi^*$ valence molecular orbitals (8 MOs) and 8 electrons. A systematic study using an increasing active space composed of $n\pi/n\pi^*$ ($n = 1–6$) MOs and their corresponding housed electrons was in fact performed, and the convergence of the excitation energies was already achieved

at the $4\pi/4\pi^*$ level. The CASSCF state-interaction (CASSI) method was employed to compute the oscillator strength with the CASPT2 excitation energies.^{70,71} CASSCF/CASPT2 calculations were carried out with the MOLCAS 7.2 package.⁷²

CT excitation energies were then calculated using standard TD-DFT methodology and the B3LYP, BHHLYP, B2PLYP, M06-2X, and M06-HF functionals. CT excitations were also evaluated using LC functionals as Coulomb-attenuated CAM-B3LYP,⁴⁹ LC-wPBE,⁴⁸ and wB97X-D. The excitation energies for B2PLYP were computed according to the Neese and Grimme procedure.⁵⁴ In this procedure and in complete analogy to the ground-state treatment, a scaled second-order perturbation correction to configuration interaction with singles (CIS(D)) wave functions developed some years ago by Head-Gordon et al.⁷³ is computed on the basis of density functional data and added to the TD-DFT excitation energies as follows:

$$\omega_{\text{corr}} = \omega_{\text{TD-B2PLYP}} + a_c \Delta_{(\text{D})} \quad (4)$$

where ω_{corr} is the corrected excitation energy, $\omega_{\text{TD-B2PLYP}}$ is the TD-B2PLYP excitation energy computed in a standard way and $\Delta_{(\text{D})}$ is the second-order perturbation term. The a_c parameter is equal to that used for the ground state. B2PLYP calculations were done using ORCA, whereas the rest of TD-DFT calculations used Gaussian 09. Vertical CT excitation energies were computed using the cc-pVDZ basis set.

RESULTS AND DISCUSSION

Intermolecular Interaction Energies. As previously discussed, the description of the ground-state potential energy curves of the 3T–TCNQ complex requires accurate quantum chemical calculations fully accounting for π – π stacking interactions. Figure 2 displays the potential energy curves calculated for the cofacial 3T–TCNQ complex at the SCS-MP2/cc-pVTZ level (taken here as a reference) and at the DFT level using different functionals in combination with the cc-pVDZ basis set. The potential energy curves in Figure 2 are computed without taking into account the BSSE correction. The importance of the

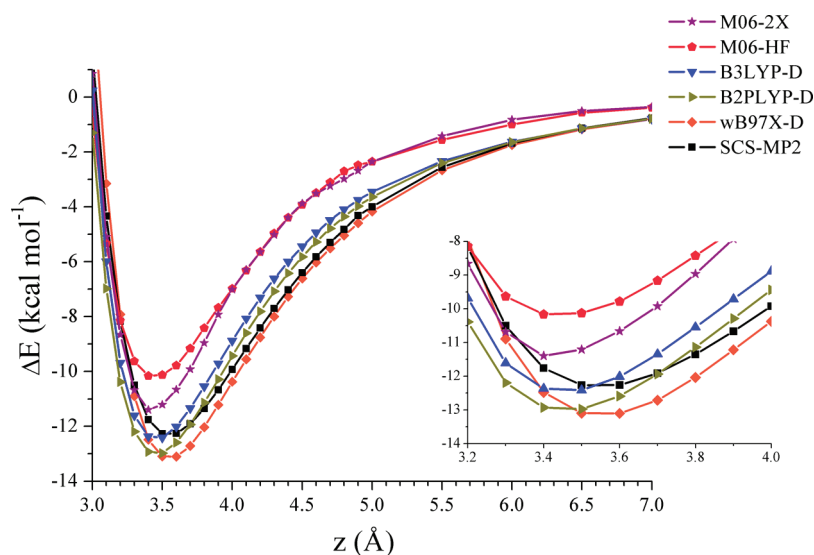


Figure 3. Potential energy curves for the cofacial 3T–TCNQ complex computed at various DFT levels with the cc-pVTZ basis set. The SCS-MP2/cc-pVTZ curve is shown as a reference. All intermolecular interaction energies (ΔE) are corrected for BSSEs by the counterpoise technique.

BSSE correction and of the size of the basis set employed with DFT calculations is later analyzed. SCS-MP2 calculations predict a minimum of $-15.97 \text{ kcal mol}^{-1}$ at a separation of 3.44 \AA between the molecular planes. Note that SCS-MP2 calculations provide a significant improvement with respect to MP2, as reported before for related D–A complexes.⁷⁴

The potential energy curves provided by DFT methods can be interpreted according to the nature of the functional employed (Figure 2). As expected, the hybrid B3LYP and BHHLYP functionals predict too shallow potential energy curves with energy minima of -1.23 and $-2.41 \text{ kcal mol}^{-1}$, respectively, and thus dramatically underestimate the interaction energies compared to SCS-MP2/cc-pVTZ results. The optimum intermolecular distances (B3LYP: 4.15 \AA , BHHLYP: 3.94 \AA) are strongly overestimated in comparison with that obtained at the SCS-MP2 level (3.44 \AA) and with those commonly found for related π – π complexes (3.4 – 3.5 \AA).⁷⁵ Though they predict a bound dimer, the hybrid functionals do not therefore properly account for π – π stacking interactions in the 3T–TCNQ complex, as previously reported for similar π – π complexes.^{75,76} The culprit for the rather unsatisfactory performance of the hybrid functionals in describing intermolecular interactions is the local nature of the correlation kernel. Therefore, in hybrid functionals the correlation energy is calculated only from the local properties of the density, and the dispersion energy, which arises from a truly nonlocal correlation effect, cannot be explicitly described.

The double hybrid B2PLYP functional partially accounts for the dispersion energy. This effect is clearly observed in the potential energy curve computed for the 3T–TCNQ complex, which exhibits a deeper energy minimum ($-6.27 \text{ kcal mol}^{-1}$) compared to B3LYP and BHHLYP results. The minimum-energy intermolecular distance is 3.63 \AA , in better agreement with that obtained at the SCS-MP2 level (3.44 \AA). The better description of the dispersion energy provided by B2PLYP arises from the incorporation of dynamical electron correlation effects through a perturbative second-order correlation term obtained from the KS orbitals and eigenvalues.⁶² Despite the partial introduction of nonlocal electron correlation effects,

the interaction energy predicted by B2PLYP in the minimum-energy region underestimates by $\sim 10.0 \text{ kcal mol}^{-1}$ the SCS-MP2 value.

In order to obtain more accurate potential energy curves, DFT calculations using hybrid meta functionals (M06-2X and M06-HF) and incorporating a dispersion term (DFT-D) were carried out. The M06-2X and M06-HF functionals yield quantitatively reliable potential energy curves with potential energy minima of -14.96 and $-15.90 \text{ kcal mol}^{-1}$, respectively. The minimum-energy intermolecular separations are predicted at 3.40 and 3.39 \AA , respectively, which are slightly shorter than that estimated at the SCS-MP2 level. Both the interaction energies and the intermolecular distances are in good agreement with the SCS-MP2 results. However, the curves generated using these functionals are narrower than that predicted at the SCS-MP2 level. The addition of the dispersion correction term to the B3LYP, B2PLYP, and wB97X functionals, denoted as B3LYP-D, B2PLYP-D, and wB97X-D, gives rise to a fully quantitative description of the potential energy curve. The three DFT-D functionals predict interaction energies at the minimum (-14.84 , -14.72 , and $-16.14 \text{ kcal mol}^{-1}$ for B3LYP-D, B2PLYP-D, and wB97X-D, respectively) very close to that computed at the SCS-MP2 level ($-15.97 \text{ kcal mol}^{-1}$) with differences of $\sim 1 \text{ kcal mol}^{-1}$. The optimum 3T–TCNQ intermolecular distances are calculated at 3.39 (B3LYP-D), 3.39 (B2PLYP-D), and 3.49 \AA (wB97X-D), slightly underestimating (B3LYP-D and B2PLYP-D) and overestimating (wB97X-D) the SCS-MP2 value (3.44 \AA). Among the hybrid meta and the DFT-D functionals employed to describe the 3T–TCNQ complex, the wB97X-D seems to yield the best performance to estimate the potential energy curve (see Figure 2).

In a further step, the potential energy curves for the cofacial 3T–TCNQ complex were recalculated including BSSE corrections and the larger cc-pVTZ basis set (Figure 3). As a consequence of BSSE corrections, the depth of the well for the SCS-MP2 curve decreases $\sim 4 \text{ kcal mol}^{-1}$ and the minimum-energy intermolecular distance undergoes a significant increase going from 3.44 to 3.55 \AA . The potential energy minimum is now predicted at $-12.31 \text{ kcal mol}^{-1}$.

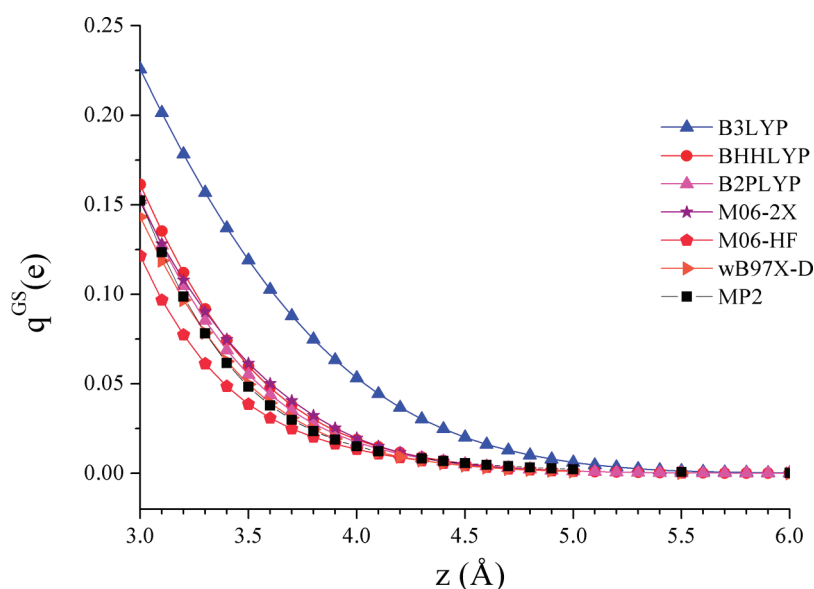


Figure 4. Mulliken charge transferred from 3T to TCNQ in the ground state (q^{GS}) calculated with different functionals as a function of the intermolecular distance (z) in the cofacial geometry.

A detailed inspection of Figure 3 reveals that DFT-D functionals (B3LYP-D, B2PLYP, and wB97X-D) yield potential energy curves that are better than those obtained by M06 functionals (M06-2X and M06-HF). The M06-2X and M06-HF functionals underestimate by more than 1 kcal mol^{-1} the intermolecular interaction energies at the minimum-energy region. The optimum intermolecular distances are found at 3.41 and 3.43 Å, respectively, which are too short compared to the SCS-MP2 value (3.55 Å). It should be noted that at longer distances, the potential energy curve produced by M06-2X deviates significantly from the SCS-MP2 curve (and the DFT-D curves), with energies that rise sharply in the range between the minimum and 4.0 Å. As a consequence, the M06-2X potential well is too narrow near the minimum. The DFT-D functionals predict interaction energies at the minimum (-12.46 , -13.02 , and $-13.16 \text{ kcal mol}^{-1}$ for B3LYP-D, B2PLYP-D, and wB97X-D, respectively), very close to that computed at the SCS-MP2 level ($-12.31 \text{ kcal mol}^{-1}$). In terms of minimum-energy intermolecular separations, wB97X-D gives the best result, with a potential energy minimum placed at 3.56 Å, which matches perfectly with that calculated at the SCS-MP2 level. The intermolecular distance afforded by B3LYP-D and B2PLYP-D functionals (3.45 Å) underestimates the SCS-MP2 value by $\sim 0.1 \text{ Å}$. It is worth to note that both B3LYP-D and wB97X-D functionals in conjunction with the cc-pVTZ basis set result in small BSSEs of $\sim 1 \text{ kcal mol}^{-1}$ (Figure S1, Supporting Information).

The results clearly point out that DFT-D functionals in combination with cc-pVTZ basis set yield the best performance to estimate the potential energy curves for the 3T–TCNQ complex. Among them, wB97X-D predicts accurate interaction energies and yields the best intermolecular distances. In addition to the dispersion term, the wB97X-D functional is built on the basis of the long-range corrected scheme which already partially takes noncovalent interactions into account.⁵¹ The LC scheme seems to be important not only for the treatment of noncovalent interactions but also for the description of CT excited states, as will be discussed below. It is also important to stress that wB97X-D in combination with the cc-pVDZ basis set and without

correction of the BSSE, although overestimates the interaction energies by about 4 kcal mol^{-1} , is able to provide relatively accurate intermolecular distances. This information is of great relevance in order to study larger π – π D–A complexes for which more extended basis sets are prohibitive.

Ground-State Charge-Transfer Analysis. π – π D–A molecular complexes are characterized by a small amount of charge transferred from the donor to the acceptor in the ground state.⁷⁷ Unfortunately, the charge distribution calculated for the 3T–TCNQ complex in the ground state strongly relies on the choice of the functional. Figure 4 shows the evolution of the Mulliken charge transferred from 3T to TCNQ in the cofacial configuration as a function of the intermolecular distance. To test the reliability of the Mulliken charges, the charge transferred from 3T to TCNQ was also analyzed according to the CHelpG scheme (Figure S2, Supporting Information).⁷⁸ The charges obtained from this scheme are similar to those computed using the Mulliken approach, and therefore, Mulliken charges will be used in the following discussion. It should be mentioned that Hobza et al. have recently shown that Mulliken analysis lead to reliable estimates of the charge transferred in a family of D–A complexes.⁷⁹

The amount of charge decays quickly with increasing separation between the 3T and TCNQ molecules (Figure 4). Although all functionals display this behavior, B3LYP clearly tends to overestimate the charge transferred in the ground state. For instance, at an intermolecular distance of 3.4 Å in the minimum-energy region, B3LYP predicts a CT of 0.14e, whereas BHHLYP, B2PLYP, and M06-2X yield a value of $\sim 0.07\text{e}$, and wB97X-D, M06-HF, and MP2⁸⁰ lead to charge transfers smaller than 0.07e. These results corroborate that increasing the percentage of HF exchange in the functional causes a decrease of the amount of charge transferred in the ground state of the 3T–TCNQ complex. This trend is in agreement with recent results reported for the tetrathiafulvalene–TCNQ D–A complex.⁷⁴ Therefore, the choice of standard hybrid functionals, such as B3LYP, for studying D–A complexes should be taken with care, since the amount of charge transferred in the ground state is artificially

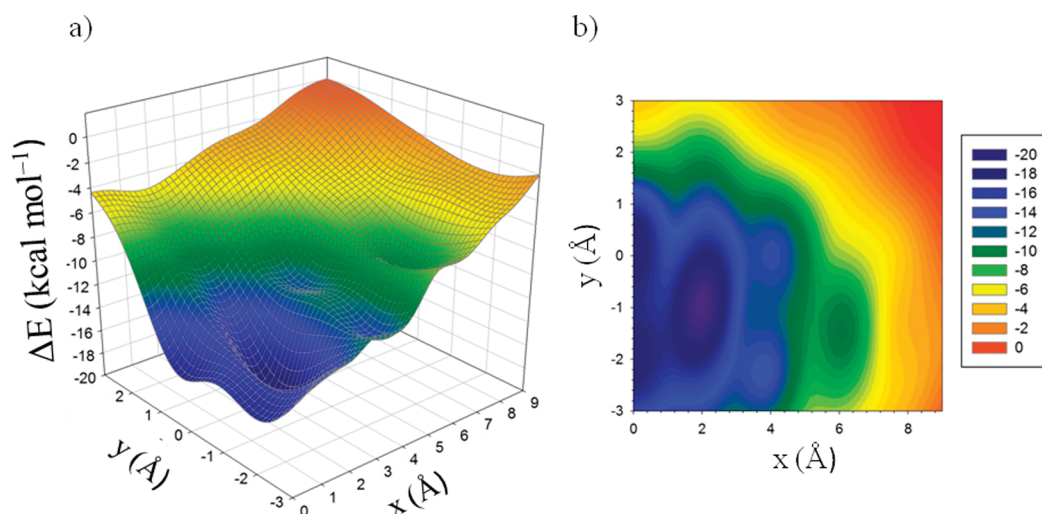


Figure 5. (a) Potential energy surface computed at the B3LYP-D/cc-pVDZ level for the 3T–TCNQ complex at a fixed intermolecular distance (z) of 3.39 Å. (b) Interaction energy contour map (kcal mol⁻¹).

Table 1. Optimized Geometries of the Three Most Stable Structures (1–3) Calculated for the 3T–TCNQ Complex^a

distances	B3LYP-D		wB97X-D	
	cc-pVDZ	cc-pVTZ	cc-pVDZ	cc-pVTZ
a	3.170	3.177	3.307	3.310
b	3.170	3.177	3.307	3.311
c	3.136	3.175	3.278	3.325
d	3.205	3.228	3.343	3.373
e	3.240	3.264	3.349	3.385
f	3.120	3.145	3.296	3.344
g	3.043	3.063	3.192	3.221
h	3.224	3.250	3.371	3.404
i	3.220	3.250	3.370	3.404
j	3.041	3.064	3.192	3.220

^a The shortest carbon–carbon intermolecular distances (in Å) are given.

high, and it can even increase upon full geometry optimization, as will be discussed next.

Optimized Geometries. With the aim of identifying the most stable structures of the 3T–TCNQ complex, a comprehensive exploration of the energy landscape along the x and y axis (see Figure 1) is required. Figure 5 depicts the potential energy surface and its respective contour map computed at the B3LYP-D/cc-pVDZ level by varying the distance along the x and y axis and by maintaining the monomers frozen at the minimum-energy intermolecular distance ($z = 3.39$ Å). The most stable structures are located in the dark-blue region of the potential energy surface. Three minimum-energy conformations were found and used subsequently as starting points for full geometry optimizations at B3LYP-D and wB97X-D levels in conjunction with cc-pVDZ and cc-pVTZ basis sets.

Geometry optimizations lead to the three minimum-energy structures (1–3) depicted in Table 1. Structures 1 and 3 differ

only in the relative displacement of the monomers along the short molecular y axis. In structure 1, the benzene ring of TCNQ is centered above the single C_{β} – C_{β} bond of the central thiophene ring of 3T, whereas in 3 the midpoint of the benzene ring is approximately placed above the central sulfur atom of 3T. In contrast, TCNQ in structure 2 is mainly displaced along the long molecular x -axis, and the benzene ring is now located above one of the inter-ring C_{α} – $C_{\alpha'}$ bonds of 3T. Geometry optimizations using cc-pVTZ provide structures almost identical to those obtained with cc-pVDZ, but with a much higher computational cost. The optimized intermolecular distances differ by less than 0.01 Å for structure 1 and by ~ 0.03 Å in structures 2 and 3 (Table 1). Although both B3LYP-D and wB97X-D functionals give rise to minimum-energy structures with identical orientations, it is noteworthy that they predict different intermolecular distances. The closest carbon–carbon intermolecular distances found at the B3LYP-D level are in the 3.0–3.2 Å range. Such values are considerably shorter than those computed at the CCSD(T) level for a benzene–hexafluorobenzene complex (3.5–3.6 Å).⁸¹ In contrast, wB97X-D leads to a larger separation between the interacting molecules with carbon–carbon distances ranging from 3.2 to 3.4 Å. Therefore, wB97X-D provides more reliable intermolecular distances than those obtained with B3LYP-D.

Table 2 collects the interaction energies computed by performing single-point calculations using the cc-pVTZ basis set and the fully relaxed cc-pVDZ geometries. Calculations using the cc-pVTZ optimized geometries lead to interaction energies that differ from those in Table 2 by less than 0.1 kcal mol⁻¹ (Table S1, Supporting Information). Calculations therefore show that the small structural changes found in passing from cc-pVDZ to cc-pVTZ optimized structures have no significant effect on the interaction energies and that cc-pVDZ optimized geometries can be safely used to calculate the interaction energies. This is of special interest when dealing with larger D–A complexes for which geometry optimizations using the cc-pVTZ basis set would be computationally prohibitive.

The CP-corrected interaction energies clearly indicate that structure 2 is the most stable with binding energies of -20.32 and -17.78 kcal mol⁻¹ at B3LYP-D and wB97X-D levels,

Table 2. Interaction Energies (kcal mol⁻¹) Calculated for the Most Stable Structures of the 3T–TCNQ Complex^a

method	1	2	3
B3LYP-D/cc-pVTZ ^b	-15.87 (-17.00)	-20.32 (-21.60)	-16.83 (-18.03)
SCS-MP2/cc-pVTZ ^b	-14.03 (-18.90)	-17.28 (-22.78)	-14.83 (-19.84)
wB97X-D/cc-pVTZ ^c	-14.88 (-15.85)	-17.78 (-18.86)	-15.66 (-16.65)
SCS-MP2/cc-pVTZ ^c	-13.95 (-18.14)	-16.65 (-21.33)	-14.62 (-18.86)

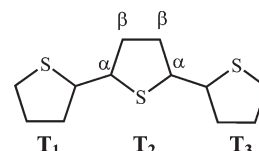
^aEnergies are corrected for BSSE using the counterpoise method (counterpoise uncorrected values are given within parentheses). ^bSingle-point energy calculations on the B3LYP-D/cc-pVDZ optimized geometries. ^cSingle-point energy calculations on the wB97X-D/cc-pVDZ optimized geometries.

respectively. Although the absolute difference between the interaction energies calculated for structures 1 and 3 using B3LYP-D and wB97X-D is almost constant (~1 kcal mol⁻¹), the relative energies of the three conformers vary with the functional. More specifically, 1 and 3 lie 4.45 and 3.49 kcal mol⁻¹ above 2 at the B3LYP-D level, whereas they are found 2.90 and 2.12 kcal mol⁻¹ higher in energy than 2 at the wB97X-D level. This arises because B3LYP-D stabilizes considerably 2 compared to wB97X-D results (see Table 2). It is interesting to note that both B3LYP-D and wB97X-D slightly overestimate the binding energies of the three minimum-energy structures when compared to SCS-MP2 energies. However, wB97X-D yields interaction energies that are closer to those obtained at SCS-MP2 level with differences around 1 kcal mol⁻¹.

Table 2 reveals that DFT-D and SCS-MP2 methods present significantly different BSSEs. B3LYP-D gives rise to errors in the 1.12–1.29 kcal mol⁻¹ range, whereas wB97X-D yields errors which range from 0.98 to 1.08 kcal mol⁻¹. The slightly smaller BSSEs found for wB97X-D are likely due to the fact that wB97X-D-optimized structures show longer intermolecular distances than B3LYP-D structures. SCS-MP2/cc-pVTZ interaction energies show the largest BSSEs, with values around 4.8 kcal mol⁻¹ for the three minimum-energy structures. These results contrast to those reported for relevant biological complexes, where SCS-MP2/cc-pVTZ was found to be almost free from BSSE due to error compensations.²¹ Our results therefore point to the importance of including BSSE corrections at the SCS-MP2 level to compute accurate interaction energies in D–A conjugated complexes as 3T–TCNQ. In contrast, DFT-D methods in combination with cc-pVTZ basis sets lead to rather small BSSEs.

Full geometry optimizations allow for a careful analysis of the changes occurred in the 3T and TCNQ monomers upon formation of the complex. The structural changes can be easily quantified by using the carbon–carbon single–double bond length alternation (BLA), which estimates the degree of aromatization/quinoidization along the conjugated backbone.⁸² The BLA parameter has been widely used to characterize the carbon skeleton in different families of oligothiophenes^{83–85} and is calculated for each thiophene ring as the difference between the length of the C_β–C_β bond and the average of the two C_α–C_β bonds (see sketch in Table 2). An aromatic ring is thus characterized by a positive BLA value, while a quinoid ring shows a negative BLA value.

Table 3 gathers the BLA values for 3T and the Mulliken charge transferred from 3T to TCNQ (q^{GS}) computed for structure 2 at the B3LYP-D/cc-pVDZ and wB97X-D/cc-pVDZ levels. The use of the cc-pVTZ basis set has almost no effect on the BLA values, and the charge distribution (Table S2). BLA data for isolated 3T are also included in Table 3 for comparison. At the B3LYP-D

Table 3. Bond Length Alternation Values (in Å) Computed at the B3LYP-D/cc-pVDZ and wB97X-D/cc-pVDZ Levels for Each Thiophene Ring of 3T in Structure 2 and for Isolated 3T^a


3T	T ₁	T ₂ ^b	T ₃ ^b	q ^{GS} (e)
B3LYP-D				
isolated	0.047	0.036	0.047	–
2	0.041	0.012	0.030	0.30
wB97X-D				
isolated	0.056	0.047	0.056	–
2	0.052	0.038	0.046	0.12

^aThe Mulliken charge transferred from 3T to TCNQ (q^{GS}) is also included. ^bTCNQ lies on top of the T₂ and T₃ thiophene rings, as illustrated in Table 1.

level, the BLA values computed for 3T in the complex are significantly smaller than those of isolated 3T. For instance, the BLA obtained for the central ring (T₂) decreases from 0.036 Å in isolated 3T to 0.012 Å in the complex. The reduction of BLA for the thiophene spine in the 3T–TCNQ complex indicates a loss of aromatic character of the conjugated backbone compared to isolated 3T. The partial quinoidization of 3T results from the charge transferred from 3T to TCNQ (0.30e, Table 2). In oligothiophenes, quinoidization of the carbon skeleton is produced upon injection of charges in oxidation/reduction processes.⁸⁴

In contrast to B3LYP-D, wB97X-D predicts more localized structures with larger BLA values, owing to the higher amount of HF-like exchange included in this functional. This behavior has been previously reported for related conjugated systems.⁸⁶ The comparison between the BLA values computed at the wB97X-D level for isolated 3T and for 3T in structure 2 reveals less marked differences than those obtained at the B3LYP-D level (Table 2). The less pronounced structural relaxation calculated for 3T at the wB97X-D level matches perfectly with the smaller amount of charge transferred (0.12e) from 3T to TCNQ.

Calculations therefore show that the B3LYP-D functional overestimates the charge transfer between the donor and the acceptor in the D–A complex. This leads to too short intermolecular distances and to too large changes on the structures of the monomers when the geometry of the complex is fully relaxed. The wB97X-D functional provides more accurate intermolecular distances and predicts smaller structural changes on the monomers forming the complex, due to the lower charge transfer, which is in better agreement with the description expected for weakly interacting D–A complexes.

Charge-Transfer Excited States. The characterization of CT excited states in D–A complexes is of utmost importance to understand the photophysical properties of these complexes and their potential use in organic solar cells. In order to gain insight into the performance of several functionals to treat CT excitations, the excitation energies of the lowest CT state were computed using the TD-DFT approach and the optimized ground-state molecular geometries of structures 1–3. Hybrid (B3LYP, BHHLYP), hybrid meta (M06-2X, M06-HF), double-hybrid (B2PLYP),

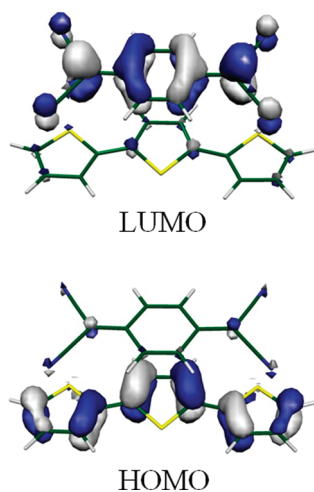


Figure 6. Electronic density isocontours (0.03 e bohr^{-3}) calculated at the CASSCF level for the frontier molecular orbitals of the 3T–TCNQ complex.

and LC-corrected (CAM-B3LYP, wB97X-D, LC-wPBE) functionals have been chosen for TD-DFT calculations. CASSCF/CASPT2 calculations have been carried out to benchmark the TD-DFT results. Figure 6 depicts the topologies of selected molecular orbitals as a guide to the description of the lowest CT excitation. Table 4 collects CT excitation energies and oscillator strengths computed using TD-DFT and CASPT2 methodologies.

According to the CASPT2 calculations, the lowest singlet excited state (S_1) is computed at 1.79, 1.72, and 1.83 eV above the ground state (S_0) for 1–3, respectively (Table 4). The intensity of the $S_0 \rightarrow S_1$ transition clearly depends on the configuration of the complex, with oscillator strength values of $f = 0.044$ for 1, $f = 0.176$ for 2, and $f = 0.059$ for 3. For the three structures, the CASSCF wave function is dominated by the promotion of an electron from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO). The HOMO is completely localized on the conjugated carbon–carbon skeleton of 3T, whereas the LUMO is located on the conjugated backbone of TCNQ (see Figure 6). Therefore, the $S_0 \rightarrow S_1$ transition corresponds to a CT excitation and is proposed to be responsible for the weak feature detected experimentally at 1.49 eV in chloroform.¹²

Note that CASPT2 calculations are performed in gas-phase while the experimental data are recorded in solution. Stein et al.⁸⁷ recently studied a family of aromatic donor–tetracyanoethylene complexes, for which optical data in gas phase and solution are available, and concluded that solvent models hardly stabilize the lowest CT states and are not able to reproduce the experimental solvent effect that lowers the CT excitation energies by 0.32 eV on average. These authors suggest that the only reasonable way to proceed in these cases is to assume that subtraction of 0.32 eV from the gas phase calculations makes them comparable to experiments in solvent. Using this assumption, the CASPT2 results match now perfectly the experimental CT excitation found at 1.49 eV for the 3T–TCNQ complex. Hence, the CASPT2 gas-phase energies can be reasonably employed to benchmark the TD-DFT results.

As can be seen in Table 4, all functionals predict larger CT excitation energies for structure 2 than for 1 and 3, which contrasts to CASPT2 results. Concerning the intensities, all

Table 4. $S_0 \rightarrow S_1$ Vertical Excitation Energies (eV) Computed for Structures 1–3 of the 3T–TCNQ Complex

method ^a	1 ^b	2 ^b	3 ^b
CASPT2	1.79 (0.044)	1.72 (0.176)	1.83 (0.059)
B3LYP	1.25 (0.070)	1.46 (0.112)	1.32 (0.075)
BHLYP	1.71 (0.067)	1.88 (0.121)	1.72 (0.060)
B2PLYP	1.30 (0.060)	1.48 (0.115)	1.29 (0.053)
M06-2X	1.73 (0.062)	1.89 (0.109)	1.73 (0.053)
M06-HF	2.63 (0.062)	2.68 (0.125)	2.65 (0.053)
LC-wPBE	2.49 (0.055)	2.54 (0.111)	2.52 (0.045)
CAM-B3LYP	1.75 (0.063)	1.91 (0.114)	1.76 (0.055)
wB97X-D	1.81 (0.062)	1.96 (0.115)	1.83 (0.055)

^a All calculations were performed with the cc-pVDZ basis set. ^b Oscillator strength (f) values are given within parentheses.

functionals predict the highest oscillator strength for structure 2 in good agreement with CASPT2 oscillator strengths. CT excitation energies using the TD-DFT approach are analyzed according to the nature of the functional (hybrid, double hybrid, hybrid meta, and long-range functionals) and the amount of HF-like exchange. B3LYP, which incorporates a small percentage (20%) of HF-like exchange, underestimates the CASPT2 CT excitation energies with deviations of up to -0.54 eV. In contrast, the CT excitations computed by BHLYP (50% of HF-like exchange) yield CT excitation energies closer to those computed at the CASPT2 level. The largest deviation is found for structure 2, for which the CT state is calculated 0.16 eV higher in energy.

The double-hybrid B2PLYP functional also underestimates the CT excitation energies with deviations from -0.24 (2) to -0.54 eV (3). The underestimation of the CT excitations results from the perturbative correction term ($\Delta_{(D)}$) computed through the CIS(D) method (see Theoretical and Computational Details Section) that stabilizes the CT state. Note that B2PLYP has a 53% of HF-like exchange, and thus, when the $\Delta_{(D)}$ term is not added, the uncorrected CT excitation energies are similar to those computed at BHLYP level.

The hybrid meta M06-2X functional predicts CT excitation energies with accuracy similar to that obtained at BHLYP level (Table 4). The resemblance in the CT excitation energies is due to the fact that M06-2X incorporates a 54% of HF-like exchange very close to the 50% included in BHLYP. In contrast, M06-HF implies a 100% of HF-like exchange and dramatically overestimates the CT excitation energies with a deviation of up to 0.96 eV with respect to CASPT2 values. These results are consistent with those recently reported by Li et al. showing that for CT transitions at intermediate interelectronic separations M06-2X performs notably, whereas M06-HF fails completely.⁸⁸

LC-wPBE, CAM-B3LYP, and wB97X-D represent a more sophisticated class of functionals based on the long-range corrected scheme. However, the behavior of LC-wPBE is quite different with respect to their homologous CAM-B3LYP and wB97X-D due to the amount of HF-like exchange. LC-wPBE has a 0% of HF-like exchange at short-range and a 100% at long-range and overestimates severely the CT excitation energies with deviations ranging from 0.69 to 0.80 eV compared to that of the CASPT2 results. CAM-B3LYP shows a 19 and 65% of HF-like exchange at short and long interelectronic separations, whereas wB97X-D includes a 22 and 100%, respectively. Both functionals predict similar CT excitation energies, the largest deviation being found for structure 2 with differences of 0.19 and 0.24 eV for

CAM-B3LYP and wB97X-D, respectively. wB97X-D performs slightly better than CAM-B3LYP for structures 1 and 3, for which it provides CT excitations in excellent accord with CASPT2 data.

Overall, wB97X-D provides the best performance with the smallest deviations in the CT excitation energies in comparison to CASPT2 results, although CAM-B3LYP, BHHLYP, and M06-2X functionals also behave accurately enough. The results evidence that the CT transition in the 3T–TCNQ complex occurs at intermediate interelectronic separations, and thus, not only long-range corrected functionals describe correctly this transition but also hybrid and hybrid meta functionals with an appropriate portion of HF-like exchange. It should be also mentioned that the size of the basis set has no special effect on the CT excitation energies. For instance, almost identical excitation energies are obtained at wB97X-D/cc-pVDZ (1.96 eV) and wB97X-D/cc-pVTZ (1.98 eV) levels for structure 2.

CONCLUSIONS

We have presented a detailed quantum chemical investigation of the terthiophene–tetracyanoquinodimethane complex as model for π – π D–A systems. The study has focused on the performance of several functionals to treat π – π intermolecular interactions, the distribution of charge in the ground state of the complex, the identification and the characterization of the most stable supramolecular structures, and the description of the CT excitation found in these D–A complexes.

Density functionals including the dispersion term (DFT-D) treat in an accurate way the π – π interactions yielding potential energy curves similar to those computed at the SCS-MP2 level. The distribution of the charge in the ground state is clearly determined by the portion of HF-like exchange in the functional. The charge transferred from the donor to the acceptor in the ground state is inversely proportional to the percentage of HF-like exchange. Hence, a too large charge transfer is predicted when using the common B3LYP functional. When characterizing the most stable structures of the complex, wB97X-D functional predicts reliable intermolecular distances, small structural changes in the monomers forming the complex, and a small amount of charge transferred from 3T to TCNQ in the ground state, in agreement with the chemically intuitive description of weak interacting D–A complexes. High-level ab initio CASSCF/CASPT2 calculations have enabled to benchmark the TD-DFT results for excited states. Long-range corrected functionals, such as CAM-B3LYP and wB97X-D, as well as the hybrid BHHLYP are able to treat accurately the CT excitations for the 3T–TCNQ complex, yielding values very similar to those obtained by the CASPT2 method.

Overall, the wB97X-D functional displays the best performance to treat π – π intermolecular interactions and CT excitations in a balanced way, for the 3T–TCNQ complex, thanks to the combination of dispersion and long-range corrected terms. Hence, the wB97X-D functional might be reliably used to further investigate π – π D–A complexes. Calculations show that the use of a cc-pVTZ basis set is needed to predict accurate interaction energies with small BSSEs, whereas reliable optimized geometries and excitation energies are obtained using the less-demanding cc-pVDZ basis set. Theoretical information for larger D–A complexes would be very useful to guide experimentalists in the design of fine-tuned D–A complexes for organic solar cells.

ASSOCIATED CONTENT

Supporting Information. B3LYP-D/cc-pVTZ and wB97X-D/cc-pVTZ counterpoise corrected and uncorrected potential energy curves; CHelpG charge transferred from 3T to TCNQ as a function of the functional and the intermolecular distance; B3LYP-D/cc-pVTZ and wB97X-D/cc-pVTZ interaction energies as a function of the optimization level; and influence of the basis set on BLA and charge transferred. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: enrique.orti@uv.es.

ACKNOWLEDGMENT

Financial support by the MICINN of Spain (projects CTQ2009-08970, CTQ2011-27253, and Consolider-Ingenio CSD2007-00010 in Molecular Nanoscience) is gratefully acknowledged. J.A. thanks the MICINN of Spain for a FPI doctoral grant. The work in Mons was supported by the Interuniversity Attraction Pole program of the Belgian Federal Science Policy Office (PAI 6/27) and FNRS-FRFC. D.B. is a FNRS Research Director.

REFERENCES

- (1) Sariciftci, N. S.; Smilowitz, L.; Heeger, A. J.; Wudl, F. *Science* **1992**, *258*, 1474.
- (2) Schmidt-Mende, L.; Fechtenkötter, A.; Mullen, K.; Moons, E.; Friend, R. H.; MacKenzie, J. D. *Science* **2001**, *293*, 1119.
- (3) Brédas, J. L.; Norton, J. E.; Cornil, J.; Coropceanu, V. *Acc. Chem. Res.* **2009**, *42*, 1691.
- (4) Dimitrakopoulos, C. D.; Malefant, P. *Adv. Mater.* **2002**, *14*, 99.
- (5) Murphy, A. R.; Fréchet, J. M. J. *Chem. Rev.* **2007**, *107*, 1066.
- (6) Mishra, A.; Ma, C.-Q.; Bäuerle, P. *Chem. Rev.* **2009**, *109*, 1141.
- (7) Perepichka, I. F.; Perepichka, D. F. *Handbook of Thiophene-Based Materials: Applications in Organic Electronics and Photonics*; Wiley: Weinheim, Germany, 2009.
- (8) Ferraris, J.; Cowan, D. O.; Walatka, V.; Perlstein, J. H. *J. Am. Chem. Soc.* **1973**, *95*, 948.
- (9) Coleman, L. B.; Cohen, M. J.; Sandman, D. J.; Yamaguchi, F. G.; Garito, A. F.; Heeger, A. J. *Solid State Commun.* **1973**, *12*, 1125.
- (10) Aziz, E. F.; Vollmer, A.; Eisebitt, S.; Eberhardt, W.; Pingel, P.; Neher, D.; Koch, N. *Adv. Mater.* **2007**, *19*, 3257.
- (11) Braun, K.-F.; Hla, S. W. *J. Chem. Phys.* **2008**, *129*, 064707.
- (12) Panda, P.; Veldman, D.; Sweelssen, J.; Bastiaansen, J. J. A. M.; Langeveld-Voss, B. M. W.; Meskers, S. C. J. *J. Phys. Chem. B* **2007**, *111*, 5076.
- (13) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, U.K., 1997.
- (14) Pitonak, M.; Riley, K. E.; Neogady, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *9*, 1636.
- (15) Sinnkrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200.
- (16) Cremer, D. *Encyclopedia of Computational Chemistry*; P. v. R. Schleyer, Ed.; Wiley: New York, 1998; Vol.3, p 1706.
- (17) Hobza, P.; Selzle, H. L.; Schkage, E. W. *J. Chem. Phys.* **1996**, *100*, 18790.
- (18) Sinnkrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887.
- (19) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. *J. Chem. Phys.* **2004**, *120*, 647.

- (20) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
- (21) Antony, J.; Grimme, S. *J. Phys. Chem. A* **2007**, *111*, 4862.
- (22) Kristyan, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175.
- (23) Hobza, P.; Sponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *11*, 1315.
- (24) Allen, M.; Tozer, D. *J. Chem. Phys.* **2002**, *117*, 11113.
- (25) Dion, M.; Rydberg, H.; Scroder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.
- (26) Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. B* **2010**, *82*, 081101.
- (27) Cooper, V. R. *Phys. Rev. B* **2010**, *81*, 161104.
- (28) Vydrov, O. A.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *132*, 164113.
- (29) Vydrov, O. A.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *133*, 244103.
- (30) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2007**, *127*, 154108.
- (31) Román-Pérez, G.; Soler, J. M. *Phys. Rev. Lett.* **2009**, *103*, 096102.
- (32) Heßelmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 14103.
- (33) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem.* **2006**, *125*, 194101.
- (34) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
- (35) Elstner, M.; Hobza, P.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149.
- (36) Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *115*, 515.
- (37) Zimmerli, U.; Parrinello, M.; Koumoutsakos, P. *J. Chem. Phys.* **2004**, *120*, 5149.
- (38) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
- (39) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.
- (40) Ducere, J.-M.; Cavallo, L. *J. Phys. Chem. B* **2007**, *111*, 13124.
- (41) Tapavicza, E.; Lin, I.-C.; von Lilienfeld, O. A.; Tavernelli, I.; Coutinho-Neto, M. D.; Rothlisberger, U. *J. Chem. Theory Comput.* **2007**, *3*, 1673.
- (42) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287.
- (43) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2010**, *6*, 107.
- (44) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 291.
- (45) Dreuw, A.; Weisman, M.; Head-Gordon, M. *J. Phys. Chem.* **2003**, *119*, 2943.
- (46) Serrano-Andrés, L.; Fülischer, M. *J. Am. Chem. Soc.* **1998**, *120*, 10912.
- (47) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2004**, *120*, 8425.
- (48) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 234109.
- (49) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51.
- (50) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. *J. Chem. Phys.* **2008**, *128*, 044118.
- (51) Chai, J.-D.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 084106.
- (52) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
- (53) Jacquemin, D.; Perpète, E. A.; Ciofini, I.; Adamo, C. *Theor. Chem. Acc.* **2011**, *128*, 127.
- (54) Grimme, S.; Neese, F. *J. Chem. Phys.* **2007**, *127*, 154116.
- (55) Goerigk, L.; Moellmann, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4611.
- (56) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397.
- (57) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
- (58) Neese, F. *ORCA 2.7.0*; University of Bonn, Bonn, Germany; <http://www.thch.uni-bonn.de/tc/orca/>.
- (59) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.
- (60) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (61) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (62) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
- (63) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13126.
- (64) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
- (65) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331.
- (66) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143.
- (67) Andersson, K.; Malmqvist, P.-Å.; Roos, B. *J. Chem. Phys.* **1992**, *96*, 1218.
- (68) Andersson, K.; Roos, B. Part I, Advanced Series in Physical Chemistry. *Modern Electron Structure Theory*; Yarkony, R., Ed.; World Scientific, Singapore, 1995; Vol. 2, p 55.
- (69) Roos, B.; Fülischer, M. P.; Malmqvist, P.-Å.; Merchán, M.; Serrano-Andrés, L. *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*; S. R. Langhoff, Ed.; Kluwer Academic, Dordrecht, The Netherlands, 1995, p 357.
- (70) Malmqvist, P.-Å. *Int. J. Quantum Chem.* **1986**, *30*, 479.
- (71) Malmqvist, P.-Å.; Roos, B. *Chem. Phys. Lett.* **1989**, *155*, 189.
- (72) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.-Å.; Pedersen, T.; Pitonak, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Veryazov, V.; Lindh, R. *J. Comput. Chem.* **2010**, *31*, 224.
- (73) Head-Gordon, M.; Rico, R. J.; Oumi, M.; Lee, T. *J. Chem. Phys. Lett.* **1994**, *219*, 21.
- (74) Sini, G.; Sears, J. S.; Brédas, J.-L. *J. Chem. Theory Comput.* **2011**, *7*, 602.
- (75) Riley, K. E.; Pitonak, M.; Cerny, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 66.
- (76) Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P. *Chem. Rev.* **2011**, *110*, 5023.
- (77) Mulliken, R. S. *J. Am. Chem. Soc.* **1950**, *72*, 600.
- (78) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
- (79) Karthikeyan, S.; Sedlak, R.; Hobza, P. *J. Phys. Chem. A*, DOI: 10.1021/jp1112476.
- (80) The charges are calculated at the MP2 level because the spin component scaled procedure is only applied to correct the energy.
- (81) Tsuzuki, S.; Uchimaru, T.; Mikami, M. *J. Phys. Chem. A* **2006**, *110*, 2027.
- (82) Brédas, J. L. *J. Chem. Phys.* **1985**, *82*, 3808.
- (83) Aragón, J.; Viruela, P. M.; Ortí, E. *J. Mol. Struct. (Theochem)* **2009**, *912*, 27.
- (84) Aragón, J.; Viruela, P. M.; Ortí, E.; Osuna, R. M.; Vercelli, B.; Zotti, G.; Hernández, V.; López Navarrete, J. T.; Henssler, J. T.; Matzger, A. J.; Suzuki, Y.; Yamaguchi, S. *Chem.—Eur. J.* **2010**, *16*, 5481.
- (85) Aragón, J.; Viruela, P. M.; Gierschner, J.; Ortí, E.; Milián-Medina, B. *Phys. Chem. Chem. Phys.* **2011**, *13*, 1457.
- (86) Sancho-García, J. C.; Pérez-Jiménez, A. *J. Phys. Chem. Chem. Phys.* **2007**, *9*, 5874.
- (87) Stein, T.; Kronik, L.; Baer, R. *J. Am. Chem. Soc.* **2009**, *131*, 2818.
- (88) Li, R.; Zheng, J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12697.

Definitive Benchmark Study of Ring Current Effects on Amide Proton Chemical Shifts

Anders S. Christensen,* Stephan P. A. Sauer, and Jan H. Jensen*

Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

S Supporting Information

ABSTRACT: The ring current effect on chemical shifts of amide protons ($\Delta\delta_{\text{RC}}$) is computed at the B3LYP/6-311++G(d,p)//B3LYP/aug-cc-pVTZ level of theory for 932 geometries of dimers of *N*-methylacetamide and aromatic amino acid side chains extracted from 21 different proteins. These $\Delta\delta_{\text{RC}}$ values are scaled by 1.074, based on MP2/cc-pVQZ//B3LYP/aug-cc-pVTZ chemical shift calculations on four representative formamide/benzene dimers, and are judged to be accurate to within 0.1 ppm based on CCSD(T)/CBS//B3LYP/aug-cc-pVTZ calculations on formamide. The 932 scaled $\Delta\delta_{\text{RC}}$ values are used to benchmark three empirical ring current models, including the Haigh–Mallion model used in the SPARTA, SHIFTX, and SHIFTS chemical shift prediction codes. Though the RMSDs for these three models are below 0.1 ppm, deviations up to 0.29 ppm are found, but these can be decreased to below 0.1 ppm by changing a single parameter. The simple point-dipole model is found to perform just as well as the more complicated Haigh–Mallion and Johnson–Bovey models.

1. INTRODUCTION

Prediction of chemical shifts in proteins based on protein structure serves many uses in structure verification or fast generation of structures in accordance with relatively inexpensive experimental nuclear magnetic resonance (NMR) data.^{1–5} The most popular protein chemical prediction software packages include the SHIFTX,⁶ SHIFTS,^{7,8} SPARTA,⁹ and PROSHIFT¹⁰ servers. These programs use empirical models that relate various features of protein structure, such as secondary structure, hydrogen bond geometry, and ring current effects, to changes in chemical shifts. The contributions from several different sources of chemical shift perturbations are in all cases assumed to work additively. These small additive terms are in many cases given as classical approximations to well-known physical interactions. For example, SPARTA, SHIFTX, and SHIFTS use the approximation due to Haigh and Mallion¹¹ to model the changes in chemical shifts due to ring current effects in the aromatic side chains of phenylalanine, tyrosine, tryptophan, and histidine residues. The Haigh–Mallion model contains a single adjustable parameter for each side chain, which must be parametrized from a data set. In the case of SHIFTX, these parameters are obtained by a data mining approach, which correlates experimental chemical shifts with corresponding, known X-ray protein structures, based on a series of predefined physical and empirical terms.⁶ In the case of SHIFTS, the parameters are obtained by fitting parameters for a set of known physical terms, which relates protein structure and chemical shifts to a data set which combines empirical chemical shift data as well as data obtained through quantum chemical calculations.¹² Ultimately, these fitting methods include uncertainties from many terms in the underlying approximations of the fit, as well as the uncertainties connected to the experimental data. It is thus unclear how accurate the obtained parameters are in reproducing the underlying physics of the system.¹³

Other methods exist to approximate the ring current effect, most notably^{11,14} the Johnson–Bovey model¹⁵ and the simpler

point-dipole model due to Pople.^{16,17} In general, the three methods describe the change in chemical shift due to a nearby aromatic ring, formally as

$$\Delta\delta_{\text{RC}} = i B G \quad (1)$$

where G is a geometric factor, which depends on the spatial orientation and distance of the ring relative to the proton, B is a “natural constant” denoting the ring current intensity for a benzene ring, and i is the ring current intensity relative to that of a benzene ring, such that $i_{\text{benzene}} \equiv 1$. A thorough description of the three models mentioned above can be found in the Supporting Information. It has previously been attempted to derive analytical expressions for the values of i and B for different functional forms of G , but these have not been successful in reproducing experimental results.¹¹ Consequently, various numerical methods have been widely used to obtain the intensity values.

The study by Case¹⁸ is one of the few that has addressed these issues using *ab initio* calculations. A methane molecule is used to probe the chemical shift perturbation due to a nearby aromatic ring, with the chemical shift calculated at the CSGT/PW91/IGLO-III level of theory. However, it is not clear whether the level of theory used at that time (1995) is sufficiently accurate, and second, it is unknown whether the ring current parameters obtained for a methane proton can directly be transferred to amide protons. The parameters obtained by Case are used in the SPARTA program.⁹

In this study, we use CCSD(T) and MP2 methods to benchmark the accuracy of $\Delta\delta_{\text{RC}}$ calculations at the B3LYP/6-311++G(d,p)//B3LYP/aug-cc-pVTZ level of theory. This level of theory is then used to compute nearly 1000 representative

Received: April 15, 2011

Published: June 01, 2011

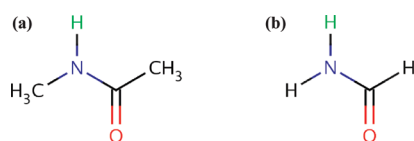


Figure 1. The molecules used as probes for the ring current effects on amide protons: *trans*-*N*-methylacetamid (NMA) (a) and formamide (FMA) (b). The probe nucleus for which the shielding constants are calculated are the amide proton of NMA and the amide proton *trans* to the C=O bond in FMA, here marked in green color.

$\Delta\delta_{RC}$ values, which, in turn, are used to obtain parameters for three empirical $\Delta\delta_{RC}$ models.

The paper is organized as follows. First, we describe the computational methodology used to isolate the ring current effect and to obtain data sets, against which the ring current intensity parameters are fitted. Then, we benchmark various levels of theory, in order to estimate error bounds on our data. This is followed by a presentation of the obtained intensities and a comparison to the intensities obtained by other authors.

2. COMPUTATIONAL METHODOLOGY

2.1. Isolating the Ring Current Effect. As a probe nucleus, for which the isotropic shielding is calculated using quantum chemical methods, the amide proton of a *trans*-*N*-methylacetamide (NMA; Figure 1a) molecule is used in order to provide a small and inflexible molecule with a high degree of resemblance to the amide group of the protein backbone. For more computationally demanding calculations, such as MP2, CCSD, and CCSD(T) calculations, the two methyl groups are removed in order to save computational time, and the probe nucleus is then the amide proton *trans* to the C=O bond in formamide (FMA; Figure 1b).

A large number of dimer systems (see next subsection), consisting of an amide probe molecule and a nearby aromatic ring in different conformations is constructed. Following the general approach of Boyd and Skrynnikov,¹⁹ the absolute shielding of the probe hydrogen in the dimer system can be written as

$$\sigma_{\text{H}}^{\text{Dimer}} = \Delta\sigma_{\text{Conformation}} + \Delta\sigma_{\text{Local}} + \Delta\sigma_{\text{RC}} \quad (2)$$

The chemical shift of the probe hydrogen atom in the dimer system will, apart from the ring current effect, also be influenced by the exact geometry and type of the probe molecule (described in the $\Delta\sigma_{\text{Conformation}}$ term) as well as any possible interactions with the aromatic moiety, such as electrostatic forces, possible hydrogen bonding, spin–spin repulsion, and other effects which can be difficult to quantify and separate (described in the $\Delta\sigma_{\text{Local}}$ term). Finally, the chemical shift perturbation due to the aromatic ring is approximated as $\Delta\sigma_{\text{RC}}$.

Reference systems for each dimer system, which have approximately identical local interactions between the molecules, apart from the ring current effect, are constructed in order to filter out these hard-to-quantify effects (see Figure 2). These are modeled as corresponding dimer systems, where the aromatic ring has been replaced by an olefinic analogue. The definition of an olefinic analogue here is an aromatic ring which, by the addition of two hydrogen atoms, has lost its aromaticity. The protonation is done such that the planar geometry of the ring is still enforced, causing deviations in the spatial positions relative to the corresponding aromatic ring dimer to be negligible. The olefinic

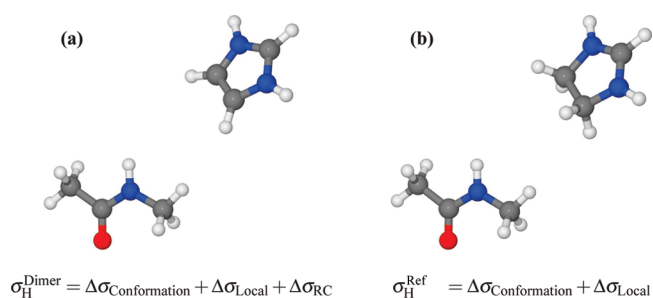


Figure 2. Two example geometries demonstrating the two different dimers used in the calculation scheme to isolate the ring current effect for one amide-ring conformation. The shown geometries correspond to the amide proton of ILE64 and the side chain of HIS69 in the HIV-1 protease, PDB-code 2I4V. In part a, the chemical shift of the probe nucleus is determined by the conformation of the NMA molecule, electrostatic and spin–spin repulsion interactions to the positively charged imidazolium molecule, and, by comparison, a small ring current interaction. In part b, the aromaticity of the imidazolium is broken, but the spatial distribution of charge as well as the internal conformation of NMA is approximately identical to those found in part a.

analogue is placed such that the ring center corresponds to the center of the aromatic ring, relative to the probe hydrogen, and the coordinates of the carbon/nitrogen atoms are matched as closely as possible. This approach ensures that $\Delta\sigma_{\text{Conformation}}$ and $\Delta\sigma_{\text{Local}}$ are largely retained, while $\Delta\sigma_{\text{RC}}$ is removed. Using this substitution scheme, the absolute shielding of the hydrogen atom in the reference system can be written as

$$\sigma_{\text{H}}^{\text{Ref}} = \Delta\sigma_{\text{Conformation}} + \Delta\sigma_{\text{Local}} \quad (3)$$

which enables us to estimate the ring current contribution to the chemical shift due to the aromatic ring as

$$\Delta\delta_{\text{RC}} = -\Delta\sigma_{\text{RC}} \approx \sigma_{\text{H}}^{\text{Ref}} - \sigma_{\text{H}}^{\text{Dimer}} \quad (4)$$

The aromatic rings studied here are equivalent to the rings found in the aromatic protein side chains. See Table 1 for an overview of the used molecules. Sketches are shown in the Supporting Information as well.

2.2. Construction of Test Systems. Dimers consisting of an amide probe and an aromatic ring were generated from a data set of 21 protein structures obtained from the RCSB Protein Data Bank (PDB),²⁰ in order to ensure that only realistic conformations were used in the QM calculations.

The structures used were (listed by PDB code): 1F94, 1GK1, 1IGD, 1JYQ, 1JYR, 1JYU, 1Q3E, 1QJP, 1VJC, 1XA5, 1ZJK, 2ACO, 2B6C, 2DS7, 2DRJ, 2ETL, 2F47, 2FZG, 2GOL, 2I4D, and 2I4V. Since the used structures were experimental X-ray structures, no hydrogen atoms were present in the structures, and PDB2PQR 1.5^{21,22} was used to protonate the structures in order to obtain hydrogen atom positions. From all of these protein structures, we selected systems where the center of an aromatic ring was within a cutoff distance of 7 Å from an amide proton. This resulted in a total of 932 different dimer conformations (see Table 1). For each of these conformations, a dimer was created with a simpler aromatic ring in place of the aromatic side chain with the ring centers at identical coordinates and in the same plane. A directional vector was used to align the rotation of the ring in the plane, in order to have closely matching coordinates for the heavy atoms. For tyrosine, the center to oxygen vector was used. For benzene, the center to C^γ to C^δ vector was used.

Table 1. List of the Side Chain Approximations Used in This Work and Their Olefinic Analogues^a

side chain	analogue	olefinic analogue	# dimers
phenylalanine	benzene	1,4-cyclohexadiene	276
tyrosine	phenol	cyclohexa-1,4-diene-1-ol	172
tryptophan	indole	2,3,5,6-tetrahydroindole	113
histidine	imidazole	4,5-dihydroimidazole	185
histidine ⁺	imidazolium	4,5-dihydroimidazolium	174

^aThe residue type is listed along with the aromatic and olefinic analogues, as well as the total number of different NMA/ring dimer conformations in each data set.

For histidine (both in the charged and neutral state), the center to C^{ε1} vector was used. The same histidine–amide group pairs were used to generate the dimers for both charged and neutral histidine conformations. The N^{ε2} nitrogen atom was in all cases of neutral histidine assumed to be the deprotonated nitrogen.

A given backbone amide group within the 7 Å range from the aromatic ring was substituted by an NMA or an FMA molecule, with the nitrogen atoms at identical coordinates. Furthermore, the N–H vector and the C(=O)–N–H plane were also aligned. See Table 1 for the number of dimers for each ring type.

If the aromatic ring corresponded to the ring of the side chain of the previous residue, with respect to the investigated amide group, the dimer construction scheme occasionally caused a clash between the extra hydrogens, where the C^β atom was previously located. Other conformations also gave rise to unphysical conformations, due to clashes between the inflexible subunits of the constructed dimers. To avoid computational artifacts from these, all dimers with a shortest intermolecular distance of 3.4 Å or less were discarded, since 3.4 Å is twice the van der Waal radius of the largest atom (carbon) in the system.²³ NMR shielding constants were then calculated for the dimer systems.

2.3. Basis Set Extrapolations and Correlation Effects. In this work, density functional theory (DFT; and the very popular B3LYP functional^{24,25}) is used to obtain NMR shielding constants. Due to the partly semiempirical nature of the approximated exchange–correlation functionals used, B3LYP data cannot in general be expected to show convergence toward experimental shielding values or values obtained at very accurate levels of theory when increasing the basis set size.²⁶ It is, however, often the case that a small error can be obtained in calculated DFT chemical shielding constants, compared to high-level correlated wave function methods, if a simple linear correction or scaling factor is applied to the DFT data.²⁷ In this work, a comparison of B3LYP to high-level correlated methods is used to obtain such a linear scaling factor.

Unfortunately, CCSD(T) calculations with an appropriate basis set are still not possible for the FMA/benzene dimer, which has 70 electrons. Instead, we benchmark the shielding constants for FMA alone at the B3LYP, MP2, CCSD, and CCSD(T) levels of theory, in order to allow us to estimate error bounds to shielding constants obtained at levels of theory less accurate than CCSD(T).

For complete basis set limit (CBS) estimates, we use the approach of Moon and Case²⁶ and Kupka et al.^{27,28} By using Dunning's correlation consistent basis sets²⁹ (cc-pVxZ; where $x \in \{D, T, Q, 5, 6, \dots\}$ is the valence orbital splitting in the basis set), it is possible to carry out calculations using a sequence of basis sets of well-defined, systematic increasing quality. Kupka et al.

extrapolate calculated NMR shielding values toward infinite basis set size with a three parameter exponential decreasing function:

$$\sigma(x) = \sigma(\infty) + A \exp(-x/B) \quad (5)$$

where $\sigma(x)$ is the shielding obtained using a basis set with the valence orbital splitting number of x and $\sigma(\infty)$, A and B are the fitting parameters, with $\sigma(\infty)$ being the estimated shielding in the complete basis limit. A nonlinear least-squares Marquardt–Levenberg algorithm^{30,31} is used to fit the parameters.

Jensen has constructed a set of basis sets for the purpose of Hartree–Fock (HF) and DFT NMR shielding calculations, called the polarization consistent pcS- n basis sets.³² For basis sets of similar valence orbital splitting, the pcS- n basis sets contain more basis functions of low angular momentum, compared to the Dunning-type basis sets. pcS-1 is a double- ζ quality basis set, pcS-2 is triple- ζ , and so forth. When estimating the complete basis limit based on the pcS- n basis sets, a value of $x = n + 1$ is thus used in eq 5.

Last, we compare the proton chemical shift of the amide proton trans to the C=O bond in FMA at the CCSD(T)/CBS level of theory to the experimental value, in order to verify that CCSD(T)/CBS is, in fact, a reliable method. Inferring the experimental gas-phase ¹H shielding values from CH₄ ($\sigma_{\text{H}} = 30.61$ ppm³³), an experimental value of $\sigma_{\text{H}} = 26.24$ ppm in the gas phase³⁴ at 483 K is obtained. At this temperature, thermal motion cause rapid switching of the two N-amide protons and the peaks are not separable, so this value has to be considered as an average over the two proton chemical shifts.³⁴

It is well-known^{35–38} that a zero-point vibrational correction (ZPVC) has to be added to equilibrium geometry *ab initio* shielding constants in order to obtain close agreement to experimental data. This vibrational averaging correction can easily be calculated using the method of Kern and Matcha.³⁹ While we had preferred to carry out a ZPVC calculation at the same level of theory as the geometry optimization of the molecules used throughout this work, no program is currently capable of automatically computing a ZPVC at the DFT level of theory with Gaussian-type basis sets. In the following, the ZPVC is calculated at the MP2/cc-pVQZ level of theory instead.

2.4. Software. All geometries were minimized at the B3LYP/aug-cc-pVTZ level of theory using Gaussian 03,⁴⁰ except when otherwise noted. All DFT calculations of NMR shielding constants were carried out using Gaussian 03. MP2/6-311++G(d,p) NMR calculations were also carried out in Gaussian 03, while the calculation of MP2 shielding constants using Dunning's correlation consistent basis sets²⁹ and the polarization consistent pcS- n and aug-pcS- n basis sets of Jensen³² were carried out with Turbomole 6.2.⁴¹ All calculations at the CCSD and CCSD(T) levels of theory were carried out using CFOUR 1.0.⁴² All NMR shielding constants are calculated using the Gauge-Including Atomic Orbital formulation.^{43–45} For the calculations of the ZVPC to the FMA chemical shifts, the equilibrium geometry of a planar FMA molecule was obtained at the MP2/cc-pVQZ level of theory with CFOUR 1.0, exploiting the C_s symmetry of the molecule. From this equilibrium geometry, the ZVPC to the NMR isotropic shielding was subsequently calculated at the MP2/cc-pVQZ level of theory using the method of Kern and Matcha³⁹ as implemented in CFOUR 1.0.

Table 2. Absolute Isotropic Chemical Shielding of the *cis-N*-amide Proton in Gas-Phase FMA at the B3LYP, MP2, CCSD, and CCSD(T) Levels of Theory Using Dunning's Correlation Consistent Basis Sets and the Polarization Consistent Shielding Basis Sets of Jensen^a

basis set	size	method			
		CCSD(T)	CCSD	MP2	B3LYP
cc-pVDZ	57	28.06	28.09	27.90	27.67
cc-pVTZ	132	27.29	27.35	27.16	27.17
cc-pVQZ	255	26.92	27.00	26.80	26.94
cc-pV5Z	438	26.78	26.86	26.65	26.83
$\sigma_{\text{cc-pVxZ}}(\infty)$		26.64	26.73	26.50	26.73
pcS-0	44	29.32	29.36	29.31	28.88
pcS-1	66	27.55	27.58	27.40	27.29
pcS-2	141	27.02	27.09	26.89	26.91
pcS-3	321	26.75	26.83	26.62	26.77
$\sigma_{\text{pcS-}n}(\infty)$		26.67	26.78	26.57	26.75
$\sigma_{\text{expt}}(\text{gas})$			26.24		

^a All values are given as ppm. The experimental value is obtained at 483 K.³⁴ $\sigma(\infty)$ is obtained using eq 5 and fitted over all values in the corresponding series of basis sets. The size indicates the number of basis functions in the system at the given basis set size. All shielding constants are given in ppm.

3. RESULTS

3.1. Correlation and Basis Set Effects on the Chemical Shift of the (N–)H Proton in Formamide. It is currently not feasible to perform a complete basis set study at the CCSD(T) level for $\Delta\delta_{\text{RC}}$ of a FMA/benzene dimer. Instead, we perform such a study of the chemical shielding of the *cis-N*-proton in FMA and use the results to benchmark more approximate methods that can be applied to FMA/benzene dimers (as described in the next subsection).

Table 2 lists CCSD(T) chemical shielding values computed using a B3LYP/aug-cc-pVTZ optimized geometry of FMA and two different, systematic series of basis sets (cc-pVxZ and pcS-*n*). Each set of calculations is used to extrapolate shielding constants to the complete basis set limit (as described in the previous section) and lead to very similar results: 26.64 and 26.67 ppm for cc-pVxZ and pcS-*n*, respectively. In the following, we will refer to 26.64 ppm as CCSD(T)/CBS, since this value is extrapolated using the largest basis set (cc-pV5Z) and since the Dunning-type basis sets are constructed for the purpose of correlated wave function calculations, whereas the pcS-*n* basis sets are constructed specifically for shielding constant calculations at the HF and DFT levels of theory.

The CCSD(T)/CBS value is 0.40 ppm higher than the experimental gas phase value obtained at 483 K of 26.24 ppm. However, this experimental value includes vibrational effects and is an average of the chemical shieldings of both amide protons. The effect of vibrations at 0 K (i.e., the zero-point vibrational correction) can be estimated relatively easily, as described in the previous section. At the MP2/cc-pVQZ level of theory, the ZPVC is -0.26 ppm, which, when used to correct the CCSD(T)/CBS value, results in a chemical shielding of 26.38 ppm—within 0.14 ppm of experiment. The ZPVC correction is unlikely to contribute significantly to $\Delta\delta_{\text{RC}}$, because it is a shielding difference between two molecular systems with very

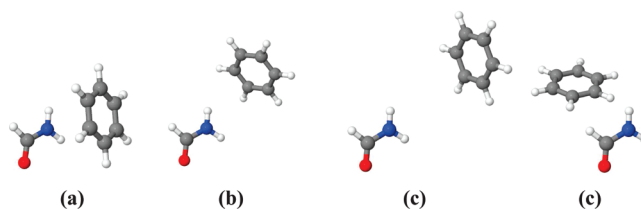


Figure 3. Pictures of the four FMA/benzene dimers used in this study. The resulting $\Delta\delta_{\text{RC}}$ calculated at various levels of theory for each conformation can be found in Table 3, where $\Delta\delta_{\text{RC}}^1$ corresponds to conformation a, $\Delta\delta_{\text{RC}}^2$ to conformation b, and so forth.

similar vibrational normal modes of FMA. Thus, we in the following focus on the electronic contribution to the chemical shielding alone.

The CBS values computed using CCSD, MP2, and B3LYP are all within 0.14 ppm of the CCSD(T)/CBS value, suggesting that the amide proton chemical shielding is relatively insensitive to electronic correlation effects in the limit of large basis sets. However, it is quite basis-set-dependent as at least the cc-pV5Z or the pcS-3 basis set is needed to get within 0.2 ppm of the CCSD(T)/CBS, with the exception of MP2/cc-pVQZ, which deviates by 0.16 ppm. We therefore choose MP2/cc-pVQZ for the $\Delta\delta_{\text{RC}}$ calculations using the FMA/benzene dimers described in the next subsection. Since $\Delta\delta_{\text{RC}}$ is a relative shielding value between two very similar systems, we expect that the MP2/cc-pVQZ results are well within 0.1 ppm of what would be computed with CCSD(T)/CBS and measured experimentally. A factor not investigated here was the dependence on the used geometry, which is known to cause deviations in calculated ¹H shielding constants on the order of ± 0.1 ppm—see for instance Rablen et al.⁴⁶

3.2. Scaling B3LYP Results to Those Obtained with Correlated Wave Function Methods. In this section, high-level correlated wave function methods are used to obtain a linear scaling correction to the chemical shift contribution due to ring current effects, obtained at the B3LYP/6-311++G(d,p)//B3LYP/aug-cc-pVTZ level of theory.

Four dimer systems were selected from the large data set of NMA/benzene dimers (see Figure 3), in such a way that the ring current contributions ($\Delta\delta_{\text{RC}}$) in the four dimer conformations cover a range from -0.72 ppm to $+0.15$ ppm, at the B3LYP/6-311++G(d,p) level of theory, in even sized steps. In these dimers, the NMA molecule was replaced with the much smaller FMA molecule, and the isotropic shielding was calculated using various methods and basis sets. Here, the chemical shift is modeled by

$$\Delta\delta_{\text{RC}}^{\text{(uncorrected)}} \approx \sigma_{\text{H}}^{\text{Probe}} - \sigma_{\text{H}}^{\text{Dimer}} \quad (6)$$

where $\sigma_{\text{H}}^{\text{Probe}}$ is the shielding of the probe nucleus in the probe molecule alone and $\sigma_{\text{H}}^{\text{Dimer}}$ is the shielding of the probe nucleus in the probe molecule in the dimer. Note that an NMR calculation for a reference dimer is not carried out, and the linear scaling factor is unaffected, whether a reference calculation is carried out, since this calculation would also have to be scaled by the same factor. The results are collected in Table 3. We observe the following:

1. Regardless of the basis set or method, the obtained $\Delta\delta_{\text{RC}}$'s have a linear correlation to B3LYP/6-311++G(d,p) data of 0.992 or better (see the Supporting Information). It is thus demonstrated that applying a linear correction based

Table 3. Shielding Constant of the FMA Probe Proton in a Vacuum for Each Method and Basis Set Used in This Section, As Well As the Chemical Shift Ring Current Interaction ($\Delta\delta_{RC}^n$) for Each of the Four Different Conformations Used^a

method	σ_{FMA}	$\Delta\delta_{RC}^1$	$\Delta\delta_{RC}^2$	$\Delta\delta_{RC}^3$	$\Delta\delta_{RC}^4$	scaling
B3LYP/6-311++G(d,p)	27.64	-0.72	-0.43	-0.21	0.15	
MP2/6-311++G(d,p)	27.70	-0.76	-0.45	-0.22	0.15	1.052
CCSD/6-311++G(d,p)	27.90	-0.75	-0.44	-0.22	0.15	1.033
CCSD(T)/6-311++G(d,p)	27.85	-0.73	-0.43	-0.21	0.15	1.012
MP2/cc-pVDZ	27.90	-0.74	-0.43	-0.22	0.17	1.033
MP2/cc-pVTZ	27.16	-0.77	-0.46	-0.23	0.17	1.076
MP2/cc-pVQZ	26.80	-0.81	-0.45	-0.22	0.13	1.074
B3LYP/pcS-0	28.88	-0.75	-0.42	-0.21	0.16	1.042
B3LYP/pcS-1	27.29	-0.76	-0.44	-0.22	0.16	1.056
B3LYP/pcS-2	26.91	-0.80	-0.47	-0.24	0.14	1.087
B3LYP/pcS-3	26.77	-0.80	-0.47	-0.24	0.16	1.097
B3LYP/pcS-4	26.76	-0.80	-0.47	-0.25	0.16	1.095
CCSD(T)/CBS	26.64					
B3LYP/6-311++G(d,p) (NMA)		-0.74	-0.43	-0.22	0.14	1.004

^a Furthermore, the resulting scaling factor relative to data obtained at the B3LYP/6-311++G(d,p) level of theory is noted. The B3LYP/6-311++G(d,p) shieldings for the identical conformation with NMA as a probe are given in the bottom row. All shielding constants and $\Delta\delta_{RC}$ values are given in ppm.

on correlated methods is a very good approximation. No constant offset (intercept) greater than 0.01 ppm was found, so the relationship between data obtained at the B3LYP level of theory and data obtained using a correlated wave function method is effectively a simple scaling factor. Thus, the fit was carried out as

$$\Delta\delta_{RC}^{\text{Other}} = k_{\text{scaling}} \Delta\delta_{RC}^{\text{B3LYP}} \quad (7)$$

where $\Delta\delta_{RC}^{\text{B3LYP}}$ is the $\Delta\delta_{RC}$ obtained at the B3LYP/6-311++G(d,p) level of theory, $\Delta\delta_{RC}^{\text{Other}}$ is the $\Delta\delta_{RC}$ obtained using another method and/or basis set, and k_{scaling} is the fitted scaling constant. Here, the $\Delta\delta_{RC}$ values are obtained via eq 6.

- All $\Delta\delta_{RC}$ values are within 0.1 ppm of one another, including B3LYP/6-311++G(d,p), which is used for the 932 dimer calculations. This supports our previous assertion that it is easier to compute $\Delta\delta_{RC}$ accurately compared to the computation of absolute shielding constants. Therefore, the $\Delta\delta_{RC}$ values listed in Table 3 are very likely within 0.1 ppm of what would be computed with CCSD(T)/CBS and measured experimentally.
- As a result, all scaling factors are within 10% and will all yield very similar results. However, on the basis of the results in Table 2, we pick the scaling factor computed at the MP2/cc-pVQZ level, where $k_{\text{scaling}} = 1.074$.
- The difference between ring current effects acting on either an FMA or an NMA probe was found at the B3LYP/6-311++G(d,p) level of theory to be a factor of 1.004 (see Table 3). This suggests that results obtained using FMA as a probe are, to a very good approximation, transferable to systems where NMA is used as a probe.

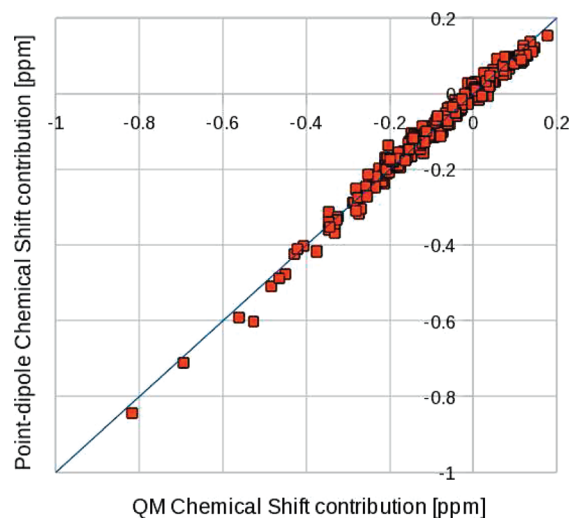


Figure 4. Correlation between the chemical shift predictions of the point-dipole model and the chemical shifts obtained by eq 4 for a set of NMA/benzene dimers using a best fit value of $B_{PD} = 30.42 \pm 0.16 \text{ ppm } \text{\AA}^3$. The blue line represents the best fit between the two methods. The linear correlation of the data set is 0.993.

3.3. Expressions for the B Factors. In the point-dipole model, the definition of $i_{\text{Benzene}} \equiv 1$ is used, and the B factor in the point-dipole model (B_{PD}) is obtained via a fit using the chemical shifts obtained for all NMA/benzene dimers to their corresponding G values in the point-dipole model ($G_{PD}(\vec{r}, \theta)$, see Supporting Information), using the following formula:

$$\Delta\delta_{RC} = B_{PD} G_{PD}(\vec{r}, \theta) \quad (8)$$

where $\Delta\delta_{RC}^{\text{QM}}$ is the scaled QM calculated chemical shifts of the amide protons using eq 4 and $G_{PD}(\vec{r}, \theta)$ is the geometric term of the NMA/benzene dimers in the point-dipole model. This gives a value of $B_{PD} = 30.42 \pm 0.16 \text{ ppm } \text{\AA}^3$. The linear correlation of this fit is $r = 0.993$. See Figure 4 for a scatter plot of the fitted data set.

In the literature, the trend has been to use formally derived B factors in the Johnson–Bovey model (B_{JB}) and scale the relative intensities accordingly.¹⁸ Following this, the analytical values of B_{JB} are used in the Johnson–Bovey model in this work. These evaluate to -3.79 ppm and -3.25 ppm for five- and six-membered rings, respectively. To facilitate an easy comparison of the ring current intensities to those found by Case,¹⁸ a B factor in the Haigh–Mallion model of $B_{HM} = 5.455 \text{ ppm } \text{\AA}$ is adopted.

3.4. Fitting the Relative i Factors. Using the B factor obtained in the previous subsection, the relative ring current intensities of all ring types in the three ring current models are obtained as the best fit of i when fitting the right-hand side of eq 4 to the right-hand side of eq 1.

The relative intensities of the two rings in tryptophan were trivially fitted using a two parameter fitting routine, although the contributions from the five- and six-membered rings were somewhat correlated. The fitted relative ring current intensities can be found in Table 4, which also features a comparison to the i factors found in other studies. A comparison is made to the values used in the SHIFTX and SHIFTS programs and to the values obtained for methane hydrogen by Case¹⁸ as used in SPARTA. The linear correlation between the B3LYP/6-311++G(d,p) ring current contributions and the predictions of the three approximations

Table 4. Relative Ring Current Intensity Factors of the Different Side Chains, As Found in This Study, Compared to the Value of Other Studies^a

model	point-dipole	Haigh–Mallion			Johnson–Bovey		
	reference	This Work	SHIFTX ⁶	SHIFTS ⁴⁷	Case ¹⁸ SPARTA ⁹	This Work	Case ¹⁸
PHE	1.00 (0.02, 0.07)	1.05 (0.05, 0.18)	1.00 (0.05, 0.17)	1.46 (0.07, 0.17)	1.18 (0.03, 0.06)	1.27 (0.03, 0.14)	1.13 (0.02, 0.06)
TYR	0.81 (0.02, 0.10)	0.92 (0.02, 0.09)	0.84 (0.02, 0.08)	1.24 (0.06, 0.22)	0.93 (0.02, 0.09)	1.10 (0.04, 0.10)	0.91 (0.02, 0.07)
HIS+	0.69 (0.02, 0.05)	0.43 (0.08, 0.29)	0.90 (0.06, 0.12)	1.35 (0.05, 0.07)	1.26 (0.03, 0.05)	1.40 (0.03, 0.08)	1.27 (0.03, 0.05)
HIS	0.68 (0.03, 0.06)	0.43 (0.08, 0.28)	0.90 (0.07, 0.11)	1.35 (0.06, 0.08)	1.22 (0.03, 0.07)	1.40 (0.04, 0.09)	1.25 (0.03, 0.06)
TRP5	0.57 (0.03, 0.08)	0.90 (0.03, 0.11)	1.04 (0.03, 0.08)	1.32 (0.04, 0.15)	0.97 (0.02, 0.10)	1.02 (0.02, 0.10)	1.06 (0.02, 0.09)
TRP6	1.02	1.04	1.02	1.24	1.18	1.27	1.18
B-factor	30.42 ppm Å ³	5.13 ppm Å	5.455 ppm Å	5.455 ppm Å	5.455 ppm Å	−3.25 ppm ^b	−3.25 ppm ^b
						−3.79 ppm ^b	−3.79 ppm ^b

^a The RMSD associated with using the given intensity factor and *B* value is given as the first entry in the parentheses, and MAD as the second entry, for each intensity factor. The RMSD and MAD are calculated over all dimer systems used in the fits to obtain intensity factors. The RMSD given for tryptophan is the RMSD for a sum of both rings with the given intensities. ^b In the Johnson–Bovey model, values of −3.25 ppm and −3.79 ppm are used for six- and five-membered rings, respectively.

were found to be $r = 0.980$ or better (see Supporting Information), so using linear fits to determine the *i* factors is evidently a very good approximation. For the data sets for each ring type, chemical shift predictions of different sets of *i* factors are compared to our QM data. We present the *root-mean-square deviation* (RMSD) and the *maximum absolute deviation* (MAD) of the data set. The RMSDs to our QM values are seemingly very small for all sets of intensity parameters. However, this is mostly due to the magnitude of the ring current effect in the dimers used in the data set being on average very small, and RMSD is thus not a very good measure in this case. Our intensities, however, do have the lowest maximum RMSD of up to just 0.03 ppm, while the competing methods have RMSDs of up to 0.08 ppm (SHIFTX), 0.07 ppm (SHIFTS), 0.07 ppm (Case, Haigh–Mallion), and 0.04 ppm (Case, Johnson–Bovey) for a residue type. A much better metric than RMSD is in this case the maximum average deviation (MAD) to QM values, which loosely corresponds to the largest error one can expect from using a certain set of intensities. In this metric, our method has a MAD of 0.05–0.10 ppm or better, while the corresponding numbers for the competing methods are 0.09–0.29 ppm (SHIFTX), 0.08–0.17 ppm (SHIFTS), 0.09–0.29 ppm (Case, Haigh–Mallion), and 0.09–0.14 ppm (Case, Johnson–Bovey), for all residue types.

We note that the SHIFTX and SHIFTS predictions are, on average, slightly lower than our prediction and those of Case. One possible explanation is that the SHIFTX and SHIFTS predictions are based on empirical parameters fitted to chemical shifts measured for solution phase structures. These structures may exhibit larger conformational fluctuations, leading to a net larger average distance between the ring and the amide proton compared to the X-ray structure as well as fluctuations in the direction of the magnetic dipole arising from the aromatic side chains and, therefore, a smaller ring current effect.

4. SUMMARY

We have presented sets of ring current intensity parameters for chemical shift predictions with the point-dipole, Haigh–Mallion, and Johnson–Bovey models. The maximum errors arising from use of the presented parameters are judged to be within ± 0.1 ppm from what would have been computed at the CCSD(T)/CBS for a set of 932 test cases. Further improvements in computational

methodology are thus not expected to yield any significant qualitative or quantitative improvement in chemical shift prediction in proteins. Preliminary calculations at the B3LYP/6-311++G(d,p) level using methane as a probe and intermolecular geometries corresponding to those in Table 3 suggest that the current parameters can be used to predict ring current effects on CH protons to within 0.2 ppm.

The presented parameters are rigorously based on the underlying physical properties of aromatic molecules. Parameters based on empirical models were found to perform worse on our amide proton test set. Our report of the superiority of a physics-based method over empirical methods is backed up by the fact that the parameters obtained by Case¹⁸ through QM methods have the same disagreements with the empirical methods as our parameters, despite the fact that the computational methodology used by Case was somewhat different than ours.

Finally, we have made a detailed numerical comparison between the point-dipole, Haigh–Mallion, and Johnson–Bovey models. The chemical shift predictions of the three models were nearly identical, and no outliers compared to our quantum mechanical calculations were found in any of the three models. Apart from reported problems with predictions of ring current effects in macrocyclic rings, such as those found in porphyrins,⁴⁸ which should be a nonissue for most uses in protein chemical shift predictions, the three methods should yield results of identical accuracy. Hence, we suggest that the point-dipole model should be used in future chemical shift prediction software, since (1) it both is computationally faster than competing models, since it does not require any integral evaluation as opposed to the Johnson–Bovey model, and contains significantly fewer geometric terms than the Haigh–Mallion model and (2) it is much easier to implement than the competing models.

■ ASSOCIATED CONTENT

Supporting Information. Thorough descriptions of the point-dipole, Haigh–Mallion, and Johnson–Bovey models and our implementations are described. The material also includes sketches of the used molecules, linear correlation values and scatter plots of the fits used to obtain the intensity values of Table 4, and an additional investigation of the equilibrium geometry dependence of FMA NMR shielding calculations.

This information is available free of charge via the Internet at <http://pubs.acs.org/>.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: andersx@nano.ku.dk; jhjensen@chem.ku.dk.

ACKNOWLEDGMENT

A.S.C. is recipient of a Ph. D. scholarship funded by the Novo Nordisk STAR program. S.P.A.S. thanks the Danish Center for Scientific Computing (DCSC), the Danish Natural Science Research Council/The Danish Council for Independent Research, and the Carlsberg Foundation for support.

REFERENCES

- (1) Raman, S.; Lange, O.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G.; Ramelot, T.; Eletsky, A.; Szyperki, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D. *Science* **2010**, *327*, 1014–1018.
- (2) Meiler, J.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15404–15409.
- (3) Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 1270–1272.
- (4) Vila, J. A.; Scheraga, H. A. *Acc. Chem. Res.* **2009**, *42*, 1545–1553.
- (5) Robustelli, P.; Kohlhoff, K. J.; Cavalli, A.; Vendruscolo, M. *Structure* **2010**, *18*, 923–933.
- (6) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Am. Chem. Soc.* **1991**, *111*, 9436–9444.
- (7) Xu, X. P.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321–333.
- (8) Xu, X. P.; Case, D. A. *Biopolymers* **2002**, *65*, 408–423.
- (9) Shen, Y.; Bax, A. *J. Biomol. NMR* **2007**, *38*, 289–302.
- (10) Meiler, J. *J. Biomol. NMR* **2003**, *26*, 25–37.
- (11) Haigh, C. W.; Mallion, R. B. *Prog. NMR Spectrosc.* **1980**, *13*, 303–344.
- (12) Moon, S.; Case, D. A. *J. Biomol. NMR* **2007**, *38*, 139–150.
- (13) Mulder, F. A.; Filatov, M. *Chem. Soc. Rev.* **2010**, *39*, 578–590.
- (14) Monya, G.; Zauhar, R. J.; Williams, H. J.; Nachman, R. J.; Scott, A. I. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 702–709.
- (15) Johnson, C. E.; Bovey, F. A. *J. Chem. Phys.* **1958**, *29*, 1012–1014.
- (16) Pople, J. A. *J. Chem. Phys.* **1956**, *24*, 1111.
- (17) Pople, J. A. *Mol. Phys.* **1958**, *1*, 175–180.
- (18) Case, D. A. *J. Biomol. NMR* **1995**, *6*, 341–346.
- (19) Boyd, J.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2002**, *124*, 1832–1833.
- (20) Berman, J.; Westbrook, H. M.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Res.* **2000**, *106*, 16972–16977.
- (21) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.* **2007**, *35*, 522–525.
- (22) Dolinsky, T. J.; Nielsen, J.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.* **2004**, *32*, 665–667.
- (23) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451.
- (24) Becke, A. D. *J. Phys. Chem.* **1993**, *98*, 5648–5652.
- (25) Stephens, P.; Devlin, F.; Chabalowski, C.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (26) Moon, S.; Case, D. A. *Nucleic Acids Res.* **2004**, *32*, 665–667.
- (27) Kupka, T.; Ruscic, B.; Botto, R. E. *J. Phys. Chem. A* **2002**, *106*, 10396–10407.
- (28) Kupka, T.; Ruscic, B.; Botto, R. E. *Solid State Nucl. Magn. Reson.* **2003**, *23*, 145–167.
- (29) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (30) Marquardt, D. W. *J. Appl. Math.* **1963**, *11*, 431–441.
- (31) Levenberg, K. Q. *J. Appl. Math.* **1944**, *2*, 164–168.
- (32) Jensen, F. *J. Chem. Theory Comput.* **2008**, *8*, 719–727.
- (33) Jameson, K. A.; Jameson, C. J. *Chem. Phys. Lett.* **1987**, *134*, 461–466.
- (34) Vaara, J.; Kaski, J.; Joksaari, J.; Diehl, P. *J. Phys. Chem. A* **1997**, *101*, 5069–5081.
- (35) Ruud, K.; Astrand, P.-O.; Taylor, P. R. *J. Am. Chem. Soc.* **2001**, *123*, 4826–4833.
- (36) Sauer, S. P. A.; Spirko, V.; Paidarová, I.; Kraemer, W. P. *Chem. Phys.* **1997**, *214*, 91–102.
- (37) Wigglesworth, R. D.; Raynes, W. T.; Sauer, S. P. A.; Oddershede, J. *Mol. Phys.* **1999**, *96*, 1595–1607.
- (38) Wigglesworth, R. D.; Raynes, W. T.; Kirpekar, S.; Oddershede, J.; Sauer, S. P. A. *J. Chem. Phys.* **2000**, *112*, 736–746.
- (39) Kern, C. W.; Matcha, R. L. *J. Phys. Chem.* **1968**, *49*, 2081–2092.
- (40) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, A. J., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.
- (41) Ahlrichs, R.; Baer, M.; Haeser, M.; Horn, H.; Koelmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
- (42) Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P.; Auer, A. A.; Bartlett, R. J.; Benedict, U.; Berger, C.; Bernholdt, D. E.; Bomble, Y. J.; Christiansen, O.; Heckert, M.; Heun, O.; Huber, C.; Jagau, T.-C.; Jonsson, D.; Juselius, J.; Klein, K.; Lauderdale, W. J.; Matthews, D. A.; Metzroth, T.; O'Neill, D. P.; Price, D. R.; Prochnow, E.; Ruud, K.; Schiffmann, F.; Stopkowitz, S.; Tajti, A.; Vazquez, J.; Wang, F.; Watts, J. D.; Almlöf, J.; Taylor, P. R.; Taylor, P. R.; Helgaker, T.; Jensen, H. J. A.; Jorgensen, P.; Olsen, J.; Mitin, A. V.; van Wüllen, C. CF04, a quantum chemical program package. For the current version, see <http://www.cfour.de> (accessed June 2010).
- (43) London, F. *J. Phys. Radium* **1937**, *8*, 397–409.
- (44) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789–807.
- (45) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, *104*, 5497–5509.
- (46) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. *J. Phys. Chem. A* **1999**, *103*, 7357–7363.
- (47) Ösapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *111*, 9436–9444.
- (48) Perkins, S. J. *Applications of ring current calculations to proton NMR of proteins and transfer RNA*; Plenum Press: New York, 1982; Vol. 4, Chapter 4, pp 193–336.

A Simple Algorithm for Determining Orthogonal, Self-Consistent Excited-State Wave Functions for a State-Specific Hamiltonian: Application to the Optical Spectrum of the Aqueous Electron

Leif D. Jacobson and John M. Herbert*

Department of Chemistry, The Ohio State University, Columbus, Ohio 43210, United States

ABSTRACT: We recently introduced a mixed quantum/classical model for the hydrated electron that includes electron/water polarization in a self-consistent fashion, using a polarizable force field for the water molecules [*J. Chem. Phys.* **2010**, *133*, 154506]. Calculation of the electronic absorption spectrum for this model is not straightforward, owing to the state-specific nature of the Hamiltonian, the high density of electronic states, and the large solvent polarization response upon electronic excitation. Together, these properties make it difficult or impossible to converge the polarizable solvent dipoles self-consistently for each excited-state wave function. Here, we overcome this problem by means of an extended Lagrangian procedure for performing constrained annealing in the space of electronic variables. By construction, this algorithm affords self-consistent, mutually orthogonal solutions for any state-specific Hamiltonian, and we illustrate this approach by computing the optical spectrum of our polarizable model for the aqueous electron. The spectrum thus obtained affords better agreement with experiment than previous, perturbative calculations of solvent dipole relaxation. Strengths, weaknesses, and possible generalizations of this procedure are discussed.

I. INTRODUCTION

First observed directly in 1962,¹ the aqueous (or hydrated) electron, $e^-(aq)$, has since that time been the subject of numerous experimental and theoretical investigations.^{2,3} Despite numerous atomistic simulations of this species over the past 25 years,³ it was not until quite recently that the Lorentzian decay on the high-energy side of the optical absorption spectrum was reproduced even qualitatively.^{4,5}

Due to the highly quantum mechanical nature of the solute (an electron), the dynamics and the bulk structure of $e^-(aq)$ have mostly been studied using one-electron pseudopotential methods,^{3,6–9} in other words, hybrid quantum mechanics/molecular mechanics (QM/MM) procedures with an one-electron QM region. The ostensible simplicity of such models (only one QM electron), combined with the importance of $e^-(aq)$ in the radiation chemistry of aqueous systems,^{2,10,11} means that these one-electron pseudopotential models have historically been used to test a variety of mixed quantum/classical simulation techniques.

We have recently developed a new one-electron pseudopotential model that incorporates self-consistent polarization between the solvent (water) and the single “excess” electron.⁵ Results from this model compare favorably to ab initio calculations in $(H_2O)_n^-$ clusters, and various properties of the bulk species, $e^-(aq)$, are also reproduced reasonably well.⁵ Our model utilizes the AMOEBA water potential,¹² which treats polarization by means of inducible point dipoles located on each MM atom. In our hydrated electron model,^{5,13} the electric field generated by the QM wave function contributes to the total electric field that polarizes these dipoles.

Because the induced dipoles represent electronic degrees of freedom, they should respond (polarize) on the time scale of electronic excitation. As such, it seems physically reasonable that the calculation of excited states in our polarizable model should

require a self-consistent calculation in which the solvent dipoles are converged with respect to each excited-state wave function. Because the QM Hamiltonian depends on the inducible dipoles, the realization of such a procedure effectively renders the Hamiltonian state specific, i.e., the nature of the Hamiltonian depends upon the particular electronic state that one is attempting to calculate.

In previous work,⁵ we encountered difficulties in obtaining self-consistent, excited-state solutions to this effective Hamiltonian, owing to the fact that the energy gaps between states are small (~ 0.1 eV), while the electronic relaxation energy of the solvent is large (e.g., 1.4 eV for vertical electron detachment in the bulk limit).⁵ This leads to frequent state switching during the wave function/dipole optimization. Even if we were able to converge the excited-state wave functions self-consistently with the induced dipoles, the wave functions thus obtained would not be mutually orthogonal, owing to the state-specific nature of the effective Hamiltonian. In view of these difficulties, we have previously resorted to the use of a perturbative correction for the solvent’s polarization response upon excitation of the wave function.^{4,5} While this approach allowed us to make progress in understanding the role of solvent polarization, it suffers from a lack of mutual orthogonality among the excited-state wave functions, owing to the state-specific nature of the perturbation. As such, one might reasonably be concerned about possible artifacts in the predicted oscillator strengths.

Here, we report a simulated annealing procedure in the space of electronic variables (wave function amplitudes and induced dipoles) by means of which the classical dipoles are converged self-consistently with respect to each wave function. In addition, our algorithm employs Lagrange multipliers to ensure that all of

Received: April 16, 2011

Published: June 13, 2011

the wave functions are orthonormal, despite the state-specific nature of the Hamiltonian. As a numerical demonstration of this procedure, we calculate the electronic absorption spectrum of the aqueous electron, using our polarizable one-electron model. The orthogonality issue is general to QM/MM methods that employ polarizable force fields, and therefore these ideas may be more broadly applicable. (However, the large polarization energies that we encounter may be unique to charged systems.)

Orthogonality is also an issue in certain self-consistent field (SCF) methods. For example, Gill and co-workers^{14,15} have recently introduced a maximum overlap method (MOM) that attempts to find excited-state solutions to the SCF equations by choosing the occupied orbitals at each SCF iteration, not in the usual aufbau way but rather by selecting those molecular orbitals that have the largest overlap with a set of user-specified guess orbitals. This situation is similar to the problem outlined above in that the effective Hamiltonian (Fock matrix) is state specific, and the excited-state solutions are not mutually orthogonal. Moreover, there is a direct correspondence between our polarizable QM/MM method and the SCF method. In the QM/MM procedure, we use the one-electron density, $|\psi|^2$, to compute induced dipoles, then use these dipoles to construct an effective Hamiltonian and finally diagonalize this Hamiltonian to obtain a new density. This process is iterated to self-consistency. In the SCF method, one uses the density to compute a new Fock matrix. We believe that our algorithm can be modified for use in the SCF procedure, in a manner that is conceptually (if not computationally) straightforward, and we hope to report on this in the future.

This paper is organized as follows: Section II provides a brief overview of our one-electron pseudopotential model for e^- (aq) and introduces the electronic annealing method. Details of the calculations are given in Section III. In Section IV, we present results for the optical absorption spectrum of e^- (aq) and draw a comparison with results obtained previously, using a perturbative treatment of the solvent's polarization response. We discuss certain formal aspects of the method, and some possible generalizations, in Section V. Section VI provides a summary.

II. THEORY

A. Polarizable QM/MM Model. We will not discuss our hydrated electron model in detail but will only highlight those aspects that are important to understand the annealing procedure. As in many polarizable QM/MM models, the total Hamiltonian in our model is a function of both the coordinates of the MM atoms, $\{\vec{R}_i\}$, as well as the induced MM dipoles, $\{\vec{\mu}_i\}$. The one-electron Hamiltonian can be written

$$\begin{aligned} \hat{H}(\{\vec{\mu}_i\}, \{\vec{R}_i\}) \\ = \hat{T} + V_{\text{elec-water}}(\{\vec{\mu}_i\}, \{\vec{R}_i\}) + V_{\text{MM}}(\{\vec{\mu}_i\}, \{\vec{R}_i\}) \end{aligned} \quad (1)$$

Here, \hat{T} is the one-electron kinetic energy operator, $V_{\text{elec-water}}$ is the electron–water pseudopotential, and V_{MM} is the molecular mechanics (MM) potential energy function for the polarizable water molecules. In our model, V_{MM} is the AMOEBA water force field.¹² The pseudopotential, $V_{\text{elec-water}}$, contains electrostatic interactions between the electron and both the permanent and the induced multipole moments of the water molecules. In addition, it contains a repulsive potential that keeps the electron from collapsing into the core molecular region.

The induced dipoles are obtained by solution of the equation:^{5,13}

$$\vec{\mu}_i = \alpha_i (\vec{F}_i^{\text{MM}} + \vec{F}_i^{\text{QM}}) \quad (2)$$

in which α_i is an (isotropic) polarizability for site i , \vec{F}_i^{MM} is the electric field produced by the MM subsystem at site i , and \vec{F}_i^{QM} is the electric field due to the wave function, also evaluated at site i . It can be shown that the induced dipoles defined by eq 2 minimize the total energy with respect to variations in $\vec{\mu}_i$.^{13,16} The one-electron wave function is determined by the solution of the Schrödinger equation:

$$\hat{H}(\{\vec{\mu}_i\}, \{\vec{R}_i\})|\psi\rangle = E|\psi\rangle \quad (3)$$

In practice, $|\psi\rangle$ is replaced by \mathbf{c} , a vector of amplitudes on a real-space grid. In order to obtain self-consistent polarization, we iterate eqs 2 and 3 to self-consistency. This procedure works well for the ground state but is difficult to converge for more than one or two excited states.

As a result of this difficulty we have, in previous work, computed approximate excited states by means of a simple perturbative scheme.^{4,5} To define the perturbation, we first calculate the ground-state wave function $|\psi_0\rangle$ and some number of excited state wave functions, $|\psi_n\rangle$, using dipoles $\{\vec{\mu}_i^{(0)}\}$ that are converged with respect to $|\psi_0\rangle$. For each excited state, we then obtain new dipoles, $\{\vec{\mu}_i^{(n)}\}$, that are converged with respect to $|\psi_n\rangle$, without relaxing $|\psi_n\rangle$. The quantity

$$\hat{W}_n = \hat{H}(\{\vec{\mu}_i^{(n)}\}, \{\vec{R}_i\}) - \hat{H}(\{\vec{\mu}_i^{(0)}\}, \{\vec{R}_i\}) \quad (4)$$

serves as the perturbation.^{4,5}

The perturbatively corrected wave functions thus obtained are not orthogonal, because the perturbation is state specific. However, they do turn out to be approximately orthogonal, with typical overlaps on the order of ~ 0.1 . Similar overlaps have been reported in MOM-SCF calculations, yet oscillator strengths in these calculations are in reasonable agreement with benchmark results.¹⁴ As such, we believe that the e^- (aq) spectra computed using the perturbative approach are at least qualitatively correct.

B. Electronic Annealing Procedure. We next describe our new algorithm to determine orthogonal excited states for state-specific effective Hamiltonians. The idea is not entirely new and is inspired by the Car–Parrinello molecular dynamics (CPMD) method,^{17,18} wherein the electronic degrees of freedom are propagated dynamically as classical variables. The CPMD approach can also be used to obtain ground-state, single-determinant wave functions by clamping the nuclei in place and “annealing” a guessed wave function.^{17,19} This amounts to a systematic removal of the fictitious kinetic energy associated with the electronic degrees of freedom. So far as we are aware, however, this technique has not been applied to the annealing of excited states. The main difference here, apart from the obvious difference of having only one QM electron in the present implementation, is that we constrain the wave function of interest to be orthogonal to each previously determined wave function. Doing this allows one to “march up” the manifold of excited states. Each excited state will then be defined as the lowest energy state that is orthogonal to all previously determined states. In a sense, this is a natural generalization of the linear variation method in elementary quantum mechanics.

Let \mathbf{c}_0 denote the vector of wave function amplitudes that we are interested in optimizing, and let $\{\mathbf{c}_i\}_{i=1}^N$ denote a set of

previously determined states. Note that \mathbf{c}_0 need not (and probably does not) represent the ground state, but the notation for the equations of motion will be simpler if we adopt a common index for all of the vectors. Only \mathbf{c}_0 is propagated in time, whereas $\mathbf{c}_1, \dots, \mathbf{c}_N$ are fixed. We also find it convenient to define a dot product

$$\mathbf{c}_i \cdot \mathbf{c}_j = \langle \psi_i | \psi_j \rangle = \sum_{\mu=1}^{N_{\text{grid}}} c_{i,\mu} c_{j,\mu} \Delta\tau \quad (5)$$

where the sum runs over grid points and $\Delta\tau$ is the volume element defined by the cubic grid.

We insist that the new state, \mathbf{c}_0 , be orthogonal to the previously determined states $\mathbf{c}_1, \dots, \mathbf{c}_N$. Our method employs a Lagrangian

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \tilde{m}_{el} \dot{\mathbf{c}}_0 \cdot \dot{\mathbf{c}}_0 + \frac{1}{2} \lambda_0 (\mathbf{c}_0 \cdot \mathbf{c}_0 - 1) + \sum_{i=1}^N \lambda_i (\mathbf{c}_i \cdot \mathbf{c}_0) \\ & - E[\mathbf{c}_0, \{\vec{\mu}_i\}, \{\vec{R}_i\}] \end{aligned} \quad (6)$$

where the λ_i are the undetermined multipliers that enforce orthonormality constraints. The parameter \tilde{m}_{el} is a fictitious electron mass, and $E[\mathbf{c}_0, \{\vec{\mu}_i\}, \{\vec{R}_i\}]$ is the energy functional. In principle, one could also propagate the induced dipoles dynamically. Because updating the Hamiltonian is far more expensive than minimizing the energy with respect to the induced dipoles, however, we choose to converge the dipoles each time \mathbf{c}_0 is updated.

From the Lagrangian in eq 6, one obtains the following equations of motion:

$$\tilde{m}_{el} \ddot{\mathbf{c}}_0 = -2\mathbf{H}\mathbf{c}_0 + \sum_{i=0}^N \lambda_i \mathbf{c}_i \quad (7)$$

Here, and in what follows, we use \mathbf{H} to denote the Hamiltonian matrix, and for convenience we omit from our notation the explicit dependence of \mathbf{H} on $\{\vec{\mu}_i\}$ and $\{\vec{R}_i\}$. In deriving eq 7, we have assumed that all quantities are real valued.

In the limit that the fictitious kinetic energy goes to zero, minimizing \mathcal{L} with respect to \mathbf{c}_0 is equivalent to solving the time-independent Schrödinger equation. Therefore, if we propagate the electronic degrees of freedom according to eq 7 and systematically remove kinetic energy, we should eventually find a local minimum where $\partial\mathcal{L}/\partial\mathbf{c}_0 = 0$, although this minimum certainly need not be the global minimum. To remove kinetic energy, we add a velocity-dependent friction term to the equations of motion. Equation 7 is thereby modified, affording

$$\tilde{m}_{el} \ddot{\mathbf{c}}_0 = -2\mathbf{H}\mathbf{c}_0 + \sum_{i=0}^N \lambda_i \mathbf{c}_i - \gamma \tilde{m}_{el} \dot{\mathbf{c}}_0 \quad (8)$$

The friction parameter, γ , has dimensions of reciprocal time. This modified equation of motion is not conservative and does not arise from any Hamiltonian.

We next develop our algorithm for propagating the equations of motion in eq 8. For this we use a modified form of the velocity Verlet (VV) algorithm²⁰ and follow closely the work and the notation of Tuckerman and Parrinello,²¹ who developed a VV-type algorithm to integrate the CPMD equations of motion. In the case of no damping ($\gamma = 0$), the appropriate VV equations for

our purpose are

$$\mathbf{c}_0(t + \delta t) = \mathbf{c}_0(t) + \delta t \dot{\mathbf{c}}_0(t) + \frac{(\delta t)^2}{2\tilde{m}_{el}} \mathbf{f}(t) + \frac{(\delta t)^2}{2\tilde{m}_{el}} \sum_{i=0}^N \lambda_i^R \mathbf{c}_i(t) \quad (9a)$$

$$\dot{\mathbf{c}}_0\left(t + \frac{1}{2}\delta t\right) = \dot{\mathbf{c}}_0(t) + \frac{\delta t}{2\tilde{m}_{el}} \mathbf{f}(t) + \frac{\delta t}{2\tilde{m}_{el}} \sum_{i=0}^N \lambda_i^R \mathbf{c}_i(t) \quad (9b)$$

$$\begin{aligned} \dot{\mathbf{c}}_0(t + \delta t) = & \dot{\mathbf{c}}_0\left(t + \frac{1}{2}\delta t\right) + \frac{\delta t}{2\tilde{m}_{el}} \mathbf{f}(t + \delta t) \\ & + \frac{\delta t}{2\tilde{m}_{el}} \sum_{i=0}^N \lambda_i^V \mathbf{c}_i(t + \delta t) \end{aligned} \quad (9c)$$

Here, δt is the time step, and $\mathbf{f}(t) = -2\mathbf{H}\mathbf{c}_0(t)$ is the force on \mathbf{c}_0 at time t . Although we have written all of the vectors \mathbf{c}_i as functions of time (in order to use a common index for \mathbf{c}_0 and \mathbf{c}_i , which facilitates a compact notation), the vectors $\{\mathbf{c}_i\}_{i=1}^N$ are fixed, and only \mathbf{c}_0 is propagated forward in time. In other words

$$\mathbf{c}_{i \neq 0}(t) = \mathbf{c}_{i \neq 0}(t + \delta t) \quad (10)$$

As in the RATTLE method,²² the undetermined multipliers in eqs 9a–9c are allowed to have two different values, λ_i^R and λ_i^V , representing coordinate and velocity constraints, respectively. This is similar to the approach used to maintain orthonormality constraints when integrating the CPMD equations of motion.²¹

Upon substituting $\mathbf{f}(t) \rightarrow \mathbf{f}(t) - \gamma \tilde{m}_{el} \dot{\mathbf{c}}_0(t)$ in eqs 9a–9c, one obtains equations for the case of finite damping. The corresponding VV algorithm can be expressed in three steps. The first step consists of both “coordinate” ($\tilde{\mathbf{c}}_0$) and half-step “velocity” ($\dot{\tilde{\mathbf{c}}}_0$) updates:

$$\tilde{\mathbf{c}}_0(t + \delta t) = \mathbf{c}_0(t) + \delta t \left(1 - \frac{1}{2}\gamma\delta t\right) \dot{\mathbf{c}}_0(t) + \frac{(\delta t)^2}{2\tilde{m}_{el}} \mathbf{f}(t) \quad (11a)$$

$$\dot{\tilde{\mathbf{c}}}_0\left(t + \frac{1}{2}\delta t\right) = \left(1 - \frac{1}{2}\gamma\delta t\right) \dot{\mathbf{c}}_0(t) + \frac{\delta t}{2\tilde{m}_{el}} \mathbf{f}(t) \quad (11b)$$

The second step consists of corrections:

$$\mathbf{c}_0(t + \delta t) = \tilde{\mathbf{c}}_0(t + \delta t) + \sum_{i=0}^N X_i \mathbf{c}_i(t) \quad (12a)$$

$$\dot{\mathbf{c}}_0\left(t + \frac{1}{2}\delta t\right) = \dot{\tilde{\mathbf{c}}}_0\left(t + \frac{1}{2}\delta t\right) + \sum_{i=0}^N \frac{X_i}{\delta t} \mathbf{c}_i(t) \quad (12b)$$

where the intermediate quantities X_i are defined below. The final step is an update and a correction:

$$\dot{\mathbf{c}}_0(t + \delta t) = \left(1 + \frac{1}{2}\gamma\delta t\right)^{-1} \left[\dot{\tilde{\mathbf{c}}}_0\left(t + \frac{1}{2}\delta t\right) + \frac{\delta t}{2\tilde{m}_{el}} \mathbf{f}(t + \delta t) \right] \quad (13a)$$

$$\dot{\mathbf{c}}_0(t + \delta t) = \dot{\mathbf{c}}_0(t + \delta t) + \sum_{i=0}^N Y_i \mathbf{c}_i(t + \delta t) \quad (13b)$$

Equations 12a, 12b, 13a and 13b employ the intermediate quantities:

$$X_i = \frac{(\delta t)^2}{2\tilde{m}_{el}} \lambda_i^R \quad (14)$$

and

$$Y_i = \frac{\delta t}{2\tilde{m}_{el}} \left(1 + \frac{1}{2}\gamma\delta t \right)^{-1} \lambda_i^V \quad (15)$$

The values of X_i and Y_i are chosen to satisfy the constraint equations:

$$\mathbf{c}_0 \cdot \mathbf{c}_0 = 1 \quad (16a)$$

$$\mathbf{c}_0 \cdot \mathbf{c}_{i \neq 0} = 0 \quad (16b)$$

We start by substituting the first update of the second step of the algorithm, eq 12a, into these constraint equations. The result of this exercise is the following pair of equations:

$$1 = X_0^2 + 2X_0[\mathbf{c}_0(t) \cdot \tilde{\mathbf{c}}_0(t + \delta t)] + [\tilde{\mathbf{c}}_0(t + \delta t) \cdot \tilde{\mathbf{c}}_0(t + \delta t)] - \sum_{i=1}^N X_i^2 \quad (17a)$$

$$X_{i \neq 0} = -\tilde{\mathbf{c}}_0(t + \delta t) \cdot \mathbf{c}_i \quad (17b)$$

Equation 17b can be solved for X_i for each $i > 0$, and then eq 17a affords X_0 . To obtain Y_i , we first obtain velocity constraints by differentiating eqs 16a and 16b with respect to t and then substitute the final velocity update, eqs 13a and 13b, into these velocity constraints. The result is

$$Y_i = -\dot{\tilde{\mathbf{c}}}_0(t + \delta t) \cdot \mathbf{c}_i(t + \delta t) \quad (18)$$

In deriving eqs 17a and 17b, we have assumed that the constraints are satisfied at time t , and in obtaining eq 18, we have assumed that the position constraints (eqs 16a and 16b) are satisfied at time $t + \delta t$. In practice, this means that the dynamics *cannot* start from a vector \mathbf{c}_0 that does not satisfy the constraints in eqs 16a and 16b. At the beginning of the annealing procedure for a particular state, the guess vector must be orthogonalized against all previously obtained vectors.

III. COMPUTATIONAL DETAILS

We compute the optical absorption spectrum of the bulk hydrated electron under periodic boundary conditions, using Ewald summation for the long-range interactions.⁵ Two hundred geometries were taken from a ground-state MD run in a periodic box that is 26.2015 Å on a side and contains 600 water molecules, corresponding to a density of 0.997 g/cm³. We solve the Schrödinger equation on a grid with a spacing of 0.93 Å, for a total of $28^3 = 21\,952$ grid points. Details of the simulation protocol can be found in ref 5.

The zeroth-order states are obtained with the Davidson–Liu method,²³ using a convergence threshold $\|(\hat{H} - E)\psi\| < 10^{-8} E_h$ as described in ref 13. We use these zeroth-order states to generate a guess for the induced dipoles, $\{\vec{\mu}_i\}$, which we use to construct a Hamiltonian matrix. We then “anneal” the state of interest, subject to the constraint that it remain normalized and orthogonal to the previously determined states, as described above. Prior to initiation of the MD procedure, we orthogonalize the state of interest against all previous states, using the Gram–Schmidt procedure, so that the constraints are satisfied initially. The initial velocities ($\dot{\mathbf{c}}_0$) are taken to be zero. The electronic degrees of freedom quickly pick up kinetic energy since the guess vector is rarely near a minimum. Annealing proceeds until the change in energy between successive MD

steps is less than $10^{-8} E_h$. (By that point, the total electronic kinetic energy is also $\sim 10^{-8} E_h$.) At this point we have an updated wave function that we use to induce new dipoles. This procedure is repeated until the energy change between successive dipole updates is less than $10^{-8} E_h$.

In a typical CPMD calculation, one has to choose the fictitious electron mass and time step in such a way that the electronic degrees of freedom are adiabatically decoupled from the nuclear dynamics. (See refs 24–26 for an interesting discussion in the context of extended-Lagrangian MD.) This is not an issue here, as we are not propagating the nuclei; rather, we are trying to *find* the Born–Oppenheimer surface, not propagate dynamics along or near it. For this reason, we simply choose a time step and an electronic mass such that the annealing is stable. We use $\delta t = 0.1$ fs and $\tilde{m}_{el} = 400$ au, but we have not attempted to optimize these parameters. (We do find that for $\delta t = 0.1$ fs, masses less than 200 au lead to a failure to maintain the constraints.) In our calculations, the position and the velocity constraints are typically satisfied to an average absolute error of 10^{-14} and 10^{-16} au, respectively.

The friction parameter, γ , is chosen according to the recommendation in ref 27, which is based on a three-point fit using energies from successive steepest-descent steps. Since the initial wave function guess may be far from the minimum, we found it helpful to generate γ several times during the MD routine; we do this every 50 time steps. We find that the annealing typically converges after 20–300 time steps if the guess is reasonable. However, in cases where the guess is poor, it may take upward of 2000 steps. The Hamiltonian is not updated during this procedure, so the annealing steps are quite inexpensive compared to inducing new dipoles and updating the potential energy at each grid point.

Below, we will compare the e^- (aq) spectrum obtained from the annealing procedure to that calculated using the perturbative scheme that was described in Section II. In the latter scheme, we do not allow the perturbed wave function to mix with the ground state, so that each perturbed state remains orthogonal to the ground state, even though the excited-state wave functions are not mutually orthogonal. (This at least ensures that the transition dipoles are translationally invariant.) An electronic spectrum is constructed from a histogram of oscillator strengths

$$f_{0 \rightarrow n} = \frac{2m_e}{3\hbar^2} (E_n - E_0) \sum_{\kappa \in \{x,y,z\}} |\langle \psi_0 | \hat{\kappa} | \psi_n \rangle|^2 \quad (19)$$

Wave functions were visualized with the Visual Molecular Dynamics program,²⁸ and isocontour values were generated with OpenCubMan.²⁹ Calculations were performed with a simulation code that is described in refs 5 and 13.

IV. RESULTS

A. Benchmark Tests Using Fixed Dipoles. Prior to applying our procedure to determine the fully relaxed excited states of the aqueous electron, we would first like to demonstrate the method’s effectiveness in the case that the induced dipoles are not updated. That is, we will first verify that the annealing procedure reproduces the lowest few eigenstates of a Hamiltonian where the induced dipoles are converged to the ground-state wave function (only), in which case there is no orthogonality problem. For this test, we first determine the ground-state wave function and induced dipoles with our standard method,^{5,13} then solve for

the lowest 30 eigenstates of \hat{H} with fixed dipoles. Next, we take a set of vectors composed of random numbers and use these as initial guess vectors for the annealing algorithm, fixing the induced dipoles at the values previously determined for the ground-state wave function.

Table I shows that the annealing procedure reproduces—with high accuracy—both the excitation energies and the oscillator strengths that are obtained by a straightforward Davidson–Liu procedure. In this particular case, where the dipoles are fixed, the states emerge from the annealing procedure in ascending order of energy, indicating that the procedure does not become trapped in any local minima and most likely locates *global* minima of the constrained optimization problem. (Of course, there is no guarantee that this will be the case once we allow the dipoles to relax.) Using the convergence thresholds specified in Section III, we can reproduce excitation energies to within $\sim 10^{-4}$ eV, which is far smaller than the error intrinsic to the pseudopotential model. Due to the completely random nature of the initial guesses, the annealing procedure takes ~ 1500 steps to converge in this example.

Table I. Excitation Energies (in eV) and Oscillator Strengths, in the Absence of Dipole Relaxation, Computed Using Two Different Algorithms

<i>n</i>	Davidson–Liu diagonalization		electronic annealing	
	$E_n - E_0$	$f_{0 \rightarrow n}$	$E_n - E_0$	$f_{0 \rightarrow n}$
1	2.13004	0.294897	2.13009	0.294889
2	2.22616	0.306758	2.22618	0.306793
3	2.45889	0.271951	2.45890	0.272044
4	2.89899	0.002009	2.89901	0.001878
5	3.35058	0.007256	3.35084	0.007513
6	3.36738	0.001328	3.36724	0.001038
7	3.42538	0.000995	3.42542	0.001048
8	3.47247	0.010379	3.47247	0.010336
9	3.57655	0.000950	3.57662	0.000979
10	3.62341	0.006859	3.62349	0.006142

Because the annealing procedure employs a larger number of constraints for higher-energy states as compared to lower-energy states, one might question whether the accuracy of the computed energies degrades as one marches up the manifold of states, adding more and more constraints as the calculation proceeds. The data in Table I suggest that this is not the case. For example, the $n = 8$ excitation energy computed by means of the annealing algorithm is closer to the Davidson–Liu result than is the $n = 1$ excitation energy. The accuracy is not degraded because the annealing algorithm does not introduce any *new* constraints beyond those imposed by linear algebra. For a fixed set of dipoles, the exact (nondegenerate) eigenvectors of the Hamiltonian are necessarily orthogonal, and obtaining them via diagonalization or via Davidson’s procedure is equivalent to minimizing the Rayleigh–Ritz quotient

$$R[\psi] = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \quad (20)$$

subject to the constraint that $|\psi_n\rangle$ must be orthogonal to all lower-lying states, $|\psi_0\rangle, \dots, |\psi_{n-1}\rangle$. Our annealing algorithm simply provides an alternative means to enforce these constraints and to carry out the Rayleigh–Ritz variational procedure in a robust way.

Unlike this benchmark test involving fixed dipoles, the “right” answer is no longer well-defined once we let the MM dipoles relax. However, the very close agreement between the annealing results and the Davidson–Liu results in this test gives us confidence that our approach is a reasonable one, if one insists (as we do here) that the relaxed wave functions should be orthogonal to one another.

The excited states need not emerge in energetic order once we allow the induced dipoles to relax. They would do so only if the annealing procedure managed to find the global minimum of the effective potential (with constraints) on each annealing cycle. The presence of inducible dipoles appears to make this quite challenging, as the states do not come out of the calculations in ascending order. This gives us some pause and calls into question the nature of our guess. We have run additional calculations in which the guess for the annealing procedure is provided by the

Table II. Excitation Energies (in eV) and Oscillator Strengths Computed by Electronic Annealing, Using Two Different Initial Guesses

<i>n</i>	ordered by n^a				ordered by energy ^b				difference ^c	
	zeroth-order guess		first-order guess		zeroth-order guess		first-order guess			
	$E_n - E_0$	$f_{0 \rightarrow n}$	$E_n - E_0$	$f_{0 \rightarrow n}$	$E_n - E_0$	$f_{0 \rightarrow n}$	$E_n - E_0$	$f_{0 \rightarrow n}$	$E_n - E_0$	$f_{0 \rightarrow n}$
1	1.73894	0.133993	1.73894	0.133927	1.73894	0.133993	1.73894	0.133927	0.00000	0.000066
2	1.95139	0.237197	1.95140	0.237188	1.93676	0.028331	1.93668	0.028360	0.00008	0.000029
3	1.93676	0.028331	1.93668	0.028360	1.95139	0.237197	1.95140	0.237188	0.00001	0.000029
4	2.10816	0.132911	2.10817	0.132922	2.10816	0.132911	2.10817	0.132922	0.00001	0.000011
5	2.11486	0.001157	2.28370	0.001407	2.11486	0.001157	2.11521	0.001104	0.00035	0.000053
6	2.28397	0.001359	2.11521	0.001104	2.15224	0.000537	2.14726	0.000641	0.00498	0.000104
7	2.46492	0.001676	2.26762	0.004268	2.26925	0.004114	2.26762	0.004268	0.00163	0.000154
8	2.26925	0.004114	2.14726	0.000641	2.28397	0.001359	2.28370	0.001407	0.00027	0.000048
9	2.42928	0.001936	2.43138	0.001780	2.36290	0.003176	2.36278	0.003178	0.00012	0.000002
10	2.15224	0.000537	2.36278	0.003178	2.42928	0.001936	2.43138	0.001780	0.00210	0.000156
11	2.36290	0.003176	2.46681	0.001478	2.46492	0.001676	2.46681	0.001478	0.00189	0.000198

^a Excitation energies listed in the order that the states are generated by the annealing procedure. ^b Excitation energies listed in ascending order of energy.

^c Difference in energies and oscillator strengths for the energy-ordered states computed using two different initial guesses.

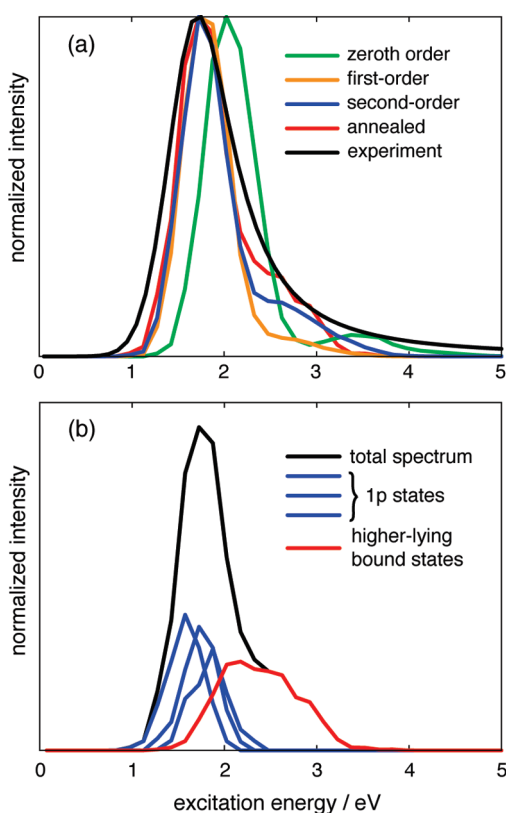


Figure 1. Absorption spectra for $e^-(aq)$ in bulk water: (a) comparison of spectra computed using zeroth-, first-, and second-order perturbative treatments of the \hat{W}_n (eq 4) to the spectrum computed using the annealing procedure proposed here; and (b) decomposition of the annealed spectrum into contributions from various types of excited states.

first-order corrected wave function from the perturbative scheme. Inspection of the energies and oscillator strengths indicates that typically, the first four states are identical and emerge in the same order for either initial guess. Table II shows a typical case. The first four excitation energies are nearly identical for either initial guess and emerge in the same order, but the ordering is different starting with $n = 5$. However, both initial guesses do find the same set of excitation energies through at least $n = 11$.

The fact that the states do not come out energetically ordered is worrisome because the constraints placed on a particular state depend upon the order in which it is determined, and this should effect the energy. In the latter columns of Table II we have reordered the states energetically, and tabulated the differences in excitation energies and oscillator strengths between the two different initial guesses. The largest discrepancy in the energies between the two initial guesses is only 0.005 eV. This is smaller than the typical energy gap between states, and we therefore find this to be a tolerable error. In principle, one could probably ensure energetic ordering by annealing the same state several times, starting from a variety of different guesses and taking the lowest energy result in an attempt to find the global minimum for each set of constraints. Another possibility would be to perform the annealing, reorder the states energetically, and repeat the entire procedure using the annealed states as guesses. We have not done so here, owing to the smallness of the discrepancies between energies obtained using different initial guesses.

B. Aqueous Electron Absorption Spectrum. Figure 1a compares the absorption spectrum obtained using perturbative techniques⁵ to that obtained using the annealing algorithm that is described here. The experimental spectrum (reproduced from the line shape parameters in ref 30) is also shown. With the exception of the annealed spectrum, which is new, these spectra have been described in detail in our previous work,^{3–5} but for completeness, we briefly summarize these results here. At zeroth-order in the perturbation, the peak intensity is blue-shifted relative to experiment, and although this zeroth-order spectrum does reproduce the main, Gaussian feature in the experimental spectrum, it exhibits a gap in intensity just below 3 eV, which is followed by a “hump” centered around 3.5 eV that is essentially a photoelectron spectrum. The first-order correction for \hat{W}_n shifts the maximum into quantitative agreement with experiment and also binds states that were (vertically) unbound at zeroth order, meaning that the excitation energies were greater than the vertical detachment energy. A second-order treatment of \hat{W}_n affords a correction to the wave function and hence the transition dipoles, and this has the effect of increasing intensity in the “blue tail”.

The spectrum obtained from electronic annealing agrees quantitatively with the second-order perturbation theory spectrum in the Gaussian region, but the annealing procedure shifts even more oscillator strength into the higher-lying bound states that comprise the blue tail. (All of the spectra in Figure 1 are normalized to unit intensity at their respective absorption maxima.) If anything, the blue tail in the annealed spectrum is in better agreement with experiment than is the second-order perturbation theory result.

Figure 1b decomposes the annealed spectrum into contributions from $1s \rightarrow 1p$ transitions versus excitations into higher-lying bound states. The 1p states are the only bright states, for an aqueous electron modeled as a particle in a spherical box,³ and indeed the $1s \rightarrow 1p$ excitations carry much of the oscillator strength in the annealed spectrum. However, the 1p band has significant energetic overlap with the higher-lying bound states, which borrow intensity from the 1p states and give rise to a significant “blue tail”. The states that comprise this tail are unbound in the zeroth-order treatment, and we have previously referred to them as “quasi-continuum, polarization-bound” excited states.⁴ These states have very little oscillator strength at zeroth order, but relaxation of the solvent dipoles allows them to mix with (and borrow intensity from) the 1p states. For the annealed spectrum, all 30 states that we calculate are vertically bound. (The average vertical binding energy for the simulation cell used in this work is 3.35 eV,⁵ well into the blue tail in the spectra shown in Figure 1.)

At zeroth-order in \hat{W}_n (what we have previously called the “unrelaxed” approximation),^{4,5,3} the states are ordered as follows. The ground state is spherical ($1s$) and resides in a roughly spherical solvent cavity, while the first three excited states are p-like ($1p$). The fourth excited state is typically more diffuse and can be identified as the $2s$ state by virtue of a radial node. Above the $2s$ state are several states that resemble $1d$ states, but above this it becomes difficult to assign particle-in-a-cavity quantum numbers to the excited states, whose wave functions are quite diffuse and contain many different lobes. The qualitative nature of these states is not altered significantly by application of second-order perturbation theory.

The annealing procedure, on the other hand, sometimes *does* alter the initial guess wave functions in a qualitative way. In

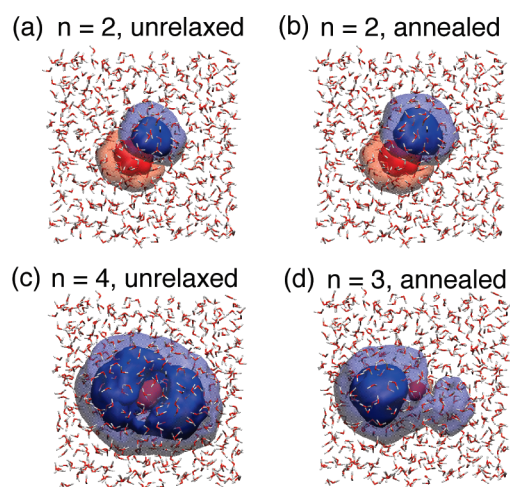


Figure 2. Examples of 1p- and 2s-like excited states of the aqueous electron. Panels (a) and (c) depict the “unrelaxed” states (zeroth order in \hat{W}_n), while panels (b) and (d) depict the wave functions that are obtained by electronic annealing. The opaque and translucent isosurfaces encapsulate 70% and 95% of $|\psi|^2$, respectively.

particular, the annealing procedure appears to have the ability to localize diffuse electronic states composed of largely disjoint lobes and in some cases may enhance the oscillator strength associated with these states, relative to the nominal bright states. In cases where we observe such localization, the nodal character of the state appears to be preserved, although this is only evident if the wave function is plotted using an isosurface that encapsulates nearly all of the electron density.

As an example, Figure 2 depicts the unrelaxed $n = 2$ and $n = 4$ wave functions from the calculation reported in Table I as well as the corresponding annealed wave functions from the calculation reported in Table II. (The states are labeled in the order that they are calculated by the annealing procedure, which need not be in energetic order.) The unrelaxed 1p state shown in Figure 2a is not altered by the annealing process in any substantive way and is nearly identical to the $n = 2$ state in the manifold of annealed excited states (Figure 2b). However, the annealed analogue (Figure 2d) of the $n = 4$ zeroth order wave function (Figure 2c) is more localized than its counterpart. The annealed function appears p-like rather than s-like, if a large isosurface contour value is used to plot the wave function. However, a smaller contour that encapsulates more of the wave function reveals s-like character. The transition from the unrelaxed to the annealed wave function (i.e., Figure 2c→d) enhances the transition dipole of the state in question, because the localized, annealed state has better overlap with the ground state and furthermore sheds some of the pseudo-s-type symmetry that causes the unrelaxed state in Figure 2c to exhibit a rather small oscillator strength.

Comparison of Tables I and II seems to indicate that the $n = 3$ state loses significant oscillator strength upon annealing, but an inspection of the wave functions reveals that the state that emerges as $n = 3$ from the annealing procedure actually corresponds to the *fourth* excited state at zeroth-order. The latter acquires significant oscillator strength upon annealing and drops below a state with p-type character to become $n = 3$. While this sort of reordering does not occur in the majority of the cases, it is also not entirely uncommon.

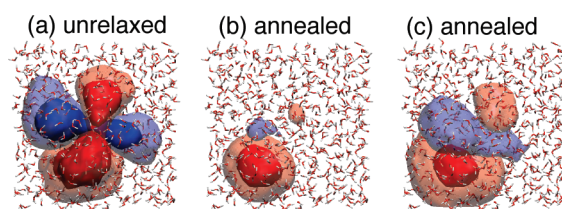


Figure 3. An excited electronic state of the aqueous electron with d-type character. Panels (a) and (b) depict the zeroth-order (unrelaxed) and annealed wave functions, respectively, using opaque and translucent isosurfaces that encapsulate 70% and 95% of $|\psi|^2$, respectively. Panel (c) depicts the same annealed wave function as in (b) but plotted using isosurfaces that encapsulate 70% and 99% of $|\psi|^2$.

From the spectrum in Figure 1b, it appears that the highest-lying 1p state carries somewhat less intensity than the two lower-lying 1p states. This is partly an artifact of the manner in which we analyzed the data, namely, we assumed in constructing Figure 1b that the first three states are the 1p states, which is always true in the perturbative approach but is occasionally *not* true following annealing. Despite this occasional reordering of states, the 1p states still carry the vast majority of the oscillator strength and are still responsible for the Gaussian feature in the absorption spectrum.

Figure 3 shows the zeroth-order and the annealed wave functions for a d-type state. Using an isosurface that encapsulates 90% of $|\psi|^2$ (Figure 3b), it appears as though the annealed state is effectively a “charge hop”, in which the electron is transferred a sizable distance away from the cavity in which the ground-state wave function is localized. However, Figure 3c depicts the same annealed wave function, plotted using an isosurface that encapsulates 99% of $|\psi|^2$. In the latter depiction, it is clear that the wave function remains d-like, but the electron has largely localized into one of the lobes. In this example, the annealed state has very little overlap with the ground state, which results in a very small transition dipole. States that have localized to such an extent as to exhibit charge-transfer or charge-hopping character exhibit very small oscillator strengths and thus do not contribute greatly to the absorption spectrum. These states are most likely not accessed in experiments that probe vertically excited states. The “blue tail” does not arise from localized charge-hopping states, such as that shown in Figure 3b and c. Rather, it arises due to higher-lying, diffuse excited states that do have reasonable overlaps with the ground-state wave function.^{4,5}

V. DISCUSSION

According to the Thomas–Reiche–Kuhn (TRK) sum rule³¹

$$\sum_{n > 0} f_{0 \rightarrow n} = N_e \quad (21)$$

where N_e is the number of electrons. By construction, $N_e = 1$ in our pseudopotential model. In previous work,⁵ we observed that $f_{0 \rightarrow 1} + f_{0 \rightarrow 2} + \dots + f_{0 \rightarrow 29} \approx 0.95$ at zeroth-order, that is, the first 29 excited states account for 95% of the total oscillator strength. A first-order correction for \hat{W}_n reduces the electronic energy gaps ($E_n - E_0$) but does not affect the wave functions, and as a result, the total oscillator strength carried by the first 29 excited states is reduced to ≈ 0.8 . At second order, the wave function is corrected, and the total oscillator strength recovers, to ≈ 0.9 . In the present treatment, however, we find that $f_{0 \rightarrow 1} + \dots + f_{0 \rightarrow 29} \approx 0.65$.

The question then arises as to whether the TRK sum rule is preserved in the case of a state-dependent Hamiltonian and whether or not the expression for $f_{0 \rightarrow n}$ in eq 19 is even valid in such a case. Here, we address these questions in the context of the proposed annealing procedure.

In principle, the annealing procedure provides a way to obtain an infinite number of mutually orthogonal states, each of which is an eigenfunction of a different Hamiltonian. For the purpose of analyzing the sum rule in eq 21, let us make the (perhaps dubious) assumption that this set of eigenfunctions forms a complete orthonormal basis. Then to derive eq 21, one employs the identity:

$$[\hat{H}, \hat{x}] = -\frac{i\hbar}{m} \hat{p}_x \quad (22)$$

In principle, \hat{H} could be any of the aforementioned Hamiltonians. Inserting eq 22 into the expression

$$\frac{1}{i\hbar} \langle 0 | [\hat{x}, \hat{p}_x] | 0 \rangle = 1 \quad (23)$$

and using a resolution of the identity, one obtains

$$\frac{m}{\hbar^2} \sum_n [\langle 0 | \hat{x} | n \rangle \langle n | [\hat{H}, \hat{x}] | 0 \rangle - \langle 0 | [\hat{H}, \hat{x}] | n \rangle \langle n | \hat{x} | 0 \rangle] = 1 \quad (24)$$

This equation is valid for any Hamiltonian and any complete orthonormal basis. However, in order to obtain the sum rule in eq 21 from eq 24, the basis states $|n\rangle$ must in addition be eigenstates of the *same* Hamiltonian. In the present case, however, each state is a solution to a different Hamiltonian so the sum rule is not preserved by the annealing procedure. (As such, nothing rests upon our dubious assumption that the states $|n\rangle$ form a complete basis; the sum rule is not preserved, whether or not this is in fact the case.)

Next, we address the question of whether or not eq 19 is a valid formula for computing absorption intensities. In what follows, we assume that the nuclei are clamped, and we consider the electronic dynamics. The oscillator strength formula in eq 19 follows from time-dependent perturbation theory.³¹ If the system is in state $|n\rangle$ at time $t = 0$, then it seems reasonable that the system evolves under the influence of the Hamiltonian for state $|n\rangle$, \hat{H}_n . That is,

$$|\Psi(t)\rangle = e^{-i\hat{H}_n t/\hbar} |n\rangle = e^{-iE_n t/\hbar} |n\rangle \quad (25)$$

where we have used the fact that $\hat{H}_n |n\rangle = E_n |n\rangle$.

We now investigate the time evolution in the presence of a time-dependent perturbation. We assume that the time-dependent wave function can be written

$$|\Psi(t)\rangle = \sum_n c_n(t) e^{-iE_n t/\hbar} |n\rangle \quad (26)$$

This expansion may seem suspicious in light of questions regarding whether the basis $\{|n\rangle\}$ is complete. However, we assume below that the system is initially in the ground state, and we are only interested in the dynamics within the finite subset of states that we have determined by means of annealing. In other words, this basis constitutes the region of interest in Hilbert space. To derive a formula for the transition probabilities, the ansatz in eq 26 should next be inserted into the time-dependent Schrödinger equation, but with which Hamiltonian? In the weak-field limit, the traditional assumption is that the system occupies the ground state at $t = 0$, $c_n(0) = \delta_{n,0}$. It therefore seems

reasonable to assume that the dynamics is governed by the ground-state Hamiltonian, so that

$$i\hbar |\dot{\Psi}(t)\rangle = (\hat{H}_0 + V(t)) |\Psi(t)\rangle \quad (27)$$

These assumptions, together with the fact that the basis is orthonormal, lead to the textbook³¹ dynamical equations for the expansion coefficients $c_n(t)$. For this reason, we would argue that eq 19 is still valid, even though the TRK sum rule is not.

The ambiguity regarding which Hamiltonian guides the dynamics of the system is clearly an artifact of the model. The inducible dipoles represent electronic degrees of freedom and should respond on the time scale of electronic motion, i.e., these degrees of freedom participate in the short-time dynamics that results in absorption of radiation, and they ought to be included in the quantum mechanical description of the system. Our decision to treat some of the electronic variables (solvent dipoles) classically leads to some ambiguity (multiple Hamiltonians) since we do not have information regarding the short-time quantum dynamics of these variables. This is to be contrasted with the MOM-SCF technique^{14,15} that was mentioned in Section I. In that method, there is a single Hamiltonian but multiple stationary points (solutions to the SCF equations). Since the SCF energy, at least in Hartree–Fock theory, is the expectation value of the true Hamiltonian, there is no ambiguity as to the quantum dynamics.

In the case of the methodology pursued here, one way around these difficulties would be to use a linear-response formalism, which has been explored in the context of time-dependent density functional theory (TD-DFT) in the presence of a polarizable medium.^{32,33} Here, however, we were interested in a self-consistent, nonperturbative approach. In the future, it might be interesting to compare results obtained from linear-response theory to those obtained from our electronic annealing procedure.

Finally, we would like to speculate that this annealing procedure might be useful for MOM-SCF calculations. The MOM-SCF method appears quite promising and avoids some problems associated with TD-DFT. However, the excited-state wave functions obtained in MOM-SCF calculations are not orthogonal, although preliminary results do not seem to exhibit any adverse effects on oscillator strengths, possibly because the deviations from orthogonality are small in cases examined so far.¹⁴ In any case, it is possible that the sort of electronic annealing that is introduced here could eliminate any concern over oscillator strengths. This technique might also be useful in the context of excited-state Kohn–Sham simulations,³⁴ nonadiabatic (surface hopping) simulations utilizing CPMD,³⁵ or “constrained” DFT calculations,^{36,37} each of which is potentially subject to nonorthogonality problems. Extensions to many-electron QM/MM methods using polarizable force fields are also worth exploring.

VI. SUMMARY

We have introduced a novel “electronic annealing” procedure that is capable of finding orthogonal solutions to a state-dependent Hamiltonian. This procedure appears to be robust and is capable of finding many such solutions. When applied to a polarizable QM/MM model of the aqueous electron in bulk water,⁵ the electronic absorption spectrum computed by means of electronic annealing is in reasonable agreement with results obtained previously^{4,5} based on a perturbative treatment of the

MM polarization response following excitation of the QM region. In fact, the annealed spectrum is in slightly better agreement with experiment, as compared to perturbative results. In any case, these computed spectra all support the hypothesis that electronic polarization (as described theoretically via atom-centered inducible dipoles) binds additional excited states of the aqueous electron and facilitates intensity borrowing from the 1p states that carry most of the oscillator strength. The “blue tail” in the optical spectrum of $e^{-}(\text{aq})$ arises from what we have termed “polarization-bound quasi-continuum states”.⁴ Here, we find that electronic reorganization of the solvent can also cause diffuse excited states of the electron to localize into “charge-hopping” states. These excitations, however, carry very little oscillator strength and do not make a substantial contribution to the optical absorption spectrum.

In the future, we plan to explore generalizations of this electronic annealing algorithm that are suitable for many-electron QM calculations.

AUTHOR INFORMATION

Corresponding Author

*E-mail: herbert@chemistry.ohio-state.edu.

ACKNOWLEDGMENT

We thank Xiaosong Li for encouraging us to complete and submit this paper. This work was supported by a National Science Foundation CAREER award (CHE-0748448). Calculations were performed at the Ohio Supercomputer Center under project no. PAS-0291. L.D.J. acknowledges a Presidential Fellowship from The Ohio State University. J.M.H. is an Arthur P. Sloan Foundation Fellow and a Camille Dreyfus Teacher–Scholar.

REFERENCES

- (1) Hart, E. J.; Boag, J. W. *J. Am. Chem. Soc.* **1962**, *84*, 4090–4095.
- (2) Mostafavi, M.; Lampre, I. In *Radiation Chemistry*; Spothem-Maurizot, M., Mostafavi, M., Belloni, J., Douki, T., Eds.; EDP Sciences: Les Ulis Cedex A, France, **2008**; Chapter 3, pp 33–52.
- (3) Herbert, J. M.; Jacobson, L. D. *Int. Rev. Phys. Chem.* **2011**, *30*, 1–48.
- (4) Jacobson, L. D.; Herbert, J. M. *J. Am. Chem. Soc.* **2010**, *132*, 10000–10002.
- (5) Jacobson, L. D.; Herbert, J. M. *J. Chem. Phys.* **2010**, *133*, 154506:1–19.
- (6) Schnitker, J.; Rosky, P. J. *J. Chem. Phys.* **1987**, *86*, 3462–3470.
- (7) Barnett, R. N.; Landman, U.; Cleveland, C. L.; Jortner, J. *J. Chem. Phys.* **1988**, *88*, 4421–4428.
- (8) Turi, L.; Gaigeot, M.-P.; Levy, N.; Borgis, D. *J. Chem. Phys.* **2001**, *114*, 7805–7815.
- (9) Sommerfeld, T.; DeFusco, A.; Jordan, K. D. *J. Phys. Chem. A* **2008**, *112*, 11021–11035.
- (10) Garrett, B. C.; Dixon, D. A.; Camaioni, D. M.; Chipman, D. M.; Johnson, M. A.; Jonah, C. D.; Kimmel, G. A.; Miller, J. H.; Rescigno, T. N.; Rosky, P. J.; Xantheas, S. S.; Colson, S. D.; Laufer, A. H.; Ray, D.; Barbara, P. F.; Bartels, D. M.; Becker, K. H.; Bowen, K. H., Jr.; Bradforth, S. E.; Carmichael, I.; Coe, J. V.; Corrales, L. R.; Cowin, J. P.; Dupuis, M.; Eisenthal, K. B.; Franz, J. A.; Gutowski, M. S.; Jordan, K. D.; Kay, B. D.; LaVerne, J. A.; Lyman, S. V.; Madey, T. E.; McCurdy, C. W.; Meisel, D.; Mukamel, S.; Nilsson, A. R.; Orlando, T. M.; Petrik, N. G.; Pimblott, S. M.; Rustad, J. R.; Schenter, G. K.; Singer, S. J.; Tokmakoff, A.; Wang, L.-S.; Wittig, C.; Zwier, T. S. *Chem. Rev.* **2005**, *105*, 355–389.
- (11) Buxton, G. V. In *Radiation Chemistry*; Spothem-Maurizot, M., Mostafavi, M., Belloni, J., Douki, T., Eds.; EDP Sciences: Les Ulis Cedex A, France, **2008**; Chapter 1, pp 3–16.
- (12) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.
- (13) Jacobson, L. D.; Williams, C. F.; Herbert, J. M. *J. Chem. Phys.* **2009**, *130*, 124115:1–18.
- (14) Gilbert, A. T. B.; Besley, N. A.; Gill, P. M. W. *J. Phys. Chem. A* **2008**, *112*, 13164–13171.
- (15) Besley, N. A.; Gilbert, A. T. B.; Gill, P. M. W. *J. Chem. Phys.* **2009**, *130*, 124308:1–7.
- (16) Dupuis, M.; Aida, M.; Kawashima, Y.; Hirao, K. *J. Chem. Phys.* **2002**, *117*, 1242–1255.
- (17) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- (18) Marx, D.; Hutter, J. In *Modern Methods and Algorithms in Quantum Chemistry*, 2nd ed.; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, Germany, 2000; Vol. 1 of NIC Series, pp 329–477.
- (19) Tassone, F.; Mauri, F.; Car, R. *Phys. Rev. B* **1994**, *50*, 10561–10573.
- (20) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: New York, 1992.
- (21) Tuckerman, M. E.; Parrinello, M. *J. Chem. Phys.* **1994**, *101*, 1302–1315.
- (22) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24–34.
- (23) Murray, C. W.; Racine, S. C.; Davidson, E. R. *J. Comput. Phys.* **1992**, *103*, 382–389.
- (24) Herbert, J. M.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 11542–11556.
- (25) Iyengar, S. S.; Schlegel, H. B.; Scuseria, G. E.; Millam, J. M.; Frisch, M. J. *J. Chem. Phys.* **2005**, *123*, 027101:1–2.
- (26) Herbert, J. M.; Head-Gordon, M. *J. Chem. Phys.* **2005**, *123*, 027102:1–2.
- (27) Probert, M. I. J. *J. Comput. Phys.* **2003**, *191*, 130–146.
- (28) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (29) Haranczyk, M.; Gutowski, M. *J. Chem. Theory Comput.* **2008**, *4*, 689–693.
- (30) Coe, J. V.; Williams, S. M.; Bowen, K. H. *Int. Rev. Phys. Chem.* **2008**, *27*, 27–51.
- (31) McHale, J. L. *Molecular Spectroscopy*; Prentice Hall: Upper Saddle River, New Jersey, 1998.
- (32) Yoo, S.; Zahariev, F.; Sok, S.; Gordon, M. S. *J. Chem. Phys.* **2008**, *129*, 144112:1–8.
- (33) Steindal, A. H.; Ruud, K.; Frediani, L.; Aidas, K.; Kongsted, J. *J. Phys. Chem. B* **2011**, *115*, 3027–3037.
- (34) Doltsinis, N. L.; Marx, D. *Phys. Rev. Lett.* **2002**, *88*, 166402:1–4.
- (35) Tapavicza, E.; Tavernelli, I.; Rothlisberger, U. *Phys. Rev. Lett.* **2007**, *98*, 023001:1–4.
- (36) Wu, Q.; Van Voorhis, T. *J. Chem. Theory Comput.* **2006**, *2*, 765–774.
- (37) Wu, Q.; Cheng, C.-L.; Van Voorhis, T. *J. Chem. Phys.* **2007**, *127*, 164110:1–9.

Theoretical Thermochemistry for Organic Molecules: Development of the Generalized Connectivity-Based Hierarchy

Raghunath O. Ramabhadran and Krishnan Raghavachari*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

S Supporting Information

ABSTRACT: A generalized, unique thermochemical hierarchy applicable for all closed shell organic molecules is developed in this paper. In this chemically intuitive, structure-based approach, the connectivity of the atoms in an organic molecule is used to construct our hierarchy called “connectivity-based hierarchy” (CBH). The hierarchy has several rungs and ascending up the hierarchy increasingly balances the reaction energy. It requires no prior knowledge of the types of molecules and hybridizations for the appropriate balancing of the bond types and the bonding environments of the atoms. The rungs can be generated by an automated computer program for any closed shell organic molecule, and the first three rungs generate the simplest reactions for the widely used isodesmic, hypohomodesmotic, and hyperhomodesmotic schemes. The generated reaction schemes are unique for each rung and are derived in a simpler manner than previous approaches, avoiding potential errors. This work also suggests that for closed shell organic molecules, the previously well-studied homodesmotic scheme does not have a fundamental structure-based origin. In a preliminary application of CBH, density functional theory has been used to calculate accurate enthalpies of formation for a test set of 20 organic molecules. The performance of the hierarchy suggests that it will be useful to predict accurate thermodynamic properties of larger organic molecules.

1. INTRODUCTION

Electronic structure theory has been widely used to make thermodynamic predictions on organic molecules for over four decades.^{1–3} Initially, in an era when sophisticated theoretical methods were not yet developed, accurate methods to compute the bond energies and the thermodynamic properties of organic compounds were not available. Subsequently, introduction of the isodesmic bond separation (IBS) scheme by John Pople in 1970^{4–6} enormously improved the accuracy of the predictions using simple theoretical models such as Hartree–Fock theory with moderate basis sets. Pople’s IBS scheme illustrated, for the first time, the significance of appropriately balancing reaction energies resulting in substantial error cancellation and yielding better calculated accuracy.

Furthering the ideas developed by Pople, George and co-workers in 1975^{7–9} proposed the hybridization-based homodesmotic scheme (HS) of reactions for organic molecules. This scheme sought to improve upon Pople’s IBS scheme and was constructed to offer a superior balance of the bond types and the hybridization of the atoms involved. Following this, a plethora of schemes, some of which are called hyperhomodesmotic,¹⁰ semi-homodesmotic,¹¹ quasihomodesmotic,^{12,13} homomolecular homodesmotic,¹⁴ isogeitonic,¹⁵ isoplesitoic,¹⁶ homoplesitoic,¹⁶ and s-homodesmotic^{17–19} have since been developed to successfully predict various properties of organic molecules.

Recently, Wheeler, Houk, Schleyer, and Allen offer a detailed account of the widespread inconsistency in the definition of the term “homodesmotic scheme”.²⁰ Recognizing the need for greater uniformity and generality in such reaction schemes, they developed a general and a systematic hybridization-based hierarchy of homodesmotic reactions for closed shell hydrocarbons. By utilizing predefined reactants and products at each level of their

hierarchy, they achieve an increased balance in the hybridization and the covalent bonding environment of the carbon atoms within the family of hydrocarbon molecules. They further rightly point out the need for a general and a systematic hierarchy that spans beyond hydrocarbons and is applicable to any organic molecule containing any heteroatom (for instance, O, N, S, etc.). Finally, acknowledging the enormous variety of functional groups that can exist in organic molecules, they clearly indicate that predefining reactants and products for various levels of a hierarchy involving organic molecules is a prominent challenge. Thus the construction of a generalized hierarchy for all closed shell organic molecules has remained an open problem.^{4–9,20}

An alternate approach to the construction of the generalized hierarchy, i.e., one based on merely the connectivity of the atoms in an organic molecule instead of utilizing predefined reactants and products in a hybridization-based hierarchy, helps in overcoming this challenge and solves this long-standing problem. Such a hierarchy is applicable to all classes of closed shell organic molecules and does not require significant effort to balance the various coefficients in the chemical equations employed in the hierarchy. Another advantage of such a connectivity-based approach is that it enormously minimizes the use of complicated terminology that is sometimes used in the definition of the different homodesmotic schemes, without any compromise in the goal of preserving the bond types and the hybridization of the atoms involved.

In this work, we develop the general and systematic connectivity-based hierarchy (CBH) for closed shell organic molecules containing various hetero atoms (e.g., N, O, F, P, S, Cl, Br), a variety of functional groups (e.g., alcohols, amines, ketones, aldehydes,

Received: April 21, 2011

Published: June 02, 2011

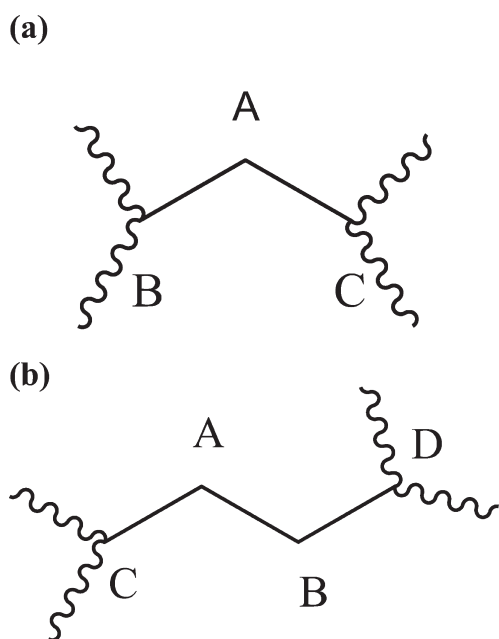


Figure 1. (a) A generic representation of the preservation of atom connectivity. Atom A's connectivity is preserved by keeping it connected with atoms B and C. The wiggly lines merely indicate the extension of the molecule. (b) A generic representation of the preservation of bond connectivity. Bond A–B's connectivity is preserved by maintaining the C–A–B–D framework. The wiggly lines merely indicate the extension of the molecule.

acids, thiols), and different molecular architectures (e.g., acyclic, cyclic, branched, linear) to show how the reaction energy converges better at each increased level in the hierarchy. Construction of the hierarchy requires a knowledge of only the connectivity and the types of bonds (single, double, triple) in the molecule, i.e., *the chemical structure is all that is necessary*, and the reaction schemes can be generated in a simple and automated manner with increasing levels of sophistication. We further go on to prove how the CBH scheme satisfies the bond type and the hybridization requirements set by Wheeler, Houk, Schleyer, and Allen in their hierarchy. Finally, we apply CBH to compute the enthalpies of formation of various closed shell organic compounds and demonstrate how the higher levels of the hierarchy help achieve accurate results using commonly used density functionals with modest double- or triple- ζ quality basis sets.

2. CONSTRUCTION OF THE GENERALIZED CBH FOR ORGANIC MOLECULES

CBH features several rungs at increasing levels of sophistication. They are all based on the local structure and bonding in the molecule under consideration. The rungs alternate between atom- and bond-centric perspectives. The simplest, atom-centric rung is the isogyric scheme^{2,4,20} where each heavy (non-hydrogen) atom is extracted in its saturated valence state (e.g., each C as CH₄) by adding appropriate number of hydrogen molecules. The number of hydrogen molecules equals the number of covalent bonds between heavy atoms (counting a double bond as two covalent bonds, etc.). This rung is named CBH-0. The next, bond-centric rung (CBH-1) of the hierarchy is obtained simply by extracting all the heavy-atom bonds in the molecule as isolated valence-satisfied molecules (e.g., a carbon–carbon double bond

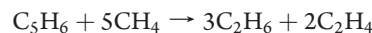
as H₂C=CH₂). This is identical to the widely used Pople's isodesmic bond separation scheme.⁴ At CBH-2, we preserve the immediate connectivity of all the atoms in the molecule, i.e., every heavy atom is extracted with its immediate bonding environment (Figure 1a). Later, we will show that this is equivalent to the simplest hypohomodesmotic reaction scheme developed by Wheeler et al.²⁰ Perhaps a more appropriate, simpler, and illuminative name is "isoatomic" scheme (vide infra). Finally, at CBH-3, we preserve the immediate connectivity of all the bonds in the molecule, i.e., every heavy-atom bond is extracted maintaining its immediate connectivity (Figure 1b). Later, we will show that this is equivalent to the simplest hyperhomodesmotic reaction scheme. Higher levels can be defined in a similar manner. For example, at CBH-4, we extract every heavy atom while maintaining two immediate bonds. While additional higher levels can be defined easily, in this paper, we restrict our discussion to the rungs most likely to be useful, viz. CBH-1 to CBH-3.

In each rung, in order to avoid overcounting of the atoms and bonds, additional molecules have to be included as reactants to balance the equations. We show in this paper (vide infra) that this can be accomplished in an elegant and automated manner. This results from the connections between the products formed in one rung and the reactants in the next higher rung of the hierarchy (Figure 2). We illustrate the construction of CBH using a simple example involving a single-ring system without any substituents, cyclopentadiene²¹ (C₅H₆, Figure 3a):

- CBH-0: based on the isogyric scheme.



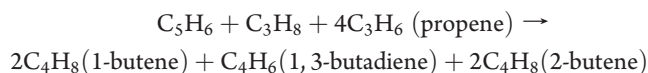
- CBH-1: based on the isodesmic bond separation scheme.



- CBH-2: based on preserving the environment of the atoms (vide infra for more details).



- CBH-3: based on preserving the environment of the bonds (vide infra for more details).



The first observation for this simple system is that the products formed in the lower hierarchy appear exactly as the reactants in the next hierarchy along with the parent molecule itself (Figure 2). This is entirely understandable since we are progressively making a larger part of the local bonding environment on both sides of the reaction as similar as possible. For example, each pair of adjacent heavy-atom bonds (products in CBH-1) shares a common heavy atom (as reactants in CBH-1 as well as products in CBH-0). Similarly, each pair of adjacent atoms preserving their immediate environment (products in CBH-2) shares a common heavy-atom bond (reactants in CBH-2 as well as products in CBH-1). Clearly, the additional reactants at each level are necessary to take care of the double counting that would otherwise be present. Since the products in any hierarchy are easily derived from the local bonding environment as noted earlier, the procedure can be easily automated giving considerable advantage to our scheme. For example,

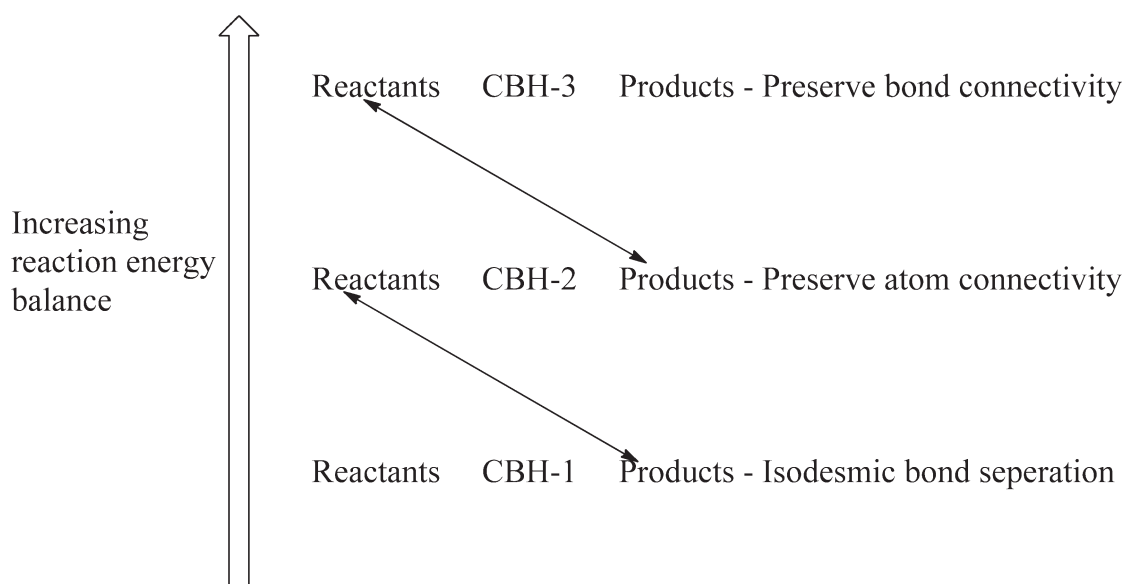
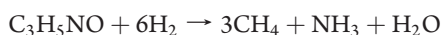


Figure 2. Showing the connection between various rungs of CBH. Reactants and products at the first rung (CBH-1) are obtained based on Pople's IBS scheme. Products of CBH-1 then become the reactants for the next rung (CBH-2). The products for CBH-2 are obtained by preserving the atom connectivity as shown in Figure 1a. Reactants for the third rung (CBH-3) are in turn the same as the products for CBH-2. Products for CBH-3 are then obtained by preserving the bond connectivity as shown in Figure 1b.

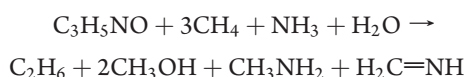
the CBH schemes can be written for other hydrocarbon ring systems, such as 1,3-cyclohexadiene, cyclohexane, etc.

The scheme can also be easily extended to molecules containing heteroatoms. For example, if we take oxazoline (Figure 3b), a simple heterocyclic ring containing oxygen and nitrogen where every heavy atom has a different bonding environment, we have the following reactions:

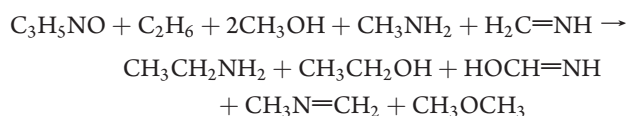
- CBH-0: based on the isogyric scheme.



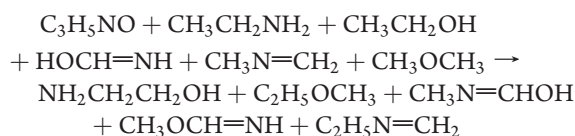
- CBH-1: based on the IBS scheme.



- CBH-2: based on preserving the environment of the atoms



- CBH-3: based on preserving the environment of the bonds



Again, the products in one hierarchy become the reactants in the next hierarchy, and the products of the balanced reaction can be written by inspection or by an automated program. It is worthwhile to observe here that, even for a simple molecule like oxazoline, using a reaction scheme involving predefined reactants and products may introduce complications in selecting and balancing uniquely defined equations at the higher rungs of the hierarchy.

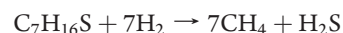
Thus far we have considered simple cyclic systems without any branching or side groups. In order for us to construct the CBH for any organic molecule, cyclic/acyclic and with/without a branching group as well, we need to understand two additional simple facets. Both are straightforward to implement and can also be automated easily.

3. CANCELLATION OF THE MOLECULES REPRESENTING THE TERMINAL MOIETIES

The first new facet that we consider involves molecules that have terminal moieties. For example, in an *n*-alkane chain, the two methyl groups at the two ends are terminal moieties. It is clear that such groups are different from the units in a simple cyclic system considered earlier. The principal difference is that such a terminal moiety does not have two (or more) heavy-atom bonds. In such cases, it is easy to see that (vide infra) the molecules representing such terminal moieties need to be treated differently and get canceled as reactants in the equations representing the next hierarchy.

Further, it is readily seen at different rungs of the hierarchy that we have different molecules representing the terminal moieties. For example, in *n*-heptanethiol (vide infra for the construction of the CBH), the terminal moieties are the methyl and the thiol groups. The molecules representing these moieties as reactants in the different hierarchies are: methane and hydrogen sulfide at CBH-1, ethane and methanethiol at CBH-2, etc.

The cancellation of the molecules that represent terminal moieties can be seen using an example involving an open chain molecule without branching, *n*-heptanethiol ($\text{C}_7\text{H}_{16}\text{S}$, Figure 3c):
CBH-0: isogyric



It is clear that the molecules representing the terminal moieties (one molecule of methane and a molecule of hydrogen sulfide) do not appear as reactants at the next level. The CBH-1 scheme is, thus:

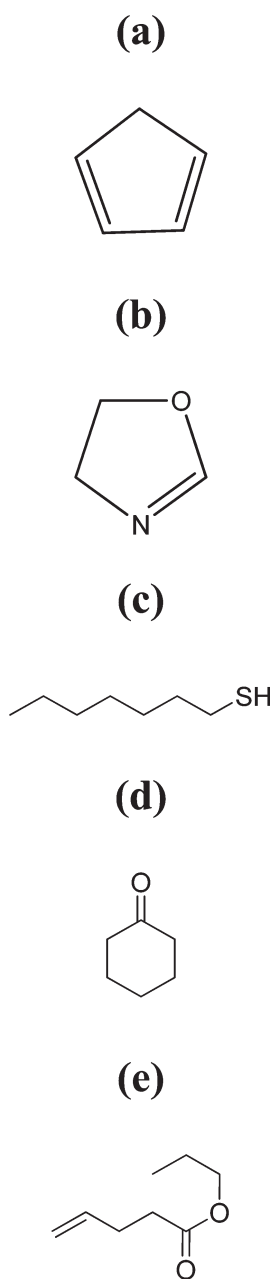


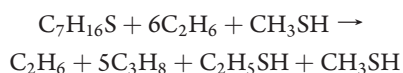
Figure 3. Various molecules used as examples to illustrate CBH: (a) cyclopentadiene, (b) *n*-heptane thiol, (c) oxazoline, (d) cyclohexanone and (e) propyl-pent-4-enoate.

CBH-1: after terminal moiety cancellation



The reaction representing the next hierarchy is:

CBH-2: preserving the environment of the atoms



Here, the products that maintain the local environments of each of the heavy atoms, share some common species with the reactants. After cancellation of common species, we get

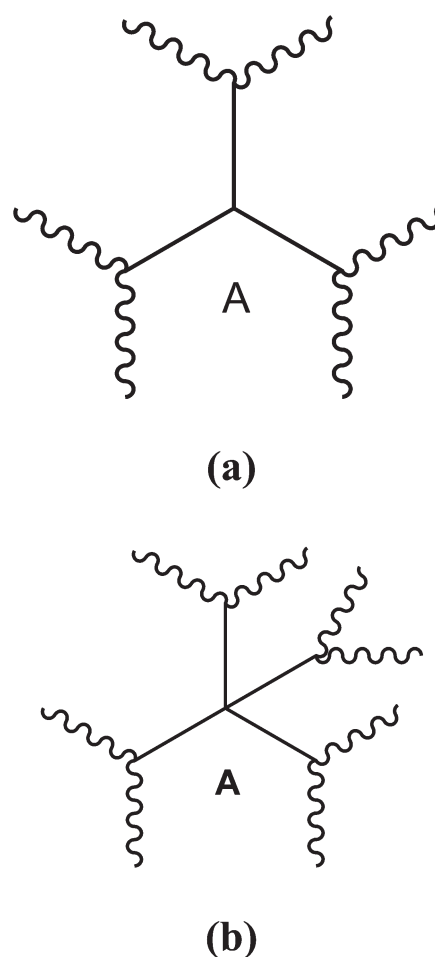


Figure 4. (a) A generic representation of a branch point. Here the branching point is the atom A. The molecule representing this branch point as a reactant is counted twice at the bond-centric rungs. (b) A generic representation of two branch points on the same atom.

Again, the molecules representing the terminal moieties (one molecule of ethane and a molecule of methanethiol) get canceled in this hierarchy.

Using the same logic, one molecule of propane and a molecule of ethanethiol gets canceled in the next hierarchy, and we get

CBH-3: after cancellation



Overall, the molecules representing the terminal moieties (which are products at a lower rung) do not appear in the next hierarchy.

4. TAKING BRANCHING INTO ACCOUNT: A MOLECULE REPRESENTING A BRANCH POINT IS COUNTED TWICE PER BRANCH

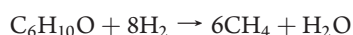
Given the enormous affinity of carbon toward catenation, branching of certain bonds in the structural framework of an organic molecule is very common. At any branching point in an organic molecule, the atom at the branching point is attached to one additional heavy atom in comparison to an atom not at the branching point, i.e., there is an additional covalent bond that needs to be considered at the branching

points. These can be taken into account by adjusting the coefficients of the defining molecules that represent the branch points. These molecules are counted twice per branch when they occur as reactants (Figure 4a). Similarly, if there is an additional branch (Figure 4b, e.g., the central carbon in neopentane), then the molecule as a reactant has to be counted thrice overall.

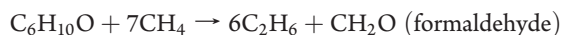
Such a coefficient takes into account the overcounting that would otherwise occur in a simplistic pairwise treatment of intersecting bonds.²² It is important to note here that branching occurs only with bonds. This is due to the fact that, while an atom can be at the intersection of multiple bonds, each covalent bond terminates at exactly two atoms. Thus, the phenomenon of counting twice is applicable only at the bond centric-rungs, i.e., at CBH-1 and CBH-3.

We illustrate this facet of CBH using cyclohexanone (C₆H₁₀O, Figure 3d) as an example. Here, the branching point occurs at the carbonyl carbon. The molecule representing this branching point at CBH-1 is methane. Therefore, methane needs to be counted twice in CBH-1. Similarly, the molecule representing the branch point at CBH-3 is acetone and hence is counted twice at that rung.

CBH-0:

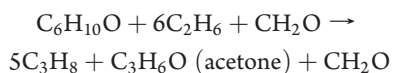


CBH-1: a combination of terminal moiety cancellation and taking the branch point into account



Due to the terminal moiety cancellation, water does not appear in the reactants. On the other hand, due to methane representing a branching point, it is counted twice.

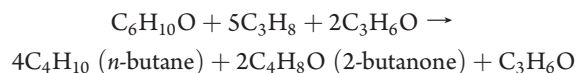
CBH-2:



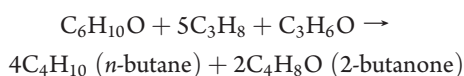
i.e.,



CBH-3: counting acetone twice



i.e.,



The same procedure is thus adopted for any molecule possessing a branched molecular architecture. The Supporting Information features another example of the construction of CBH for a more complex branched molecule.

5. STRAIGHTFORWARD BALANCING OF THE CHEMICAL EQUATIONS AND AUTOMATED

CBH starts with the simple isogyric scheme at the lowest rung (CBH-0). It preserves the atom and bond connectivities such that products of a lower rung occur as the reactants at higher rungs—with two minor modifications: First, the terminal moieties get canceled out. At bond-centric rungs (CBH-1 and CBH-3), molecules representing a branch

position are counted twice per branch. Since these are the only features of CBH, balancing of the chemical equations involved in the hierarchy naturally follows the construction of the hierarchy. Further, since the CBH is entirely automated, given any molecule, the reaction schemes are generated with ease.

6. GENERAL AND UNIQUE FOR ANY CLOSED SHELL ORGANIC MOLECULE AND AVOIDS COMPLICATED TERMINOLOGY

Our hierarchy is based on the connectivity of the atoms in an organic molecule. Hence, it is independent of the elements present in the organic molecule. This makes it general for all organic molecules. Since the connectivity of the atoms in a molecule is fixed, there is only one way of constructing the hierarchy. Thus, CBH is unique for any organic molecule. CBH therefore does not offer the scope for definition-based inconsistencies noted by Wheeler, Houk, Schleyer, and Allen in the homodesmotic schemes (vide supra).

Moreover, the hierarchy does not involve any complicated terminology. The only terms used in the hierarchy are: (a) CBH, connectivity-based hierarchy; (b) molecule representing a terminal moiety; and (c) molecule representing the branch point.

7. HOW DOES CBH PRESERVE HYBRIDIZATION AND BOND TYPES?

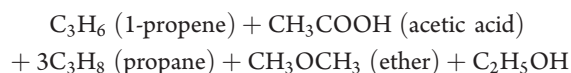
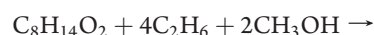
In this section, we show how CBH, in addition to being general and unique for all closed shell organic molecules, preserves the hybridization and the bond types as prescribed by Wheeler, Houk, Schleyer, and Allen in their homodesmotic hierarchy for hydrocarbons as well. Propyl-pent-4-enoate (C₈H₁₄O₂, Figure 3e), having both carbon atoms as well as heteroatoms in different states of hybridizations, is chosen as an example to demonstrate the preservation of hybridization and the bond types. Since CBH-0 corresponds to the trivial, isogyric scheme, which is known to have considerable imbalances in the reaction energies,^{4–6,20} we henceforth start only from the CBH-1 rung.

CBH-1:



This rung of CBH, utilizing Pople's IBS scheme, corresponds to the RC2 level of the homodesmotic hierarchy for closed shell hydrocarbons developed by Wheeler, Houk, Schleyer, and Allen.

CBH-2:



This rung maintains the numbers of all heavy atoms in their different states of hybridization as well as the numbers of all heavy atoms (regardless of the hybridization state) attached to the appropriate number of hydrogens (0–3). Thus, the CBH-2 rung corresponds to the RC3 (hypohomodesmotic) level of the *n*-homodesmotic hierarchy for closed shell hydrocarbons. Due to the confusion in the existing literature on the names of thermochemical schemes, we would like to suggest a more appropriate, much simpler and illuminative name “isoatomic” scheme for this useful rung, since it

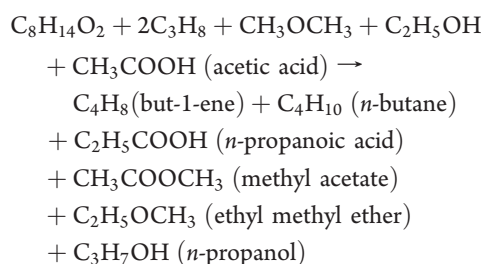
Table 1. Zero Point and Thermal Corrected Mean Absolute Reaction Energies (kcal/mol) Computed Using CBH^{a,b}

density functional	CBH-1	CBH-2	CBH-3
BPW91	26.51	1.38	0.93
BMK	28.79	1.13	0.52
B3LYP	26.15	1.30	0.65
M05-2X	31.71	1.15	0.59
M06-2X	31.00	1.53	0.70
TPSSH	25.88	1.08	0.62
B2PLYP	29.52	0.95	0.51

^a All the geometries were optimized at the B3LYP/6-31G(2df,p) level. A scale factor of 0.9854 was applied to the harmonic frequencies obtained.
^b 6-311++G(3df,2p) basis set was used throughout.

preserves the immediate environment of all the heavy atoms in an organic molecule.

CBH-3:



CBH-3 maintains the number of bond types for all heavy-atom bonds (e.g., H₃C–O, H₂C–O, HC–O, C–O, etc.) as well as the numbers of each type of heavy atoms in the same hybridization state (sp³, sp², sp) attached to appropriate number of hydrogens (0–3). This is as prescribed at the RC5 (hyperhomodesmotic) level of the homodesmotic hierarchy for closed shell hydrocarbons.

The same logic can be extended to all the closed shell organic molecules, and it can be easily shown that CBH preserves the appropriate hybridizations as well as the bond types. It is important to note here that, throughout the construction of the hierarchy, we present a general, physical picture of an organic molecule, solely based on the connectivity of the atoms in the molecule. Terms such as hybridization and bond multiplicities are never used herein. Yet, the hierarchy preserves (vide supra) both the bond multiplicities as well as hybridization.

8. WHAT IS THE REASON FOR THE WIDESPREAD DEFINITION-BASED INCONSISTENCY?

Since our physically motivated, chemical structure-based method does not correspond with the homodesmotic scheme but instead maps appropriately with the hypohomodesmotic and hyperhomodesmotic schemes developed by Wheeler et al.,²⁰ this work enables us to point out that, in general, for any closed shell organic molecule, *the homodesmotic scheme does not have a fundamental structure-based origin*. This is the principal reason for the widespread definition-based inconsistencies noted in the homodesmotic scheme. Wheeler and co-workers had previously noted that multiple definitions of “homodesmotic reactions” are available for the family of closed shell hydrocarbons but simply observed that “homodesmotic reactions must be considered a special case.”²⁰

A final important point to note is that many different hypo- and hyperhomodesmotic schemes can be written for a given molecule. However, by merely examining the chemical structure

Table 2. Computed Reaction Energies using CBH^a

molecular formula	chemical name	CBH-1	CBH-2	CBH-3
C ₄ H ₁₀ N ₂	piperazine	21.77	4.22	−3.16
C ₅ H ₁₀ O	3-pentanone	26.95	0.64	−0.04
C ₅ H ₁₀ O ₂	isopropyl acetate	59.68	−1.76	0.57
C ₅ H ₁₂ S	ethyl propyl sulfide	10.76	−0.26	0.07
C ₆ H ₁₃ NO ₂	6-aminohexanoic acid	57.48	−1.27	−0.10
C ₆ H ₁₂ O	cyclohexanol	23.13	−1.43	−1.53
C ₆ H ₁₀ O	cyclohexanone	30.77	−3.15	−3.67
C ₆ H ₁₂ S	cyclohexanethiol	19.05	−1.40	−0.62
C ₆ H ₁₁ Cl	cyclohexyl chloride	21.88	−1.26	−1.58
C ₆ H ₁₃ Br	n-hexyl bromide	14.01	−0.28	−0.14
C ₇ H ₁₂ O	cyclohexanal	28.18	−2.61	−0.92
C ₇ H ₁₆ S	1-heptanethiol	15.18	−0.27	0.06
C ₈ H ₁₅ N	octanenitrile	26.63	−0.78	0.11
C ₈ H ₁₉ N	dibutylamine	22.17	−0.97	−0.53
C ₈ H ₁₄ O ₂	propyl pent-4-enoate	65.94	−2.25	0.19
C ₈ H ₁₆ O	t-butyl isopropyl ketone	37.06	−2.99	0.04
C ₁₀ H ₁₈	2-decyne	30.40	−2.56	−0.08
C ₁₀ H ₁₉ N	caprinitrile	31.58	−0.90	0.06
C ₁₁ H ₂₁ N	1-cyanodecane	33.92	−1.10	−0.10
C ₁₂ H ₂₄ O	decyl methyl ketone	43.53	−0.53	−0.57

^a At the M06−2X/6-311++G(3df,2p) // B3LYP/6-31G(2df,p) level of theory. The reaction energies include zero-point and thermal corrections (see text).

of any organic molecule, the simplest and unique reaction schemes are the ones generated using our CBH-2 and CBH-3 rungs. Overall, the *n*-homodesmotic reaction scheme for closed shell hydrocarbons indicated by Wheeler et al.²⁰ is similar in spirit to our hierarchy. However, our structure-based formalism is much more general, applicable to a substantially wider variety of molecules and avoids any definition-based inconsistencies noted previously in the homodesmotic hierarchy. Further, it is easily constructed without using predefined reactants and products, resulting in substantial simplicity in the implementation.²¹

9. TEST SET

In our efforts to duly represent the enormous structural variety of organic compounds, we have chosen organic molecules containing the following features: a variety of functional groups (for instance, alcohols, amines, ketones, aldehydes, acids, thiols), hetero atoms (N, O, F, S, Cl, Br), and different molecular architectures (for instance, acyclic, cyclic, branched, linear). Our test set consists of 20 molecules containing between 6 and 12 heavy atoms, consistent with our goal of applying these methods to larger and more general classes of organic molecules. The complete test set is provided in the Supporting Information as well as in Table 2 (vide infra).

10. COMPUTATIONAL METHODS

All the computations have been performed using the Gaussian 09²³ suite of programs. In this initial evaluation study, we have used seven different density functional methods, possessing different rungs of exchange and correlation energy functionals. They are: (a) the BPW91 generalized gradient approximation (GGA) functional,^{24–26} (b) the kinetic energy density (τ) dependent BMK functional,²⁷ (c) the popular B3LYP^{28,29} hybrid

functional, (d and e) Truhlar's meta-exchange–correlation functionals, M05-2X³⁰ and M06-2X,³¹ (f) Perdew and Scuseria's hybrid meta-GGA functional TPSSH,³² and (g) Grimme's double-hybrid functional B2PLYP.³³

To obtain reliable geometries, we use the same level of theory as employed by the popular G4 method,^{34,35} i.e., B3LYP/6-31G(2df,p) level. A scale factor of 0.9854 was applied in the calculation of the zero point and thermal corrections to the enthalpy. The optimized geometries were then used in single point calculations using the six different density functionals (vide supra). Six different modest-sized basis sets, three with Pople-style (6-31+G(d,p), 6-31G(2df,p), and 6-311++G(3df,2p)) and three with Dunning-style (aug-cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ) have been used in these single point computations to study the effect of the basis sets. While much larger basis sets are needed to get converged results, we use these more practical, modest basis sets to evaluate if they are sufficient to have significant error cancellations. As expected from the systematic construction of the CBH,

Table 3. Zero Point and Thermal Corrected Mean Absolute Reaction Energies (kcal/mol) for Different Pople and Dunning Style double- and triple- ζ Basis Sets^{a,b}

basis set	CBH-1	CBH-2	CBH-3
6-31+G(d,p)	30.76	1.66	0.79
6-31 g(2df,p)	30.41	1.20	0.91
6-311++g(3df,2p)	31.00	1.53	0.70
aug-cc-pVDZ	29.67	1.88	0.79
cc-pVTZ	29.93	1.40	0.76
aug-cc-pVTZ	30.02	1.42	0.72

^aAll the geometries were optimized at the B3LYP/6-31G(2df,p) level. A scale factor of 0.9854 was applied to derive the zero point and thermal energies. ^bThe M06–2X density functional was used throughout.

we do indeed observe that they lead to considerable cancellation (vide infra).

11. RESULTS AND DISCUSSION

We evaluate the performance of CBH in two steps: In the first step, we evaluate the reaction energies for a test set of molecules at the different rungs of the hierarchy. Since there is an increasingly better matching of the bond types, the reaction energies should get smaller at the higher hierarchies. This is similar to the approach used by Wheeler et al. In the second step, we apply the CBH to evaluate the enthalpies of formation for the set of molecules at different hierarchies and compare to the available experimental data.

A. Performance of CBH: Reaction Energy Convergence

Table 1 provides the mean absolute values for the reaction energies computed at the different rungs of CBH. It can be readily seen that, for all the density functionals, as we go from CBH-1 to CBH-3, there is a greater balance of the bond types leading to a decrease in the calculated reaction energies. At CBH-1, the mean absolute value of the reaction energies varies from 25 to 31 kcal/mol (Table 1). It is consistently observed that, both at CBH-2 and CBH-3, the mean absolute reaction energies are less than 2 kcal/mol. In fact, the numbers obtained at CBH-3 level indicate convergence to the subkcal/mol level. These results suggest that chemical accuracy (typically taken to be $\pm 1-2$ kcal/mol) may be possible even with such simple theoretical models.

A glance at Table 2 reveals the performance of the individual molecules at the three rungs (data for the M06-2X functional are given in Table 2, and data for the rest of the functionals are given in the Supporting Information). For all the molecules, irrespective of the functional used, reaction energies at CBH-1 are found to be endothermic. At CBH-2 and CBH-3, significant improvement is seen: the smaller reaction energies indicating better

Table 4. Difference between Experimentally Determined Enthalpies of Formations and Computed Enthalpies of Formation (298 K, expt–theory, kcal/mol) at the Various Rungs of CBH {M06-2X/6-311++G(3df,2p)//B3LYP/6-31G(2df,p)}

molecular formula	chemical name	expt ΔH_f	\pm	CBH-1	CBH-2	CBH-3
C ₄ H ₁₀ N ₂	piperazine	6.0 ^a	1.5	−3.61	−6.38	0.98
C ₅ H ₁₀ O	3-pentanone	−61.7 ^b	0.2	0.04	1.12	0.33
C ₅ H ₁₀ O ₂	isopropyl acetate	−115.1 ^b	0.1	−2.62	−0.66	0.11
C ₅ H ₁₂ S	ethyl propyl sulfide	−25.03 ^a	0.19	−1.71	−0.28	0.28
C ₆ H ₁₃ NO ₂	6-aminohexanoic acid	−115.28 ^a	0.72	−3.24	−2.25	0.29
C ₆ H ₁₂ O	cyclohexanol	−69.3 ^b	0.2	−2.99	−1.02	−0.93
C ₆ H ₁₀ O	cyclohexanone	−55.23 ^a	0.21	−3.32	−1.96	−1.61
C ₆ H ₁₂ S	cyclohexanethiol	−22.88 ^a	0.19	−3.19	−1.38	−0.95
C ₆ H ₁₁ Cl	cyclohexyl chloride	−39.79 ^a	0.46	−3.14	−2.53	−0.72
C ₆ H ₁₃ Br	<i>n</i> -hexyl bromide	−35.88 ^a	0.49	−2.77	−1.28	−1.21
C ₇ H ₁₂ O	cyclohexanal	−56.2 ^b	NA	−4.70	−2.57	−1.29
C ₇ H ₁₆ S	1-heptanethiol	−35.73 ^a	0.23	−2.01	−0.37	0.72
C ₈ H ₁₅ N	octanenitrile	−12.1 ^a	0.36	−1.72	−1.06	0.71
C ₈ H ₁₉ N	dibutylamine	−40.89 ^a	0.76	−5.61	−6.18	−2.88
C ₈ H ₁₄ O ₂	propyl pent-4-enoate	−95.1 ^a	0.8	−2.40	0.05	1.84
C ₈ H ₁₆ O	<i>t</i> -butyl isopropyl ketone	−80.86 ^a	0.29	−2.58	0.19	−0.55
C ₁₀ H ₁₈	2-decyne	5.63 ^a	0.82	−0.73	−1.61	0.47
C ₁₀ H ₁₉ N	caprinitrile	−21.9 ^a	0.43	−2.20	−1.14	0.86
C ₁₁ H ₂₁ N	1-cyanodecane	−27.10 ^a	0.48	−2.88	−1.62	0.50
C ₁₂ H ₂₄ O	decyl methyl ketone	−96.62 ^a	0.59	−3.15	−0.97	−0.09

^aExperimental enthalpies of formation taken from the NIST Web site. ^bExperimental enthalpies of formation taken from ref 36.

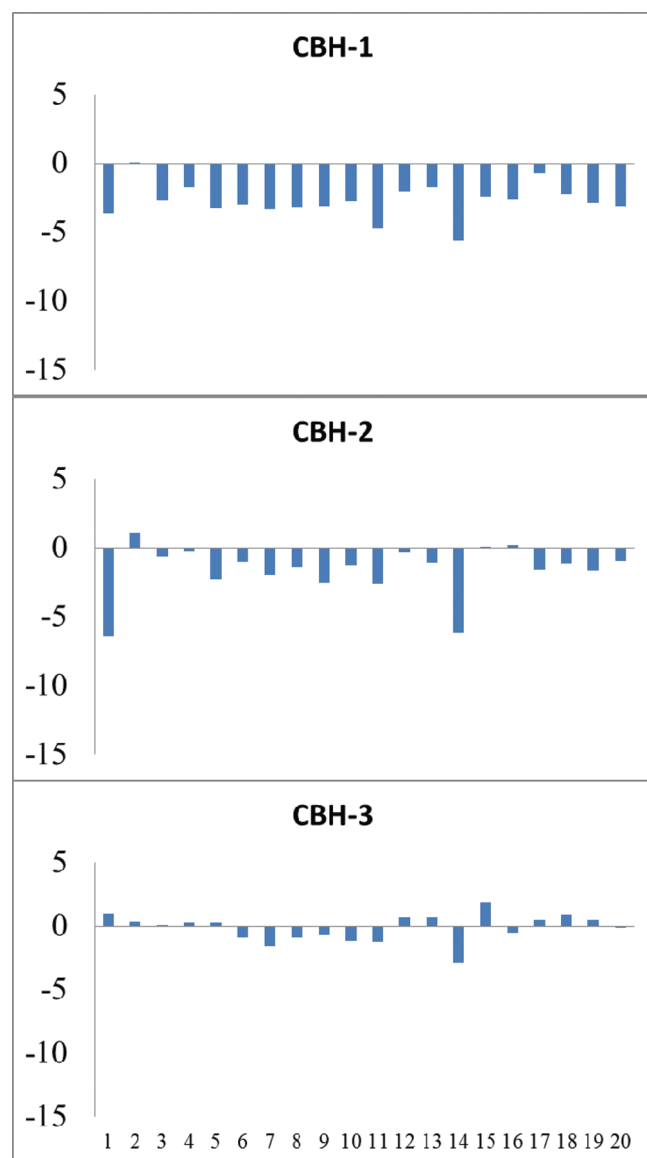


Figure 5. Pictorially depicting the errors (experiment–theory, vertical axis, kcal/mol; also see Table 4) for all the molecules used in the test set. The molecules listed in the horizontal axis follow the same order as listed in Table 4. Data are provided here at the M06-2X/6-311++G(3df,2p) // B3LYP/6-31G(2df,p) level of theory. Supporting Information provides the data for rest of the density functionals used.

balance for the various heteroatom substituents as well as for different molecular architectures.

B. Cases When the Reaction Energies Do Not Appropriately Get Balanced. When CBH is constructed for an aromatic molecule, such as azulene, it is clear that the aromaticity is not preserved. This leads to imbalances in the reaction energies at various rungs (see Supporting Information). Similarly, for a strained bicyclic organic compound such as camphor, the extent of the ring strain is not exactly balanced upon the construction of CBH. Thus, the reaction energies do not converge for such strained molecules at the different rungs (see Supporting Information).

A molecule such as hexafluoro-2,4-hexadiyne nicely illustrates the significance of appropriately preserving the electronic

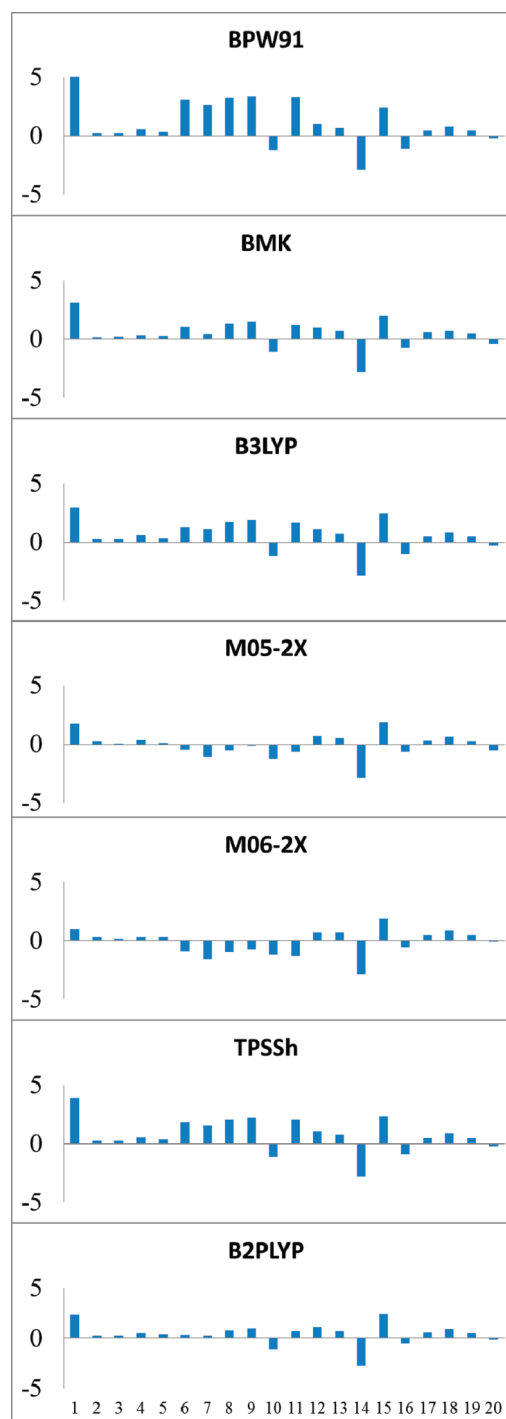


Figure 6. A comparison of the errors (experiment–theory, vertical axis, kcal/mol) in the enthalpies of formation of all the molecules used in our test set at the CBH-3 rung using different density functionals. The molecules listed in the horizontal axis follow the same order as listed in Table 4 (and Tables S8–S13 in the Supporting Information).

environment in leading to a balance in the reaction energies. For this diyne, (see Supporting Information) at CBH-2 there is a considerable imbalance in the reaction energies. However, at CBH-3 the reaction energy is well balanced. A possible reason might be that, in the construction of CBH-2 for hexafluoro-2,4-hexadiyne, the products side does not involve any molecule which has a fluorine atom as well as a triple bond in the same molecule, i.e., the electronic

environment of hexafluoro-2,4-hexadiyne is not appropriately preserved. At CBH-3, the products side does involve a molecule which has a fluorine atom as well as a triple bond in the same molecule, and hence the reaction energy gets balanced at CBH-3.

C. Effect of the Basis Sets. CBH works robustly with both Pople-style as well as Dunning-style basis sets. Table 3 demonstrates how the mean absolute values for the reaction energies converge better as we move up the hierarchy. CBH performs well for the double- ζ as well as triple- ζ basis sets, once again illustrating the enhanced balancing of reaction energies at the higher rungs of the hierarchy. It is also observed that, for each rung, the mean absolute reaction energies do not oscillate much (Table 3) upon changing the basis sets. At CBH-3, for all the basis sets employed, convergence of the unsigned averages for the reaction energies toward subkcal/mol is noticed as well. Thus, even by merely using double- and triple- ζ basis sets routinely employed by even nonexperts in theory these days, a very high degree of convergence in the reaction energies is noted, thereby illustrating the practical utility of CBH.

D. Performance of CBH: Application to Computing Enthalpies of Formations. Table 4 contains the difference between the experimentally determined enthalpies of formations and the computed enthalpies of formation (at 298 K, kcal/mol) at the various rungs of CBH [M06-2X/6-311++G(3df,2p)//B3LYP/6-31G(2df,p) shown with the results for the rest of the functionals provided in the Supporting Information]. At each rung, the computed reaction energies have been used in conjunction with the experimentally determined enthalpies of formation for all the molecules that appear as reactants and products (at that particular rung). Experimental enthalpies of formation have been obtained from the NIST Web site and reference.³⁶³⁶ Errors in the computed enthalpies of formation for all the molecules are shown pictorially in Figure 5 (as well as in the Supporting Information) for the different rungs of the hierarchy.

For most of the molecules (Table 4, Figure 5, and Supporting Information), the difference between the experimentally observed enthalpies of formations and the corresponding computed values (kcal/mol) decreases dramatically as we go from CBH-1 to CBH-3. For the M06-2X functional (Table 4), the largest error at the CBH-3 level (-2.88 kcal/mol) occurs for dibutylamine. It is interesting to note that a nearly identical error for dibutylamine occurs for all seven density functionals (Table 4, Supporting Information). Figure 6 shows the errors in the computed enthalpies of formation for all the molecules at the CBH-3 rung for each of the seven functionals. There is clearly some correlation in the performance of the different density functionals. This may be useful in the future to separate the contributions of systematic vs random errors in such calculated enthalpies of formation.

Table 5 lists the mean absolute errors in the calculated enthalpies of formation for the 20 molecule test set for the 7 different density functionals used. Substantial improvement is noted on ascending from CBH-1 to CBH-2 in all cases with the best results for the B2PLYP functional. As expected, (based on the smaller changes in reaction energies, Table 1), going up from CBH-2 to CBH-3 leads to a less dramatic improvement. For all except the lowest rung BPW91 functional, some improvement is seen on going from CBH-2 to CBH-3. The most prominent improvement is noted with Truhlar's M05-2X and M06-2X functionals, though Grimme's B2PLYP double hybrid functional also has a mean absolute deviation of less than 1 kcal/mol. The mean absolute deviation from experiment is less than 2 kcal/mol at both CBH-2 and CBH-3 levels for all the density functionals,

Table 5. Mean Absolute Errors in the Calculated Enthalpies of Formation (298 K, kcal/mol) using CBH^{a,b}

density functional	CBH-1	CBH-2	CBH-3
BPW91	7.22	1.59	1.68
BMK	4.93	1.29	1.00
B3LYP	7.57	1.56	1.17
M05-2X	2.07	1.37	0.75
M06-2X	2.73	1.73	0.86
TPSSH	7.85	1.30	1.29
B2PLYP	4.21	1.15	0.89

^a 6-311++G(3df,2p) basis set was used throughout. ^b Experimental enthalpies of formation taken from the NIST Web site or from ref 36 (see Table 4 for more details).

remarkable considering the size of the molecules and the modest levels of theory used (DFT with DZP or TZP basis sets).

Overall, at the higher rungs of the hierarchy, sub kcal/mol accuracy is readily achieved in the enthalpies of formation as well as the reaction energies, thus indicating the widespread utility of CBH.

12. CONCLUSIONS AND SCOPE

In this work, we have developed the connectivity-based hierarchy (CBH), general and unique for all closed shell organic molecules. Construction of the hierarchy is straightforward, and it does not involve any complicated terms. The hierarchy preserves the hybridization and the bond types for all closed shell organic molecules, thus generalizing the homodesmotic hierarchy for closed shell hydrocarbons developed by Wheeler, Houk, Schleyer, and Allen to all closed shell organic molecules.

Generalization of their hierarchy to all organic molecules, constructed utilizing predefined reactants and products, presents the challenge of encompassing the massive structural variety possible with organic molecules. By employing a physical picture based merely on the atom connectivity in an organic molecule, we are able to generalize our CBH to all closed shell organic molecules, thus solving a long-standing problem in theoretical thermochemistry. Ascending up the rungs of CBH offers a greater reaction balance. We have demonstrated using different density functionals here, that, with merely using double- or triple- ζ basis sets, satisfactory reaction energy balance can be obtained. The success of CBH, however, depends on how well the environment of a molecule is preserved. Thus in certain cases, such as in aromatic molecules and strained molecules, when the aromaticity or the ring strain is not appropriately preserved, the reaction energies do not get balanced, even at higher rungs of the hierarchy.

Application of CBH to compute enthalpies of formations of various organic molecules shows that the computed values agree well with respect to experimentally observed values. We are currently working on theoretical predictions of thermochemical properties of a variety of organic molecules using CBH with density functional theory as well as wave function-based methods.

■ ASSOCIATED CONTENT

S Supporting Information. The performance of all the molecules in the test set using different functionals are provided. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: kraghava@indiana.edu.

ACKNOWLEDGMENT

The authors acknowledge support from NSF grant CHE-0911454 and Indiana University.

REFERENCES

- (1) *Theory and Applications of Computational Chemistry The First Forty years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds; Elsevier: Amsterdam, The Netherlands, 2005; pp 1–7.
- (2) For one of the earliest papers that recognize the need for conserving the correlation energy to predict thermodynamic properties, see: Snyder, L. C. *J. Chem. Phys.* **1967**, *46*, 3602–3606.
- (3) For one of the earliest use of hydrogenation, isogyric schemes to predict thermodynamic properties, see: Snyder, L. C.; Basch, H. *J. Am. Chem. Soc.* **1969**, *91*, 2189–2198.
- (4) Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. *J. Am. Chem. Soc.* **1970**, *92*, 4796–4801.
- (5) Radom, L.; Hehre, W. J.; Pople, J. A. *J. Am. Chem. Soc.* **1971**, *93*, 289–300.
- (6) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley-Interscience: New York, 1986.
- (7) George, P.; Trachtman, M.; Bock, C. W.; Brett, A. M. *Theor. Chem. Acc.* **1975**, *38*, 121–129.
- (8) George, P.; Trachtman, M.; Bock, C. W.; Brett, A. M. *Tetrahedron* **1976**, *32*, 317–323.
- (9) George, P.; Trachtman, M.; Bock, C. W.; Brett, A. M. *J. Chem. Soc., Perkin Trans. 2* **1976**, 1222–1227.
- (10) Hess, B. A., Jr.; Schaad, L. J. *J. Am. Chem. Soc.* **1983**, *105*, 7500–7505.
- (11) Nyulaszi, L.; Vamai, P.; Veszpremi, T. *J. Mol. Struct. THEO-CHEM* **1995**, *358*, 55–61.
- (12) Vianello, R.; Liebman, J. F.; Maksic, Z. B. *Chem.—Eur. J.* **2004**, *10*, 5751–5760.
- (13) Vianello, R.; Liebman, J. F.; Maksic, Z. B.; Muller, T. J. *J. Phys. Chem. A* **2005**, *109*, 10594–10606.
- (14) Chestnut, D. B.; Davis, K. M. *J. Comput. Chem.* **1997**, *18*, 584–593.
- (15) El-Nahas, A. M.; Bozelli, J. W.; Simmie, J. M.; Navarro, M. V.; Black, G.; Curran, H. J. *J. Phys. Chem. A* **2006**, *110*, 13618–13623.
- (16) George, P.; Bock, C. W.; Trachtman, M. *Theor. Chim. Acta* **1987**, *71*, 289–298.
- (17) Gimarc, B. M.; Zhao, M. *J. Phys. Chem.* **1994**, *98*, 1596–1600.
- (18) Zhao, M.; Gimarc, B. M. *J. Phys. Chem.* **1993**, *97*, 4023–4030.
- (19) Warren, D. S.; Gimarc, B. M. *J. Phys. Chem.* **1993**, *97*, 4031–4035.
- (20) Wheeler, S. E.; Houk, K. N.; Schlyer, P. v. R.; Allen, W. D. *J. Am. Chem. Soc.* **2009**, *131*, 2547–2560.
- (21) In the Supporting Information of ref 20, the authors construct their *n*-homodesmotic schemes for cyclopentadiene and 1,3-cyclohexadiene. Inspection of their scheme for the RC5 level in their hierarchy (the hyperhomodesmotic level) for both these molecules shows that their reactions do not match the criteria set by them in their definition of the RC5 (hyperhomodesmotic) level. For instance at the RC5 level, in the case of cyclopentadiene, there are only two H₃C–CH bonds in the reactants side, whereas there are four H₃C–CH bonds in the products side. This example clearly illustrates how even for a simple hydrocarbon like cyclopentadiene, using predefined reactants and products in the construction of a thermochemical hierarchy, can lead to minor oversights. Consequently, the balance in the reaction energy is slightly affected. In fact, this may be responsible for the slight deterioration of the performance observed by Wheeler et al. at the RCS level for cyclic molecules. A physical, connectivity-based approach, naturally following the chemical structure of an organic compound, precludes the occurrence of such minor issues.
- (22) In combinatorics, this phenomenon is often called the “inclusion–exclusion principle” or the “sieve principle”.
- (23) Frisch, M. J.; G. W. T., Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; ; Raghavachari, K.; A. R., J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, Fox, D. J.; *Gaussian 09*, Gaussian, Inc.: Wallingford CT, 2009.
- (24) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiollhais, C. *Phys. Rev. B* **1992**, *46*, 6671–6687.
- (25) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiollhais, C. *Phys. Rev. B* **1998**, *48*, 4978.
- (26) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533–16539.
- (27) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405–3416.
- (28) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (29) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (30) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.
- (31) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (32) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- (33) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
- (34) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127*, 124105.
- (35) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127*, 084108.
- (36) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data* **1988**, *17S1*, 1–861.

Extraordinary Difference in Reactivity of Ozone (OOO) and Sulfur Dioxide (OSO): A Theoretical Study

Yu Lan,[†] Steven E. Wheeler,[‡] and K. N. Houk^{*,†}

[†]Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095-1569, United States

[‡]Department of Chemistry, Texas A&M University, College Station, Texas 77843, United States

 Supporting Information

ABSTRACT: Ozone and sulfur dioxide are valence isoelectronic yet show very different reactivity. While ozone is one of the most reactive 1,3-dipoles, SO₂ does not react in this way at all. The activation energies of dipolar cycloadditions of sulfur dioxide with either ethylene or acetylene are predicted here by B3LYP, M06-2X, CBS-QB3, and CCSD(T) to be much higher than reactions of ozone. The dipolar cycloaddition of ozone is very exothermic, while that of sulfur dioxide is endothermic. The prohibitive barriers in the case of SO₂ arise from large distortion energies as well as unfavorable interaction energies in the transition states. This arises in part from the HOMO–LUMO gap of sulfur dioxide, which is larger than that of ozone. Valence bond calculations also show that while ozone has a high degree of diradical character, SO₂ does not, and is better characterized as a ditterion.

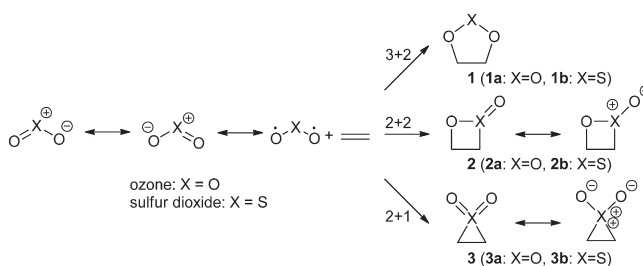
INTRODUCTION

We have recently computed activation barriers and reaction energies for 1,3-dipolar cycloaddition reactions (3 + 2 cycloadditions) of 48 1,3-dipoles¹ as part of a general investigation of the factors that control reactivities in 1,3-dipolar cycloadditions. We were struck by the very great difference between the reactivity of ozone, O₃, which reacts with virtually all CC multiple bonds, and SO₂, which is so unreactive that it is not even considered a 1,3-dipole. This prompted a more detailed investigation of the various cycloadditions, such as those outlined in Scheme 1, that these species can undergo, as well as an exploration of the differences in reactivity.

The 1,3-dipolar cycloaddition² is a powerful synthetic tool used in materials chemistry,³ drug discovery,⁴ and chemical biology.⁵ Quantum chemistry calculations⁶ and molecular dynamics investigations⁷ have shown that the 3 + 2 cycloaddition of 1,3-dipoles to C–C multiple bonds occurs via a concerted transition state, as proposed by Huisgen.⁸

The factors controlling reactivity in 1,3-dipolar cycloadditions have recently been elucidated.^{9,12} Frontier molecular orbital (FMO) theory¹⁰ is a successful qualitative reactivity model for 1,3-dipolar cycloadditions. This model commonly involves the use of approximate quantum mechanical methods and the calculation of electronic properties of isolated reactants. To evaluate the cycloaddition reactivity, conceptual density functional theory and configuration mixing also have been employed.¹¹ We have found that most of the activation energy of a 1,3-dipolar cycloaddition is due to the distortion of the reactants.¹¹ Interaction energies, such as FMO effects, slightly reduce these barriers. The distortion energies of 1,3-dipolar cycloadditions of ethylene and acetylene and for substituted alkenes and alkynes correlate rather quantitatively with activation energies.¹² Recently, Hiberty demonstrated that the diradical character of 1,3-dipoles is related to computed activation barriers,¹³ with lower barriers correlated with more substantial diradical character, but the correlation is not quantitative.

Scheme 1. Possible Cycloaddition Reactions of Ozone or Sulfur Dioxide with Ethylene



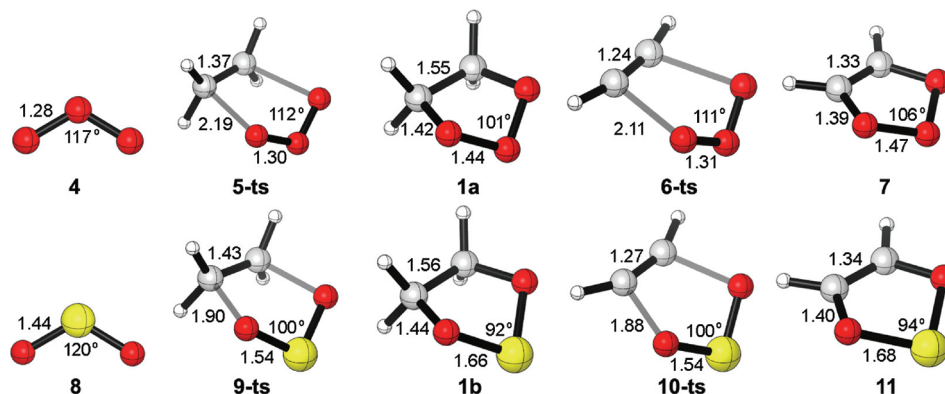
Ozone (O₃), which like all 1,3-dipoles is represented in an all-octet valence bond structure as a zwitterionic (dipolar) species,¹⁴ readily reacts with alkenes and alkynes.¹⁵ This reaction is used for the synthesis of ketones, aldehydes, and epoxides or derivatives.¹⁶ Early work employing natural orbital occupation and generalized valence bond computations showed the significant diradical character of ozone,¹⁷ and ab initio studies have shown that the 1,3-dipolar cycloaddition of ozone occurs via a concerted transition state.¹⁸ Barriers for these reactions are difficult to compute using standard computational approaches, leading to predicted activation enthalpies for the cycloaddition of O₃ with ethylene and acetylene ranging from 2 to 18 kcal/mol and 5 to 23 kcal/mol, respectively.¹⁹ Recent high-accuracy focal point extrapolations yielded activation enthalpies for the cycloadditions with ethylene and acetylene of 5.3 and 9.4 kcal/mol, respectively.²⁰

Sulfur dioxide (SO₂), in which the central oxygen atom of ozone is replaced by sulfur, has the same Lewis structures as ozone.²¹ The geometries of SO₂ and O₃ are similar, with bond angles of 120° for SO₂ and 117° for O₃. The bond length of SO₂ (1.43 Å) is longer than the bond length of O₃ (1.27 Å).²²

Received: April 27, 2011

Published: May 27, 2011

Scheme 2. Geometries of Ozone, Sulfur Dioxide, and the Transition States and Cycloadducts of 1,3-Dipolar Cycloadditions with Ethylene and Acetylene Calculated by the CCSD(T)/cc-pV(T+d)Z Method



Although many 3 + 2 cycloadditions of ozone are known, there is no experimental report of a 3 + 2 cycloaddition of sulfur dioxide (SO₂), and when sulfur dioxide and an alkene or alkyne are mixed, no dioxathiolane is formed.²³

As shown in Scheme 1, there are three possible cycloaddition pathways for the reaction between SO₂ and ethylene. The 3 + 2 cycloaddition generates the 1,3,2-dioxathiolane **1b**. A previous theoretical investigation²⁴ showed that this 3 + 2 cycloaddition can proceed on a graphite surface at 900 °C. Although the reaction energy, ring strain of products,²⁵ and reactivity toward carbon surfaces²⁶ of SO₂ have been studied, the activation energy of the 3 + 2 cycloaddition of SO₂ with simple alkenes or alkynes has not been reported. The endothermic 2 + 2 cycloaddition generates 1,2-oxathietene 2-oxide (**2b**).^{24,27} The reverse reaction of the thermal extrusion of SO₂ from **2b** is used to generate a double bond via a proposed concerted pathway.²⁸ The generation of 2 + 1 cycloaddition product thiirane 1,1-dioxide (**3b**) is also endothermic.²⁷ Oxidation of thiirane by oxone gives **3b**, which is a general method to prepare thiirane 1,1-dioxide.²⁹ The thermal extrusion of SO₂ from **3b** gives the alkene.³⁰ The methanolysis of complex **3b** is used to generate methyl 2-methoxyethanesulfinate.³¹ In the present work, we focus on the different reactivities of ozone and sulfur dioxide in 3 + 2 cycloadditions.

COMPUTATIONAL METHODS

Optimized geometries, reaction barriers, and reaction energies were computed at the B3LYP/6-311+G(d), M06-2X/6-311+G(d), CBS-QB3, and CCSD(T)/cc-pV(T+d)Z levels of theory. B3LYP³² is the most widely used hybrid generalized gradient approximation (GGA) functional.³³ M06-2X is a recently developed hybrid meta GGA functional parametrized for nonmetal atoms.³⁴ It has been benchmarked against a number of main-group energetic databases and is suitable for main-group thermochemistry and kinetics,³⁵ provided that adequate integration grids are used.^{35c} The complete basis set (CBS) methods³⁶ strive to eliminate errors that arise from basis set truncation in quantum mechanical calculations by extrapolating to the CBS limit by exploiting the N¹⁻ asymptotic convergence of MP2 pair energies calculated from pair natural orbital expansions. In particular, CBS-QB3 uses B3LYP/6-311G(2d,d,p) geometries and frequencies³⁷ followed by CCSD(T), MP4(SDQ), and MP2 single-point calculations and a CBS

extrapolation.³⁶ While ordinarily reliable, CBS-QB3 has previously been shown to predict anomalous activation energies in some cases,³⁸ including for 1,3-dipolar cycloadditions of ozone to ethylene and acetylene.²⁰ These errors have been shown²⁰ to arise from the exclusive reliance on MP2 to recover basis set effects. Consequently, in this work, geometries and energies were also computed using CCSD(T)³⁹ paired with the cc-pV(T+d)Z basis set.^{40a} At this level of theory, benchmark values²⁰ for the reaction barriers of O₃ + C₂H₂ and C₂H₄ are reproduced within 0.5 kcal/mol, and the corresponding reaction energies are within 1.0 kcal/mol. Use of this modified form of the standard cc-pVTZ basis set of Dunning^{40b} has been shown to be necessary to achieve qualitatively correct geometries and energies for some sulfur-containing molecules.^{40c} All B3LYP, M06-2X, and CBS-QB3 calculations were performed here using the Gaussian 09 package,⁴¹ while the CCSD(T) calculations were performed using CFOUR.⁴²

Reported molecular orbital energies were evaluated at the RHF/6-311++G(2d,p) level of theory to avoid the poor estimates for ionization potentials of small molecules provided by Kohn–Sham orbitals. The large basis set also can give better estimates of unoccupied orbital eigenvalues.⁴³

Finally, the diradical character of various species was quantified using valence bond (VB) theory,⁴⁴ a powerful tool to calculate the weights of VB structures for 1,3-dipoles.^{13,45} Here the D-BOVB⁴⁶ approach was used to calculate the weights of VB structures for ozone and sulfur dioxide. The VB calculations were carried out with the Xiamen VB (XMVB) package.⁴⁷

RESULTS AND DISCUSSION

1. Competition of the Dipolar Cycloaddition Reactivity between Ozone and Sulfur Dioxide. The geometries of the transition states and products for the dipolar cycloaddition of ozone and sulfur dioxide with ethylene and acetylene are shown in Scheme 2. Ozone (**4**) reacts with ethylene to form the primary ozonide **1a** via transition state **5-ts**. In this transition state, the O–O bond length increases by 0.02 Å, and the O–O–O bond angle is reduced about 5°, compared to isolated O₃. The forming C–O bond is 2.19 Å. In product **1a**, the O–O bond is increased to 1.44 Å, and the bond angle is reduced to 101°. The activation energy for the dipolar cycloaddition of ozone and ethylene is known to be only 3.4 kcal/mol,²⁰ and as seen in Table 1, all but the CCSD(T) method underestimate this value. The reaction is

Table 1. Activation Energies and Reaction Energies for the 1,3-Dipolar Cycloaddition of Ozone and Sulfur Dioxide to Ethylene and Acetylene

	B3LYP	CBS-QB3	M06-2X	CCSD(T)
5-ts	-1.5	0.3	-1.3	3.6
1a	-57.7	-55.1	-74.9	-57.3
6-ts	2.7	4.7	2.9	8.0
7	-65.6	-60.9	-79.2	-64.1
9-ts	39.9	49.4	39.3	51.0
1b	2.0	17.1	-12.2	13.7
10-ts	43.2	52.6	43.6	54.7
11	-10.0	8.6	-20.7	4.1

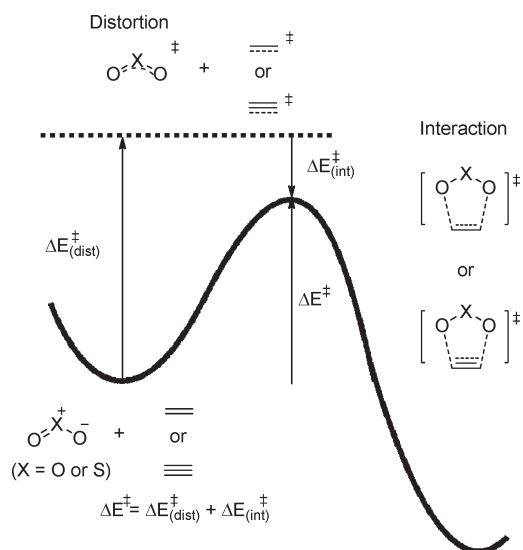
highly exothermic ($\Delta E = -56.4$ kcal/mol),²⁰ and CBS-QB3 and CCSD(T) predict values in qualitative agreement with this benchmark result. B3LYP yields a similar prediction (-57.7 kcal/mol), while the M06-2X reaction energy is about 20 kcal/mol too exothermic.

Sulfur dioxide reacts with ethylene to form cycloadduct **1b** via transition state **9-ts**. The O–S bond length increases 0.10 Å in the transition state, and the O–S–O angle is reduced by 20°. The forming C–O bond is 1.90 Å. These geometric differences demonstrate that the transition state **9-ts** is significantly later than **5-ts**. B3LYP and M06-2X predict an activation energy of 40 kcal/mol, and the CBS-QB3 and CCSD(T) data are higher still by about 10 kcal/mol. In the geometry of dipolar cycloaddition product **1b**, the O–S bond length increased to 1.66 Å, and the O–S–O bond angle is reduced to 92°. These geometric distortions going from reactant to product are more severe than for ozone. CBS-QB3 and CCSD(T) predict that the dipolar cycloaddition of sulfur dioxide with ethylene is endothermic by 17.1 and 13.7 kcal/mol, respectively. The endothermicity of this reaction is consistent with the late transition state. M06-2X, on the other hand, predicts that this reaction is exothermic by 12.2 kcal/mol.

The dipolar cycloadditions of ozone and sulfur dioxide with acetylene follow the same trends as ethylene. Transition state **10-ts** is significantly later than **6-ts**, and the activation energy of the dipolar cycloaddition between sulfur dioxide and acetylene is about 41–48 kcal/mol higher than between ozone and acetylene. The predicted reaction energy for the dipolar cycloaddition of O₃ to acetylene is 56–70 kcal/mol more exothermic than for SO₂.

2. Distortion and Interaction of the Dipolar Cycloaddition Reactivity between Ozone and Sulfur Dioxide. For bimolecular reactions, distortion–interaction energy analysis is a powerful tool to explain reactivity trends.^{12,48} In this analysis, the total activation energy (ΔE^\ddagger) is decomposed into the sum of distortion energy ($\Delta E_{\text{dist}}^\ddagger$) and interaction energy ($\Delta E_{\text{int}}^\ddagger$) between distorted reactants (see Figure 1).

Table 2 shows the distortion energies and interaction energies of the transition states (**5-ts**, **6-ts**, **9-ts**, and **10-ts**). The CCSD(T) calculations show that for the dipolar cycloaddition of ozone with ethylene (**5-ts**), the distortion energy is 4.0 kcal/mol and the interaction energy is -0.3 kcal/mol. For the dipolar cycloaddition between sulfur dioxide and ethylene (**9-ts**), the distortion energy is 30 kcal/mol larger and the interaction energy is repulsive by 16.5 kcal/mol. The results for the cycloaddition to acetylene follow the same trend; both the distortion and interaction energies of the dipolar cycloaddition between sulfur dioxide and acetylene are larger than between ozone and acetylene. Although

**Figure 1.** The relationship between activation, distortion, and interaction energies.**Table 2.** Distortion Energies and Interaction Energies of the Transition States **5-ts**, **6-ts**, **9-ts**, and **10-ts**

	B3LYP		CBS-QB3		M06-2X		CCSD(T)	
	E_{dist}	E_{int}	E_{dist}	E_{int}	E_{dist}	E_{int}	E_{dist}	E_{int}
5-ts	3.6	-5.1	1.4	-1.1	3.7	-5.0	4.0	-0.3
6-ts	6.6	-3.8	3.5	1.2	5.9	-3.0	6.3	1.7
9-ts	33.0	6.9	36.1	13.3	29.8	9.5	34.5	16.5
10-ts	33.2	10.0	35.2	17.4	32.1	11.5	35.7	19.1

the other theoretical methods (B3LYP, M06-2X, and CBS-QB3) give somewhat different distortion and interaction energies, the overall trends hold across each of these levels of theories.

To shed further light on the disparate behavior of O₃ and SO₂ in dipolar cycloadditions, the distortion and interaction energies have been computed along the reaction pathway at the B3LYP level of theory, using the forming C–O bond distances as a reaction coordinate (see Figure 2). As shown in Figure 2, as long as the forming C–O bonds are longer than 2.0 Å, the distortion energies for O₃ and SO₂ are essentially the same. As the C–O bond decreases below 2.0 Å, the distortion energy of the reaction of sulfur dioxide climbs higher than for ozone. In contrast to SO₂, the interaction energy for the reaction of ozone is negative throughout the reaction pathway and drops off rapidly as the C–O bond shortens. The transition state, which occurs when $d(\Delta E_{\text{dist}})/dr + d(\Delta E_{\text{int}})/dr = 0$, occurs at a C–O bond length of 2.30 Å. The interaction between sulfur dioxide and ethylene is unfavorable at long range. As the C–O distance decreases, the repulsion between sulfur dioxide and ethylene increases, until the forming C–O bond reaches 2.15 Å. The maximum of the interaction energy is about 23 kcal/mol at the B3LYP level, and this value is expected to be larger at the CCSD(T) level of theory. The transition state occurs at a C–O distance of 1.94 Å, at which point the interaction energy is still highly unfavorable. This positive interaction energy, along with the large distortion energy associated with a late transition state, leads to the prohibitive reaction barrier for the dipolar cycloaddition of SO₂ to ethylene.

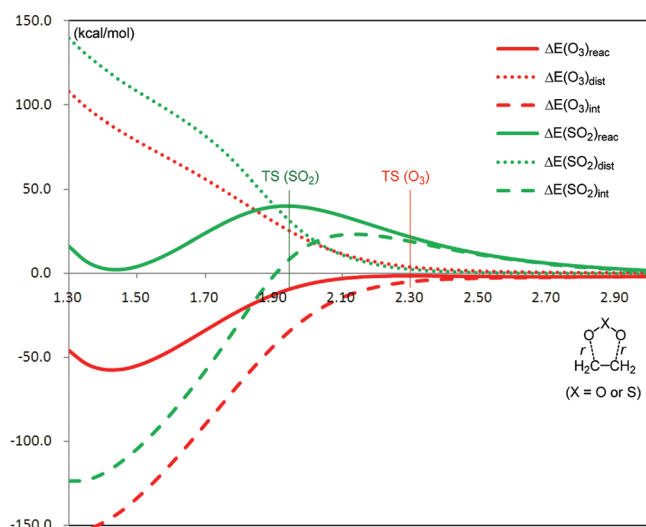


Figure 2. Distortion, interaction, and total energies along the reaction pathways of the dipolar cycloadditions between ozone or sulfur dioxide and ethylene. The solid lines are the reaction energies. The dotted lines are the distortion energies. The dashed lines are the interaction energies.

3. Molecular Orbital Analysis of Ozone and Sulfur Dioxide.

There is no simple way to explain the distortion–interaction energies and the reactivity in dipolar cycloadditions, but molecular orbital theory provides some insight.^{12a} The molecular orbitals and orbital energies of ozone and sulfur dioxide are shown in Figure 3. The gap between the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) for the ground state of ozone is 13.8 eV at the RHF/6-311++G(2d,p) level. The HOMO of ground state SO₂ is 1.0 eV lower than ozone and the LUMO is 0.8 eV higher, leading to a HOMO–LUMO gap for SO₂ that is 1.8 eV larger than that of ozone. These differences in HOMO–LUMO gaps help explain the differences in interaction energies displayed in Figure 2 and Table 2. The high energy level of the HOMO is responsible for the pseudoradical character of ozone and increases the reactivity. As shown below, both distortion and interaction energies are influenced by the HOMO–LUMO gaps of reactants.

As shown in Figure 4, a schematic diagram of the molecular orbital of ozone and sulfur dioxide explains why the LUMO of ozone is lower than that of sulfur dioxide, and the HOMO of ozone is higher than that of sulfur dioxide. For both ozone and sulfur dioxide, the delocalized π bond can be separated into two parts: the p orbitals of the two terminal oxygen atoms (Figure 4, column c) and the p orbital of the center oxygen (Figure 4, column a) or sulfur (Figure 4, column e) atom. The p orbital of the two terminal oxygen atoms can combine into symmetric or antisymmetric combination to form orbital π_{O-O} and π^*_{O-O} . In ozone, the p orbital of the center oxygen atom has the same symmetry and similar energy as the π_{O-O} orbital, and thus it forms the π_{1a} and π_{3a} orbital with the π_{O-O} orbital in ozone (Figure 4, column b), and the π^*_{O-O} orbital forms the π_{2a} orbital. For sulfur dioxide, the energy level of the 3p orbital of sulfur atom is higher than the 2p orbital of the oxygen atom. The combination of p_s orbital and π_{O-O} orbital forms the π_{1b} and π_{3b} orbitals (Figure 4, column d). The energy level of π_{3b} is higher than that of π_{3a} because the p_s orbital is higher than the p_o orbital. In Figure 3, it can be seen that the weight of the sulfur

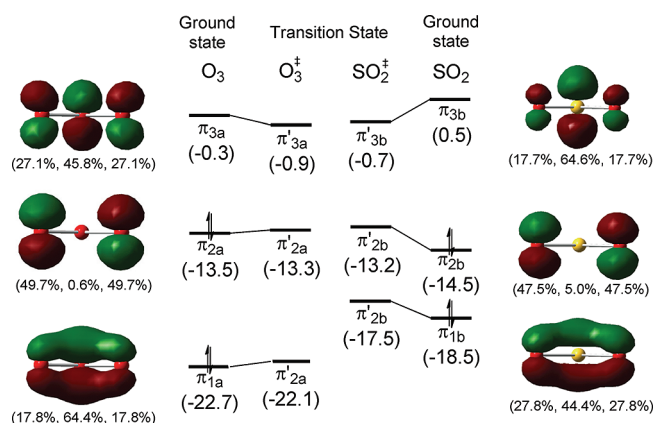


Figure 3. The molecular orbitals of ozone and sulfur dioxide at the reactant and transition state geometries. The values in parentheses are the orbital levels in eV. The percentages are the weight of the atomic orbitals in the ground state molecular orbitals.

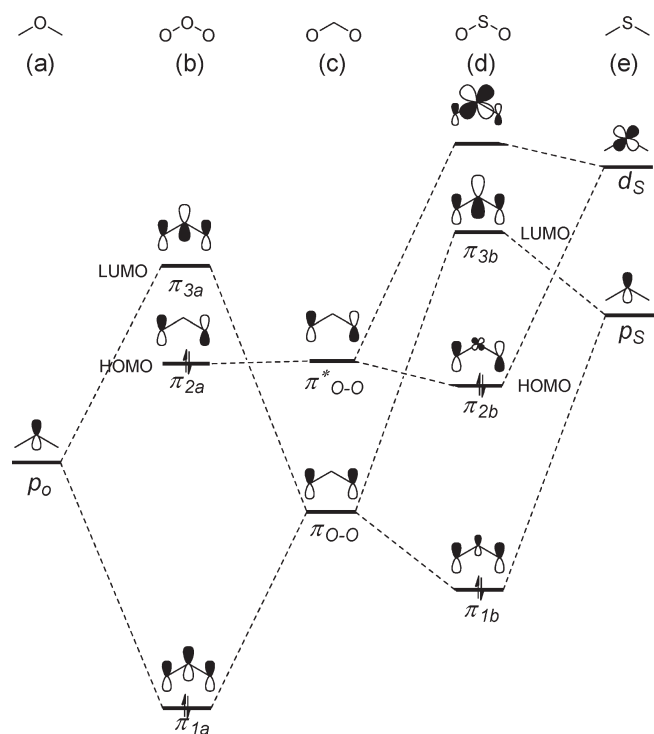


Figure 4. Schematic diagram of the molecular orbital of ozone and sulfur dioxide.

orbital (64.6%) in the π_{3b} orbital is higher than the weight of the oxygen orbital (45.8%) in π_{3a} , and the weight of the sulfur orbital (44.4%) in the π_{1b} orbital is lower than the weight of the oxygen orbital (64.4%) in π_{1a} . Atomic sulfur has an unoccupied d orbital, although the energy level of it is high.⁴⁹ The symmetry of one d orbital of sulfur atom is the same as the π^*_{O-O} orbital, and they form the π_{2b} orbital. Consequently, the energy level of the π_{2b} orbital is lower than that of π_{2a} orbital. In molecular orbital π_{2b} , the weight of the d orbital of sulfur is 5.0%.

4. The Singlet–Triplet Gap of Ozone and Sulfur Dioxide.

Previous theoretical investigations of 1,3-dipolar cycloadditions showed that the activation energy is determined by the avoided crossing of the singlet ground state of the 1,3-dipole and

Table 3. Optimized (ΔE_o) and Vertical (ΔE_v) Singlet–Triplet Gap of O₃ and SO₂^a

	B3LYP	CBS-QB3	M06-2X	CCSD(T)
$\Delta E_o(\text{O}_3)$	21.6	30.9	24.2	34.3
$\Delta E_v(\text{O}_3)$	28.6	41.5	32.4	40.3
$\Delta E_o(\text{SO}_2)$	61.0	70.6	66.3	73.2
$\Delta E_v(\text{SO}_2)$	67.6	76.6	73.4	79.0

^aThe values of the O–O and O–S bond lengths in the ground state and excited state are also given in Å.

Table 4. Hydrogenation Energies of Ozone and Sulfur Dioxide

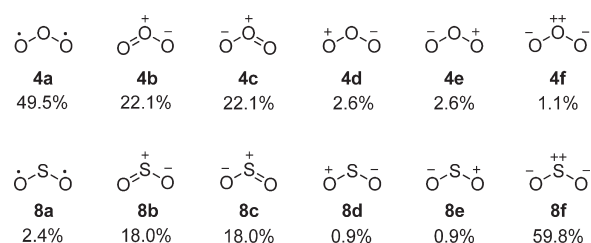
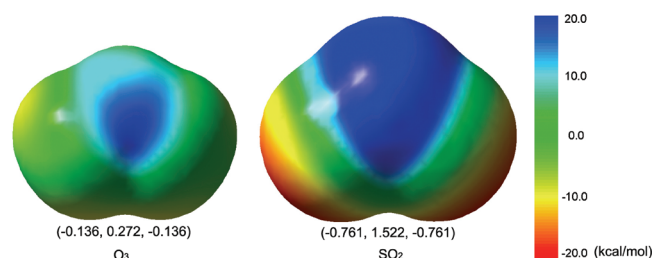
	B3LYP	CBS-QB3	M06-2X	CCSD(T)
$\Delta E_1(\text{O}_3)$	–61.5	–56.3	–74.3	–63.6
$\Delta E_2(\text{O}_3)$	–63.4	–66.6	–66.3	–70.3
$\Delta E_1(\text{SO}_2)$	–12.7	4.6	–21.1	–4.0
$\Delta E_2(\text{SO}_2)$	–18.6	–17.2	–18.7	–20.2

dipolarophile leading to the triplet excited state of the cycloadduct and the excited state triplet of the reactants leading to the ground state singlet of the product.^{5c,5d} For the dipolar cycloaddition with simple dipolarophiles, the transition state is typically correlated with the singlet–triplet energy gap of the 1,3-dipole. Lower singlet–triplet energy gaps for the 1,3-dipole correlates with an earlier transition state, lower activation energy, and a more exothermic reaction energy.⁵⁰

Optimized (ΔE_o) and vertical (ΔE_v) singlet–triplet gaps for O₃ and SO₂ are given in Table 3. Although the different theoretical methods give different values, the relative singlet–triplet energy gap of ozone and sulfur dioxide is almost the same across all four levels of theory considered. ΔE_o for ozone is 34.3 kcal/mol at the CCSD(T) level of theory. ΔE_o for sulfur dioxide is 38.9 kcal/mol higher, explaining the earlier transition state, lower activation energy, and greater exothermicity for cycloadditions of ozone versus SO₂.

5. The Hydrogenation Energy of Ozone and Sulfur Dioxide. To investigate the strength of the π -bonds in ozone and SO₂, hydrogenation energies were computed. As shown in Table 4, the first hydrogenation (ΔE_1) yields H–O–X–O–H (X = O or S), which no longer has a delocalized π bond. ΔE_1 for ozone is about 50 kcal/mol more exothermic than for sulfur dioxide. The hydrogenation of H–O–X–O–H (ΔE_2) breaks one of the O–X (X = O or S) bonds. ΔE_2 for ozone is again about 50 kcal/mol more exothermic than for sulfur dioxide. The hydrogenation energies show that both the delocalized π bond and O–X σ bond of sulfur dioxide is significantly stronger than those of ozone. This enhanced stability of sulfur dioxide contributes to its reduced reactivity compared to ozone.

6. The Diradical Character of Ozone and Sulfur Dioxide. In the VB theory study of Hiberty, the reactivity of 1,3-dipoles in dipolar cycloaddition was qualitatively related to the diradical character of the 1,3-dipoles.¹³ The six VB structures of ozone and sulfur dioxide are shown in Figure 5. For ozone, the dominant VB

**Figure 5.** Weight of VB structures for ozone and sulfur dioxide.**Figure 6.** Electrostatic potential of ozone and sulfur dioxide. The natural population analysis (NPA) charges are given in parentheses.**Table 5.** Reaction Energies (kcal/mol) for 3 + 2 (1), 2 + 2 (2), and 2 + 1 (3) Cycloadditions of Ozone (a) and Sulfur Dioxide (b) to Ethylene

	B3LYP	CBS-QB3	M06-2X	CCSD(T)
1a	–57.7	–55.1	–74.9	–57.3
2a	–17.3	–19.1	–19.4	–
3a	67.3	72.5	52.1	54.8
1b	2.0	17.1	–12.2	13.7
2b	5.0	10.6	–4.2	9.1
3b	30.1	21.3	19.4	21.0

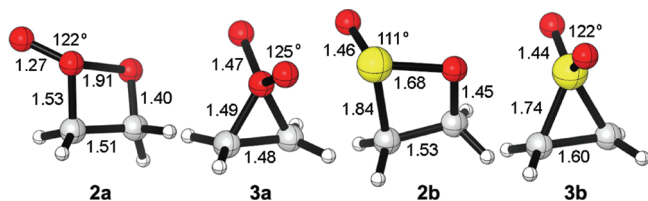
structure is the diradical structure **4a**, and ozone shows 49.5% diradical character. The weight of the sum of the traditional all-octet dipolar VB structures **4b** and **4c** is 44.2%. In other words, ozone has about half diradical and half zwitterionic character. For sulfur dioxide, the weight of the diradical VB structure **8a** is only 2.4%. The weight of the sum of the traditional all-octet dipolar VB structures **8b** and **8c** is 36.0%. The weight of dicationic sulfur VB structure **8f** is largest. Thus, although sulfur dioxide has both zwitterionic and dicationic character, the diradical character of sulfur dioxide is very weak, providing further explanation for the reduced reactivity of sulfur dioxide compared to ozone.

The characterization of SO₂ as dominated by a VB structure with a central dication is corroborated by NPA calculations (see Figure 6). In particular, the NPA charge on sulfur is 1.522, with the two terminal oxygen atoms carrying charges of –0.761. This accumulation of negative charge on the oxygen atoms, which is significantly reduced in O₃, is one source of the repulsive electrostatic interaction between SO₂ and an approaching dipolarophile.

7. The Stability of 2 + 2 and 2 + 1 Cycloaddition Products.

Finally, to assess the thermodynamic viability of other cycloadditions of ozone and SO₂ to ethylene and acetylene, the 2 + 2 and 2 + 1 cycloaddition products have been predicted (see Table 5). The 2 + 2 cycloaddition between ozone and ethylene forms complex **13**. At the CCSD(T)/cc-pV(T+d)Z level of theory, a

Scheme 3. Geometries of 3 + 2, 2 + 1, and 2 + 2 Cycloadducts of Ozone and SO₂ with Ethylene Calculated at the CCSD(T)/cc-pV(T+d)Z Level of Theory^a



^aFor **2a**, the B3LYP/6-311+G(d) geometry is given; see the text.

stationary point corresponding to this cycloadduct could not be located, and all optimizations gave a ring-opened diradical species. This structure is a minimum on the B3LYP and M06-2X potential energy surfaces, and the B3LYP geometry is presented in Scheme 3. At all levels of theory considered, these 2 + 1 and 2 + 2 cycloadducts are less favorable energetically than the corresponding 3 + 2 cycloadditions, except for the 2 + 2 cycloaddition of SO₂ to ethylene. When the center atom of the 1,3-dipole is an oxygen atom, the 2 + 2 and 2 + 1 cycloaddition species (**2a** and **3a**) are 35 and 112 kcal/mol less stable than the 3 + 2 cycloaddition product **1a**. When the center atom of the 1,3-dipole is a sulfur atom, the 2 + 2 cycloaddition species **2b** is 4.6 kcal/mol more stable than 3 + 2 cycloaddition product **1b**, and the 2 + 1 cycloaddition species **3b** is only 7.3 kcal/mol less stable than **1b**. Therefore, on thermodynamic grounds, adduct **2b** is the expected intermediate in cycloadditions of sulfur dioxide and ethylene; indeed, such species have been established as intermediates in reactions of SO₂ with dienes by Vogel and co-workers.²⁷

CONCLUSIONS

Ozone and sulfur dioxide show very different reactivities in 1,3-dipolar cycloadditions. In particular, while O₃ undergoes facile addition to ethylene and acetylene, the barriers for the corresponding reactions of SO₂ are 40 kcal/mol larger. Moreover, the highly exothermic nature of the cycloaddition of ozone to C–C multiple bonds is not mimicked by SO₂, for which accurate CCSD(T) calculations predict reactions with ethylene and acetylene that are uphill by 14 and 4 kcal/mol, respectively. These stark differences in reactivity can be understood from a number of perspectives. In particular, we have shown that unfavorable interaction energies between SO₂ and an approaching dipolarophile, which emanate in part from the large charge separation in SO₂, paired with large distortion energies in the cycloaddition transition state lead to the large predicted reaction barriers. Alternatively, these differences in reactivity can be understood in terms of the enhanced stability of both the π - and σ -bonds of SO₂ compared to O₃, the reduced diradical character of SO₂ compared to O₃, or the increased HOMO–LUMO gap in SO₂ compared to O₃. At the CBS-QB3 and CCSD(T)/cc-pV(T+d)Z levels of theory, the 2 + 2 cycloadduct of SO₂ with ethylene (**3b**) is thermodynamically favored over the 1,3-dipolar cycloaddition product and is the expected intermediate in cycloadditions of SO₂ with ethylene.

ASSOCIATED CONTENT

Supporting Information. Cartesian coordinates and energies of all reported structures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: houk@chem.ucla.edu.

ACKNOWLEDGMENT

We are grateful to the National Science Foundation and ACS PRF 50645-DNI6 for financial support of this research. Computations were performed on the National Science Foundation Terascale Computing System at the National Center for Supercomputing Applications (NCSA), on the California Nano Systems Institute clusters and UCLA Hoffman2 cluster, and using resources provided by the Texas A&M Supercomputing Facility.

REFERENCES

- (1) Lan, Y.; Zou, L.; Cao, Y.; Houk, K. N. Computational Methods to Calculate Accurate Activation and Reaction Energies of 1,3-Dipolar Cycloadditions of 24 1,3-Dipoles. Manuscript in preparation.
- (2) *Synthetic Applications of 1,3-Dipolar Cycloaddition Chemistry Toward Heterocycles and Natural Products*; Padwa, A., Pearson, W. H., Eds.; Wiley: New York, 2002.
- (3) (a) Collman, J. P.; Devaraj, N. K.; Chidsey, C. E. D. *Langmuir* **2004**, *20*, 1051. (b) Speers, A. E.; Adam, G. C.; Cravatt, B. F. *J. Am. Chem. Soc.* **2003**, *125*, 4686.
- (4) Krasinski, A.; Radic, Z.; Manetsch, R.; Raushel, J.; Taylor, P.; Sharpless, K. B.; Kolb, H. C. *J. Am. Chem. Soc.* **2005**, *127*, 6686.
- (5) Seo, T. S.; Bai, X.; Ruparel, H.; Li, Z.; Turro, N. J.; Ju, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5488.
- (6) (a) Ess, D. H.; Houk, K. N. *J. Phys. Chem. A* **2005**, *109*, 9542. (b) Karadakov, P. B.; Cooper, D. L.; Gerratt, J. *Theor. Chem. Acc.* **1998**, *100*, 222. (c) Nguyen, M. T.; Chandra, A. K.; Sakai, S.; Morokuma, K. *J. Org. Chem.* **1999**, *64*, 65. (d) Su, M.-D.; Liao, H.-Y.; Chung, W.-S.; Chu, S.-Y. *J. Org. Chem.* **1999**, *64*, 6710. (e) Domingo, L. R.; Sáez, J. A. *J. Org. Chem.* **2011**, *76*, 373.
- (7) (a) Xu, L.; Doubleday, C. E.; Houk, K. N. *Angew. Chem., Int. Ed.* **2009**, *48*, 2746. (b) Xu, L.; Doubleday, C. E.; Houk, K. N. *J. Am. Chem. Soc.* **2010**, *132*, 3029.
- (8) (a) Huisgen, R. *1,3-Dipolar Cycloaddition Chemistry*; Padwa, A., Ed.; Wiley: New York, 1984; Vol. 1, pp 1–176. (b) Huisgen, R. *J. Org. Chem.* **1968**, *33*, 2291. (c) McDouall, J. J. W.; Robb, M. A.; Niazi, U.; Bernardi, F.; Schlegel, H. B. *J. Am. Chem. Soc.* **1987**, *109*, 4642.
- (9) Engels, B.; Christl, M. *Angew. Chem., Int. Ed.* **2009**, *48*, 7968.
- (10) (a) Salem, L. *J. Am. Chem. Soc.* **1968**, *90*, 543. (b) Salem, L. *J. Am. Chem. Soc.* **1968**, *90*, 553. (c) Fukui, K. *Acc. Chem. Res.* **1971**, *4*, 57. (d) Houk, K. N. *Acc. Chem. Res.* **1975**, *8*, 361. (e) Houk, K. N. In *Pericyclic Reactions*; Marchand, A. P., Lehr, R. E., Eds.; Academic Press: New York, 1977; Vol. 2, p 181. (f) Fukui, K. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 801.
- (11) (a) Ess, D. H.; Jones, G. O.; Houk, K. H. *Adv. Synth. Catal.* **2006**, *348*, 2337. (b) Pross, A.; Shaik, S. S. *Acc. Chem. Res.* **1983**, *16*, 363.
- (12) (a) Ess, D. H.; Houk, K. N. *J. Am. Chem. Soc.* **2008**, *130*, 10187. (b) Ess, D. H.; Houk, K. N. *J. Am. Chem. Soc.* **2007**, *129*, 10646.
- (13) Braid, B.; Walter, C.; Engels, B.; Hiberty, P. C. *J. Am. Chem. Soc.* **2010**, *132*, 7631.
- (14) (a) Huisgen, R. *Angew. Chem., Int. Ed. Engl.* **1963**, *2*, 565. (b) Huisgen, R. *Angew. Chem., Int. Ed. Engl.* **1963**, *2*, 633. (c) Huisgen, R. *Angew. Chem., Int. Ed. Engl.* **1968**, *7*, 321.
- (15) Ornum, S. G. V.; Champeau, R. M.; Pariza, R. *Chem. Rev.* **2006**, *106*, 2990.
- (16) (a) Bailey, P. S.; Chang, Y.-G.; Kwie, W. W. I. *J. Org. Chem.* **1961**, *27*, 1198. (b) DeMore, W. B.; Lin, C.-L. *J. Org. Chem.* **1973**, *38*, 985. (c) DeMore, W. B. *Int. J. Chem. Kinet.* **1969**, *1*, 209. (d) Miller, D. J.; Nemo, T. E.; Hull, L. A. *J. Org. Chem.* **1975**, *40*, 2675. (e) Jackson, S.; Hull, L. A. *J. Org. Chem.* **1976**, *41*, 3340. (f) Griesbaum, K.; Dong, Y.

- J. Prakt. Chem.* **1997**, 339, 575. (g) Horie, O.; Moortgat, G. K. *Acc. Chem. Res.* **1998**, 31, 387.
- (17) (a) Walch, S. P.; Goddard, W. A., III *J. Am. Chem. Soc.* **1975**, 97, 5319. (b) Kahn, S. D.; Hehre, W. J.; Pople, J. A. *J. Am. Chem. Soc.* **1987**, 109, 1871. (c) Hiberty, P. C.; Leforestier, C. *J. Am. Chem. Soc.* **1978**, 100, 2012.
- (18) (a) Harding, L. B.; Goddard, W. A., III *J. Am. Chem. Soc.* **1978**, 100, 7180. (b) Hiberty, P. C. *J. Am. Chem. Soc.* **1976**, 98, 6088. (c) Yamaguchi, K.; Takahara, Y.; Fueno, T.; Houk, K. N. *Theor. Chim. Acta* **1988**, 73, 337. (d) Ruoff, P.; Almlöf, J.; Saebo, S. *Chem. Phys. Lett.* **1980**, 72, 489.
- (19) (a) Cremer, D.; Crehuet, R.; Anglada, J. *J. Am. Chem. Soc.* **2001**, 123, 6127. (b) Cremer, D.; Kraka, E.; Crehuet, R.; Anglada, J. M.; Gräfenstein, J. *Chem. Phys. Lett.* **2001**, 347, 268. (c) Gillies, J. Z.; Gillies, C. W.; Lovas, F. J.; Matsumura, K.; Suenram, R. D.; Kraka, E.; Cremer, D. *J. Am. Chem. Soc.* **1991**, 113, 6408. (d) Chan, W.-T.; Weng, C.; Goddard, J. D. *J. Phys. Chem. A* **2007**, 111, 4792. (e) Anglada, J. M.; Crehuet, R.; Bofill, J. M. *Chem.—Eur. J.* **1999**, 5, 1809. (f) Li, L.-C.; Deng, P.; Xu, M.-H.; Wong, N.-B. *Int. J. Quantum Chem.* **2004**, 98, 309. (g) Chan, W. T.; Hamilton, I. P. *J. Chem. Phys.* **2003**, 118, 1688. (h) Ljubic, I.; Sabljic, A. *J. Phys. Chem. A* **2002**, 106, 4745. (i) Gillies, J. Z.; Gillies, C. W.; Lovas, F. J.; Matsumura, K.; Suenram, R. D.; Kraka, E.; Cremer, D. *J. Am. Chem. Soc.* **1991**, 113, 2412. (j) Olzmann, M.; Kraka, E.; Cremer, D.; Gutbrod, R.; Andersson, S. *J. Phys. Chem. A* **1997**, 101, 9421.
- (20) Wheeler, S. E.; Ess, D. H.; Houk, K. N. *J. Phys. Chem. A* **2008**, 112, 1798.
- (21) Purser, G. H. *J. Chem. Educ.* **1999**, 76, 1013.
- (22) (a) Tanaka, T.; Morino, Y. *J. Mol. Spectrosc.* **1970**, 33, 538. (b) Elliott, R.; Compton, R.; Levis, R.; Matsika, S. *J. Phys. Chem. A* **2005**, 109, 11304.
- (23) (a) Suárez, D.; González, J.; Sordo, T. L.; Sordo, J. A. *J. Org. Chem.* **1994**, 59, 8058. (b) Manoharan, M.; Venuvanalingam, P. *J. Phys. Org. Chem.* **1997**, 10, 769.
- (24) Pliego, J. R.; Resende, S. M.; Humeres, E. *Chem. Phys.* **2005**, 314, 127.
- (25) Liebman, J. F.; Petersen, K. N.; Skancke, P. N. *Acta Chem. Scand.* **1999**, 53, 1003.
- (26) (a) Humeres, E.; Peruch, M. G. B.; Moreira, R. F. P. M.; Schreiner, W. *Int. J. Mol. Sci.* **2005**, 6, 130. (b) Humeres, E.; Castro, K. M.; Moreira, R. F. P. M.; Peruch, M. G. B.; Schreiner, W. H.; Aliev, A. E.; Canle, M.; Santaballa, J. A.; Fernández J. *Phys. Chem. C* **2008**, 112, 581.
- (27) Varela-Álvarez, A.; Marković, D.; Vogel, P.; Sordo, J. A. *J. Am. Chem. Soc.* **2009**, 131, 9547.
- (28) Durst, T.; Gimbarzevsky, B. P. *J. Chem. Soc. Chem. Comm.* **1975**, 724.
- (29) Johnson, P.; Taylor, R. J. K. *Tetrahedron Lett.* **1997**, 38, 5873.
- (30) (a) Ando, W.; Sonobe, H.; Akasaka, T. *Tetrahedron Lett.* **1986**, 37, 4473. (b) Harel, T.; Amir, E.; Rozen, S. *Org. Lett.* **2006**, 8, 1213.
- (31) Jensen, F.; Foote, C. S. *J. Am. Chem. Soc.* **1987**, 109, 1478.
- (32) (a) Becke, A. D. *Phys. Rev. A* **1988**, 38, 3098. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785. (c) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, 98, 11623. (d) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, 110, 6158.
- (33) (a) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, 104, 4811. (b) Zhao, Y.; Tishchenko, O.; Truhlar, D. G. *J. Phys. Chem. B* **2005**, 109, 19046. (c) Tsuzuki, S.; Luthi, H. P. *J. Chem. Phys.* **2001**, 114, 3949. (d) Duncan, J. A.; Spong, M. C. *J. Phys. Org. Chem.* **2001**, 114, 3949. (d) Duncan, J. A.; Spong, M. C. *J. Phys. Org. Chem.* **2005**, 18, 462. (e) Wodrich, M. D.; Corminboeuf, C.; Schleyer, P. R. *Org. Lett.* **2006**, 8, 3631.
- (34) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, 120, 215.
- (35) (a) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. C* **2008**, 112, 6860. (b) Rokob, T. A.; Hamza, A.; Pápai, I. *Org. Lett.* **2007**, 9, 4279. (c) Wheeler, S. E.; McNeil, A. J.; Müller, P.; Swager, T. M.; Houk, K. N. *J. Am. Chem. Soc.* **2010**, 132, 3304. (c) Wheeler, S. E.; Houk, K. N. *J. Chem. Theory Comput.* **2010**, 6, 395.
- (36) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, 110, 2822.
- (37) (a) Nyden, M. R.; Petersson, G. A. *J. Chem. Phys.* **1981**, 75, 1843. (b) Al-Laham, M. A.; Petersson, G. A. *J. Chem. Phys.* **1991**, 94, 6081. (c) Petersson, G. A.; Tensfeldt, T. G.; Montgomery, J. A. *J. Chem. Phys.* **1991**, 94, 6091. (d) Petersson, G. A.; Malick, D. K.; Wilson, W. G.; Ochterski, J. W.; Montgomery, J. A.; Frisch, M. J. *J. Chem. Phys.* **1998**, 109, 10570. (e) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, 112, 6532.
- (38) (a) Ess, D. H.; Houk, K. N. *J. Phys. Chem. A* **2005**, 109, 9542. (b) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A., Jr. *J. Chem. Phys.* **1996**, 104, 2598.
- (39) (a) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, 157, 479. (b) Bartlett, R. J.; Watts, J. D.; Kucharski, S. A.; Noga J. *Chem. Phys. Lett.* **1990**, 165, 513. (c) Bartlett, R. J.; Watts, J. D.; Kucharski, S. A.; Noga J. *Chem. Phys. Lett.* **1990**, 167, 609. (d) Gauss, J.; Lauderdale, W. J.; Stanton, J. F.; Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1991**, 182, 207. (e) Watts, J. D.; Gauss, J.; Bartlett, R. J. *Chem. Phys. Lett.* **1992**, 200, 1. (f) Watts, J. D.; Gauss, J.; Bartlett, R. J. *J. Chem. Phys.* **1993**, 98, 8718.
- (40) (a) Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, 114, 9244. (b) Dunning, T. H. *J. Chem. Phys.* **1989**, 90, 1007. (c) Wheeler, S. E.; Schaefer, H. F. *J. Phys. Chem. A* **2009**, 113, 6779.
- (41) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
- (42) (a) *CFOUR, Coupled-Cluster Techniques for Computational Chemistry*, is a quantum-chemical program package by Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G., with contributions from Auer, A. A.; Bartlett, R. J.; Benedikt, U.; Berger, C.; Bernholdt, D. E.; Bomble, Y. J.; Cheng, L.; Christiansen, O.; Heckert, M.; Heun, O.; Huber, C.; Jagau, T.-C.; Jonsson, D.; Jusélius, J.; Klein, K.; Lauderdale, W. J.; Matthews, D. A.; Metzroth, T.; O'Neill, D. P.; Price, D. R.; Prochnow, E.; Ruud, K.; Schiffrmann, F.; Schwabach, W.; Stopkowitz, S.; Tajti, A.; Vázquez, J.; Wang, F.; Watts, J. D., and the integral packages MOLECULE (Almlöf, J.; Taylor, P. R.), PROPS (Taylor, P. R.), and ABACUS (Helgaker, T.; Jensen, H. J. Aa.; Jørgensen, P.; Olsen, J.) and ECP routines (Mitin, A. V.; van Wüllen, C.). For the current version, see <http://www.cfour.de>. (b) Harding, M. E.; Metzroth, T.; Gauss, J.; Auer, A. A. *J. Chem. Theor. Comput.* **2008**, 4, 64.
- (43) (a) Politzer, P.; Abu-Awwad, F. *Theor. Chem. Acc.* **1998**, 99, 83. (b) Kar, T.; Ángyán, J. G.; Sannigrahi, A. B. *J. Phys. Chem. A* **2000**, 104, 9953. (c) Zhang, G.; Musgrave, C. B. *J. Phys. Chem. A* **2007**, 111, 1554.
- (44) (a) Hoffmann, R.; Shaik, S.; Hiberty, P. C. *Acc. Chem. Res.* **2003**, 36, 750. (b) Shaik, S.; Hiberty, P. C. *A Chemist's Guide to Valence Bond Theory*; Wiley-Interscience: New York, 2007, p 1.
- (45) Lan, Y.; Houk, K. N. *J. Am. Chem. Soc.* **2010**, 132, 17921.
- (46) (a) Hiberty, P. C.; Flament, J. P.; Noizet, E. *Chem. Phys. Lett.* **1992**, 189, 259. (b) Hiberty, P. C.; Humbel, S.; Byrman, C. P.; van Lenthe, J. H. *J. Chem. Phys.* **1994**, 101, 5969. (c) Hiberty, P. C.; Shaik, S. *Theor. Chem. Acc.* **2002**, 108, 255.
- (47) (a) Song, L.; Wu, W.; Mo, Y.; Zhang, Q. *XMVB: An Ab Initio Nonorthogonal Valence Bond Program*; Xiamen University: Xiamen, People's Republic of China, 1999. (b) Song, L.; Mo, Y.; Zhang, Q.; Wu, W. *J. Comput. Chem.* **2005**, 26, 514. (c) Song, L.; Song, J.; Mo, Y.; Wu, W. *J. Comput. Chem.* **2009**, 30, 399.

- (48) (a) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325. (b) Nagase, S.; Morokuma, K. *J. Am. Chem. Soc.* **1978**, *100*, 1666. (c) Houk, K. N.; Gandour, R. W.; Strozier, R. W.; Rondan, N. G.; Paquette, L. A. *J. Am. Chem. Soc.* **1979**, *101*, 6797. (d) Froese, R. D. J.; Coxon, J. M.; West, S. C.; Morokuma, K. *J. Org. Chem.* **1997**, *62*, 6991. (e) Koga, N.; Ozawa, T.; Morokuma, K. *J. Phys. Org. Chem.* **1990**, *3*, 519. (f) Coxen, J. M.; Grice, S. T.; Maclagan, R. G. A. R.; McDonald, D. Q. *J. Org. Chem.* **1990**, *55*, 3804. (g) Coxon, J. M.; Rroese, R. D. J.; Ganguly, B.; Marchand, A. P.; Morokuma, K. *J. Synnlett.* **1999**, *11*, 1681. (h) Avalos, M.; Babiano, R.; Bravo, J. L.; Cintas, P.; Jiménez, J.; Palacios, J.; Silva, M. A. *J. Org. Chem.* **2000**, *65*, 6613. (i) Geetha, K.; Dinadayalane, T. C.; Sastry, G. N. *J. Phys. Org. Chem.* **2003**, *16*, 298. (j) Manoharan, M.; Venuvanalingam, P. *J. Chem. Soc., Perkin Trans. 2* **1997**, 1799. (k) Kavitha, K.; Manoharan, M.; Venuvanalingam, P. *J. Org. Chem.* **2005**, *70*, 2528. (l) Kavitha, K.; Venuvanalingam, P. *Int. J. Quantum Chem.* **2005**, *104*, 67. (m) Blowers, P.; Ford, L.; Masel, R. *J. Phys. Chem. A* **1998**, *102*, 9267.
- (49) (a) Stromberg, A.; Wahlgren, U.; Pettersson, L.; Siegbahn, E. M. *Chem. Phys.* **1984**, *89*, 323. (b) Keeton, M.; Santry, D. P. *Chem. Phys. Lett.* **1970**, *7*, 105.
- (50) (a) Liao, H.-Y.; Su, M.-D.; Chung, W.-S.; Chu, S.-Y. *Int. J. Quantum Chem.* **2001**, *83*, 318. (b) Sakai, S.; Nguyen, M. T. *J. Phys. Chem. A* **2004**, *108*, 9169. (c) Sakai, S.; Nguyen, M. T. *J. Phys. Chem. A* **2000**, *104*, 922.

The Remarkable Nb₂(CO)₁₂ with Seven-Coordinate Niobium: Decarbonylation to Nb₂(CO)₁₁ and Nb₂(CO)₁₀

Lihong Tang,[†] Qiong Luo,^{*,†} Qian-shu Li,^{†,‡} Yaoming Xie,[§] R. Bruce King,^{*,†,§} and Henry F. Schaefer, III[§][†]The School of Sciences, Beijing Institute of Technology, Beijing 100081, P. R. China[‡]School of Chemistry and Environment, South China Normal University, Guangzhou, Guangdong 510631, P. R. China[§]Department of Chemistry and Center for Computational Chemistry, University of Georgia, Athens, Georgia 30602, United States Supporting Information

ABSTRACT: The dissociation of Nb₂(CO)₁₂ into two Nb(CO)₆ units is predicted to require ~13 kcal/mol so that Nb₂(CO)₁₂ rather than Nb(CO)₆ is the anticipated initial oxidation product of the known Nb(CO)₆[−] anion. This differs from the corresponding vanadium carbonyl chemistry where V(CO)₆ rather than V₂(CO)₁₂ is found experimentally to be the oxidation product of V(CO)₆[−]. The lowest energy Nb₂(CO)₁₂ structure consists of two Nb(CO)₆ fragments joined by a Nb–Nb bond of ~3.4 Å length so that each niobium atom is heptacoordinate, counting the metal–metal bond. These niobium coordination polyhedra can be approximated by capped octahedra. Among unsaturated binuclear niobium carbonyls the lowest energy Nb₂(CO)₁₁ structure has a formal four-electron donor bridging η²-μ-CO group and a formal Nb–Nb single bond rather than only two-electron donor carbonyl groups and a formal Nb=Nb double bond. The Nb₂(CO)₁₁ structures with formal Nb=Nb double bonds and exclusively two-electron donor carbonyl groups lie more than 13 kcal/mol above this low-energy Nb₂(CO)₁₀(η²-μ-CO) structure. However, Nb₂(CO)₁₁ is predicted to be thermodynamically disfavored, owing to disproportionation into Nb₂(CO)₁₂ + Nb₂(CO)₁₀, a slightly exothermic process by ~4 kcal/mol. The Nb₂(CO)₁₀ structures with formal Nb=Nb triple bonds and all two-electron donor carbonyl groups appear to be particularly favorable, as suggested by high CO dissociation energies and viability toward disproportionation. Such structures are isolobal with Cp₂Mo₂(CO)₄, which was the first stable metal carbonyl to be discovered with a short metal–metal distance, corresponding to a formal triple bond. Considerably higher energy Nb₂(CO)₁₀ structures (by more than 20 kcal/mol) have two four-electron donor bridging carbonyl groups and long niobium–niobium distances. Such structures can be considered to consist of a bidentate Nb(CO)₆ “ligand” coordinating to a Nb(CO)₄ unit through the two η²-μ-CO groups.

1. INTRODUCTION

The 18-electron rule has played an important role in the structure and bonding in simple metal carbonyls.^{1,2} Thus, the well-known simple binary carbonyls of the first row transition metals from chromium to nickel, inclusive, namely, Cr(CO)₆, Mn₂(CO)₁₀, Fe(CO)₅, Co₂(CO)₈, and Ni(CO)₄, all obey the 18-electron rule. However, difficulties arise when applying the 18-electron rule to neutral binary carbonyls of the early transition metals of groups 4 and 5 because of the large numbers of carbonyl groups required to give the central metal the favored 18-electron configuration. Thus, the only binary carbonyl of vanadium which has been isolated³ is the 17-electron complex V(CO)₆. The hypothetical binuclear complex V₂(CO)₁₂ with a V–V bond has seven-coordinate vanadium atoms and is unstable with respect to dissociation to mononuclear V(CO)₆, as suggested by both theory⁴ and experiment.⁵ No neutral binary carbonyls of titanium have been isolated, although the anion Ti(CO)₆^{2−} with an 18-electron configuration for the central titanium has been isolated as a potassium-2,2,2-cryptate salt.⁶ However, recent theoretical studies suggest that the titanium carbonyl species observed spectroscopically in low temperature matrices^{7,8} is the heptacoordinate Ti(CO)₇ in which the titanium atom has the favored 18-electron configuration. However, Ti(CO)₇ appears to be unstable under ambient conditions.

The existence of binary carbonyls of the early second and third row transition metals might be expected to be more favorable than that of the corresponding first row transition metals. Thus, the larger sizes of the second and third row metals might be expected to favor the larger coordination numbers required to accommodate enough carbonyl groups to give the central metal atoms the favored 18-electron configurations. However, no neutral binary carbonyl derivatives of zirconium, niobium, hafnium, or tantalum are known despite the fact that they are potentially accessible by mild oxidation of the known anions M(CO)₆^{2−} (M = Zr,⁹ Hf¹⁰) and M(CO)₆[−] (M = Nb,¹¹ Ta¹²), all of which have the favored 18-electron configuration of the central metal atom. Such a synthetic method is analogous to the known method³ for synthesizing the neutral V(CO)₆ from the anion V(CO)₆[−]. In this case, the hydride HV(CO)₆ initially formed by acidification of V(CO)₆[−] with a nonoxidizing acid such as phosphoric acid spontaneously evolves hydrogen to give directly the neutral V(CO)₆. However, neither neutral Nb(CO)₆ nor its dimer Nb₂(CO)₁₂ has yet been reported as an oxidation product of the known Nb(CO)₆[−]. The experimental conditions tried so far¹³ have had halide present so that the products are the niobium

Received: May 25, 2011

Published: May 31, 2011

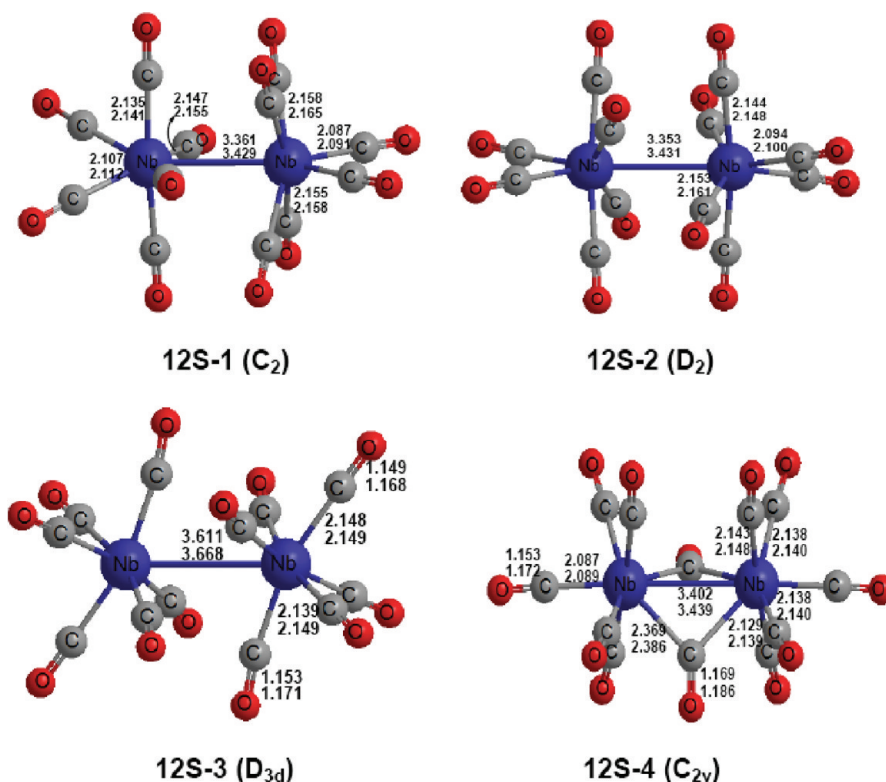


Figure 1. Optimized geometries of $\text{Nb}_2(\text{CO})_{12}$. The distances are given in Å. The upper numbers were determined by the MPW1PW91 method and the lower distances by the BP86 method.

Table 1. Total Energies (E in hartrees), Relative Energies (ΔE in kcal/mol), Numbers of Imaginary Vibrational Frequencies (Nimag), and Nb–Nb Distances (in Å) for the Lowest-Energy $\text{Nb}_2(\text{CO})_{12}$ Structures

		12S-1 (C_2)	12S-2 (D_2)	12S-3 (D_{3d})	12S-4 (C_{2v})
MPW1PW91	E	−1474.01069	−1474.01055	−1473.99678	−1473.98066
	ΔE	0.0	0.09	8.7	18.8
	Nimag	0	1 (17i)	2 (19i, 19i)	3 (200i, 17i, 11i)
	Nb–Nb	3.361	3.353	3.611	3.402
BP86	E	−1474.58661	−1474.58652	−1474.57691	−1474.56636
	ΔE	0.0	0.06	6.1	12.7
	Nimag	0	1 (19i)	2 (10i, 10i)	2 (145i, 13i)
	Nb–Nb	3.429	3.431	3.668	3.439

carbonyl halide anions $[\text{Nb}_2(\mu\text{-X})_3(\text{CO})_8]^-$ ($X = \text{Cl}, \text{Br}, \text{I}$). However, binary niobium carbonyl derivatives have been generated in low-temperature neon matrices from laser ablated niobium atoms and carbon monoxide.¹⁴

This paper reports theoretical studies on the binuclear niobium carbonyl derivatives $\text{Nb}_2(\text{CO})_n$ ($n = 12, 11, 10$) with particular focus on the following points:

- (1) Does the larger size of niobium relative to vanadium favor the binuclear species $\text{Nb}_2(\text{CO})_{12}$ with at least seven-coordinate niobium and an 18-electron configuration for the niobium over the mononuclear $\text{Nb}(\text{CO})_6$ with a 17-electron configuration for the niobium? The tendency for niobium to favor larger coordination numbers than vanadium in metal carbonyls is shown by a recent gas phase study of the $\text{M}(\text{CO})_n^+$ cations generated in a molecular beam by laser vaporization.¹⁵ The vanadium system forms exclusively the hexacoordinate $\text{V}(\text{CO})_6^+$ with a 16-electron configuration. However, the

niobium system forms both the hexacoordinate $\text{Nb}(\text{CO})_6^+$ with a 16-electron configuration and the heptacoordinate $\text{Nb}(\text{CO})_7^+$ with an 18-electron configuration.

- (2) Are there viable structures for $\text{Nb}_2(\text{CO})_{10}$ where the possibility of a formal $\text{Nb} \equiv \text{Nb}$ triple bond can lead to a structure with the favored 18-electron configuration with a possible metal coordination number of six? There is some evidence,⁴ for the formation of an analogous vanadium derivative $\text{V}_2(\text{CO})_{10}$ in the photolysis of gas-phase $\text{V}(\text{CO})_6$.¹⁶

2. THEORETICAL METHODS

Density functional theory (DFT) appears to be a powerful and effective computational tool to study organometallic compounds.^{17–32} In this connection, two different density functional theory (DFT) methods were used in the present study. The first

Table 2. Infrared $\nu(\text{CO})$ Vibrational Frequencies (cm^{-1}) Predicted by the BP86 Method for the $\text{Nb}_2(\text{CO})_{12}$ Structures^a

	BP86
12S-1 (C_2)	1937 (307), 1944 (374), 1945 (300), 1959 (574), 1960 (714), 1963 (419) 1973 (185), 1983 (1783), 1984 (38), 1990 (2433), 2019 (1730), 2067 (4)
12S-2 (D_2)	1945 (13), 1949 (1), 1955 (842), 1960 (0), 1961 (891), 1969 (416) 1970 (0), 1972 (0), 1977 (2169) 1985 (2817), 2019 (1709), 2065 (0)
12S-3 (D_{3d})	1937 (0), 1937 (0), 1945 (703), 1961 (0), 1968 (1366), 1968 (1367) 1973 (1867), 1973 (1867), 1977 (0), 1977 (0), 2019 (3122), 2069 (0)
12S-4 (C_{2v})	1829 (402), 1836 (118) , 1946 (246), 1959 (1208), 1962 (19), 1968 (0) 1973 (13), 1977 (2761), 1981 (2715), 1983 (171), 2019 (2226), 2064 (15)

^a Infrared intensities are in parentheses (in km/mol). The bridging CO frequencies are in **bold**.

DFT method is the BP86 method, which uses Becke's 1988 exchange functional (B) with Perdew's 1986 gradient corrected correlation functional method (P86).^{33,34} The second DFT method is a newer generation functional, MPW1PW91, which is a combination of the modified Perdew–Wang exchange functional with Perdew–Wang's 91 gradient-correlation functional.³⁵ This MPW1PW91 functional has been shown to be better than the first generation functionals for some heavy transition metal compounds.³⁶

The Stuttgart/Dresden double- ζ (SDD) basis set with an effective core potential (ECP)^{37,38} was used for the niobium atoms. In this basis set, the 28 core electrons for the niobium atoms are replaced by an effective core potential (ECP), and the valence basis set is contracted from (8s7p6d) primitive sets to (6s5p3d). The effective core approximation includes scalar relativistic contributions, which become significant for the heavy transition metal atoms. For the C and O atoms, the all-electron DZP basis sets are used. They are Huzinaga–Dunning contracted double- ζ sets^{39,40} plus a set of spherical harmonic d polarization functions with the orbital exponents $\alpha_d(\text{C}) = 0.75$ and $\alpha_d(\text{O}) = 0.85$. The DZP basis sets for C and O atoms may be designated as 9s5p1d/4s2p1d.

The geometries of all structures were fully optimized using the MPW1PW91 and BP86 methods, and the vibrational frequencies were determined by evaluating analytically the second derivatives of the energy with respect to the nuclear coordinates. The corresponding infrared intensities were also evaluated analytically. All of the computations were carried out with the Gaussian 03 and Gaussian 09 programs,⁴¹ exercising the fine grid option (75 radial shells, 302 angular points) for evaluating integrals numerically, while the tight designation is the default for the self-consistent field (SCF) convergence. For structures with small imaginary frequencies, the finer grid (120, 974) is used for the further confirmation.

In the search for minima, low magnitude imaginary vibrational frequencies are suspect, because of significant limitations in the numerical integration procedures used in standard DFT computations. Thus, imaginary vibrational frequencies with a magnitudes less than $50i \text{ cm}^{-1}$ are considered questionable. In less critical cases, we do not follow the eigenvectors corresponding to imaginary frequencies less than $50i \text{ cm}^{-1}$ in search of another minimum.⁴²

In the present study, the MPW1PW91 and BP86 methods agree with each other fairly well in predicting the structural characteristics of the $\text{Nb}_2(\text{CO})_n$ derivatives ($n = 12, 11, 10$). However, the BP86 method was used in previous work^{20,43} to predict $\nu(\text{CO})$ frequencies closer to the experimental values.

3. RESULTS

3.1. $\text{Nb}_2(\text{CO})_{12}$. A total of 11 structures for $\text{Nb}_2(\text{CO})_{12}$ were investigated. However, only four of these structures lie within

20 kcal/mol of the global minimum **12S-1** and are discussed in detail in this paper. All of these structures are singlet ($S = 0$) electronic states. Attempts to optimize triplet $\text{Nb}_2(\text{CO})_{12}$ structures led instead to mononuclear fragments $\text{Nb}(\text{CO})_6$.

The global minimum **12S-1** for $\text{Nb}_2(\text{CO})_{12}$ (Figure 1 and Tables 1 and 2) is a C_2 structure with 12 terminal CO groups. This structure is predicted to have all real vibrational frequencies by both the MPW1PW91 and BP86 methods. The Nb–C distances fall in the range of 2.087 to 2.158 Å (MPW1PW91) or 2.091 to 2.165 Å (BP86). The inward Nb–C distances of structure **12S-1** are all slightly longer than outward Nb–C distances, probably caused by steric hindrance between the inward Nb–C bonds. The six unique Nb–C–O angles are in the range of 170.8° to 179.4° (MPW1PW91) or 169.3° to 178.9° (BP86), which is close to linearity except for the two inward Nb–C–O bonds. All CO distances in **12S-1** are close to 1.150 Å, which is within the normal range. The Nb–Nb distance in **12S-1** is predicted to be 3.361 Å (MPW1PW91) or 3.429 Å (BP86), consistent with the Nb–Nb single bond required to give each niobium atom the favored 18-electron configuration.

The second stationary point of $\text{Nb}_2(\text{CO})_{12}$ is **12S-2**, which is also an unbridged structure, but with higher symmetry than **12S-1**, i.e., D_2 for **12S-2** versus C_2 for **12S-1** (Figure 1 and Table 1). Structure **12S-2** lies only 0.09 kcal/mol (MPW1PW91) or 0.06 kcal/mol (BP86) above **12S-1** (Figure 1 and Table 1). Furthermore, **12S-2** has a very small-magnitude imaginary vibrational frequency of $17i \text{ cm}^{-1}$ (MPW1PW91) or $19i \text{ cm}^{-1}$ (BP86). Following the corresponding normal mode leads to **12S-1** (C_2 symmetry). The Nb–C distances in **12S-2** fall in the range of 2.100–2.161 Å. For the eight inward CO groups (two equivalent sets), the Nb–Nb–C angles are 75.2° and 85.8° (MPW1PW91) or 73.5° and 86.8° (BP86). The Nb–Nb distance of **12S-2** is predicted to be 3.353 Å (MPW1PW91) or 3.431 Å (BP86). This Nb–Nb distance is close to that of **12S-1** and can also be interpreted as the formal Nb–Nb single bond required to give each niobium atom the favored 18-electron configuration.

The third energetically low-lying $\text{Nb}_2(\text{CO})_{12}$ structure **12S-3** lies 8.7 kcal/mol (MPW1PW91) or 6.1 kcal/mol (BP86) above **12S-1** (Figure 1 and Table 1). Structure **12S-3** is another unbridged $(\text{CO})_6\text{Nb}–\text{Nb}(\text{CO})_6$ structure, but with D_{3d} symmetry. Structure **12S-3** is predicted to have negligible doubly degenerate imaginary vibrational frequencies at $19i \text{ cm}^{-1}$ (MPW1PW91) or $10i \text{ cm}^{-1}$ (BP86). Following one of the corresponding normal modes leads first to the D_2 structure **12S-2** and then to the C_2 structure **12S-1**. The Nb–Nb distance in **12S-3** is 3.611 Å (MPW1PW91) or 3.668 Å (BP86), which is ~ 0.3 Å longer than that of **12S-1** and thus suggests a relatively weak single bond. For the six inward CO groups in **12S-3**, the Nb–C distances are 2.139 Å

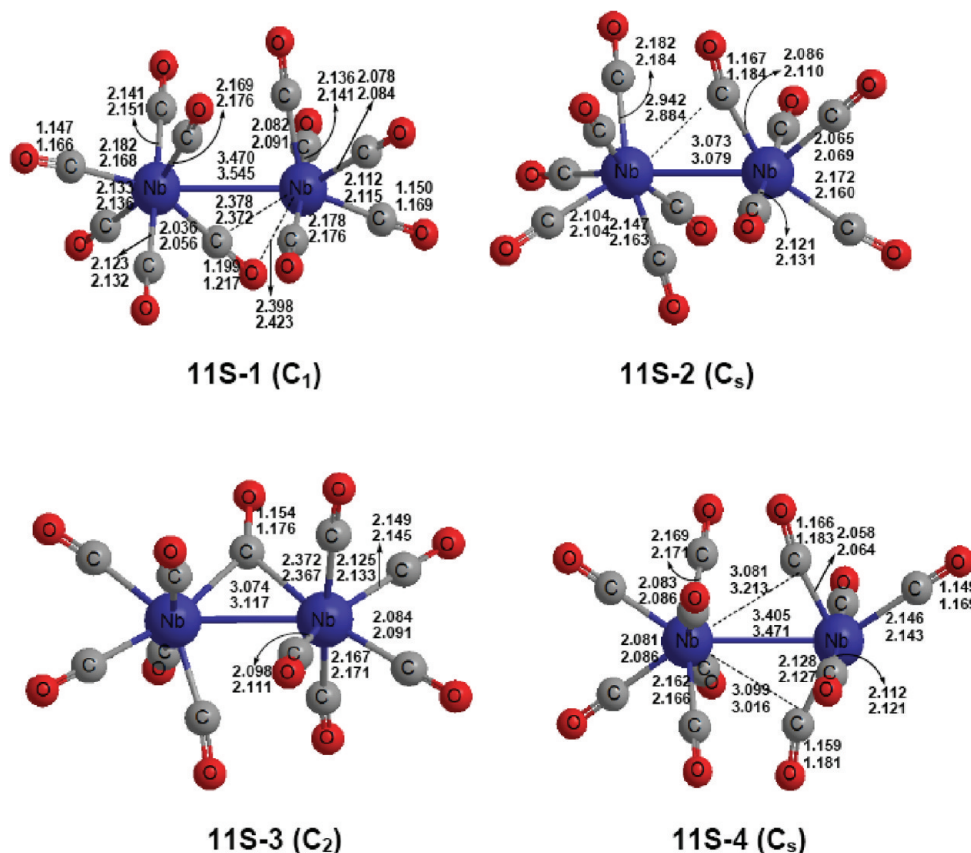


Figure 2. The four optimized singlet $\text{Nb}_2(\text{CO})_{11}$ structures.

(MPW1PW91) or 2.149 Å (BP86), and the corresponding Nb–Nb–C angles are 69.7° (MPW1PW91) or 69.9° (BP86). For the six outward CO groups, the Nb–C distances are 2.148 Å (MPW1PW91) or 2.149 Å (BP86), and the relevant Nb–Nb–C angles are 125.9° (MPW1PW91) or 126.3° (BP86). The local coordination of each niobium atom in **12S-3** (including the Nb–Nb bond) approximates a C_{3v} capped octahedron. The Nb–C–O angles for the six outward CO groups are nearly linear, i.e., 178.6° (MPW1PW91) or 178.1° (BP86), while the Nb–C–O angles for the six inward CO groups are slightly bent, i.e., 172.4° (MPW1PW91) or 171.6° (BP86), probably owing to steric interactions.

The fourth $\text{Nb}_2(\text{CO})_{12}$ structure, namely, the C_{2v} structure **12S-4** (Figure 1 and Table 1), lies 18.8 kcal/mol (MPW1PW91) or 12.7 kcal/mol (BP86) above the global minimum **12S-1**. Structure **12S-4** is a doubly bridged structure, $(\text{OC})_5\text{Nb}(\mu\text{-CO})_2\text{Nb}(\text{CO})_5$. The Nb–C distances to the two bridging CO groups are 2.369 Å (MPW1PW91) or 2.386 Å (BP86), which is significantly longer than the Nb–C distances to the terminal CO groups ranging from 2.087 to 2.143 Å (MPW1PW91) or from 2.089 to 2.148 Å (BP86). The Nb–Nb distance in **12S-4** is 3.402 Å (MPW1PW91) or 3.439 Å (BP86), which is somewhat longer than that of **12S-1** but shorter than that of **12S-3**. This is consistent with the formal Nb–Nb single bond required to give each niobium atom the favored 18-electron configuration. Structure **12S-4** has one large ($>100i\text{ cm}^{-1}$) imaginary vibrational frequency. Following the corresponding normal mode leads to **12S-1**.

Table 2 lists the infrared active $\nu(\text{CO})$ frequencies predicted by the BP86 method for the four $\text{Nb}_2(\text{CO})_{12}$ structures. Realistically, matrix isolation IR spectroscopy will probably be

the first experimental method to observe $\text{Nb}_2(\text{CO})_{12}$. Structures **12S-1**, **12S-2**, and **12S-3** exhibit only terminal $\nu(\text{CO})$ frequencies, which fall in the typical range of $1937\text{--}2069\text{ cm}^{-1}$ (BP86). For **12S-1**, the three frequencies with high IR intensities (larger or close to 2000 km/mol) are 1983 , 1990 , and 2019 cm^{-1} , while those for **12S-2** are very similar at 1977 , 1985 , and 2019 cm^{-1} . For **12S-3**, there are two frequencies, 1973 cm^{-1} (doubly degenerate) and 2019 cm^{-1} , with very large IR intensities ($>3000\text{ km/mol}$), while there is another, 1968 cm^{-1} (doubly degenerate), with large IR intensities ($>2000\text{ km/mol}$). Structure **12S-4** has two bridging CO groups, which exhibit lower $\nu(\text{CO})$ frequencies at 1829 and 1836 cm^{-1} (BP86; Table 2 in bold face). However, these frequencies do not have high IR intensities. The three frequencies in **12S-4** with high IR intensities ($>2000\text{ km/mol}$) are 1977 , 1981 , and 2019 cm^{-1} .

3.2. $\text{Nb}_2(\text{CO})_{11}$. A total of seven $\text{Nb}_2(\text{CO})_{11}$ structures (four singlets and three triplets) are found within ~ 20 kcal/mol of the global minimum. The global minimum **11S-1** (Figure 2 and Table 3) is a C_1 structure with one bridging CO group and 10 terminal CO groups. The bridging CO group is a four-electron donor $\eta^2\text{-}\mu\text{-CO}$ group, as indicated by the short Nb–O distance of 2.398 Å (MPW1PW91) or 2.423 Å (BP86) as well as the extremely low $\nu(\text{CO})$ frequency of 1701 cm^{-1} (BP86). The relatively long C–O distance of 1.199 Å (MPW1PW91) or 1.217 Å (BP86) suggests a relatively low C–O bond order for this $\eta^2\text{-}\mu\text{-CO}$ group. The short Nb–C distance to this bridging $\eta^2\text{-}\mu\text{-CO}$ group is 2.036 Å (MPW1PW91) or 2.056 Å (BP86), and the long Nb–C distance is 2.378 Å (MPW1PW91) or 2.372 Å (BP86). The Nb–Nb distance in **11S-1** is predicted to be 3.470 Å (MPW1PW91) or 3.545 Å (BP86), which is close to

Table 3. Total Energies (E in hartrees), Relative Energies (ΔE in kcal/mol), Numbers of Imaginary Vibrational Frequencies (Nimag), and Nb–Nb Distances (in Å) for the Singlet Nb₂(CO)₁₁ Structures

		11S-1 (C ₁)	11S-2 (C _s)	11S-3 (C ₂)	11S-4 (C _s)
MPW1PW91	E	–1360.68601	–1360.66366	–1360.65329	–1360.65167
	ΔE	0.0	14.0	20.5	21.5
	Nimag	0	1 (22i)	2 (384i, 90i)	0
	Nb–Nb	3.470	3.073	3.074	3.405
BP86	E	–1361.22921	–1361.20769	–1361.20125	–1361.19442
	ΔE	0.0	13.5	17.5	21.8
	Nimag	0	2 (20i, 6i)	2 (355i, 107i)	0
	Nb–Nb	3.545	3.079	3.117	3.471

that in **12S-1**. This comparison suggests a formal Nb–Nb single bond in **11S-1**. Such an analysis gives each Nb atom the favored 18-electron configuration after considering the four-electron donor η^2 - μ -CO group. The Nb₂(CO)₁₁ global minimum structure **11S-1** appears to be a very favorable structure since it lies ~13 kcal/mol below any of the other Nb₂(CO)₁₁ structures found in this work.

The second singlet Nb₂(CO)₁₁ stationary point **11S-2** is a C_s structure with one semibridging carbonyl group and 10 terminal carbonyl groups (Figure 2 and Table 3). Structure **11S-2** lies 14.0 kcal/mol (MPW1PW91) or 13.5 kcal/mol (BP86) above **11S-1** (Figure 2 and Table 3). The short Nb–C distance to the semibridging CO group in **11S-2** is 2.086 Å (MPW1PW91) or 2.110 Å (BP86), whereas the long Nb–C distance is 2.942 Å (MPW1PW91) or 2.884 Å (BP86). The ν (CO) frequency at 1877 cm^{–1} predicted for **11S-2** by the BP86 method can be assigned to this semibridging CO group. The six unique Nb–C distances to the 10 terminal carbonyls in **11S-2** range from 2.065 Å to 2.182 Å (MPW1PW91) or from 2.069 Å to 2.184 Å (BP86). The Nb=Nb distance of 3.073 Å (MPW1PW91) or 3.079 Å (BP86) in **11S-2** is significantly shorter than the Nb–Nb single bond distances in the Nb₂(CO)₁₂ structures (Figure 1) and in the Nb₂(CO)₁₀(η^2 - μ -CO) structure **11S-1**. Thus, the Nb=Nb bond in **11S-2** appears to be the formal double bond required to give each niobium atom the favored 18-electron configuration.

The next singlet Nb₂(CO)₁₁ structure **11S-3** is a singly bridged C₂ structure, lying 20.5 kcal/mol (MPW1PW91) or 17.5 kcal/mol (BP86) in energy above the global minimum **11S-1** (Figure 2 and Table 3). The Nb–C distances to the symmetrically bridging CO group in **11S-3** are 2.372 Å (MPW1PW91) or 2.367 Å (BP86), while the five unique Nb–C distances to the 10 terminal carbonyls range from 2.084 Å to 2.167 Å (MPW1PW91) or from 2.091 Å to 2.172 Å (BP86). The Nb=Nb distance is predicted to be 3.074 Å (MPW1PW91) or 3.117 Å (BP86), consistent with the formal double bond needed to give both niobium atoms the favored 18-electron configuration. Structure **11S-3** has a significant imaginary vibrational frequency (>300i cm^{–1}). Following the normal mode of this imaginary frequency leads to the global minimum **11S-1**.

The other singlet Nb₂(CO)₁₁ structure **11S-4** has C_s symmetry and lies 21.5 kcal/mol (MPW1PW91) or 21.8 kcal/mol (BP86) above **11S-1** (Figure 2 and Table 3). Structure **11S-4** has two very weakly semibridging carbonyls with short Nb–C distances of 2.058 Å and 2.128 Å (MPW1PW91) or 2.064 Å and 2.127 Å (BP86) and long Nb–C distances of 3.081 Å and 3.099 Å (MPW1PW91) or 3.213 Å and 3.016 Å (BP86). The Nb–O distances to these two unique semibridging carbonyls are very long (>3.5 Å), indicating that both of these carbonyls are

two-electron donor carbonyl groups. However, despite their weakly semibridging nature, these two carbonyl groups exhibit relatively low ν (CO) frequencies of 1877 cm^{–1} and 1883 cm^{–1} (BP86). The Nb–Nb bond distance of 3.405 Å (MPW1PW91) or 3.471 Å (BP86) in **11S-4** is close to that in **11S-1** and thus consistent with a formal single bond. This gives one niobium atom in **11S-4** the favored 18-electron configuration but the other niobium atom only a 16-electron configuration. The niobium atom in **11S-4** with only a 16-electron configuration is the niobium atom bonded to only five carbonyl groups rather than six carbonyl groups, namely, the “right” niobium atom depicted in Figure 2. The **11S-4** structure can be derived from one of the unbridged Nb₂(CO)₁₂ structures (**12S-1** through **12S-3**) by removal of a terminal carbonyl group.

The triplet Nb₂(CO)₁₁ structure **11T-1** has C_s symmetry with a semibridging CO group (Figure 3 and Table 4) and lies 12.7 kcal/mol (MPW1PW91) or 12.4 kcal/mol (BP86) above the global minimum **11S-1**. Structure **11T-1** is predicted by both methods to have a very small imaginary vibrational frequency ($\leq 15i$ cm^{–1}, Table 4). This imaginary frequency arises from numerical integration errors, since it becomes real when a finer integration grid (120, 974) is used for the optimization. Structure **11T-1** has one semibridging carbonyl and 10 terminal carbonyls. In **11T-1**, the short Nb–C distance to the semibridging carbonyl is 2.152 Å (MPW1PW91) or 2.201 Å (BP86), whereas the long Nb–C distance is 2.570 Å (MPW1PW91) or 2.495 Å (BP86). The seven unique Nb–C distances to the terminal carbonyls in **11T-1** fall in the range from 2.118 to 2.158 Å (MPW1PW91) or from 2.117 to 2.156 Å (BP86). The Nb=Nb distance of 3.129 Å (MPW1PW91) or 3.150 Å (BP86) is significantly shorter than the Nb–Nb single bonds of lengths 3.3–3.4 Å found in the Nb₂(CO)₁₂ structures (Figure 1 and Table 1). This gives each niobium atom in **11T-1** the favored 18-electron configuration. The triplet spin multiplicity in **11T-1** can arise from the Nb=Nb double bond being a $\sigma + 2/2 \pi$ bond with the two unpaired electrons in two π “half bonds.” This is similar to dioxygen or the Fe=Fe double bond in the organometallic (η^5 -C₅H₅)₂Fe₂(μ -CO)₃, which is stable enough to be isolated in the pure state and structurally characterized by X-ray diffraction.^{44–46}

The second triplet Nb₂(CO)₁₁ structure **11T-2** is a singly symmetrically bridged C₂ structure (Figure 3 and Table 4), which is geometrically similar to the singlet structure **11S-3**. Structure **11T-2** lies 13.2 kcal/mol (MPW1PW91) or 12.8 kcal/mol (BP86) above the global minimum **11S-1**. However, structure **11T-2** is a transition state with a large imaginary vibrational frequency at 165i cm^{–1} (MPW1PW91) or 101i cm^{–1} (BP86). Following the corresponding normal mode leads to **11T-1**. The Nb–C distances to the bridging CO group in **11T-2** are 2.312 Å

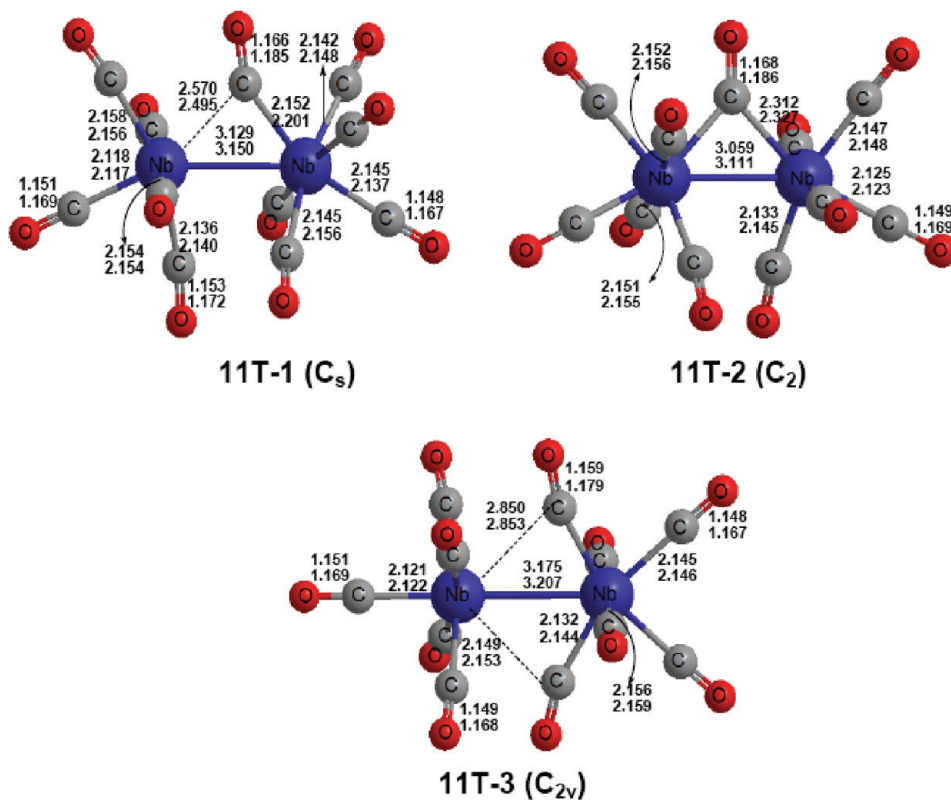


Figure 3. The triplet $Nb_2(CO)_{11}$ structures.

Table 4. Total Energies (E in hartrees), Relative Energies (ΔE in kcal/mol), Numbers of Imaginary Vibrational Frequencies (Nimag), and Nb–Nb Distances (in Å) for the Triplet $Nb_2(CO)_{11}$ Structures

		11T-1 (C_s)	11T-2 (C_2)	11T-3 (C_{2v})
MPW1PW91	E	−1360.66572	−1360.66492	−1360.66489
	ΔE	12.7	13.2	13.3
	Nimag	1 (15i)	1 (165i)	2 (43i, 10i)
	Nb–Nb	3.129	3.059	3.175
BP86	E	−1361.20948	−1361.20889	−1361.20483
	ΔE	12.4	12.8	15.3
	Nimag	1 (11i)	1 (101i)	3 (58i, 25i, 17i)
	Nb–Nb	3.151	3.111	3.207

(MPW1PW91) or 2.327 Å (BP86), whereas the five unique Nb–C distances for the 10 terminal CO groups fall in the range from 2.125 to 2.152 Å (MPW1PW91) or from 2.123 to 2.155 Å (BP86). The Nb=Nb distance in 11T-2 of 3.059 Å (MPW1PW91) or 3.111 Å (BP86) is close to that in 11T-1 and likewise can correspond to a formal double bond. This gives each niobium atom in 11T-2 the favored 18-electron configuration. The triplet spin multiplicity in 11T-2 arises from the Nb=Nb double bond being a $\sigma + 2/2 \pi$ bond with two unpaired electrons similar to that in 11T-2 discussed above.

The other low-lying triplet $Nb_2(CO)_{11}$ structure 11T-3 is a C_{2v} structure with two semibridging carbonyls and nine (four unique) terminal carbonyls (Figure 3 and Table 4). Structure 11T-3 lies 13.3 kcal/mol (MPW1PW91) or 15.3 kcal/mol (BP86) above 11S-1. The MPW1PW91 method predicts 11T-3 to have two small imaginary vibrational frequencies (43i and 10i cm^{-1}), while

the BP86 method predicts 11T-3 to have three small imaginary vibrational frequencies (58i, 25i, and 17i cm^{-1}). Following the normal mode corresponding to the largest frequency leads to 11T-1. For the two equivalent semibridging carbonyls in 11T-3, the short Nb–C bond distances are 2.132 Å (MPW1PW91) or 2.144 Å (BP86), and the long Nb–C distances are 2.850 Å (MPW1PW91) or 2.853 Å (BP86). The $\nu(CO)$ frequencies at 1890 and 1900 cm^{-1} (BP86) can be assigned to these two semibridging CO groups. The Nb=Nb distance of 11T-3 of 3.175 Å (MPW1PW91) or 3.207 Å (BP86) is similar to that in the triplets 11T-1 and 11T-2 and can likewise be interpreted as a $\sigma + 2/2 \pi$ bond containing the two unpaired electrons of the triplet spin state.

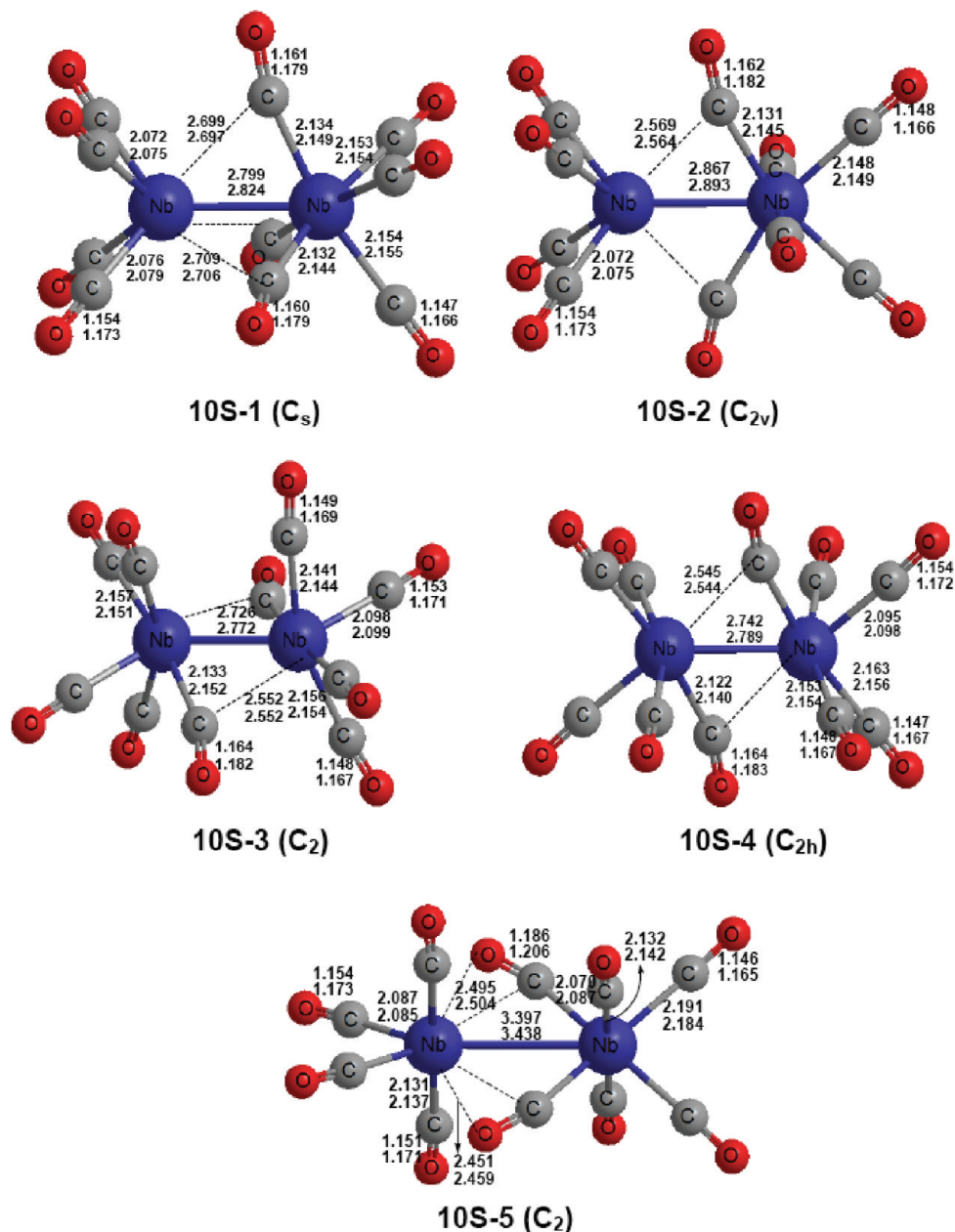
Table 5 lists the $\nu(CO)$ frequencies for the seven $Nb_2(CO)_{11}$ structures predicted by the BP86 method. The $\nu(CO)$ frequencies for the terminal CO groups fall in the range 1912–2065 cm^{-1} (Table 5). The $\nu(CO)$ frequencies for the two-electron donor bridging CO groups are significantly lower in the range 1831–1900 cm^{-1} (**bold face** in Table 5). The four-electron donor bridging $\eta^2-\mu$ -CO group in 11S-1 exhibits an extremely low $\nu(CO)$ frequency at 1701 cm^{-1} (BP86).

3.3. $Nb_2(CO)_{10}$. A total of nine low-lying structures (five singlets and four triplets) were found for $Nb_2(CO)_{10}$ (Figures 4 and 5 and Tables 6 and 7). These nine structures can be classified into three general types: (a) structures with a Nb=Nb distance of around 2.8 Å, suggesting a formal triple bond. This type includes the singlets 10S-1, 10S-2, 10S-3, and 10S-4; (b) structures with a Nb=Nb distance of around 3.0 Å, suggesting a formal double bond—this type includes the triplets 10T-1, 10T-2, and 10T-3; (c) structures with a long Nb–Nb distance (>3.4 Å), namely, 10S-5 and 10T-4. In such structures, a $Nb(CO)_6$ unit acts as a bidentate chelating ligand to a $Nb(CO)_4$

Table 5. Infrared $\nu(\text{CO})$ Vibrational Frequencies (cm^{-1}) Predicted by the BP86 Method for the $\text{Nb}_2(\text{CO})_{11}$ Structures^a

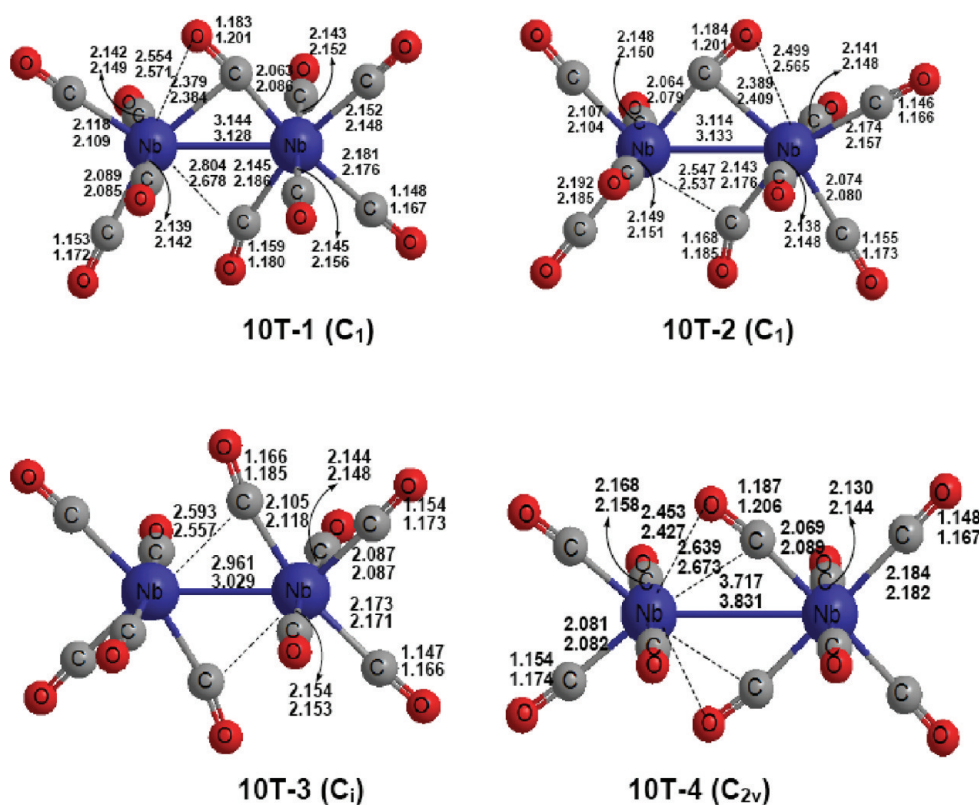
	BP86
11S-1 (C_1)	1701 (523), 1939 (188), 1952 (413), 1953 (454), 1968 (542), 1970 (1445), 1972 (754), 1983 (2544), 1986 (229), 2024 (1822), 2062 (27)
11S-2 (C_s)	1877 (269), 1912 (51), 1929 (499), 1942 (1431), 1948 (502), 1966 (988), 1967 (461), 1985 (1569), 2001 (565), 2008 (2587), 2060 (127)
11S-3 (C_2)	1896 (261), 1926 (102), 1936 (13), 1946 (261), 1953 (445), 1958 (1305), 1965 (1117), 1974 (660), 1976 (2154), 2000 (3327), 2058 (0)
11S-4 (C_s)	1877 (783), 1883 (103), 1913 (1024), 1946 (388), 1957 (633), 1973 (587), 1991 (956), 1992 (1191), 1998 (1912), 1999 (753), 2065 (83)
11T-1 (C_s)	1844 (263), 1936 (96), 1942 (2), 1961 (544), 1965 (1268), 1972 (733), 1972 (2843), 1974 (78), 1976 (556), 2004 (3637), 2062 (17)
11T-2 (C_2)	1831 (212), 1935 (56), 1943 (19), 1966 (980), 1966 (389), 1969 (653), 1971 (45), 1973 (2878), 1975 (971), 2003 (3672), 2062 (3)
11T-3 (C_{2v})	1890 (396), 1900 (66) , 1948 (8), 1963 (1333), 1964 (834), 1968 (0), 1973 (3012), 1976 (10), 1978 (1113), 2009 (3017), 2062 (7)

^aInfrared intensities are in parentheses (in km/mol). The bridging CO frequencies are in **bold**.

Figure 4. Optimized singlet $\text{Nb}_2(\text{CO})_{10}$ geometries.

unit through its doubly bridging CO groups acting as four-electron donors.

The global minimum for $\text{Nb}_2(\text{CO})_{10}$ is the asymmetric singlet C_s structure 10S-1 with one Nb atom bonded to six CO groups

Figure 5. Optimized triplet $\text{Nb}_2(\text{CO})_{10}$ geometries.Table 6. Total Energies (E , in Hartree), Relative Energies (ΔE , in kcal/mol), Numbers of Imaginary Vibrational Frequencies (Nimag), and Nb–Nb Bond Distances (Å) for the Singlet $\text{Nb}_2(\text{CO})_{10}$ Structures

		10S-1 (C_s)	10S-2 (C_{2v})	10S-3 (C_2)	10S-4 (C_{2h})	10S-5 (C_2)
MPW1PW91	E	−1247.36824	−1247.35463	−1247.35234	−1247.35149	−1247.33452
	ΔE	0.0	8.5	10.3	10.5	21.2
	Nimag	0	1 (54i)	1 (12i)	1 (18i)	0
	Nb–Nb	2.799	2.867	2.726	2.742	3.397
BP86	E	−1247.87732	−1247.86496	−1247.86293	−1247.86192	−1247.84404
	ΔE	0.0	7.8	9.0	9.7	20.9
	Nimag	0	1 (53i)	1 (20i)	1 (22i)	0
	Nb–Nb	2.824	2.893	2.772	2.789	3.438

Table 7. Total Energies (E , in Hartree), Relative Energies (ΔE , in kcal/mol), Numbers of Imaginary Vibrational Frequencies (Nimag), and Nb–Nb Bond Distances (Å) for the Triplet $\text{Nb}_2(\text{CO})_{10}$ Structures

		10T-1 (C_1)	10T-2 (C_1)	10T-3 (C_i)	10T-4 (C_{2v})
MPW1PW91	E	−1247.33730	−1247.33678	−1247.33308	−1247.32399
	ΔE	19.4	19.7	22.1	27.8
	Nimag	0	0	1 (5i)	0
	Nb–Nb	3.144	3.114	2.960	3.717
BP86	E	−1247.84553	−1247.84649	−1247.84456	−1247.83232
	ΔE	19.9	19.3	20.6	28.2
	Nimag	0	0	1 (10i)	0
	Nb–Nb	3.128	3.133	3.029	3.831

(the “right” Nb atom in Figure 4) and the other Nb atom bonded to only four CO groups (the “left” Nb atom in Figure 4). The

inward three CO groups of the $\text{Nb}(\text{CO})_6$ group are actually semibridging carbonyls to the $\text{Nb}(\text{CO})_4$ group. In these three

semibridging carbonyls, the short Nb–C distances fall in the range from 2.132 to 2.134 Å (MPW1PW91) or from 2.144 to 2.149 Å (BP86), while the long Nb–C distances fall in the range from 2.699 to 2.709 Å (MPW1PW91) or from 2.697 to 2.706 Å (BP86). The C–O distances of ~ 1.16 Å (MPW1PW91) or ~ 1.18 Å (BP86) in the semibridging CO groups of **10S-1** are slightly longer than the C–O distances in the terminal CO groups. The Nb–C–O angles for the three semibridging CO groups deviate significantly from linearity (165° and 164°), while those for the terminal CO groups are almost linear. The predicted Nb=Nb distances of 2.799 Å (MPW1PW91) or 2.824 Å (BP86) in **10S-1** correspond to a formal triple bond, thereby giving each niobium atom the favored 18-electron configuration.

The next Nb₂(CO)₁₀ structure **10S-2** is a C_{2v} doubly semibridged structure lying ~ 8 kcal/mol above the global minimum **10S-1** (Figure 4 and Table 6). In **10S-2**, one Nb atom (the “right” Nb atom in Figure 4) is bonded to six CO groups, whereas the other Nb atom (the “left” Nb atom in Figure 4) is bonded to only four CO groups, similar to **10S-1**. The two inward CO groups of the Nb(CO)₆ unit in **10S-2** are bonded to the Nb(CO)₄ unit as semibridging carbonyls with short Nb–C distances of 2.131 Å (MPW1PW91) or 2.145 Å (BP86) and long Nb–C distances of 2.569 Å (MPW1PW91) or 2.564 Å (BP86). The Nb–C–O angles for these semibridging CO groups deviate significantly from linearity, i.e., 162° (MPW1PW91) or 161° (BP86), while those for the terminal CO groups are almost linear. The Nb=Nb distance in structure **10S-2** is 2.867 Å (MPW1PW91) or 2.893 Å (BP86), which is slightly longer than that of **10S-1** but still consistent with the formal triple bond required to give each Nb atom the favored 18-electron configuration. Structure **10S-2** has an imaginary vibrational frequency of $54i$ cm⁻¹ (MPW1PW91) or $53i$ cm⁻¹. By following the corresponding normal mode, structure **10S-2** collapses to **10S-1**.

The next two singlet Nb₂(CO)₁₀ structures in terms of energy are **10S-3** (C₂) and **10S-4** (C_{2h}). These structures are both doubly semibridged (CO)₄Nb(μ -CO)₂Nb(CO)₄ structures, differing only in the orientations of the CO groups (Figure 4 and Table 6). Structure **10S-3** lies 10.3 kcal/mol (MPW1PW91) or 9.0 kcal/mol (BP86) above the global minimum **10S-1**, whereas structure **10S-4** lies 10.5 kcal/mol (MPW1PW91) or 9.7 kcal/mol (BP86) above **10S-1**. In structures **10S-3** and **10S-4**, the Nb–C distances to the terminal carbonyl groups (four unique ones for **10S-3**, and three for **10S-4**) fall in the range from ~ 2.09 to ~ 2.16 Å. The short Nb–C distances to the semibridging carbonyl groups are ~ 2.14 Å, and the long Nb–C distances to the semibridging carbonyl groups are ~ 2.55 Å. The bridging Nb–C–O angles for both structures **10S-3** and **10S-4** deviate significantly from linearity ($\sim 162^\circ$). The Nb=Nb distances for structure **10S-3** of 2.726 Å (MPW1PW91) or 2.772 Å (BP86) and for structure **10S-4** of 2.742 Å (MPW1PW91) or 2.789 Å (BP86) correspond to the formal triple bonds required to give each niobium atom the favored 18-electron configuration. Both structures **10S-3** and **10S-4** are predicted to have small imaginary vibrational frequencies at $\sim 20i$ cm⁻¹ (Table 6), which cannot be removed by using a finer integration grid (120, 974). By following the corresponding normal modes, both **10S-3** and **10S-4** collapse to **10S-1**.

Three triplet Nb₂(CO)₁₀ structures, **10T-1**, **10T-2**, and **10T-3**, without bidentate Nb(CO)₆ ligands were found having Nb–Nb distances of ~ 3.1 Å (Figure 5 and Table 7). Structure **10T-1** is a doubly bridged C₁ structure. For one of the bridging CO groups, the Nb–O distance is very short, namely, 2.55 Å (MPW1PW91)

or 2.57 Å (BP86), indicating a four-electron donor bridging η^2 - μ -CO group. The two Nb–C distances to this carbonyl are 2.379 Å (MPW1PW91) or 2.384 Å (BP86) and 2.804 Å (MPW1PW91) or 2.678 Å (BP86), respectively, and its Nb–C–O angle is $\sim 174^\circ$, deviating only $\sim 6^\circ$ from linearity. The corresponding C–O distance is relatively long at 1.183 Å (MPW1PW91) or 1.201 Å (BP86), and the ν (CO) frequency is very low at 1773 cm⁻¹ (BP86). The Nb \cdots O distance to the other bridging CO group in **10T-1** is very long, namely, 3.52 Å (MPW1PW91) or 3.49 Å (BP86). This indicates no direct Nb–O interaction and thus a normal two-electron donor bridging CO group. The Nb–C–O angle to this bridging carbonyl group of $\sim 163^\circ$ deviates significantly from linearity. The C–O distance in this bridging CO group is shorter, i.e., 1.159 Å (MPW1PW91) or 1.180 Å (BP86), than that to the η^2 - μ -CO group. This is consistent with its predicted ν (CO) frequency of 1878 cm⁻¹ (BP86). The Nb=Nb distance in **10T-1** of 3.144 Å (MPW1PW91) or 3.128 Å (BP86) can be interpreted as a formal double bond, thereby giving each niobium atom the favored 18-electron configuration for a Nb₂(CO)₁₀ structure with a single four-electron donor η^2 - μ -CO group. This Nb=Nb double bond can be a $\sigma + 2/2 \pi$ double bond similar to that in **11T-1**, **11T-2**, and **11T-3** discussed above and thus contains the two unpaired electrons for the triplet spin state.

The triplet Nb₂(CO)₁₀ structure **10T-2** is a doubly bridged C₁ structure with a geometry and energy very close to **10T-1** (Figure 5 and Table 7). The main difference between these two Nb₂(CO)₁₀ structures is the orientation of the bridging CO groups, suggesting that the two structures can interconvert by motion of the bridging carbonyls. For one of the bridging CO groups in **10T-2**, the Nb–O distance is very short, i.e., 2.50 Å (MPW1PW91) or 2.56 Å (BP86); the C–O distance is relatively long at 1.184 Å (MPW1PW91) or 1.201 Å (BP86); and the predicted ν (CO) frequency of 1769 cm⁻¹ is very low. All of these characteristics indicate a four-electron donor bridging η^2 - μ -CO group. In contrast, for the other bridging CO group in **10T-2**, the Nb–O distance is very long, namely, 3.28 Å (MPW1PW91) or 3.31 Å (BP86), indicating a typical two-electron donor bridging CO group. The related C–O distance is 0.02 Å shorter, i.e., 1.168 Å, and its corresponding ν (CO) frequency is 1847 cm⁻¹ (BP86). The structure **10T-2** (C₁) is a genuine minimum with all real vibrational frequencies (Table 7). The Nb=Nb distance in **10T-2** is 3.114 Å (MPW1PW91) or 3.133 Å (BP86), which is consistent with the formal double bond required to give each niobium atom the favored 18-electron configuration. However, as for many of the other triplet Nb₂(CO)₁₁ and Nb₂(CO)₁₀ structures, this Nb=Nb double bond can be of the $\sigma + 2/2 \pi$ type, thereby containing the two unpaired electrons for the triplet spin state.

The Nb₂(CO)₁₀ structure **10T-3** is a C_i structure with two equivalent semibridging carbonyls and eight (four unique) terminal carbonyls (Figure 5 and Table 7). Structure **10T-3** lies 22.1 kcal/mol (MPW1PW91) or 20.6 kcal/mol (BP86) above **10S-1**. Structure **10T-3** has a negligible imaginary vibrational frequency ($5i$ cm⁻¹ by MPW1PW91 or $10i$ cm⁻¹ by BP86). The former ($5i$ cm⁻¹) becomes real when a finer integration grid (120, 974) is used. The predicted ν (CO) frequencies of 1850 and 1862 cm⁻¹ (BP86) in **10T-3** correspond to the semibridging carbonyl groups. The short Nb–C bond distances to the identical semibridging carbonyl groups in **10T-3** are 2.105 Å (MPW1PW91) or 2.118 Å (BP86), whereas the long Nb–C distances are 2.593 Å (MPW1PW91) or 2.557 Å (BP86). The four unique Nb–C distances to the terminal carbonyls in **10T-3**

Table 8. Infrared $\nu(\text{CO})$ Vibrational Frequencies (cm^{-1}) Predicted by the BP86 Method for the $\text{Nb}_2(\text{CO})_{10}$ Structures (Infrared Intensities in Parentheses Are in km/mol)

BP86	
10S-1 (C_s)	1900 (514), 1900 (464), 1916 (373) , 1950 (733), 1951 (1612), 1952 (881), 1992 (714), 1992 (736), 2015 (1901), 2053 (285)
10S-2 (C_{2v})	1877 (756), 1885 (404) , 1946 (1481), 1947 (287), 1949 (0), 1967 (2471), 1982 (113), 1991 (724), 2017 (1686), 2051 (286)
10S-3 (C_2)	1874 (57), 1890 (804) , 1958 (63), 1960 (2), 1964 (2257), 1978 (163), 1979 (1537), 1983 (1190), 2012 (2669), 2058 (4)
10S-4 (C_{2h})	1867 (0), 1888 (840) , 1957 (0), 1960 (2116), 1966 (0), 1979 (323), 1985 (2856), 1987 (0), 2013 (2616), 2060 (0)
10S-5 (C_2)	1729 (650), 1754 (430) , 1934 (234), 1942 (563), 1943 (543), 1967 (2851), 1979 (217), 1995 (779), 2015 (1976), 2052 (188)
10T-1 (C_i)	1773 (368), 1878 (235) , 1935 (205), 1943 (687), 1947 (481), 1969 (2672), 1976 (36), 1987 (863), 1998 (3499), 2054 (138)
10T-2 (C_i)	1769 (333), 1847 (353) , 1939 (21), 1949 (724), 1956 (552), 1968 (2919), 1969 (849), 1980 (71), 2003 (3467), 2053 (8)
10T-3 (C_i)	1850 (0), 1862 (876) , 1939 (0), 1948 (1615), 1950 (0), 1968 (981), 1970 (2598), 1975 (0), 2001 (3396), 2053 (0)
10T-4 (C_{2v})	1720 (695), 1735 (70) , 1922 (667), 1943 (66), 1951 (886), 1969 (3016), 1970 (240), 1982 (877), 1998 (3644), 2049 (56)

Table 9. Formal Nb–Nb Single Bonds in Binuclear Niobium Carbonyls and Related Compounds

compound	bridges	Nb–Nb distance, Å
$\text{Nb}_2(\text{CO})_{12}$		
12S-1		3.40
12S-2		3.39
12S-3		3.64
12S-4	2 CO	3.42
$\text{Nb}_2(\text{CO})_{11}$		
11S-1	$\eta^2\text{-}\mu\text{-CO}$	3.51
11S-4		3.44
$\text{Nb}_2(\text{CO})_{10}$		
10S-5	2 $\eta^2\text{-}\mu\text{-CO}$	3.42
$\text{Cp}_2\text{Nb}_2(\text{CO})_7$ (ref 49)		3.40
$\text{Cp}_2\text{Nb}_2(\text{CO})_5(\eta^2\text{-}\mu\text{-CO})$ (ref 49)	$\eta^2\text{-}\mu\text{-CO}$	3.21
$\text{Cp}_3\text{Nb}_3(\text{CO})_6(\eta^2\text{-}\mu_3\text{-CO})$ (theory) ⁴⁸	variable	3.08, 3.24, 3.37
$\text{Cp}_3\text{Nb}_3(\text{CO})_6(\eta^2\text{-}\mu_3\text{-CO})$ (exptl.) ⁴⁷	variable	3.04, 3.18, 3.32

range from 2.087 to 2.173 Å (MPW1PW91) or 2.087 to 2.171 Å (BP86). The Nb=Nb distance in **10T-3** is 2.961 Å (MPW1PW91) or 3.029 Å (BP86). This is consistent with the formal double bond required to give each niobium atom the 17-electron configuration for a binuclear $\text{Nb}_2(\text{CO})_{10}$ triplet structure with only two-electron donor carbonyl groups.

The $\text{Nb}_2(\text{CO})_{10}$ structures with a chelating bidentate $\text{Nb}(\text{CO})_6$ "ligand" bonded to the $\text{Nb}(\text{CO})_4$ unit are **10S-5** (C_2) and **10T-4** (C_{2v}). These structures are characterized by very long Nb–Nb distances (>3.4 Å), suggesting very weak, if any, direct interactions between the two Nb atoms. Thus, the $\text{Nb}(\text{CO})_6$ and $\text{Nb}(\text{CO})_4$ units in these structures are linked mainly by the two bridging CO groups, which are four-electron donor CO groups. These two four-electron donor CO groups are characterized by very short Nb–O distances (<2.460 Å) and $\nu(\text{CO})$ frequencies below 1800 cm^{-1} (BP86). For **10S-5**, the Nb–Nb distance of 3.397 Å (MPW1PW91) or 3.438 Å (BP86) suggests the weak single bond required to give both niobium atoms the favored 18-electron configuration. The bridging CO groups in **10S-5** have short Nb–C distances of 2.070 Å (MPW1PW91) or 2.087 Å (BP86) to one niobium atom and long Nb–C distances of 2.495 Å (MPW1PW91) or 2.504 Å (BP86) to the other niobium atom. For the triplet structure **10T-4**, the Nb \cdots Nb distance is even longer, namely, 3.717 Å (MPW1PW91) or 3.831 Å (BP86). This suggests a lack of niobium–niobium bonding, thereby giving each

Table 10. Formal Nb=Nb Double Bonds in Binuclear Niobium Carbonyls and Related Compounds

compound	bridges	Nb–Nb distance, Å
$\text{Nb}_2(\text{CO})_{11}$		
11S-2		3.08
11S-3	CO	3.09
11T-1		3.14
11T-2	CO	3.08
11T-3		3.19
10T-1	$\eta^2\text{-}\mu\text{-CO}$	3.13
10T-2	$\eta^2\text{-}\mu\text{-CO}$	3.12
$\text{Nb}_2(\text{CO})_{10}$ (10T-3)		2.99
$\text{Cp}_2\text{Nb}_2(\text{CO})_6$ (ref 49)		2.99
$\text{Cp}_2\text{Nb}_2(\text{CO})_2(\eta^2\text{-}\mu\text{-CO})_2$ (ref 49)	2 $\eta^2\text{-}\mu\text{-CO}$	2.97

Table 11. Formal Nb≡Nb Triple Bonds in Binuclear Niobium Carbonyls and Related Compounds

compound	bridges	Nb–Nb distance, Å
$\text{Nb}_2(\text{CO})_{10}$		
10S-1		2.81
10S-2		2.88
10S-3		2.75
$\text{Cp}_2\text{Nb}_2(\text{CO})_5$ (ref 49)		2.74
$\text{Cp}_2\text{Nb}_2(\text{CO})(\eta^2\text{-}\mu\text{-CO})_2$ (ref 49)	2 $\eta^2\text{-}\mu\text{-CO}$	2.84

niobium atom a 17-electron configuration, consistent with a binuclear triplet. Both **10S-5** and **10T-4** are genuine minima, having all real vibrational frequencies. However, these two structures lie at relatively high energies, namely, ~ 21 kcal/mol above **10S-1** for **10S-5** and ~ 28 kcal/mol above **10S-1** for **10T-4**.

Table 8 lists the infrared $\nu(\text{CO})$ frequencies for the nine $\text{Nb}_2(\text{CO})_{10}$ structures, predicted by the BP86 method. The $\nu(\text{CO})$ frequencies for the terminal CO groups fall between 1922 and 2060 cm^{-1} , whereas those for the two-electron donor bridging CO groups are appreciably lower between 1850 and 1916 cm^{-1} (**bold face** in Table 8). The four-electron donor bridging $\eta^2\text{-}\mu\text{-CO}$ groups in **10S-5**, **10T-1**, **10T-2**, and **10T-4** are predicted to exhibit even lower $\nu(\text{CO})$ frequencies between 1720 and 1773 cm^{-1} (BP86; **bold face** in Table 8).

3.4. Niobium–Niobium Bonding. Tables 9, 10, and 11 compare formal Nb–Nb single bonds, Nb=Nb double bonds,

Table 12. Atomic Charges and Wiberg Bond Indices for the Nb₂(CO)_n Structures Using the BP86 Method

structure	number of CO bridges	Nb–Nb distance, Å	WBI	formal bond order	natural charge on Nb(left)/Nb(right)
Nb ₂ (CO) ₁₂					
12S-1	0	3.40	0.27	1	−1.92/−1.80
12S-2	0	3.39	0.27	1	−1.86/−1.86
12S-3	0	3.64	0.38	1	−1.85/−1.85
12S-4	2	3.42	0.14	1	−1.90/−1.90
Nb ₂ (CO) ₁₁					
11S-1	η ² -μ-CO	3.51	0.29	1	−1.90/−1.37
11S-2	1 (semi)	3.07	0.42	2	−1.71/−1.56
11S-3	1	3.10	0.47	2	−1.70/−1.70
11S-4	0	3.44	0.27	1	−1.79/−1.19
11T-1	1	3.14	0.45	2	−1.57/−1.80
11T-2	1	3.08	0.47	2	−1.70/−1.70
11T-3	2 (semi)	3.19	0.37	2	−1.44/−1.89
Nb ₂ (CO) ₁₀					
10S-1	3 (semi)	2.81	0.64	3	−0.99/−2.01
10S-2	2 (semi)	2.88	0.58	3	−0.88/−2.09
10S-3	3 (semi)	2.75	0.86	3	−1.54/−1.54
10S-4	2 (semi)	2.76	0.88	3	−1.55/−1.55
10S-5	2η ² -μ-CO	3.42	0.39	1	−1.05/−1.84
10T-1	η ² -μ-CO	3.13	0.45	2	−1.09/−1.84
10T-2	η ² -μ-CO	3.12	0.46	2	−1.45/−1.44
10T-3	2 (semi)	2.99	0.52	2	−1.44/−1.44
10T-4	2η ² -μ-CO	3.77	0.19	0	−0.89/−1.86

and Nb≡Nb triple bonds, respectively, in various Nb₂(CO)_n derivatives (*n* = 12, 11, 10) with similar bonds in other niobium carbonyl species. However, the only niobium carbonyl derivative for which experimental data are available on Nb–Nb distances⁴⁷ is the unusual trinuclear species Cp₃Nb₃(CO)₆(η²-μ₃-CO) (Cp = η⁵-C₅H₅) with an unusual six-electron donor η²-μ₃-CO group bridging all three niobium atoms in the central Nb₃ triangle. Because of this limited amount of experimental data, niobium–niobium distances in the binuclear cyclopentadienyl-niobium carbonyl derivatives Cp₂Nb₂(CO)_n (*n* = 7, 6, 5, 4, 3) predicted in previous theoretical studies^{48,49} are also included in the tables. However, only the Nb–Nb distances for the lowest energy predicted Cp₂Nb₂(CO)_n structures of a given type are included. The theoretical values of the Nb–Nb distances in the tables are the means of the values predicted by the MPW1PW91 and BP86 methods.

The Nb–Nb single bond distances in singlet Nb₂(CO)_n structures are seen to be very close to 3.4 Å (Table 9). A similar Nb–Nb single bond distance is predicted⁴⁹ for the unknown Cp₂Nb₂(CO)₇, which also has a singlet spin state.

The Nb=Nb double bond distances in both singlet and triplet Nb₂(CO)_n structures are consistently shorter than the formal Nb–Nb single bonds. Thus, the Nb=Nb double bond distances fall in the narrow range of 2.99–3.19 Å (Table 10). This range is similar to that previously predicted⁴⁹ for the Nb=Nb distances in the lowest energy singlet Cp₂Nb₂(CO)₆ and Cp₂Nb₂(CO)₂(η²-μ-CO)₂ structures.

The only binuclear niobium carbonyls found in this work with formal Nb≡Nb triple bonds have singlet spin states. The predicted Nb≡Nb triple bond lengths in such Nb₂(CO)_n structures are consistently shorter than formal Nb=Nb double bond lengths and fall in the narrow range of 2.75–2.81 Å (Table 11). This range is bracketed by the predicted⁴⁹ Nb≡Nb triple bond lengths of 2.74 Å in Cp₂Nb₂(CO)₅ and 2.84 Å in Cp₂Nb₂(CO)(η²-μ-CO)₂.

Table 12 lists the Wiberg bond indices (WBIs)⁵⁰ and bond distances for the Nb–Nb bonds in the Nb₂(CO)_n derivatives. The results with the BP86 method are used, since less spin contamination was found in the triplet structures, relative to the MPW1PW91 method. The WBIs are seen to correlate reasonably with the formal bond orders estimated from the niobium–niobium distances and electron counting. In this connection, previous studies on the WBIs in metal–metal bonded derivatives suggest typical values of 0.2–0.3 for unbridged formal metal–metal single bonds.⁵¹ The WBIs for the single Nb–Nb bonds in the unbridged Nb₂(CO)₁₂ structures fall in a similar range from 0.27 to 0.39. The likewise long Nb–Nb formal single bonds of ~3.5 Å in the Nb₂(CO)₁₁ structures 11S-1 and 11S-4, which are clearly derived from the lowest energy Nb₂(CO)₁₂ structure 12S-1 by loss of a terminal carbonyl group, have similar WBIs of 0.27–0.29. The Nb=Nb double bonds in the Nb₂(CO)₁₁ and Nb₂(CO)₁₀ structures 11S-2, 11T-1, 10S-5, 10T-1, and 10T-2 have significantly higher WBIs from 0.42 to 0.52. The Nb≡Nb triple bonds in the singlet Nb₂(CO)₁₀ structures 10S-1, 10S-2, 10S-3, and 10S-4 have still higher WBIs falling in the range from 0.58 to 0.88. The WBI for the nonbonding Nb atoms in 10T-4 is very small, i.e., 0.19.

The major factor affecting the natural charges on the niobium atoms in the Nb₂(CO)_n derivatives is the number of carbonyl groups to which they are bonded (Table 12). This suggests that the π back-bonding of the niobium atom to the antibonding orbitals of the carbonyl groups is not sufficient to remove all of the negative charge arising from the σ forward bonding. Thus, the natural charges on the niobium atoms bearing six carbonyl groups such as the unbridged Nb₂(CO)₁₂ structures fall in the narrow range −1.80 to −1.92. Niobium atoms bonded to five terminal carbonyl groups are less negative with natural charges in the range −1.19 to −1.57. Similarly for the two Nb₂(CO)₁₀ structures, in which one of the niobium atoms is bonded to only four terminal carbonyl groups, namely, 10S-1 and 10S-5 (Figure 4), the natural charges on the niobium atoms are even less negative, from −0.88 to −0.99.

3.5. Thermochemistry. Table 13 lists the dissociation energies for the reactions Nb₂(CO)_n → Nb₂(CO)_{n−1} + CO (*n* = 12, 11, 10) considering the lowest energy structures of Nb₂(CO)_m, e.g., 12S-1, 11S-1, and 10S-1. The dissociation energies for the loss of one carbonyl group from Nb₂(CO)₁₂ and Nb₂(CO)₁₁ are relatively low, namely, ~18 kcal/mol and ~15 kcal/mol, respectively. However, the predicted energy for the loss of one CO group from Nb₂(CO)₁₀ is much higher at ~37 kcal/mol. For comparison, the experimental BDEs⁵² for Ni(CO)₄, Fe(CO)₅, and Cr(CO)₆ are 27 kcal/mol, 41 kcal/mol, and 37 kcal/mol, respectively.

Table 13 also lists the energies of the disproportionation reactions 2Nb₂(CO)_n → Nb₂(CO)_{n+1} + Nb₂(CO)_{n−1}. The disproportionation of Nb₂(CO)₁₁ into Nb₂(CO)₁₂ + Nb₂(CO)₁₀ is slightly exothermic by ~4 kcal/mol, suggesting that Nb₂(CO)₁₁ might not be a viable species. However, the disproportionation of Nb₂(CO)₁₀ into Nb₂(CO)₉ + Nb₂(CO)₁₁ is endothermic by ~22 kcal/mol.

Table 13. Dissociation Energies (kcal/mol) for the Successive Removal of Carbonyl Groups, Dissociation Energies into Mononuclear Fragments (kcal/mol), and Disproportionation Energies (kcal/mol) for $\text{Nb}_2(\text{CO})_n$ ($n = 12, 11, 10$)^a

	MPW1PW91	BP86
$\text{Nb}_2(\text{CO})_{12} \rightarrow \text{Nb}_2(\text{CO})_{11} + \text{CO}$	18.7	18.9
$\text{Nb}_2(\text{CO})_{11} \rightarrow \text{Nb}_2(\text{CO})_{10} + \text{CO}$	14.4	15.5
$\text{Nb}_2(\text{CO})_{10} \rightarrow \text{Nb}_2(\text{CO})_9 + \text{CO}$	36.5	37.4
$2\text{Nb}_2(\text{CO})_{11} \rightarrow \text{Nb}_2(\text{CO})_{12} + \text{Nb}_2(\text{CO})_{10}$	-4.3	-3.5
$2\text{Nb}_2(\text{CO})_{10} \rightarrow \text{Nb}_2(\text{CO})_{11} + \text{Nb}_2(\text{CO})_9$	22.1	21.9
$\text{Nb}_2(\text{CO})_{12} \rightarrow 2\text{Nb}(\text{CO})_6$	12.2	13.9
$\text{Nb}_2(\text{CO})_{11} \rightarrow \text{Nb}(\text{CO})_5 + \text{Nb}(\text{CO})_6$	29.5	32.2
$\text{Nb}_2(\text{CO})_{10} \rightarrow 2\text{Nb}(\text{CO})_5$	51.3	53.8
$\text{Nb}_2(\text{CO})_{10} \rightarrow \text{Nb}(\text{CO})_4 + \text{Nb}(\text{CO})_6$	51.1	54.3

^a All results reported refer to the lowest energy structures of $\text{Nb}_2(\text{CO})_n$.

The dissociation of $\text{Nb}_2(\text{CO})_n$ into mononuclear fragments is also shown in Table 13. In order to obtain these data, the global minima for the mononuclear $\text{Nb}(\text{CO})_m$ ($m = 6, 5, 4$) were optimized by the same DFT methods (Figure 6). The dissociation reactions of all $\text{Nb}_2(\text{CO})_n$ structures ($n = 12, 11, 10$) are all endothermic. The dissociation energy of the saturated $\text{Nb}_2(\text{CO})_{12}$ into the $\text{Nb}(\text{CO})_6$ fragments is found to be ~ 13 kcal/mol, that for $\text{Nb}_2(\text{CO})_{11}$ into $\text{Nb}(\text{CO})_5$ and $\text{Nb}(\text{CO})_6$ is ~ 30 kcal/mol, and that for unsaturated $\text{Nb}_2(\text{CO})_{10}$ is predicted to reach ~ 50 kcal/mol. Thus, the binuclear $\text{Nb}_2(\text{CO})_n$ complexes, even including $\text{Nb}_2(\text{CO})_{12}$, are viable with respect to dissociation into mononuclear fragments. The dissociation energies of $\text{Nb}_2(\text{CO})_n$ monotonically increase with decreasing n , in accord with the increasing formal Nb–Nb bond orders.

A single $\nu(\text{CO})$ frequency at 1984.1 cm^{-1} assigned to $\text{Nb}(\text{CO})_6$ was observed by Zhou and Andrews¹⁴ in the mixture of niobium carbonyls obtained by co-deposition of laser-ablated niobium atoms and carbon monoxide molecules in excess neon. Our theoretical structure for the doublet $\text{Nb}(\text{CO})_6$ (Figure 6) has D_{3d} symmetry rather than ideal O_h symmetry, owing to the Jahn–Teller effect. For the six CO groups, the infrared active $\nu(\text{CO})$ frequencies are only the e_u and a_{2u} modes. Both of these $\nu(\text{CO})$ frequencies are predicted to occur at 1961 cm^{-1} using the BP86 method (Table 14). This is in good agreement with the experimental result of Zhou and Andrews and supports their assignment of this $\nu(\text{CO})$ frequency to $\text{Nb}(\text{CO})_6$.

4. DISCUSSION

The stable binary vanadium carbonyl is the 17-electron mononuclear $\text{V}(\text{CO})_6$ rather than its dimer $\text{V}_2(\text{CO})_{12}$, consisting of two $\text{V}(\text{CO})_6$ units linked by a V–V single bond. In $\text{V}(\text{CO})_6$, the vanadium atom is clearly six-coordinate, whereas in $\text{V}_2(\text{CO})_{12}$ the vanadium atom is necessarily seven-coordinate. For the larger niobium atom, the binuclear $\text{Nb}_2(\text{CO})_{12}$, with a seven-coordinate niobium atom, is predicted to be the stable species, requiring an energy of ~ 13 kcal/mol for dissociation into two $\text{Nb}(\text{CO})_6$ units.

The $\text{Nb}_2(\text{CO})_{12}$ structures found in this work (Figure 1 and Table 1) provide an interesting illustration of how symmetry breaking can reduce the energy of the optimized structures. Structure **12S-3** has relatively high D_{3d} symmetry with the CO groups occupying the octahedral positions and the Nb–Nb bond occupying the capping positions of the capped octahedral niobium coordination polyhedron. Structure **12S-3** has a pair of

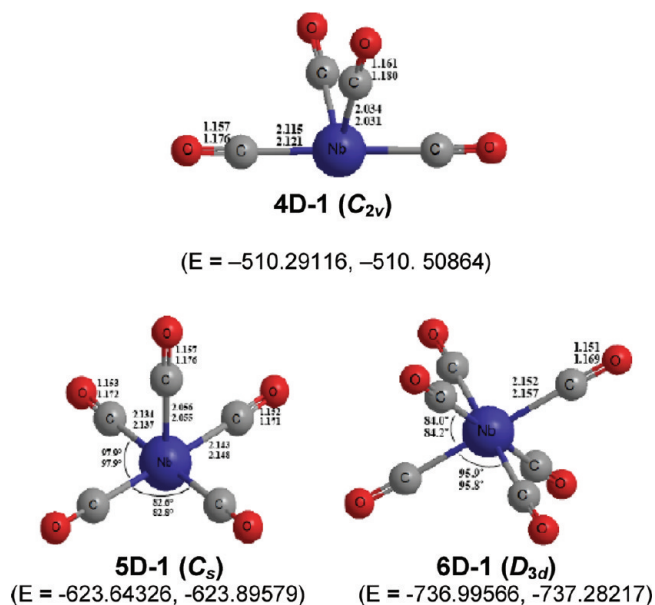


Figure 6. Optimized geometries for the mononuclear $\text{Nb}(\text{CO})_n$ ($n = 6, 5, 4$). Total energies predicted by the MPW1PW91 and BP86 methods, respectively, are listed under each structure.

small imaginary vibrational frequencies ($19i \text{ cm}^{-1}$ by MPW1PW91 or $10i \text{ cm}^{-1}$ by BP86), which might be considered to be negligible. However, following the corresponding normal mode reduces the symmetry from D_{3d} to D_2 and the energy by 8.7 kcal/mol (MPW1PW91) or 6.1 kcal/mol (BP86) to give structure **12S-2**. The collapse of **12S-3** to **12S-2** is accompanied by a shortening of the unusually long Nb–Nb bond of $\sim 3.62 \text{ \AA}$ in **12S-3** to a more normal length of $\sim 3.39 \text{ \AA}$ in **12S-2**. The $\text{Nb}_2(\text{CO})_{12}$ structure **12S-2** still has a residual imaginary vibrational frequency of $17i \text{ cm}^{-1}$ (MPW1PW91) or $19i \text{ cm}^{-1}$ (BP86). Following the corresponding normal mode reduces the symmetry from D_2 to C_2 to give **12S-1**. However, the energy changes in going from **12S-2** are minimal (<0.1 kcal/mol) and the structural changes very subtle. These structures, closely spaced in energy, suggest $\text{Nb}_2(\text{CO})_{12}$ to be a highly fluxional system.

The 18-electron rule requires $\text{Nb}_2(\text{CO})_{11}$ to have a formal Nb=Nb double bond if all of the carbonyl groups are the usual two-electron donors. However, the lowest energy $\text{Nb}_2(\text{CO})_{11}$ structure by ~ 12 kcal/mol, namely, **11S-1** (Figure 2), has a relatively long Nb–Nb distance of $\sim 3.5 \text{ \AA}$ and a four-electron donor bridging $\eta^2-\mu$ -CO group. A closely related $\text{V}_2(\text{CO})_{10}$ ($\eta^2-\mu$ -CO) structure is predicted⁴ to be the lowest energy structure for $\text{V}_2(\text{CO})_{11}$, indicating the similarity of the vanadium and niobium carbonyl systems in this respect. However, $\text{Nb}_2(\text{CO})_{11}$ is found to be disfavored with respect to disproportionation to give $\text{Nb}_2(\text{CO})_{12} + \text{Nb}_2(\text{CO})_{10}$ (Table 12).

The lowest energy $\text{Nb}_2(\text{CO})_{10}$ structure, namely, **10S-1** (Figure 4), has an unsymmetrical distribution of carbonyl groups and a short Nb=Nb distance indicative of a formal triple bond. A closely related structure is predicted⁴ to be the lowest energy structure of $\text{V}_2(\text{CO})_{10}$. Also, if the isolobal nature of $\text{Nb}(\text{CO})_3$ and CpMo moieties is recognized, where Cp is a pentahapto cyclopentadienyl ring, then $\text{Nb}_2(\text{CO})_{10}$ is isolobal with the stable $\text{Cp}_2\text{Mo}_2(\text{CO})_4$, obtained by thermal reactions of $\text{Mo}(\text{CO})_6$ with cyclopentadiene derivatives.^{53,54} Thermochemistry suggests $\text{Nb}_2(\text{CO})_{10}$ to be a viable species. Thus, the CO dissociation

Table 14. Infrared $\nu(\text{CO})$ Vibrational Frequencies (cm^{-1}) Predicted for $\text{Nb}(\text{CO})_6$ (Infrared Intensities Are in Parentheses, in km/mol)^a

$\text{Nb}(\text{CO})_6$	MPW1PW91	BP86
16S-1 (D_{3d})	2082 (0), 2084 (5035), 2085 (1784), 2196 (0)	1959 (0), 1961 (4086), 1961 (1496), 2065 (0)

^aThe IR intensities are doubled for the doubly degenerate modes.

energy of $\text{Nb}_2(\text{CO})_{10}$ is reasonably high at ~ 37 kcal/mol, which is essentially the same as that of the very stable $\text{Cr}(\text{CO})_6$.⁵² Furthermore, the energy required for disproportionation of $\text{Nb}_2(\text{CO})_{10}$ into $\text{Nb}_2(\text{CO})_{11} + \text{Nb}_2(\text{CO})_9$ is also reasonably high at ~ 22 kcal/mol.

A second type of $\text{Nb}_2(\text{CO})_{10}$ structure is found in which a $\text{Nb}(\text{CO})_6$ unit acts as a chelating ligand toward a $\text{Nb}(\text{CO})_4$ unit by using the oxygen atoms of two of its carbonyl groups. The lowest energy such $\text{Nb}_2(\text{CO})_{10}$ structure, namely, **10S-5**, lies ~ 21 kcal/mol above the lowest energy $\text{Nb}\equiv\text{Nb}$ triply bonded $\text{Nb}_2(\text{CO})_{10}$ structure **10S-1** with only two-electron donor carbonyl groups (Figure 4). In the $\text{Nb}_2(\text{CO})_{10}$ structures with a chelating $\text{Nb}(\text{CO})_6$ "ligand", both niobium atoms are hexa-coordinate, not counting any niobium–niobium interactions. The involvement of the oxygen atoms of the chelating carbonyl groups of the $\text{Nb}(\text{CO})_6$ "ligand" in these $\text{Nb}_2(\text{CO})_{10}$ structures makes the chelating carbonyl groups four-electron donors so that relatively low Nb–Nb bond orders are required to give each niobium atom a suitable electronic configuration.

The neutral homoleptic 17-electron vanadium carbonyl $\text{V}(\text{CO})_6$ is synthesized by very mild oxidation of the 18-electron $\text{V}(\text{CO})_6^-$ anion.³ Such an oxidation is very delicate since $\text{V}(\text{CO})_6$ is very sensitive toward further oxidation and indeed is spontaneously flammable in air under ambient conditions. In fact, the standard synthesis of $\text{V}(\text{CO})_6$ involves acidification of $\text{V}(\text{CO})_6^-$, effectively using the proton as an oxidizing agent via an unstable $\text{HV}(\text{CO})_6$ intermediate. A homoleptic niobium carbonyl has not yet been obtained by an analogous oxidation of the known $\text{Nb}(\text{CO})_6^-$ anion, possibly because a suitably mild and selective oxidant has not yet been found. Studies by Ellis and co-workers¹¹ suggests that the use of coordinating solvents for $\text{Nb}(\text{CO})_6^-$ oxidations prevents the oxidation from stopping at the zerovalent niobium oxidation state. This is consistent with the synthesis of $[\text{Nb}_2(\mu\text{-X})_3(\text{CO})_8]^-$ ($\text{X} = \text{Cl}, \text{Br}, \text{I}$) performed by Calderazzo and Pampaloni by the oxidation of $\text{Nb}(\text{CO})_6^-$ in the presence of halide sources.¹³ Thus, possible mild oxidation reactions of $\text{Nb}(\text{CO})_6^-$ to neutral zerovalent homoleptic niobium carbonyl derivatives must be limited to noncoordinating solvents.

The theoretical study reported here suggests that the neutral homoleptic niobium carbonyls potentially obtainable by oxidation of $\text{Nb}(\text{CO})_6^-$ are likely to be singlet binuclear derivatives with the favored 18-electron configuration for the niobium atoms rather than mononuclear $\text{Nb}(\text{CO})_6$ analogous to the known $\text{V}(\text{CO})_6$.³ The most obvious such oxidation product would be $\text{Nb}_2(\text{CO})_{12}$ with a formal Nb–Nb single bond and total niobium coordination numbers of seven, e.g., structure **12S-1** (Figure 1). However, the relatively low CO dissociation energies of $\text{Nb}_2(\text{CO})_{12}$ and $\text{Nb}_2(\text{CO})_{11}$ as well as the instability of $\text{Nb}_2(\text{CO})_{11}$ toward disproportionation into $\text{Nb}_2(\text{CO})_{12} + \text{Nb}_2(\text{CO})_{10}$ suggests $\text{Nb}_2(\text{CO})_{10}$ as another neutral homoleptic niobium carbonyl as a possible oxidation product of $\text{Nb}(\text{CO})_6^-$. The lowest energy $\text{Nb}_2(\text{CO})_{10}$ structure **10S-1** (Figure 4) has a formal $\text{Nb}\equiv\text{Nb}$ triple bond and is related to the known^{53,54}

$\text{Cp}_2\text{Mo}_2(\text{CO})_4$ by replacement of the CpMo units in the latter structure with isoelectronic and isolobal $\text{Nb}(\text{CO})_3$ units. Also, there is some experimental evidence for the production of the analogous $\text{V}_2(\text{CO})_{10}$ from the excimer laser photolysis of $\text{V}(\text{CO})_6$ vapor at 308 nm.¹⁶

5. CONCLUSIONS

Our theoretical studies predict the binuclear unbridged $\text{Nb}_2(\text{CO})_{12}$ rather than the mononuclear $\text{Nb}(\text{CO})_6$ to be a possible oxidation product of the known $\text{Nb}(\text{CO})_6^-$ anion under conditions avoiding overoxidation or the formation of side products. Thus, the dissociation of $\text{Nb}_2(\text{CO})_{12}$ to $2\text{Nb}(\text{CO})_6$ is predicted to require an energy of ~ 13 kcal/mol. This binuclear $\text{Nb}_2(\text{CO})_{12}$ structure is particularly interesting since the niobium atoms are seven-coordinate, counting the Nb–Nb bond. The niobium coordination polyhedra in these structures can be approximated by capped octahedra.

The lowest energy $\text{Nb}_2(\text{CO})_{11}$ structure is predicted to have a formal four-electron donor bridging $\eta^2\text{-}\mu\text{-CO}$ group and a formal Nb–Nb single bond rather than only two-electron donor carbonyl groups and a formal $\text{Nb}=\text{Nb}$ double bond. The $\text{Nb}_2(\text{CO})_{11}$ structures with formal $\text{Nb}=\text{Nb}$ double bonds and exclusively two-electron donor carbonyl groups lie more than 13 kcal/mol above this $\text{Nb}_2(\text{CO})_{10}(\eta^2\text{-}\mu\text{-CO})$ structure. However, $\text{Nb}_2(\text{CO})_{11}$ is predicted to be thermodynamically disfavored, owing to disproportionation into $\text{Nb}_2(\text{CO})_{12} + \text{Nb}_2(\text{CO})_{10}$, a slightly exothermic process by ~ 4 kcal/mol.

The $\text{Nb}_2(\text{CO})_{10}$ structures with formal $\text{Nb}\equiv\text{Nb}$ triple bonds and all two-electron donor groups appear to be particularly favorable, as suggested by high CO dissociation energies and viability toward disproportionation. Such structures are isolobal with $\text{Cp}_2\text{Mo}_2(\text{CO})_4$, which was the first stable metal carbonyl to be discovered with a short metal–metal distance corresponding to a formal triple bond. Considerably higher energy $\text{Nb}_2(\text{CO})_{10}$ structures by more than 20 kcal/mol have two four-electron donor bridging carbonyl groups and long niobium–niobium distances. Such structures can be considered to consist of a chelating bidentate $\text{Nb}(\text{CO})_6$ "ligand" coordinating to a $\text{Nb}(\text{CO})_4$ unit through the two $\eta^2\text{-}\mu\text{-CO}$ groups.

■ ASSOCIATED CONTENT

S Supporting Information. Tables S1–S26. Harmonic vibrational frequencies and infrared intensities for $\text{Nb}_2(\text{CO})_{12}$ (4 structures), $\text{Nb}_2(\text{CO})_{11}$ (7 structures), and $\text{Nb}_2(\text{CO})_{10}$ (9 structures); Cartesian coordinates of the optimized structures; complete Gaussian reference (reference⁴¹). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: rbking@chem.uga.edu (R.B.K.); kelly.luo@126.com (Q.L.).

ACKNOWLEDGMENT

We are indebted to the 111 Project (B07012) and the National Natural Science Foundation (20873045) of China as well as the U.S. National Science Foundation (Grants CHE-0749868 and CHE-0716718) for support of this research.

REFERENCES

- (1) Hartwig, J. *Organotransition Metal Chemistry: From Bonding to Catalysis*; University Science Books: Sausalito, CA, 2009; pp 13–26; 219–243.
- (2) Crabtree, R. H. *The Organometallic Chemistry of the Transition Metals*, 5th ed.; Wiley: New York, 2009; pp 25–32.
- (3) Natta, G.; Ercoli, R.; Calderazzo, F.; Alberola, A.; Corradini, P.; Allegra, G. *Rend. Accad. Naz. Lincei* **1959**, *27*, 107.
- (4) Liu, Z.; Li, Q.; Xie, Y.; King, R. B.; Schaefer, H. F. *Inorg. Chem.* **2007**, *46*, 1803.
- (5) Ford, T. A.; Huber, H.; Klotzbücher, W.; Moskovits, M.; Ozin, G. A. *Inorg. Chem.* **1976**, *15*, 1666.
- (6) Chi, K. M.; Frerichs, S. R.; Philson, S. B.; Ellis, J. E. *J. Am. Chem. Soc.* **1988**, *110*, 303.
- (7) Busby, R.; Klotzbücher, W.; Ozin, G. A. *Inorg. Chem.* **1977**, *16*, 822.
- (8) Luo, Q.; Li, Q.-S.; Yu, Z. H.; Xie, Y.; King, R. B.; Schaefer, H. F. *J. Am. Chem. Soc.* **2008**, *130*, 7756.
- (9) Chi, K.-M.; Frerichs, S. R.; Philson, S. B.; Ellis, J. E. *Angew. Chem., Int. Ed.* **1987**, *26*, 1190.
- (10) Ellis, J. E.; Chi, K.-M. *J. Am. Chem. Soc.* **1990**, *112*, 6022.
- (11) Barybin, M. V.; Ellis, J. E.; Pomije, M. K.; Tinkham, M. L.; Warmock, G. F. *Inorg. Chem.* **1998**, *37*, 6518.
- (12) Ellis, J. E.; Warnock, G. F.; Barybin, M. V.; Pomije, M. K. *Chem.—Eur. J.* **1995**, *1*, 521.
- (13) Calderazzo, F.; Pampaloni, G. *Chem. Commun.* **1984**, 1249.
- (14) Zhou, M.; Andrews, L. J. *Phys. Chem. A* **1999**, *103*, 7785.
- (15) Rhicks, A. M.; Reed, Z. D.; Duncan, M. A. *J. Am. Chem. Soc.* **2009**, *131*, 9176.
- (16) Ishikawa, Y.; Hackett, P. A.; Rayner, D. M. *J. Am. Chem. Soc.* **1987**, *109*, 6644.
- (17) Ehlers, A. W.; Frenking, G. *J. Am. Chem. Soc.* **1994**, *116*, 1514.
- (18) Delley, B.; Wrinn, M.; Lüthi, H. P. *J. Chem. Phys.* **1994**, *100*, 5785.
- (19) Li, J.; Schreckenbach, G.; Ziegler, T. *J. Am. Chem. Soc.* **1995**, *117*, 486.
- (20) Jonas, V.; Thiel, W. *J. Chem. Phys.* **1995**, *102*, 8474.
- (21) Barckholtz, T. A.; Bursten, B. E. *J. Am. Chem. Soc.* **1998**, *120*, 1926.
- (22) Jemmis, E. D.; Giju, K. T. *J. Am. Chem. Soc.* **1998**, *120*, 6952.
- (23) Niu, S.; Hall, M. B. *Chem. Rev.* **2000**, *100*, 353.
- (24) Cotton, F. A.; Gruhn, N. E.; Gu, J.; Huang, P.; Lichtenberger, D. L.; Murillo, C. A.; Van Dorn, L. O.; Wilkinson, C. C. *Science* **2002**, *298*, 1971.
- (25) Macchi, P.; Sironi, A. *Coord. Chem. Rev.* **2003**, *238*, 383.
- (26) Siegbahn, P. E. M. *J. Am. Chem. Soc.* **2005**, *127*, 17303.
- (27) Ziegler, T.; Autschbach, J. *Chem. Rev.* **2005**, *105*, 2695.
- (28) Mota, A. J.; Dedieu, A.; Bour, C.; Suffert, J. *J. Am. Chem. Soc.* **2005**, *127*, 7171.
- (29) Bühl, M.; Kabrede, H. *J. Chem. Theory Comput.* **2006**, *2*, 1282.
- (30) Brynda, M.; Gagliardi, L.; Widmark, P. O.; Power, P. P.; Roos, B. O. *Angew. Chem., Int. Ed.* **2006**, *45*, 3804.
- (31) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *124*, 224105.
- (32) Strickland, N. S.; Harvey, J. N. *J. Phys. Chem. B* **2007**, *111*, 841.
- (33) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- (34) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.
- (35) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.
- (36) Zhao, S.; Wang, W.; Li, Z.; Liu, Z. P.; Fan, K.; Xie, Y.; Schaefer, H. F. *J. Chem. Phys.* **2006**, *124*, 184102.
- (37) Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1993**, *85*, 441.
- (38) Bergner, A.; Dolg, M.; Kuechle, W.; Stoll, H.; Preuss, H. *Mol. Phys.* **1993**, *80*, 1431.
- (39) Dunning, T. H. *J. Chem. Phys.* **1970**, *53*, 2823.
- (40) Huzinaga, S. *J. Chem. Phys.* **1965**, *42*, 1293.
- (41) (a) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004. (b) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
- (42) Xie, Y.; Schaefer, H. F.; King, R. B. *J. Am. Chem. Soc.* **2000**, *122*, 8746.
- (43) Silaghi-Dumitrescu, I.; Bitterwolf, T. E.; King, R. B. *J. Am. Chem. Soc.* **2006**, *128*, 5342.
- (44) Caspar, J. V.; Meyer, T. J. *J. Am. Chem. Soc.* **1980**, *102*, 7794.
- (45) Hooker, R. H.; Mahmoud, K. A.; Rest, A. *J. Chem. Commun.* **1983**, 1022.
- (46) Hepp, A. F.; Blaha, J. P.; Lewis, C.; Wrighton, M. S. *Organometallics* **1984**, *3*, 174.
- (47) Herrmann, W. A.; Biersack, H.; Ziegler, M. L.; Weidenhammer, K.; Siegel, R.; Rehder, D. *J. Am. Chem. Soc.* **1981**, *103*, 1692.
- (48) Bin, P.; Li, Q.-S.; Xie, Y.; King, R. B.; Schaefer, H. F. *Dalton Trans.* **2009**, 3748.
- (49) Zhang, X.; Li, Q.-S.; Xie, Y.; King, R. B.; Schaefer, H. F. *Organometallics* **2009**, *28*, 6410.
- (50) Weinhold, F.; Landis, C. R. *Valency and Bonding: A Natural Bond Order Donor-Acceptor Perspective*; Cambridge University Press: Cambridge, England, 2005; pp 32–36.
- (51) Wang, H.; Xie, Y.; King, R. B.; Schaefer, H. F. *J. Am. Chem. Soc.* **2006**, *128*, 11376.
- (52) Sunderlin, L. S.; Wang, D.; Squires, R. R. *J. Am. Chem. Soc.* **1993**, *115*, 12060.
- (53) King, R. B.; Bisnette, M. B. *J. Organomet. Chem.* **1967**, *8*, 129.
- (54) Huang, J. S.; Dahl, L. F. *J. Organomet. Chem.* **1983**, *243*, 57.

A New Efficient Method for Generating Conformations of Unfolded Proteins with Diverse Main-Chain Dihedral-Angle Distributions

Yasutaka Seki,[†] Yudai Shimbo,^{‡,§} Takamasa Nonaka,[†] and Kunitsugu Soda^{*,§}

[†]School of Pharmacy, Iwate Medical University, Yahaba, Iwate 028-3694, Japan

[‡]Department of Bioengineering, Nagaoka University of Technology, Nagaoka, Niigata 940-2188, Japan

[§]Laboratory for Computational Molecular Design, Center for Computational Life Science, RIKEN, Kobe, Hyogo 650-0047, Japan

ABSTRACT: A new method for generating polypeptide-chain conformations has been developed for studying structural characteristics of unfolded proteins. It enables us to generate a large number of conformations very rapidly by avoiding atomic collisions efficiently with the use of main-chain dihedral-angle distributions derived from a crystal-structure database of proteins. In addition, combining main-chain dihedral-angle distributions for the amino acid residues incorporated in different secondary structures, we can obtain diverse conformational ensembles with different structural features. Structural characteristics of proteins denatured in high-concentration denaturant solution were analyzed by comparing predictions from this method with results from solution X-ray scattering (SXS) measurement. Analysis of the dependence of the mean square radius (R_{sq}) of protein on the number of residues and the shape of its Kratky profile has confirmed that the highly denaturing solvent serves as a good solvent in accordance with previous reports. It was also found that, in order for a conformational ensemble to reproduce experimental data, the percentage in which main-chain dihedral angles are found in the α region must be in the range of 20–40%. It agrees with studies on the $^3J_{HN\alpha}$ coupling constant using the multidimensional NMR method. These results confirm that our method for generating diverse conformations of polypeptide chains is very useful to the conformational analysis of unfolded protein, because it enables us to analyze comprehensively both of the local structural features obtained from NMR and the global ones obtained from SXS.

INTRODUCTION

The detailed molecular mechanism of protein folding and the physical design principle of the 3D structure of protein have been the most important targets to be studied in protein science. In order to solve the problem, it is necessary to clarify structural characteristics of protein in the completely or partially unfolded state as well as those in the natively folded state, because the non-native state appears in the initial or intermediate state of folding processes.^{1,2} In recent years, attention has been given to the function and its expression mechanism of proteins that are natively unfolded or have an intrinsically disordered region.³ Presumably, these proteins do not have a definite conformation but have diverse ones in their isolated state not bound to another protein and/or a ligand molecule. Hence, it is necessary to analyze detailed structural characteristics of unfolded proteins also for elucidating the molecular mechanism of expressing their functions. Under these circumstances, studies for analyzing structural features of proteins in the unfolded state have become increasingly important.^{4,5}

Solution X-ray and neutron scatterings (SXS and SNS) and spectroscopic techniques such as multidimensional NMR are useful methods that give mutually complementary information on the structure of unfolded proteins in solution. Application of the SXS method to a protein in solution yields information of its global structure such as the mean square radius (R_{sq}) from a Guinier analysis, the molecular shape from a Kratky plot, and the distance distribution function.² Especially for unfolded proteins, the statistical feature of an ensemble of chain-like conformations has been analyzed on the basis of the scaling law that characterizes the dependence of R_{sq} on the number of residues, N_r .^{6,7}

From analyses of SXS data, many studies have yielded an exponent of about 0.6 on the scaling law for the denatured state of proteins in high-concentration denaturant solution.^{6,7} It is assumed from comparison of the exponents estimated theoretically and experimentally that the conformational characteristic of the denatured state of a protein under highly denaturing conditions can be well approximated by a random-flight chain with a finite excluded-volume effect. Kratky-profile analysis also supports qualitatively the view that the conformation of a highly denatured protein can be virtually described by a random-flight chain.^{8,9} Wang et al.¹⁰ took into account the solvation effect in their method of modeling unfolded protein structures by assuming that the solvation energy of each conformation is proportional to its accessible surface area (ASA). They showed that the ensemble of conformations generated by their method reproduces well the experimental scaling-law exponent and SXS profile. The scaling-law analysis was also applied to the intermediate observed experimentally at the initial stage of the refolding process, which indicated that the statistical-structure analysis of chain molecules is useful for examining the molecular mechanism of protein folding.¹¹ On the other hand, the probability distribution function was estimated for the main-chain dihedral angle φ in the fully unfolded state from measurements of the $^3J_{HN\alpha}$ coupling constant by multidimensional NMR.^{12,13} In addition, the content of residual secondary structures was evaluated from chemical shifts of proteins in the partially unfolded state¹⁴ or peptides.¹⁵ The information on these local

Received: December 9, 2010

Published: June 02, 2011

structures is complementary to the information on global structures obtained from the SXS method. In recent years, NMR and CD (circular dichroism) measurements revealed that short peptides can form a polyproline II helix in aqueous solution.^{16,17} It was also found that residual dipolar couplings (RDCs) of NMR for staphylococcal nuclease,¹⁸ apomyoglobin,¹⁹ and some other proteins²⁰ under high-concentration denaturant conditions show amino acid sequence dependence clearly different from that of a simple bell-shaped form from the calculation for a homopolypeptide.²⁰ Initially, controversial interpretations were presented for these RDC data. However, it is now accepted²¹ that (a) proteins under such conditions basically form no specific residual structure such as secondary structures and (b) unique RDC data result from the fact that the occurrence probability of main-chain dihedral angles varies with amino acid species and sequences.

We can see from the above that it is essential to elucidate characteristics of unfolded-protein structures consistent with the information on both of the local and global structures. So, it would be very effective to use a molecular modeling method where both kinds of information can be accurately taken into account with a high-performance computer. Several analyses have already been made from a viewpoint such as above. Previous works^{22,23} revealed that the R_{sq} of unfolded proteins with residual secondary structures shows dependence on N_r , similar to that of completely denatured proteins under high-concentration denaturant conditions. Jha et al.²⁰ reproduced experimental data of RDCs and the exponent of the dependence of R_{sq} on N_r , but the calculated R_{sq} differs significantly from the experimental one. The analysis¹⁰ reproduced well experimental SXS data, but no comparative study with local structural analysis as NMR has been performed yet. Thus, no molecular modeling analysis of unfolded proteins in which both of the global SXS properties and the local NMR ones are reproduced has been carried out yet. One of the reasons for the difficulty of such an analysis is that conventional methods have not yet enabled us to generate a sufficiently large number of model structures with both kinds of information and reasonable conformational energies being incorporated.

To overcome the problem, we developed a new method for rapidly generating model structures of unfolded protein, where polypeptide conformations are produced by using various probability distribution functions for main-chain dihedral angles determined with the crystal structures of native globular proteins. Using this method, we can generate the structure of unfolded protein very rapidly. Moreover, in most cases, it is possible to reach a nearby local minimum-energy structure after a small number of computational steps, because interatomic collisions are eliminated in all of the initially generated structures. Diverse ensembles of polypeptide conformations with different structural characteristics can be generated by combining various main-chain dihedral-angle distribution functions derived from the ensembles of residues incorporated in different local structures of native proteins. We carried out an SXS analysis of the ensemble of conformations generated by using this method and considering the solvent effect in the ASA approximation. From our analysis, we could elucidate the relations among the scaling-law exponent of R_{sq} , the SXS profile, the distribution function of main-chain dihedral angles, and the solvent effect. On the basis of the results, it is demonstrated that detailed physical properties of unfolded protein can be analyzed by combining our modeling method with SXS measurement.

METHODS

Method for Generating Conformations of Unfolded Proteins. A new computer program was developed for generating conformations of unfolded proteins. This program is written in Fortran90 on a Linux computer. It accepts protein–structure data with any of the formats of Protein Data Bank, AMBER,²⁴ and GROMACS,²⁵ while it reads in the structural parameters such as bond lengths and bond angles from an input structure file. In this study, we use structural data with the GROMACS format and structural parameters of the AMBER99²⁶ package. Atomic collisions are decided using the set of atomic radii from Tsai et al.²⁷ Generated conformations are energy-minimized using a package in GROMACS 4.0²⁸ and the force field of AMBER 99. Solvent is implicitly considered by assuming its dielectric constant to be that of liquid water in evaluating conformational energies. The bond length of the hydrogen atom is held constant using the LINCS algorithm. All of the van der Waals and Coulomb interactions were evaluated within a cutoff distance of 1.2 nm. The steepest descent method was used for energy minimization.

We analyzed 12 proteins: horse cytochrome *c* (104 aa), bovine α -lactalbumin (123), bovine ribonuclease A (124), hen egg white lysozyme (129), sperm whale myoglobin (153), avian sarcoma virus integrase core (162), dihydrofolate reductase (191), MutY catalytic domain (225), triosephosphate isomerase (250), *EcoRI* endonuclease (276), UDP-galactose 4-epimerase (338), and creatine kinase (379). It is assumed here that cytochrome *c* and myoglobin are of the apo form and all proteins are free of disulfide bonds, because the covalent structures of the same protein must be the same between computational and experimental⁷ analyses to compare the results on the dependence of R_{sq} on the number of residues, N_r .

Determination of the Dihedral-Angle Distributions for Main and Side Chains. The main-chain dihedral-angle distributions (MCDAD) necessary for generating unfolded protein conformations were determined through the following procedure: First, the 379 native structures satisfying the three requirements below are chosen from 526 folds in the *Dali Domain Dictionary 2.0*:^{29,30} (a) The number of residues is larger than 50. (b) The resolution of X-ray structure analysis is better than 0.24 nm. (c) The probability of occurrence of main-chain dihedral angles in the “core region³¹” is higher than 0.94. Next, all of the main-chain dihedral angles are calculated, and the secondary structures in all of the structures of target proteins are determined by using the DSSP algorithm.³² Resulting data of the main-chain dihedral angles are classified by amino acid species and secondary-structure types to obtain the six MCDAD models denoted as (all), ($\alpha + t + c$), ($\beta + t + c$), ($t + c$), ($\beta + c$) and (c) or (coil) for each of the 20 residues. Here, the four letters, i.e. α , β , t and c (or coil), refer to all of the helices (H, G, and I in DSSP), β strands (B and E in DSSP), turns with a H bond (T and S in DSSP), and coils other than those above (all except secondary structures in DSSP). For example, the ($\beta + t + c$) model denotes a MCDAD model constructed from the MCDADs of the residues not incorporated in any helical structure. All MCDAD models are assumed to contain the coil model, and all of the other combinations are taken for analysis, though the ($\alpha + \beta + c$) and ($\alpha + c$) models are excluded because they show MCDADs very similar to those of the (all) and ($\alpha + t + c$) models, respectively.

As an example, the 3D plot of MCDAD for the (all) model is shown in Figure 1. Around the peak of distribution in the

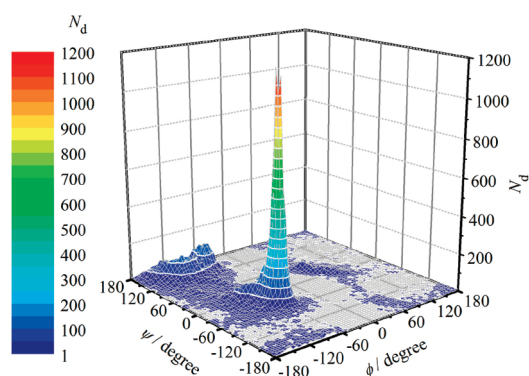


Figure 1. Example of the 3D distribution of occurrence frequency on the (φ, ψ) plane for a MCDAD model, (all). A total of 55 012 sets of (φ, ψ) data are obtained from 379 native-protein structures. All pairs of (φ, ψ) data are sorted with a division of $\Delta\varphi = \Delta\psi = 5^\circ$, and the number of data, N_d , in each unit area is 3D plotted. The color of the 3D surface is changed with a division of $\Delta N_d = 50$, while the areas with $N_d = 0$ are not colored.

α region with the highest occurrence frequency, the number of occurrences in the unit area with $\Delta\varphi = \Delta\psi = 5^\circ$ exceeds 1000. Hence, the relative standard error of the number of occurrences in this area is 3.2%, showing high accuracy in this region. The definition of the α region as well as those of β and p regions are given in the literature.³³ On the other hand, the number of occurrences is only slightly larger than 100 around the peak of distribution in the β or p region, indicating about 10% error or 3 times lower accuracy than that in the α region. However, the occurrence frequency in the β or p region is more widely distributed than that in the α region. The percentage with its standard error that a pair of main-chain dihedral angles occurs in either the β or p region is $43.6 \pm 0.3\%$. It is comparable with that for the α region, $44.0 \pm 0.3\%$. For the other regions, the number of occurrences in a unit area is smaller than 50, and the relative error is still larger. However, the total occurrence probability for these regions is not so high. For example, the total percentage that a pair of dihedral angles occurs in some area having a number of occurrences lower than 10 is only 11%. Thus, MCDADs obtained in this study acquire enough accuracy to consider the probability of occurrence in the α region or the β and p regions.

The side-chain dihedral angles are chosen as follows: The dihedral angle of rotation around a bond between two sp^3 atoms is randomly chosen at the three angles of 60, 180, and 300° with a range of $\pm 15^\circ$. The dihedral angle for sp^3 and sp^2 atoms is randomly chosen from the whole range of $0-360^\circ$. That for two sp^2 atoms is fixed to their optimum value. Exceptionally, those of χ_2 , χ_3 , and χ_4 for lysine and χ_2 , χ_3 , and χ_5 for arginine are randomly chosen within a range of $180 \pm 15^\circ$.

Calculation of the SXS Profile, R_{sq} , and ASA. The SXS profile of a protein molecule is calculated from the atomic coordinates derived from its generated conformation using the following equation:

$$I(K) = \sum_{i=1}^{N_d} \{f_i(K)\}^2 + 2 \times \sum_{i>j} f_i(K) f_j(K) \frac{\sin Kr_{ij}}{Kr_{ij}} \quad (1)$$

where K is the magnitude of the scattering vector, $f_i(K)$ is the scattering factor of atom i , r_{ij} is the distance between atoms i and j , and N_d is the number of atoms. Strictly, eq 1 is an expression for calculating the SXS profile of a molecule in a vacuum. An experimental SXS profile includes both the contrast effect of

the solvent with finite electron density and the hydration effect due to differences in spatial distribution between hydration water and bulk water. As these effects of solvent water are not taken into account in eq 1, this formula cannot be applied to compare quantitatively with an experimental profile in the range of K where they have significant contributions. However, we could confirm that, for a protein with unfolded conformations generated in this study, the SXS profiles calculated by using eq 1 virtually agree with those obtained when including the two solvent effects^{34,35} in the range of $K < 1.5 \text{ nm}^{-1}$ (data not shown). Hence, we can assume that eq 1 gives a good approximation for the SXS profile of unfolded protein. In this study, it is essential to calculate SXS profiles for protein conformations of a large number of 1 million per model. So, we limited the K range of analysis to the above and adopted eq 1, which requires the shortest time for estimating SXS profiles. For the same reason, the mean square radius (R_{sq}) is calculated using the following equation:

$$R_{sq} = \sqrt{\frac{\sum_{i=1}^N n_{e,i} r_i^2}{\sum_{i=1}^N n_{e,i}}} \quad (2)$$

where $n_{e,i}$ is the number of electrons in atom i and r_i is the position vector of atom i from the center of gravity of the protein.

The accessible surface area (ASA) of the protein is calculated analytically³⁶ using an originally developed program, "cgp". Details on the program cgp will be published elsewhere. We employed the set of atomic radii by Tsai et al.²⁷ and assumed the radius of probe water to be 0.14 nm.

RESULTS

Development of the Program for Generating Conformations of Unfolded Proteins. Previous methods for generating conformations of unfolded proteins are roughly classified into two types by the way of choosing its main-chain dihedral angles. In one method, main-chain dihedral angles are randomly chosen from the whole region of (φ, ψ) .^{37,38} After that, a possible high-energy conformation due to atomic collisions is relaxed and then brought into energy minimization. In the other method, some main-chain dihedral-angle distribution (MCDAD) derived from the database of native structures is utilized for generating unfolded conformations.²⁰ The former is inefficient for generating conformations, because the structural relaxation and energy minimization of a conformation with many atomic collisions requires a fairly long computational time, and its conformational energy often does not converge. The latter enables us to effectively obtain conformations with no intraresidue atomic collisions by excluding pairs of φ and ψ that lead to atomic collisions such as a pair of nearly zero φ and ψ values. As a result, the energy minimization in the latter is carried out far more easily compared with the former, because the generated conformation can have only inter-residue atomic collisions. Taking these into account, we have employed the following criteria for our new method of generating unfolded-protein conformations: (1) The all-atom model is adopted where all atomic species are explicitly considered. (2) All of the bond lengths and bond angles are fixed to their respective optimum values. (3) The occurrence probability of a main-chain dihedral angle is assumed to be given by a MCDAD derived from an ensemble of native-protein structures, except that all ω 's are fixed to 180° and the φ of proline is fixed to -75° . (4) The side-chain dihedral angle is randomly chosen from the conformations near its stereochemically most stable structure (see the Methods

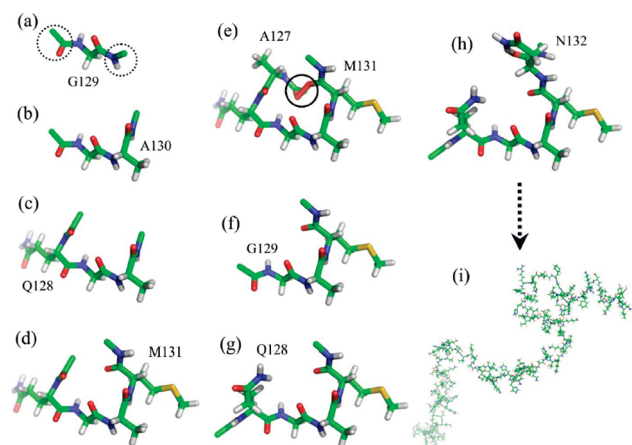


Figure 2. Algorithm for generating unfolded protein structures. An example for apomyoglobin is shown in Figure 2a–i. (a) G129 is chosen as the first residue. The coordinates of main-chain atoms of the residues neighboring the first residue, i.e., C_{α} , C, and O on the N-terminal side and N, H, and C_{α} on the C-terminal side, are calculated simultaneously, as indicated by dotted circles. These atoms determine the dihedral angles of rotation, ω , around the peptide bonds on both sides. They are generated also when the neighboring residues are generated. The neighboring residues are connected by superimposing the doubly generated atoms. (b) A130 is chosen as the next residue and added to the first residue. (c and d) Q128 and M131 are added at steps of c and d, respectively. (e) The addition of A127 has failed due to the atomic collision between the two main-chain carbonyl oxygen atoms of D126 and M131, as indicated by a circle. If the number of regeneration trials exceeds a limit, the process is restarted with returning to the step before adding several residues. In this study, the former limit is set at 1000 trials, and the latter limit is set at three residues. We confirmed that changes in the two technical parameters affect the rate of generating conformations but have practically no effect on the structural features of generated conformations. As the number of regeneration trials reached 1000 in the case of e, G129, Q128, and A127 were deleted, as shown in f. (f, g, and h) The conformations of G129 and Q128 are regenerated, and N132 is added at the respective steps. (i) The resultant whole structure is shown.

section for details). (5) All atomic collisions are eliminated between the two non-hydrogen atoms separated from each other by more than three chemical bonds. Criteria 1 and 2 are reasonable and have also been employed in many previous methods. We propose a new algorithm for eliminating atomic collisions to satisfy criterion 5.

The details of our algorithm are described below, and some examples of generated conformations are shown in Figure 2a–i. First, the first residue is randomly chosen from among all of the residues in the given sequence with equal probability. Next, main- and side-chain dihedral angles of the first residue are chosen by using the method described above. Then, the Cartesian coordinates of atoms are calculated from these dihedral angles, bond lengths, and angles. Here, the conformation of the first residue is examined for atomic collisions. In the case of some atomic collisions, the step of choosing dihedral angles is repeated again. An example of the conformation thus obtained is shown in Figure 2a. Next, a residue on either of the N- and C-terminal sides is chosen for adding to the first residue. Coordinates of the atoms in the added residue are calculated and examined for intraresidue atomic collisions. The generated residue is connected to the first residue so as to form a peptide bond (see the caption of Figure 2). Possible atomic collisions between the added and first residues are examined. If no atomic collision is found, we move on to

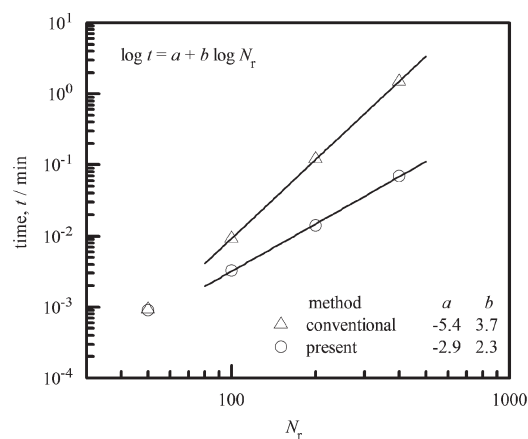


Figure 3. Comparison of dependences of the mean time, t , required for generating one conformation of an unfolded alanine-homopolypeptide on the number of residues, N_r , between the present method (\circ) and the conventional one (Δ). A function of the form of $\log t = a + b \log N_r$ is least-squares fitted to each set of data for $N_r = 100, 200,$ and 400 . The values of the parameters obtained are shown in the figure. The standard errors of b are 0.02 and 0.05 for the conventional and present methods, respectively. R^2 is greater than 0.999 for both methods. MCDAD model, “all”; CPU, Intel Xeon processor (2.8 GHz).

choosing a new residue to be added. As shown in Figure 2b–d, the peptide chain is grown alternately toward either the N-terminal or C-terminal side. When the newly added residue is either the N- or C-terminal residue of the protein, the next residue is naturally chosen so as to grow the chain only in the direction of the segment not yet generated. In this way, the conformation of a polypeptide chain is generated so that it can grow toward both termini. As shown in Figure 2e, when there are atomic collisions between the added residue and the existing chain, the process is restarted from the step of choosing dihedral angles of the added residue. In the case when atomic collisions can never be eliminated by the above means, the process is restarted with returning to the step before adding several residues. Eventually, as shown in Figure 2i, the conformation of the unfolded protein is obtained, and it meets the five criteria described above.

A unique feature of the present algorithm is that the peptide chain grows by residue from an arbitrarily chosen residue. Another feature is that only a small number of residues are remodeled when encountering atomic collisions. As a result, it is expected that the computational cost for examining atomic collisions and remodeling chain conformations is minimized. To test the performance of our method, we examined how much the computational time for generating conformations of unfolded protein is decreased by using this algorithm. A conventional method not using the growing-chain algorithm was adopted as a reference method. Specifically, the mean computational times were compared between the cases where the respective methods were employed for generating unfolded chains of polyalanine with four different chain lengths of $N_r = 50, 100, 200,$ and 400 . In the conventional method, the process of eliminating atomic collisions is carried out after main-chain dihedral angles of the whole chain are chosen. Except for this point, there is no distinction between the two methods. The results obtained are shown in Figure 3. The two computational times are nearly the same for a chain of 50 residues. The computational time with our method, however, is much shorter than that with the conventional one for the other longer chains: Those with the

Table 1. Fractions (%) of the α and β Regions in MCDADs from the Database and Generated Conformations

MCDAD	α region ^a /%		β region ^a /%	
	database ^b	model ^c	database ^b	model ^c
($\alpha + t + c$)	55.1	48.8	30.0	36.3
(all) or ($\alpha + \beta + t + c$)	44.0	37.5	43.6	50.5
($t + c$)	24.4	24.0	52.2	57.0
($\beta + t + c$)	17.3	16.8	65.5	69.1
(coil) or (c)	6.9	7.6	79.5	80.1
($\beta + c$)	4.5	4.7	86.9	87.2

^a The definition of the (φ, ψ) regions is given in ref 33. The β region in this paper is a combination of the β and p regions in ref 33. ^b These are estimated from 55 012 sets of (φ, ψ) values obtained from 379 native-protein structures. ^c These are estimated from all of the conformations generated with 12 unfolded proteins. One million conformations are generated for each set of a MCDAD model and a protein species.

former and the latter increase with an increase in the number of residues, N_r , with exponents of 2.2 and 3.6, respectively. For example, generating 1 million conformations of 400-residue polyalanine with the conventional method requires a 20 times longer computational time than that with our method. Hence, the difference in the process avoiding atomic collisions results in a great difference in the speed of generating conformations.

We confirmed that (a) almost all of the potential energies of generated conformations reach their local minima after steps smaller than 500 and (b) the average difference in conformation between those before and after energy minimization is smaller than 0.025 nm in RMSD. These small conformational changes lead to practically no significant changes in the conformational parameters such as R_{sq} and ASA in this study. Hence, we analyzed the conformation of unfolded proteins, skipping the step of energy minimization. Consequently, our newly developed method can generate reasonable conformations without energy minimization. In conclusion, it enables us to generate a large number of conformations of an unfolded protein with reasonable potential energy very rapidly by eliminating atomic collisions efficiently.

Structural Features of Unfolded Proteins. In the following, we will call each of the different MCDADs a MCDAD model. To analyze the structural features of chains derived from different MCDAD models, we generated 1 million unfolded conformations for each of the 12 proteins with different N_r 's using six MCDAD models. As a result, we obtained 72 different ensembles, each of which consists of 1 million unfolded conformations. The ensemble of conformations generated with a MCDAD model will hereafter be called briefly the MCDAD ensemble. The fractions of α and β regions in each MCDAD and those for the ensemble of conformations generated from 12 unfolded proteins are shown in Table 1. Here, the α -region fraction is defined as the probability that main-chain dihedral angles are chosen in the α region.³³ The β -region fraction is defined similarly to the α -region fraction, but the β region in this study is defined as the sum of the β and p regions in the literature.³³ The α -region fraction for generated unfolded proteins is hereafter denoted as η_α (in percent). Each of the generated MCDADs reproduces well the α -region fraction for the corresponding model derived from PDB except for the ($\alpha + t + c$) and (all) models. These two MCDADs having a high η_α give about a 6% lower α -region fraction than those from the database. Nevertheless, they have a significantly higher η_α 's than the other

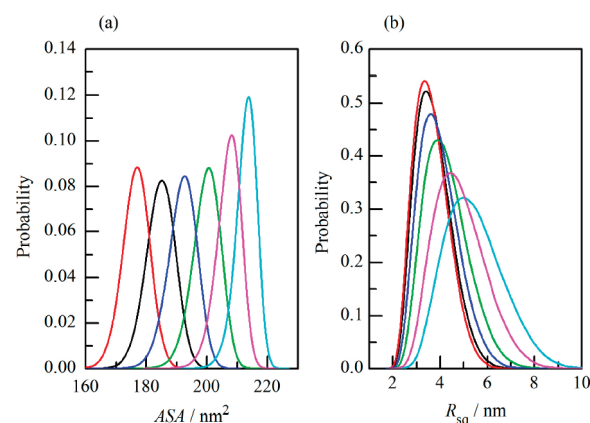


Figure 4. Probability distribution of the (a) ASA and (b) R_{sq} of unfolded apomyoglobin. MCDAD model: red, ($\alpha + t + c$); black, (all); blue, ($t + c$); green, ($\beta + t + c$); purple, (coil); light blue, ($\beta + c$). Each of the six MCDAD ensembles consists of 1 million conformations. Division of data: (a) $\Delta ASA = 0 \text{ nm}^2$, (b) $\Delta R_{sq} = 0.2 \text{ nm}$. Estimated error of probability density: (a) 3×10^{-4} , (b) 2×10^{-3} . The curves in the figure are drawn using the cubic spline interpolation.

MCDAD models for structural analysis. Such consequences for these two models are caused by the fact that atomic collisions occur more frequently when main-chain dihedral angles are chosen in the α region than in the other regions. As a result, the number of dihedral angles chosen from the other regions increases for avoiding atomic collisions. To cope with this problem, an algorithm for further exploring a dihedral angle without atomic collisions in the α region must be introduced into our method. There is a trade-off between the effectiveness of this means and the speed of generating conformations. In this study, we have taken no precautions to meet the problem, taking account of the speed of generating conformations. To minimize the effect of this problem, not the α -region fraction from the database but the value of η_α from the generated ensemble is used for comparing calculations with experimental results. Thus, the value of η_α obtained from analysis correctly reflects structural features of target proteins. The ASA (accessible surface area), R_{sq} (mean square radius), and SXS (solution X-ray scattering) profile were calculated for all of the conformations in each MCDAD ensemble.

To see structural features of the generated unfolded conformations, probability distributions of the ASA and R_{sq} in the respective MCDAD ensembles for unfolded apomyoglobin are shown in Figure 4.

We can see from Figure 4a that the bell-shaped ASA distributions for different models have their respective different peak positions, and their tails mutually overlap with those of nearby models. The values of ASA for all of the MCDAD models are distributed all over the range of 170–220 nm^2 . The ASA of fully extended apomyoglobin is 213 nm^2 , which is found in its range of the ($\beta + c$) ensemble. To compare our estimate with the previous one for the ASA of unfolded protein, we evaluated the ASA of unfolded horse apomyoglobin using the method of estimating its upper and lower limits described by Creamer et al.^{39,40} It yielded estimates for the lower and upper limits of 158 nm^2 and 203 nm^2 , respectively. This lower limit is found in the minimum ASA region of the ($\alpha + t + c$) ensemble that gives the smallest ASA among our conformational ensembles. Considering that the lower limit is estimated from the secondary-structure segment of the crystal structure in their method, we can

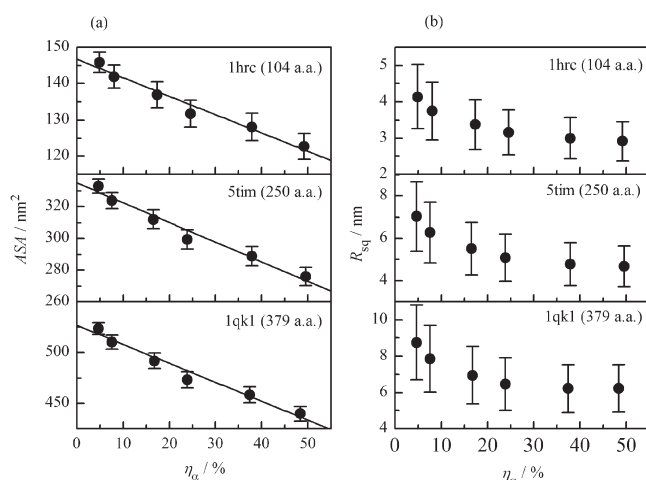


Figure 5. Dependences of the (a) mean ASA and (b) mean R_{sq} on the α -region fraction, η_{α} , of the protein. A least-squares fit of a linear function, $ASA(\eta_{\alpha}) = ASA(0) - a \times \eta_{\alpha}$, is made on each data point of a. The two obtained parameters, $ASA(0)$ and a , their standard errors, and R^2 are listed in Table 2.

safely assume that the unfolded conformation generated by us gives the lower limit of ASA for the unfolded protein having no local structure such as secondary structures. On the other hand, we can see that our ensemble covers broader conformational space than theirs, as the upper limit is found near the peak of the (coil) ensemble.

Evidently, the R_{sq} also covers conformational space extensively, as seen from its distribution shown in Figure 4b. However, the dependence on the MCDAD ensemble of the R_{sq} distribution differs greatly from that of the ASA distribution: All of the MCDAD ensembles have a common range of R_{sq} between 3 and 6 nm, and with an increase in the R_{sq} at the peak position, the peak height of the R_{sq} distribution decreases and the width of distribution increases. It is also notable that the two R_{sq} distributions for the $(\alpha + t + c)$ and the (all) ensembles are very similar to each other, though their ASA distributions are completely separated. It is strongly suggested that the lower limit of R_{sq} for unfolded apomyoglobin is near 2 nm because all of the ensembles show almost the same lower limit of about 2 nm.

To see the relation of the global-structure parameters of ASA and R_{sq} to the MCDAD in more detail, dependences of the mean ASA and R_{sq} for the six MCDAD ensembles on their α -region fraction, η_{α} , are shown in Figure 5. As seen from Figure 5a, the mean ASA decreases monotonically with an increase in η_{α} . Applying the least-squares fit of a linear function to this plot, we can see that the standard errors of the fitting parameters for each protein are sufficiently small, as shown in the caption of Figure 5. It confirms that the dependence of the mean ASA on η_{α} is well approximated by a linear function. In addition, the slope of this line, i.e., the reduction rate of ASA with increasing η_{α} , is well proportional to N_r , and the determination coefficient, R^2 , obtained from the least-squares fit of this relation is greater than 0.999. We can see from the above that the mean ASA of the unfolded protein decreases by $0.49 \pm 0.06 \text{ nm}^2$ per 100 residues with a 1% increase in η_{α} . On the other hand, we found that, though the mean R_{sq} decreases with increasing η_{α} for the MCDAD ensemble having a η_{α} lower than 30%, it is hardly changed with an η_{α} higher than 30%. The same dependence was observed for all sizes of proteins taken in this study. This result

Table 2. $ASA(0)$ and a , Their Standard Errors, and R^2

PDB ID	$a/\text{nm}^2 \text{ \%}^{-1}$	$ASA(0)/\text{nm}^2$	R^2
1hrc	-0.49 ± 0.04	146 ± 1	0.963
5tim	-1.19 ± 0.10	333 ± 3	0.966
1qk1	-1.81 ± 0.15	524 ± 4	0.966

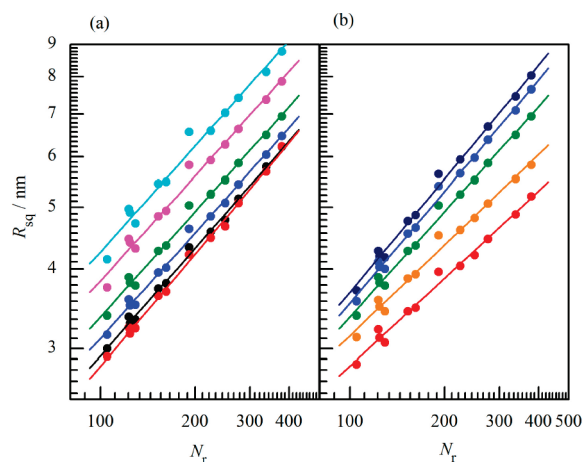


Figure 6. Dependence of the mean R_{sq} of unfolded proteins on N_r . (a) Variation of the mean R_{sq} with the MCDAD ensemble. The correspondence of ensembles to colors is the same as in Figure 4. (b) Variation of the weighted mean R_{sq} for the $(\beta + t + c)$ ensemble with solvation effects. The weighted mean R_{sq} is obtained using the Boltzmann factor at $T = 298.15 \text{ K}$ including solvation free energy, where it is assumed that the solvation free energy is proportional to the ASA. Proportional constants, σ : -10 (deep blue), -5 (blue), 0 (green), $+5$ (orange), $+10$ (red) $\text{J mol}^{-1} \text{ \AA}^{-2}$. The straight lines are best fitted to respective data by using eq 3. The obtained parameters, standard errors, and R^2 in a and b are listed in Table 3. Results of a fitting analysis are shown in Figure 8 for the combinations of MCDAD and σ , including those above.

strongly suggests that there is a lower limit of R_{sq} for each unfolded protein.

The scaling law, i.e., the dependence on N_r of R_{sq} is a very important relation for characterizing the structure of an unfolded protein. For random-coil chains consisting of $N_r (= N + 1)$ units (N is the number of virtual bonds), the following equation holds with a sufficiently large N :⁴¹

$$R_{sq} = R_0 N^{\nu} \approx R_0 N_r^{\nu}, N_r \gg 1 \quad (3)$$

where ν is the scaling exponent. The value is 0.5 for ideal random-flight chains without an excluded volume effect. The value of ν is theoretically estimated to be about 0.6 for random-coil chains with an excluded volume effect.^{42,43} The R_0 is a quantity with a dimension of length.

To examine the scaling parameters for our MCDAD ensembles, dependences of the mean R_{sq} on N_r are shown in Figure 6. As seen from Figure 6a, the values of ν for the respective models are near the range of 0.55–0.57. These values are intermediate between those for ideal random-flight chains ($\nu = 0.5$) and random-coil chains with an excluded-volume effect ($\nu \approx 0.6$). As no significant correlation is observed between the values of η_{α} and ν , we can see that the value of ν does not depend effectively on MCDAD. On the other hand, the value of R_0 varies significantly with MCDAD and decreases monotonically with increasing η_{α} .

These results agree completely with the prediction from polymer theory.^{41,42}

The dependence of R_{sq} on N_r is shown in Figure 6b, where the values of R_{sq} for the respective conformations are averaged by weighting them with the solvation effect in the ASA approximation. It is assumed here that the solvation energy is proportional to the ASA of the protein with a proportional constant, σ . The statistical-thermodynamically averaged R_{sq} is obtained by weighting the occurrence probability of each generated conformation with the Boltzmann factor at $T = 298.15$ K for the solvation energy. As seen from Figure 6b, the exponent ν , which is given by the slope of the line, decreases from 0.6 to 0.45 with the above increase in σ . It shows clearly that solvation significantly affects the scaling exponent ν in our MCDAD model, which agrees with the result of Wang et al.¹⁰ on the scaling law analysis.

The calculated Kratky profiles of SXS, which provide important information on the molecular shape, are shown in Figure 7.

Table 3. Parameters, Standard Errors, and R^2 for (a) Variation of the Mean R_{sq} with the MCDAD Ensemble and (b) Variation of the Weighted Mean R_{sq} for the $(\beta + t + c)$ Ensemble with Solvation Effects

(a) Variation of the Mean R_{sq} with the MCDAD Ensemble			
MCDAD	ν	R_0/nm	R^2
$(\alpha + t + c)$	0.58 ± 0.02	0.20 ± 0.02	0.996
(all)	0.56 ± 0.02	0.21 ± 0.02	0.996
$(t + c)$	0.55 ± 0.02	0.24 ± 0.02	0.994
$(\beta + t + c)$	0.55 ± 0.02	0.31 ± 0.03	0.991
(coil)	0.57 ± 0.03	0.48 ± 0.05	0.983
$(\beta + c)$	0.57 ± 0.04	0.64 ± 0.06	0.976
(b) Variation of the Weighted Mean R_{sq} for the $(\beta + t + c)$ Ensemble with Solvation Effects			
$\sigma/J \text{ mol}^{-1} \text{ \AA}^{-2}$	ν	R_0/nm	R^2
-10	0.59 ± 0.02	0.24 ± 0.03	0.994
-5	0.58 ± 0.02	0.24 ± 0.03	0.993
0	0.55 ± 0.02	0.26 ± 0.03	0.991
5	0.48 ± 0.02	0.35 ± 0.05	0.985
10	0.46 ± 0.02	0.34 ± 0.04	0.987

As seen from Figure 7a, the difference in MCDAD has considerable effects on the Kratky profile of the unfolded protein. There are significant effects not only in the small K region of $K < 0.5 \text{ nm}^{-1}$ essential for R_{sq} but also in the intermediate K region of $0.5 < K < 2.0 \text{ nm}^{-1}$. In these K regions, a MCDAD ensemble with a higher value of η_α yields a higher scattering intensity. On the other hand, the solvation effect on the Kratky profile is very different from the effect of MCDAD, as shown in Figure 7b. Introduction of the solvation effect considerably changes the estimate of scattering intensity at $0.5 < K < 1.0 \text{ nm}^{-1}$. Especially, there appears a peak in the profile under poor solvent conditions of $\sigma > 0$. On the other hand, the difference in scattering intensity is fairly small at $1.5 < K < 2.0 \text{ nm}^{-1}$. All of the Kratky profiles for apomyoglobin nearly coincide with each other at $K = 1.8\text{--}2.0 \text{ nm}^{-1}$. Naturally, we can see that the changes of MCDAD and solvation give qualitatively different effects on the Kratky profile.

Structural Analysis of Highly Denatured Proteins. As described above, we could confirm that the differences among various MCDAD models adopted here are strongly reflected in the value of R_0 in the scaling law and the shape of the Kratky profiles, to which attention had hardly been paid previously. It is expected that the true MCDAD model can be predicted from comparing quantitatively the calculated Kratky profiles with the experimental ones by using the above result. As a test for the possibility, we will compare quantitatively experimental data on a protein in the highly unfolded state in an aqueous concentrated solution of denaturant with our predictions from various MCDAD models. The solid and the broken lines in Figure 8 indicate the scaling-law parameter and the range of estimated errors respectively obtained from an analysis of the reported experimental values of R_{sq} for proteins in the highly unfolded state.⁷ As evident from Figure 8a, only the ensembles weighted with negative σ , i.e., assuming good solvent conditions, can reproduce the experimental value of ν . In addition, the experimental value of ν can be reproduced independently of the values of η_α . On the other hand, as shown in Figure 8b, not only the solvation effect but also the value of η_α must be limited to reproduce the experimental value of R_0 . The prediction for scaling parameters from an ensemble of generated conformations agrees with experimental results within experimental error only when the η_α of its MCDAD is higher than about 20%. In other words,

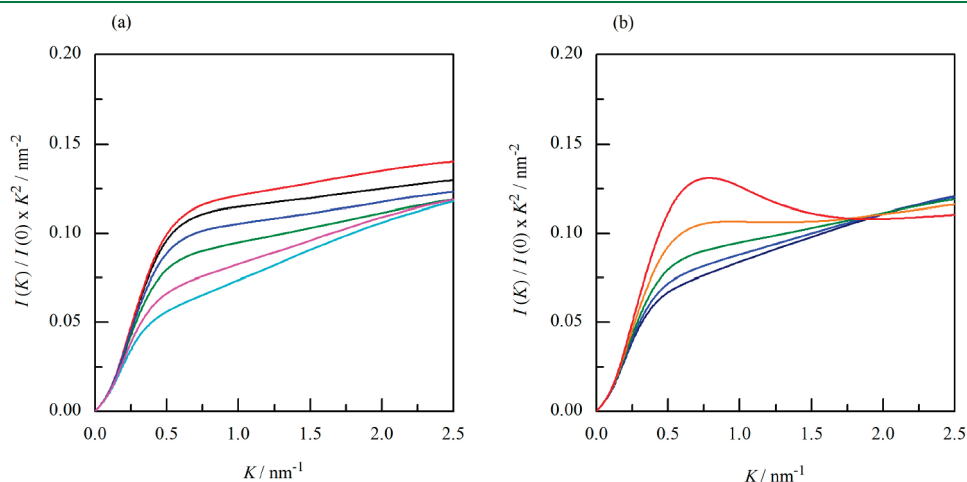


Figure 7. Kratky profiles for random chain models of apomyoglobin. The colors of lines in a and b correspond to those in Figure 6a and b, respectively.

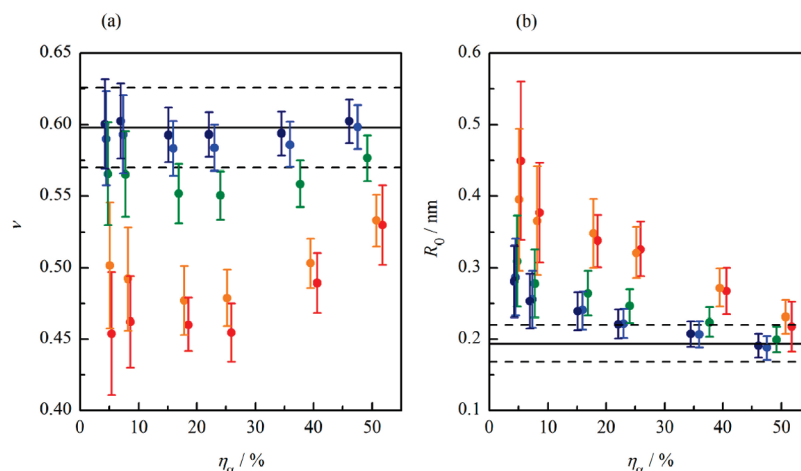


Figure 8. Comparison of experimental and calculated scaling parameters. Variation of the dependence of (a) ν and (b) R_0 on the α -region fraction, η_α , with the solvation effect, where ν and R_0 are obtained from the scaling-law analysis of MCDAD ensembles. The color of the data points corresponds to each value of σ characterizing the solvation effect, as shown in Figure 6b. The experimental values⁷ of ν and R_0 and their respective standard deviations for highly unfolded proteins are shown by solid and broken lines: $\nu = 0.598 \pm 0.028$ and $R_0 = 0.192^{+0.271}_{-0.238}$ Å.

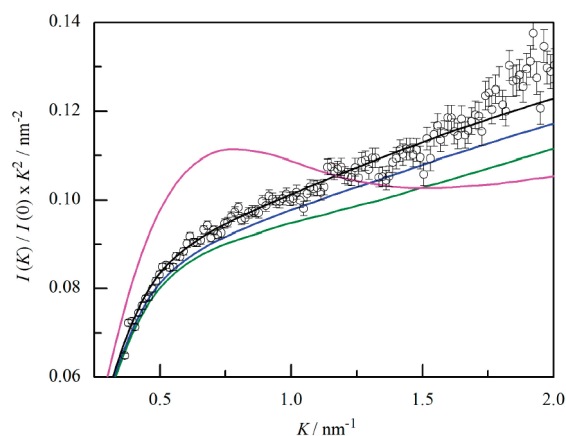


Figure 9. Comparison of the experimental and calculated SAXS profiles for unfolded apomyoglobin. Open circles indicate the experimental profile under 5 M urea conditions,⁹ and solid lines indicate the profiles estimated with various MCDAD ensembles. The MCDAD model and solvation energy for each estimated profile are as follows: black, (all) and $\sigma = -10$; blue, (t + c) and $\sigma = -5$; green, ($\beta + t + c$) and $\sigma = 0$; purple, (coil) and $\sigma = 10$ J mol⁻¹ Å⁻².

the ensembles with lower values of η_α such as the ($\beta + c$), (coil), and ($\beta + t + c$) models cannot reproduce experimental results.

To obtain more detailed information about the protein structure in the highly unfolded state, the Kratky profile for the urea unfolded state of apomyoglobin⁹ and those for the generated ensembles of the conformation are compared in Figure 9. The figure shows only profiles for the four ensembles that yield R_{sq} close to the experimental one among the 30 ensembles generated by combining six MCDAD models with five kinds of solvation effects. The experimental profile lies between those for the (all $_{\sigma=-10}$) and ($t = c_{\sigma=-5}$) ensembles at $K < 1.5$ nm⁻¹. We cannot make a decision on the validity of a model by comparing experimental and computational profiles at $K > 1.5$ nm⁻¹, because the present method of analysis cannot estimate the solvent effect on SAXS intensity with enough accuracy. The Kratky profile for the ($\beta + t + c_{\sigma=0}$) ensemble, which has an η_α lower than those of the two ensembles above, has a similar shape to the

experimental profile but deviates downward in the whole K region as its R_{sq} is larger than the experimental one. Adding a weakly poor solvation effect of $\sigma > 0$ to decrease the R_{sq} of the ensemble, the shape of the Kratky profile deviates from the experimental one to approach that of the (coil $_{\sigma=10}$) ensemble. To summarize the above, the ensemble reproducing the experimental Kratky profile for urea unfolded apomyoglobin is a MCDAD ensemble for either the (all) or (t + c) model under good solvent conditions of $\sigma = -10$ to -5 kJ mol⁻¹ Å⁻². These results indicate that the possible MCDAD model needs to be further limited to reproduce the SAXS profile in addition to the scaling law. Thus, we could predict not only the value of R_{sq} but also the solvation effect and MCDAD model of an unfolded protein more definitely by comparing the estimated Kratky profile with the experimental one.

DISCUSSION

In the algorithm developed in this study for generating unfolded-protein conformations, we make use of the probability distribution of main- and side-chain dihedral angles in native proteins and completely eliminate the atomic collisions between neighboring residues. As a result, it is expected that the conformational distribution of amino acid residues and the conformational correlation among them are both adequately considered. It is for this reason that we could obtain good agreement between the computation and experimental results for the parameters reflecting the structural characteristics of a protein at a short distance. On the other hand, the effect of cooperative interactions between residues at long distances, which is essential to the higher-order structure formation in a protein, cannot be taken into account. For example, an α helix is generated by its nucleation accompanying a large entropy loss followed by the cooperative incorporation of helical residues accumulating small enthalpy gains. However, such a structure can hardly be generated by our method. It assumes implicitly that the target is limited to an ensemble of conformations having practically no cooperative local structure, such as those of unfolded proteins in a concentrated denaturant solution. Even if the solvation effect is included, we can hardly generate any

cooperative structure by this method. It is almost inapplicable to modeling transiently denatured structures such as those in the intermediate state of protein folding, because they are formed only under the solution condition that the equilibrium is strongly biased toward higher-order structure formation. Similar situations are also expected with denatured structures at high temperatures, high pressures, and extreme pH. To generate conformations for a protein denatured under these conditions, it will be necessary to develop an algorithm that can explicitly consider the formation of higher-order structures.⁴⁴ Conversely, it is expected that this method can be applied effectively to the structural analysis of natively unfolded proteins that can hardly form a higher-order structure by itself.

While both ASA and R_{sq} are important parameters characterizing the structure of the unfolded protein, the α -region fraction, η_α is a quantity reflecting the characteristics of each MCDAD. Analyzing the interrelation between the former two and the latter one, we found that the change in η_α is reflected sensitively in the change in ASA. As shown in Figure 5a, the mean ASA decreases with an increase in η_α . We will discuss the reason why this relation holds in the following. As the bond length and bond angle of a peptide chain, and the dihedral angle of its peptide bond, ω , are nearly constant, the angle formed by two neighboring virtual bonds is predominantly determined by a pair of main-chain dihedral angles, φ and ψ . When a pair of main-chain dihedral angles of a residue is in the α region, the virtual-bond angle formed by the neighboring residues is about 90° , which is smaller than 120° , i.e., the corresponding angle for a residue in either the β or p region. Therefore, the magnitude of η_α is reflected in the degree that the main-chain is sharply bent. It is easily understood that a residue with its main-chain sharply bent tends to approach neighboring residues and decrease the mean ASA of the protein. In fact, we can see from Figure 5a that the decrease in ASA with an increase in η_α is in very good proportion to the N_r of the protein. We can conclude from the above that the difference in mean ASA between different MCDAD ensembles in the highly unfolded state is mainly brought about by local effects of the difference in the degree of spatial proximity between neighboring residues.

On the other hand, the relation between R_{sq} and η_α is not so straightforward. A change in MCDAD directly brings about a change in the local structure, but it must affect indirectly the values of global structural parameters such as R_{sq} . Interestingly, with regard to the scaling law of R_{sq} , the change in η_α does not affect the value of ν only to change the value of R_0 . As shown in Figure 8b, it decreases monotonically with increasing η_α under the same solvent conditions. This dependence of R_0 on η_α will reasonably be explained as follows: As an increase in the η_α of a polypeptide chain increases, the degree of its chain bending, the R_{sq} for an ensemble of chains with a higher probability of sharply bent conformations is smaller than those with lower probability. This change in R_{sq} only affects the value of R_0 in the low scaling, because ν hardly depends on η_α under the same solvent conditions. As the chains in any MCDAD ensemble have a value of ν larger than 0.5 and are not ideal random-flight ones, it is difficult to have more detailed, quantitative discussions about the R_0 for MCDAD ensembles. Thus, the effect of differences in MCDAD appears only in R_0 on the scaling law of R_{sq} . This is in contrast to the incorporation of solvation effects by the ASA approximation: It modifies the occurrence probability of each conformation according to the value of its ASA, which changes the distribution of the spatial extension of chains to result in a change in the value of ν .¹⁰

We have analyzed the structure of a highly unfolded protein to confirm the usefulness of conformational ensembles generated from diverse MCDADs. In practice, comparison has been made between experimental and computational estimates for the parameters, ν and R_0 , of the scaling law of R_{sq} for highly unfolded proteins and the SXS profile for urea unfolded apomyoglobin. As a result, it has become evident that the highly denaturing solvent is a good solvent with $\sigma = -10$ to -5 kJ mol⁻¹ Å⁻², and the η_α for the MCDAD of highly unfolded proteins is in the range of 20–40%. The result of our analysis on the solvation effect agrees qualitatively with that of Wang et al.¹⁰ It is also in accord with the widely accepted view that denaturant molecules interact directly with the peptide groups of a protein molecule.^{45,46} However, the present method incorporates solvation effects only through the ASA of the whole protein molecule not taking into account the variety of its polar atomic groups. Hence, with this method, we cannot discuss the detailed molecular mechanism of the interactions involved in protein denaturation such as the hydrogen bonding of main-chain polar groups³⁸ and the hydrophobic interaction of side-chain nonpolar groups⁴⁷ with water and denaturant molecules. In addition, even if the difference in the magnitude of σ between the polar and nonpolar groups is explicitly considered, it is very difficult to estimate their respective magnitudes of σ with high accuracy, because the differences in the proportion of their respective surface areas are fairly small among the target proteins taken in this study.

On the other hand, it is necessary to discuss more carefully the α -region fraction. The interrelation between the $^3J_{HN\alpha}$ coupling constant and the MCDAD has been analyzed for unfolded proteins. Smith et al.¹² determined the correlation coefficient, γ , between two $^3J_{HN\alpha}$ coupling constants of each residue, which were measured experimentally and estimated using MCDAD. They concluded that the structural ensemble of the “COIL” MCDAD ($\gamma = 0.92$) reproduces experimental coupling constants for short peptides having no secondary structure better than the ensemble of the “ALL” MCDAD ($\gamma = 0.81$). Their MCDAD models are constructed from the probability distribution of dihedral angles of native proteins similarly to our models. They reported that the values of η_α for the COIL and ALL models are 28% and 45%, respectively. These estimates for the η_α agree fairly well with our estimates of 20–40% in this study. Analyzing ubiquitin denatured in aqueous 8 M urea solution, Peti et al.¹³ confirmed that the ensemble for the COIL model reproduces experimental results with a high correlation coefficient of $\gamma = 0.96$. Thus, the value of η_α estimated from our analysis for proteins highly unfolded in concentrated denaturant solution agrees with the result of analysis of NMR coupling constants. These results confirm that our analysis gives reasonable estimates for the η_α of the unfolded protein. However, the information on MCDAD derived from our analysis and NMR coupling constants is inadequate in particular with respect to the dependence on amino acid sequence. It will be necessary to verify the validity of MCDAD models by using more efficient methods such as the residual dipolar couplings method.^{19,20}

From the viewpoint of molecular biophysics, the structure of a biomolecule can be described not by mechanics based on the minimum potential energy but by statistical thermodynamics based on the minimum free energy, including both enthalpic and entropic contributions. It is a well-known fact that all of the real, physical structures are continually fluctuating by various degrees. In this sense, the structure of a protein molecule consists of an ensemble of multiple structures not only in the unfolded state but

also in the intermediate and native states. Especially, an unfolded protein molecule, which is the target of this study, has remarkable structural characteristics such as those above. It seems to be impossible to characterize the structure of proteins in the highly unfolded state. In spite of it, our study revealed that the value of η_α for seemingly random-flight polypeptide chains in high-concentration denaturant solution is confined to a fairly narrow range. This result shows clearly that even the highly unfolded protein exhibits structural features inherent in proteins. So, it is expected that this type of study will serve as a step for clarifying the local structural characteristics of proteins in the unfolded state.

Our method of analysis using the conformational ensemble of various MCDAD models has the potential to develop into a method for more detailed analysis of unfolded proteins. For realizing the possibility, it is necessary to study various factors involved in determining the SXS profile in the intermediate and large K region of $0.5 < K < 2.0 \text{ nm}^{-1}$. To expand the possibility of comparing computation with experimental results quantitatively, it is necessary to obtain the following: (1) precise SXS data in the wide range of scattering angles including the forward scattering intensity⁹ and (2) a method of accurately estimating SXS profiles that is applicable to unfolded proteins and incorporates solvent effects explicitly.^{34,35}

AUTHOR INFORMATION

Corresponding Author

*Tel.: +81-78-940-5645. Fax: +81-78-304-4958. E-mail: soda@riken.jp.

ACKNOWLEDGMENT

Y.S. and K.S. express sincere thanks to Drs. Nobuhiro Go and Mitiko Go for their continuous support for our research work. Y.S. is very grateful to Dr. Shun-ichi Kidokoro for his support. This work was supported by a Grant-in-Aid for Young Scientists (B) (number 21700322) and for Scientific Research on Innovative Areas "Intrinsically Disordered Proteins" (number 22113507) of the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

REFERENCES

- (1) Dyson, H.; Wright, P. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv. Protein Chem.* **2003**, *62*, 311.
- (2) Kataoka, M.; Goto, Y. X-ray solution scattering studies of protein folding. *Fold. Des.* **1996**, *1*, R107.
- (3) Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **2005**, *579*, 3346.
- (4) Mittag, T.; Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3.
- (5) Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23.
- (6) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry* **1999**, *38*, 16424.
- (7) Kohn, J.; Millett, I.; Jacob, J.; Zagrovic, B.; Dillon, T.; Cingel, N.; Dothager, R.; Seifert, S.; Thiyagarajan, P.; Sosnick, T.; Hasan, M.; Pande, V.; Ruczinski, I.; Doniach, S.; Plaxco, K. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12491.
- (8) Semisotnov, G.; Kihara, H.; Kotova, N.; Kimura, K.; Amemiya, Y.; Wakabayashi, K.; Serdyuk, I.; Timchenko, A.; Chiba, K.; Nikaido, K.; Ikura, T.; Kuwajima, K. Protein globularization during folding. A study by synchrotron small-angle X-ray scattering. *J. Mol. Biol.* **1996**, *262*, 559.
- (9) Seki, Y.; Tomizawa, T.; Hiragi, Y.; Soda, K. Global structure analysis of acid-unfolded myoglobin with consideration to effects of intermolecular Coulomb repulsion on solution X-ray scattering. *Biochemistry* **2007**, *46*, 234.
- (10) Wang, Y.; Trehwella, J.; Goldenberg, D. P. Small-Angle X-ray Scattering of Reduced Ribonuclease A: Effects of Solution Conditions and Comparisons with a Computational Model of Unfolded Proteins. *J. Mol. Biol.* **2008**, *377*, 1576.
- (11) Uzawa, T.; Kimura, T.; Ishimori, K.; Morishima, I.; Matsui, T.; Ikeda-Saito, M.; Takahashi, S.; Akiyama, S.; Fujisawa, T. Time-resolved Small-angle X-ray Scattering Investigation of the Folding Dynamics of Heme Oxygenase: Implication of the Scaling Relationship for the Submillisecond Intermediates of Protein Folding. *J. Mol. Biol.* **2006**, *357*, 997.
- (12) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. Analysis of Main Chain Torsion Angles in Proteins: Prediction of NMR Coupling Constants for Native and Random Coil Conformations. *J. Mol. Biol.* **1996**, *255*, 494.
- (13) Peti, W.; Hennig, M.; Smith, L.; Schwalbe, H. NMR spectroscopic investigation of ψ torsion angle distribution in unfolded ubiquitin from analysis of $^3J(\text{C}\alpha, \text{C}\alpha)$ coupling constants and cross-correlated $\Gamma_{\text{HNN}, \text{C}\alpha\text{H}\alpha}$ relaxation rates. *J. Am. Chem. Soc.* **2000**, *122*, 12017.
- (14) Eliezer, D.; Yao, J.; Dyson, H.; Wright, P. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* **1998**, *5*, 148.
- (15) Weinstock, D. S.; Narayanan, C.; Baum, J.; Levy, R. M. Correlation between $^{13}\text{C}\alpha$ chemical shifts and helix content of peptide ensembles. *Protein Sci.* **2008**, *17*, 950.
- (16) Shi, Z.; Olson, C.; Rose, G.; Baldwin, R.; Kallenbach, N. Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190.
- (17) Shi, Z.; Chen, K.; Liu, Z.; Kallenbach, N. Conformation of the backbone in unfolded proteins. *Chem. Rev.* **2006**, *106*, 1877.
- (18) Shortle, D.; Ackerman, M. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **2001**, *293*, 487.
- (19) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J.; Dyson, H.; Wright, P. E. Structural Characterization of Unfolded States of Apomyoglobin using Residual Dipolar Couplings. *J. Mol. Biol.* **2004**, *340*, 1131.
- (20) Jha, A.; Colubri, A.; Freed, K.; Sosnick, T. Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099.
- (21) Jensen, M. R.; Markwick, P. R.; Meier, S.; Griesinger, C.; Zweckstetter, M.; Grzesiek, S.; Pau, B.; Blackledge, M. Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure* **2009**, *17*, 1169.
- (22) Fitzkee, N.; Rose, G. Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12497.
- (23) Wang, Z.; Plaxco, K.; Makarov, D. Influence of local and residual structures on the scaling behavior and dimensions of unfolded proteins. *Biopolymers* **2007**, *86*, 321.
- (24) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668.
- (25) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701.
- (26) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049.
- (27) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* **1999**, *290*, 253.

- (28) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GRO-MACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435.
- (29) Holm, L.; Sander, C. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **1998**, *26*, 316.
- (30) Holm, L.; Sander, C. Dictionary of recurrent domains in protein structures. *Proteins* **1998**, *33*, 88.
- (31) Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. Stereochemical quality of protein structure coordinates. *Proteins* **1992**, *12*, 345.
- (32) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577.
- (33) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. Intrinsic phi, psi propensities of amino acids, derived from the coil regions of known structures. *Nat. Struct. Biol.* **1995**, *2*, 596.
- (34) Soda, K.; Miki, Y.; Nishizawa, T.; Seki, Y. New method for incorporating solvent influence into the evaluation of X-ray scattering intensity of Proteins in Solution. *Biophys. Chem.* **1997**, *65*, 45.
- (35) Seki, Y.; Tomizawa, T.; Khechinashvili, N. N.; Soda, K. Contribution of solvent water to the solution X-ray scattering profile of proteins. *Biophys. Chem.* **2002**, *95*, 235.
- (36) Richmond, T. J. Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* **1984**, *178*, 63.
- (37) Goldenberg, D. Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.* **2003**, *326*, 1615.
- (38) Gong, H.; Rose, G. D. Assessing the solvent-dependent surface area of unfolded proteins using an ensemble model. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3321.
- (39) Creamer, T.; Srinivasan, R.; Rose, G. Modeling unfolded states of peptides and proteins. *Biochemistry* **1995**, *34*, 16245.
- (40) Creamer, T.; Srinivasan, R.; Rose, G. Modeling unfolded states of proteins and peptides II. Backbone solvent accessibility. *Biochemistry* **1997**, *36*, 2832.
- (41) Flory, P. *Statistical Mechanics of Chain Molecules*; Interscience: New York, 1969.
- (42) Flory, P. *Principles of Polymer Chemistry*; Cornell Uni. Press: New York, 1953.
- (43) Le Guillou, J. C.; Zinn-Justin, J. Critical Exponents for the n-Vector Model in Three Dimensions from Field Theory. *Phys. Rev. Lett.* **1977**, *39*, 95.
- (44) Kamatari, Y. O.; Ohji, S.; Konno, T.; Seki, Y.; Soda, K.; Kataoka, M.; Akasaka, K. The compact and expanded denatured conformations of apomyoglobin in the methanol-water solvent. *Protein Sci.* **1999**, *8*, 873.
- (45) Tirado-Rives, J.; Orozco, M.; Jorgensen, W. L. Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. *Biochemistry* **1997**, *36*, 7313.
- (46) Bennion, B.; Daggett, V. Counteraction of urea-induced protein denaturation by trimethylamine N-oxide: A chemical chaperone at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6433.
- (47) Ikeguchi, M.; Nakamura, S.; Shimizu, K. Molecular Dynamics Study on Hydrophobic Effects in Aqueous Urea Solutions. *J. Am. Chem. Soc.* **2001**, *123*, 677.

Coarse-Grain Model for Glucose, Cellobiose, and Cellotetraose in Water


Antti-Pekka Hynninen,[†] James F. Matthews,[‡] Gregg T. Beckham,^{†,§} Michael F. Crowley,[†] and Mark R. Nimlos^{*,†}

[†]National Bioenergy Center

[‡]Biosciences Center

National Renewable Energy Laboratory, Golden, Colorado 80401, United States

[§]Department of Chemical Engineering, Colorado School of Mines, Golden, Colorado 80401, United States

 Supporting Information

ABSTRACT: We present a coarse-grain (CG) simulation model for aqueous solutions of β -D-glucose, cellobiose, and cellotetraose, based on atomistic simulation data for each system. In the model, three spherical beads are used to represent glucose, and a single bead is used to represent water. For glucose, the force field is calculated using force matching by minimizing the sum of the square differences between forces calculated from atomistic and CG simulations. For cellobiose and cellotetraose, we use a hybrid method where the nonbonded interactions are obtained using force matching and the bonded interactions are obtained using Boltzmann inversion. We demonstrate excellent agreement in the structural properties between the atomistic simulations and the CG simulations. This model represents the first step in developing a CG force field for cellulose, as it is of significant interest to study cellulose behavior at much longer time and length scales relative to atomistic simulations.

INTRODUCTION

Carbohydrate polymers are the most abundant organic materials on Earth because of their prevalence in plants (cellulose, hemicellulose, pectin) and structural materials in fungi and insects (chitin). Cellulose is the linear polymer of β -D-glucose, and a significant component of the global carbon cycle and potential source of sugars for the production of transportation fuels. Most enzymatic strategies for deconstructing cellulose involve the hydrolysis of accessible glycosidic bonds between sugars.¹ However, accessing the individual linkages between sugars is a significant challenge because cellulose packs into dense crystalline fibrils. Understanding the structural nature of cellulose is thus important for determining how enzymes deconstruct plants in the biosphere and in designing enhanced properties to convert cellulose to glucose more efficiently for bioenergy processes.

To date, molecular simulation has been used extensively to understand how glucose, short polymers of cellulose, and cellulose crystals behave in solution. Most studies to date have treated cellulose and cellodextrin chains with fully atomistic models.^{2–18} A few studies to date have developed coarse-grained (CG) models of varying resolutions for studying monosaccharides and cellulose.^{19–23} However, there still remains a significant need to develop improved CG models for cellulose (and other carbohydrate polymers) because of the large number of atoms and the long time scales required to study these polymers. As pointed out recently for simulations of cellulose crystals, it is unlikely that any MD simulations to date have been conducted long enough (hundreds of nanoseconds) to equilibrate.^{10,24} It is quite likely that CG models will be necessary to study interesting problems such as (1) phase transitions in cellulose,²⁵ (2) microfibril dissolution and behavior in nonaqueous solvents,^{26–28} (3) material properties

of long cellulose microfibrils and cellulose chains in nonaqueous solvents,²⁹ (4) the nucleation of cellulose crystals during biosynthesis,^{30,31} (5) microfibril aggregation, (6) interactions of cellulase and cellosomal systems with the plant cell wall during biological deconstruction,^{32–34} and (7) the interactions of cellulose with hemicellulose, pectin, and lignin as a first step in developing models of the whole plant cell wall.

There are multiple methods for developing CG models in the literature.³⁵ One class includes methods based on the inverse Boltzmann rule where the probability distribution (or the radial distribution function) obtained from the atomistic simulations is matched using the CG model. This class includes the inverse Monte Carlo³⁶ and the iterative inverse Boltzmann methods.³⁷ In another class of methods, called force-matching methods, the forces calculated from the atomistic model are matched by the CG model.^{38,39} The force-matching method was originally developed by Ercolessi and Adams for producing atomistic interaction potentials from *ab initio* simulations.⁴⁰ Typically, the matching is done by minimizing the sum of squared differences in forces generated by atomistic and CG simulations. Izvekov and Voth termed this method multiscale coarse-graining (MS-CG) because the CG model is constructed using information from the atomistic scale.

Force matching has emerged as a powerful method to develop the force fields necessary to conduct CG MD simulations. However, many of these studies⁴¹ focus on systems that only require two-body interactions. Going toward more complex systems with angles and dihedrals, Noid et al. used force matching to model ethylmethylimidazolium (EMIM⁺)/NO[−]₃

Received: February 8, 2011

Published: June 07, 2011

ionic liquid.⁴² In their work, the EMIM⁺ molecule was modeled using four CG beads, and the NO₃⁻ molecule was modeled using a single CG bead. In addition, Izvekov and Voth have recently applied the force-matching method on lipid bilayers.^{43,44} In particular, the work in ref 43 shows that the force-matching method can be used to coarse-grain out the solvent to obtain a solvent-free model. Application of the force-matching method to ion channels shows that the method can be used to calculate the interactions for a mixed system with atomistic and CG parts.⁴⁵ In a recent article, Lu et al.⁴⁶ discuss the various options for the implementation of the force-matching method, which has been extended for transfer between temperatures,⁴⁷ the isothermal–isobaric ensemble,⁴⁸ and three-body potentials.⁴⁹

Using their own method that is based on reproduction of atomistic structural and thermodynamic properties, Molinero and Goddard developed a three-bead model to describe α -D-glucose and mapped the forces from atomistic simulations on the CG model.^{20,50} They used this approach to model glucose in aqueous solution and in glasses. Their CG model yields good agreement with the atomistic simulations for density, structural properties, and cohesive energy. Molinero and Goddard later used their CG model to study water diffusion in glucose glasses.⁵⁰ Liu et al.²¹ used a model similar to that used by Molinero and Goddard to coarse-grain α -D-glucose. In their method, the bonded interactions were fitted to harmonic potential functions, whereas the nonbonded interactions were tabulated and calculated using force matching. Liu et al. obtained good structural agreement between the CG and the atomistic simulations. Furthermore, they showed that their CG model could be applied to other temperatures and pressures than where it was originally derived.

Additionally, our group recently developed a preliminary CG force field to study the hydrophobic face of crystalline cellulose in the presence of an atomistic carbohydrate-binding module.^{23,51} The original model for cellulose was based on the glucose CG model from Molinero and Goddard,²⁰ and Boltzmann inversion was used to fit the bond, angles, and dihedral parameters. For the nonbonded terms, rescaling factors were applied to a standard 6–12 Lennard-Jones potential to represent the directionality in a cellulose I β crystal from atomistic simulations. The solvent was represented by the generalized Born model with switching (GBSW).^{52,53} The CG model was found to agree well with the atomistic simulations, except in the lattice parameter (or layer spacing) for cellulose I β . In our previous studies, we used a multiscale model to study enzyme behavior on the hydrophobic face of cellulose I β , and thus this model was adequate for the intended purposes. We used this model to examine the behavior of a Family 1 carbohydrate-binding module (CBM) on cellulose and showed that the CBM has regions of thermodynamic stability along a cellulose chain on the surface every 1 nm, which was later confirmed with an atomistic study.⁵⁴

Recently, Wohlert and Berglund developed a coarse-grained force field for native cellulose within the MARTINI force field suite.⁵⁵ They parametrized their model with partition free energy calculations for cellodextrin oligomers between water and cyclohexane and then extended the parameters to model cellulose I β . They found that to maintain the staggered conformation of origin and center chains in cellulose I β ,⁵⁶ an additional repulsive term was needed in the force field between certain beads in cellulose. From this, they conducted MD simulations of a cellulose microfibril and compared the lattice parameters to the experimentally determined structure.⁵⁶ They found good agreement down the length of the cellulose chains but observed

substantial deviation from the a and b lattice parameters relative to the experimental structure. The authors used their model to study the surface diffusion of the Family 1 Cel7A CBM and found good agreement with the experimentally measured diffusion coefficient for a Family 2 CBM. However, the authors pointed out that, similar to the model of Bu et al.,²³ their model is not useful for studying cellulose microfibril properties beyond surface properties.

In this article, we describe the development of coarse grain models of β -D-glucose, cellobiose, and cellotetraose as a first step to developing a CG model for cellulose. Glucose is the smallest repeat unit of a cellulose chain, and a CG model of this molecule serves as a logical starting point toward developing a model of cellulose. Cellobiose (1,4-O- β -D-glucopyranosyl-D-glucose) is the conformational monomer in crystalline cellulose, and cellotetraose is the smallest cello-oligomer that has an internal cellobiose unit. Thus, we anticipate that the force field for cellotetraose will be a reasonable first iteration for a CG force field of cellulose. The method of Noid et al.^{42,57} was originally applied directly for each of these molecules. However, we found that this approach is only useful for glucose, and a hybrid of force matching and Boltzmann inversion was needed for cellobiose and cellotetraose.

METHODS

Atomistic Simulations. All of the molecular dynamics (MD) simulations were performed using CHARMM.⁵⁸ For all three atomistic systems (glucose, cellobiose, and cellotetraose), the molecules were initially placed on a cubic grid and solvated with TIP3P water.⁵⁹ The system was then equilibrated using constant pressure simulations. The production runs were conducted in the canonical ensemble where the size of the simulation box was set to be equal to the average density from the constant pressure simulations. The time step in the atomistic MD simulations was 2 fs, and the nonbonded interaction cutoff distance was 12 Å. The electrostatic interactions were calculated using the particle mesh Ewald method with sixth-order spline, $\kappa = 0.32$, and a $32 \times 32 \times 32$ grid.⁶⁰ The SHAKE algorithm was used to fix the bond lengths to hydrogen atoms.⁶¹ The carbohydrates were modeled using the C35 force field.^{62,63} The atomistic MD simulations of glucose consisted of 64 glucose molecules and 551 TIP3 water molecules (total 3189 atoms) in a cubic box with side length of 30.79 Å. The length of the production run was 120 ns. The atomistic MD simulations of cellobiose included 32 cellobiose molecules and 1169 water molecules (total 4947 atoms). The production run length was 240 ns. The atomistic MD simulations of cellotetraose consisted of 32 cellotetraose molecules and 4954 water molecules (total 17 646 atoms). The production run length was 60 ns.

Coarse-Grain Model. In our coarse-grain model, each glycan is represented with three beads (AT, BB, C), and each water molecule is replaced with one bead (W) as originally used by Molinero and Goddard²⁰ and later by Liu et al.²¹ for modeling glucose and Bu et al.²³ for modeling cellulose. Figure 1a shows the arrangement of the coarse grain beads in glucose and the corresponding atoms.

Generally, the CG force field consists of nonbonded pair interactions, bonded pair interactions, angular three-body interactions, and dihedral four-body interactions and is represented numerically using lookup tables. This allows the interaction potentials to be of any shape, and no functional form is assumed. As described in the Results section, this feature is essential in capturing the nonharmonic nature of the bonded interactions and the shape of the nonbonded pair interactions. In all three

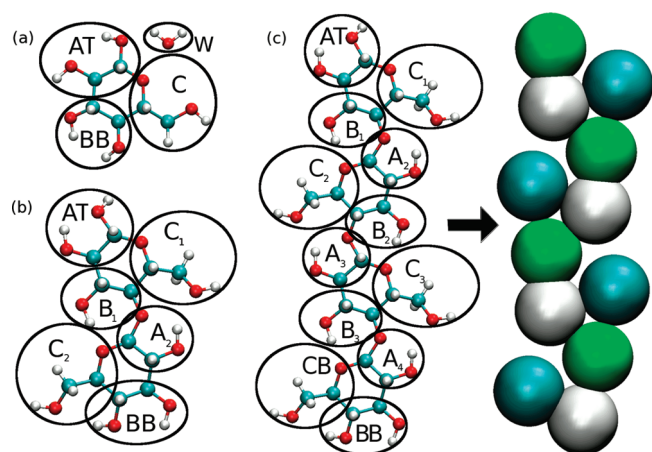


Figure 1. Atomistic structure of (a) glucose and water, (b) cellobiose, and (c) cellotetraose with definition of the CG beads. For cellotetraose, the mapping to CG beads is indicated by the arrow.

models, each CG bead is positioned at the center-of-mass of the atoms that correspond to the beads, and the mass is set to be the sum of the masses of these atoms. We did not include explicit point charges in the CG model. Instead, the Columbic and van der Waals interactions are joined together in the CG force field. Throughout the description of the beads in the model, we use subscripts to differentiate specific beads of the same type (e.g., C_1 and C_2 are the same bead type but have different locations in cellobiose and cellotetraose).

In the CG model for glucose, there are bonded interactions between all three beads and nonbonded interactions between all beads in *separate* glucose molecules. There are no nonbonded interactions between beads in the same CG glucose molecule because hydrogen bonding interactions between the hydroxyl groups on a glucose residue are incorporated into the bonded interactions. The nonbonded interactions between each glucose bead and water were also calculated.

Figure 1b shows the atomistic structure of cellobiose and the definition of the CG beads. To differentiate between the reducing and nonreducing ends of cellobiose, we use bead names AT, or “A top”, and BB, or “B bottom”. We note that the B_1 bead in cellobiose has one oxygen atom compared to glucose, where it has two. In the CG cellobiose model, there are bonded interactions between all neighboring beads, i.e., $AT-C_1$, $AT-B_1$, B_1-C_1 , B_1-A_2 , C_2-A_2 , C_2-BB , and A_2-BB . In addition, the model contains a bonded interaction between C_1 and C_2 beads to mimic the nonbonded interaction. The model also has an angular potential between beads $C_1-B_1-A_2$ and $C_2-A_2-B_1$ and dihedral potentials $B_1-C_1-C_2-A_2$, $AT-B_1-C_1-A_2$, and $BB-A_2-C_2-B_1$. In the cellobiose CG model, water is also represented using a single bead, and the nonbonded interactions are only included between beads in different cellobiose units.

The CG model for cellotetraose is shown in Figure 1c, and the definition of the beads is similar to the cellobiose model. The only difference is that we introduce an additional bead type, CB or “C bottom”. That is, for cellotetraose, we distinguish CB from C_1 , C_2 , and C_3 because we noticed that this model gave slightly better agreement with the atomistic results than where all four C beads were of the same type. Note that cellotetraose has two glycans (2 and 3 in Figure 1c) in the core of the molecule that will not have reducing end or nonreducing end interactions. Thus,

the force field for these middle glycans may be more similar to what is found in the core of cellulose chains.

In setting up the bonded interactions for the CG models, one must be careful to choose the set of bonds, angles, and dihedrals that retain the correct atomistic molecular configuration; i.e., if one or more bonds are missing, the CG model will have degrees of freedom that are not constrained by the force field. Such “underdetermined” systems have internal degrees of freedom that are not constrained in the same way as they are in the atomistic system. For example, if we were to take away the $B-C-C'-A$ dihedral interaction from the cellobiose model, the system would freely rotate around its dihedral. Force matching would still give us potentials that look reasonable, but the behavior of the system would not correspond to the atomistic model for cellobiose. One must also be careful not to overdetermine the interactions. Interactions are overdetermined if one of the bonds, angles, and dihedrals can be left out and the structure can still be uniquely identified. Overdetermined sets of bonded interactions typically give rise to interaction potentials that do not exhibit minima even though the total potential energy still has a minimum. In the case of glucose, setting up the bonded interactions is simple. Clearly, the three bonds uniquely define the molecular configuration, and none of them can be left out. One way to overdetermine the CG glucose model would be to add an angle between A, B, and C beads. In the case of cellobiose, setting up the bonded interactions is more complicated, and the set of bonded interactions we use is not unique. We developed this set by trying seven different combinations of bonded interactions. In this trial-and-error process, we calculated the bonded interaction potentials using the force-matching method. Some of the sets were discarded because they were overdetermined and yielded interaction potentials that do not have a minimum. Some others were discarded because they were underdetermined.

Force-Matching Method. The CG force field is represented by a lookup table where the forces are defined on a linear grid with a fixed bin width. We use N_D to denote the total number of bins in all lookup tables. That is, N_D includes all lookup tables for both bonded and nonbonded potentials.

In the force-matching method, the forces between CG beads are matched to the force data from atomistic MD simulations using a least-squares fit. As shown by Noid et al.,⁴² the least-squares problem reduces to a linear equation of the form

$$\mathbf{g}\mathbf{x} = \mathbf{f} \quad (1)$$

where \mathbf{g} is a matrix of size $3Nn_t \times N_D$, \mathbf{x} is a vector of length N_D , and \mathbf{f} is a vector of length $3Nn_t$. Here, N is the number of CG beads, and n_t is the number of frames in the atomistic trajectory. Vector \mathbf{f} consists of forces acting on each CG bead in each frame, which is obtained from the atomistic trajectory. Vector \mathbf{x} is the unknown, consisting of the table lookup values. That is, once eq 1 is solved, \mathbf{x} contains the CG force field. To explain the structure of matrix \mathbf{g} , we consider a simple case of two particles A and B connected by a bond. In this case, the elements of \mathbf{g} are given

$$g_{A,d}^t = \frac{c_A - c_B}{r_{AB}} \theta(r_{AB} - r_d) \quad (2)$$

where c is one of the Cartesian coordinates $c = x, y, z$; t is the frame $t = 1, \dots, n_t$; r_{AB} is the distance between particles A and B; $r_d = r_{\min} + \delta x(d - 1)$ is the bin location, $d = 1, \dots, N_D$; and θ is a δ function with width δx . To reduce the size of matrix \mathbf{g} , one can normalize eq 1 by multiplying it on both sides with \mathbf{g}^T .

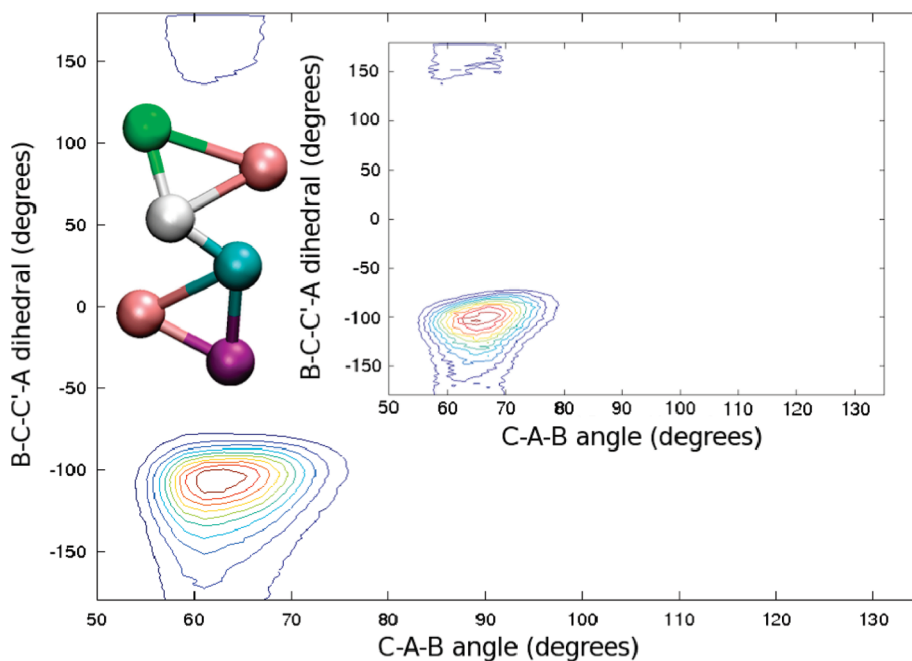


Figure 2. Contour plots of the probability distribution of the $B_1-C_1-C_2-A_2$ dihedral angle versus the $C_2-A_2-B_1$ angle from MD simulations of cellobiose. The main figure shows the result for the CG model where bonded interactions are obtained using Boltzmann inversion. The inset shows the result for the atomistic model. The snapshot shows the configuration of the cellobiose corresponding to the basin of attraction at $(60^\circ, -100^\circ)$.

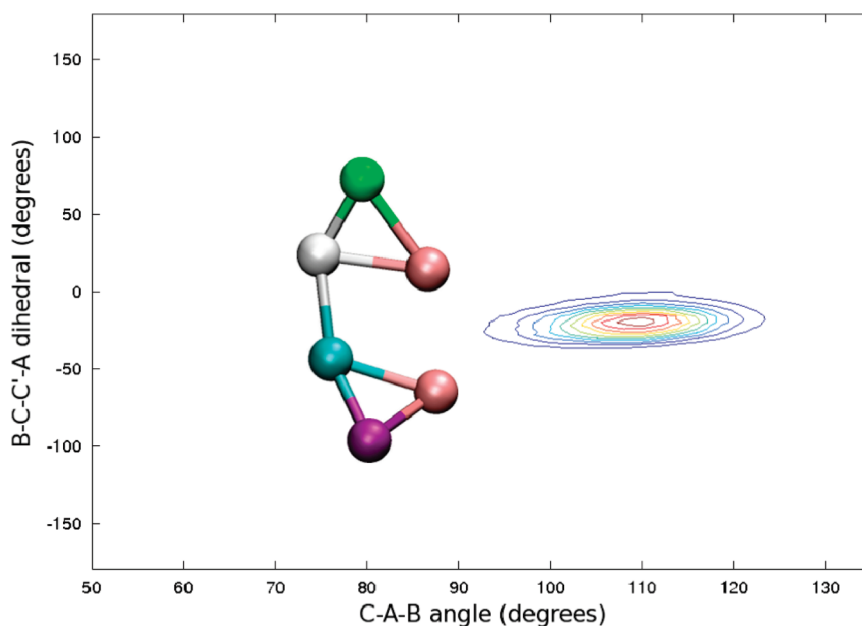


Figure 3. Contour plots of the probability distribution of the $B_1-C_1-C_2-A_2$ dihedral angle versus the $C_2-A_2-B_1$ angle from MD simulations of cellobiose showing the result for the CG model where the bonded interactions are obtained from force matching. The snapshots correspond to the basin of attraction at $(110^\circ, -25^\circ)$.

Using the notation $\mathbf{G} = \mathbf{g}^T \mathbf{g}$ and $\mathbf{b} = \mathbf{g}^T \mathbf{f}$, we obtain

$$\mathbf{G}\mathbf{x} = \mathbf{b} \quad (3)$$

where now \mathbf{G} is an $N_D \times N_D$ matrix and \mathbf{b} is a vector of length N_D . Noid et al.⁴² showed how to construct \mathbf{G} and \mathbf{b} without constructing the (usually very large) matrix \mathbf{g} . The advantage of using eq 3 versus eq 1 is that the size of matrix \mathbf{G} is usually much

smaller than that of matrix \mathbf{g} . This is because the size of matrix \mathbf{g} increases linearly with the size of the atomistic trajectory.

Finally, once \mathbf{x} is solved from eq 3, the force between particles A and B at distance r is given by

$$\mathbf{f}_{AB}(\mathbf{r}) = \sum_d x_d \theta(r - r_d) \frac{\mathbf{r}}{r} \quad (4)$$

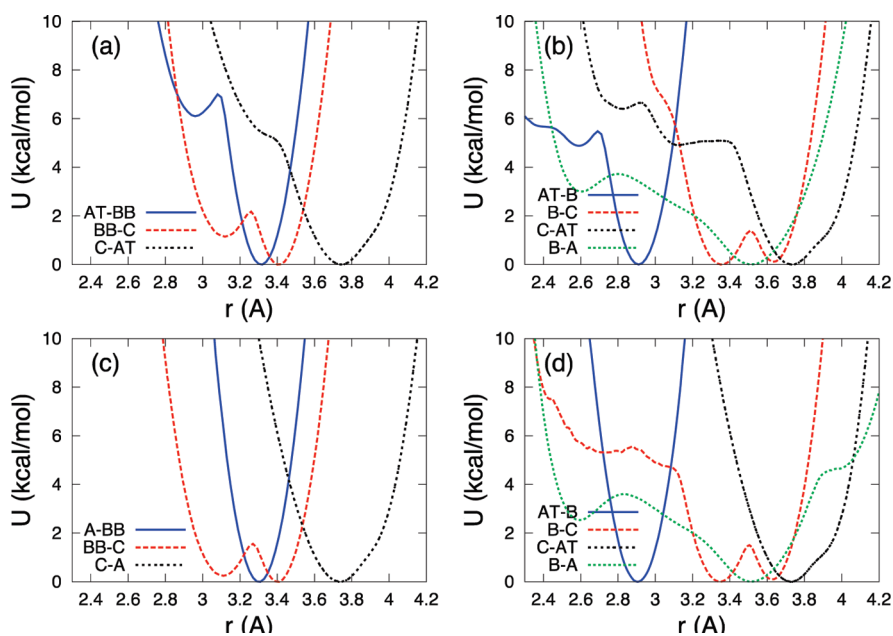


Figure 4. Bonded pair potentials for (a) glucose, (b and c) cellobiose, and (d) cellotetraose CG models. In b, “B–A” refers to the “ B_1-A_2 ” bond between glucose residues, as shown in Figure 1 (b). Similarly, in d, “B–A” refers to the “ B_1-A_2 ”, “ B_2-A_3 ”, and “ B_3-A_4 ” bonds between glucose residues, as shown in Figure 1c.

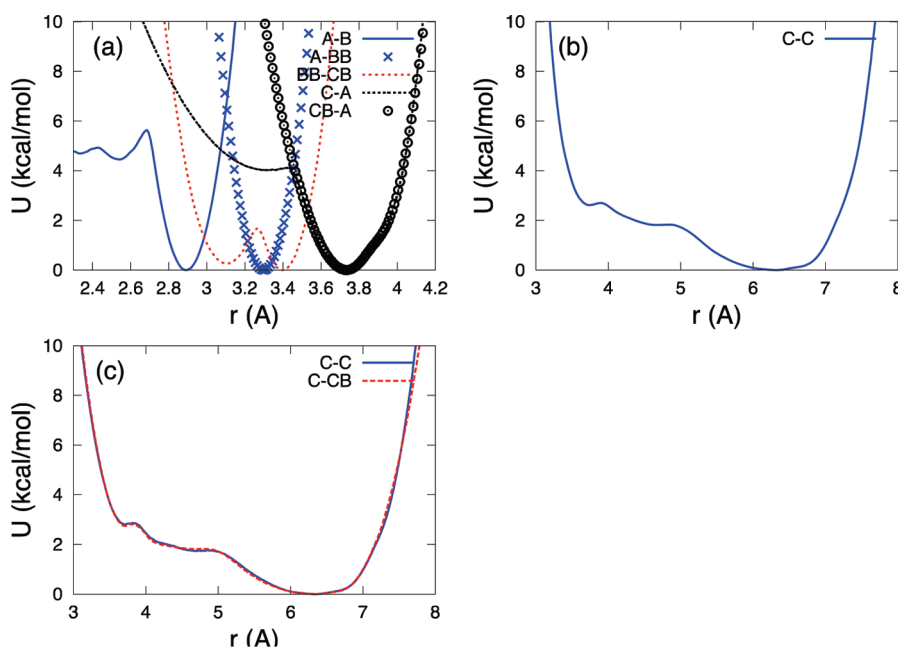


Figure 5. Bonded pair potentials for (a) cellotetraose, (b) cellobiose, and (c) cellotetraose CG models. In a, “A–B” refers to bonds “ A_2-B_2 ” and “ A_3-B_3 ” within the glucose residue, as shown in Figure 1c.

where x_d is the d th element of vector \mathbf{x} , and the role of the δ function θ is to select the contribution at the correct distance range $r = r_d - \delta x/2, \dots, r_d + \delta x/2$. A more detailed description on how to construct \mathbf{G} and \mathbf{b} is given in the original article by Noid et al.⁴²

We wrote a C-language program to perform the force matching. The program takes as input the atomistic PSF (CHARMM protein structure file) file, atomistic trajectory file, CG PSF file, and a file defining which atom belongs to which CG bead. The program then goes through the trajectory and builds the \mathbf{G} matrix and \mathbf{b} vector.

This part of the program is by far the most time-consuming. Therefore, we parallelized the program such that it splits up the trajectory evenly for all CPUs, and each CPU builds \mathbf{G} and \mathbf{b} for their part of the trajectory. At the end, the program combines the \mathbf{G} and \mathbf{b} matrices and vectors from all CPUs into the final result. Finally, the linear equation in eq 1 is solved using singular value decomposition (SVD), which was implemented using the CLAPACK library.⁶⁴

Boltzmann Inversion. We found that the direct application of force matching for cellobiose and cellotetraose produces bonded

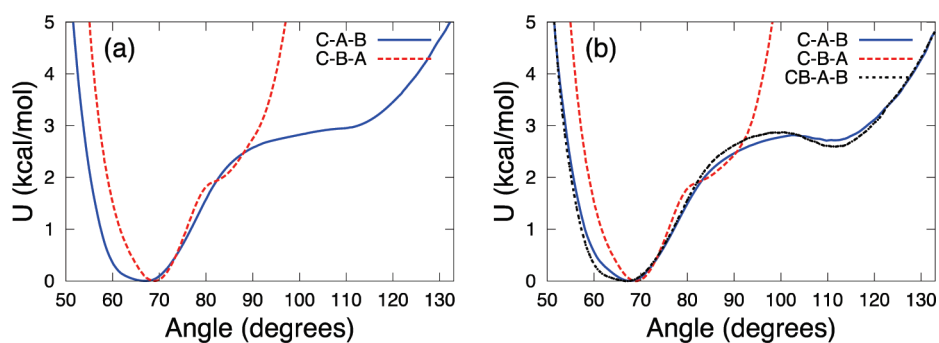


Figure 6. Angular potential for (a) cellobiose and (b) cellotetraose CG models.

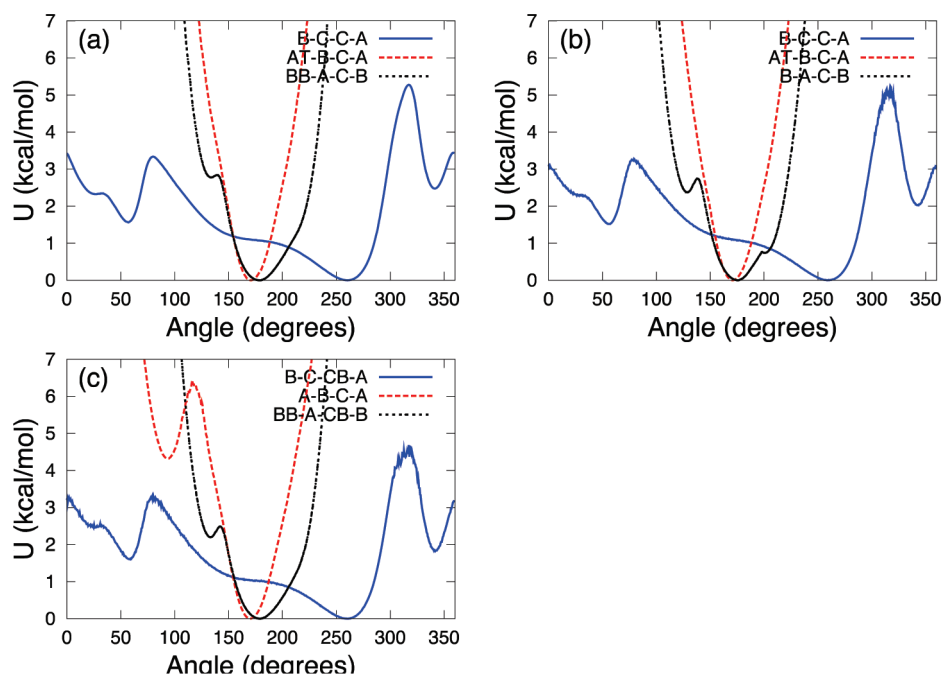


Figure 7. Dihedral potentials for (a) cellobiose and (b and c) cellotetraose CG models.

interactions that give rise to nonphysical conformations. The reason for this is that the force-matching method retains the correlations between bonded interactions, but these correlations are not taken into account in the subsequent MD simulations of the CG model. This problem is highlighted in Figure 2, where we plot contours of the probability distribution of the $B_1-C_1-C_2-A_2$ dihedral angle as a function of the $C_2-A_2-B_1$ angle from MD simulations of cellobiose. The inset in Figure 2 shows the contour plot from atomistic MD simulations. Comparing the inset in Figure 2 with Figure 3, which shows the contour plot from CG MD simulations where the bonded interactions are obtained using the force-matching method, we clearly see that CG simulations do not agree with the atomistic simulations. Instead, the CG simulations favor the “flipped” configuration shown in the snapshot in Figure 3.

To work around this issue, we used a hybrid method where the bonded interactions are obtained from Boltzmann inversion and the nonbonded interactions are obtained from force matching. This approach has been used previously by Liu et al.²¹ in the coarse-graining of glucose. In Boltzmann inversion, we first calculate the histograms $h(r)$ for each bond length, angle, and

dihedral angle and then apply the following equation:

$$U(r) = -k_B T \ln[h(r)] \quad (5)$$

to obtain the interaction potential $U(r)$, where k_B is the Boltzmann constant and T is the temperature. Finally, we use numerical derivatives to calculate the force from the potential in eq 5. In the hybrid method, we first calculate both the bonded and the nonbonded interactions using force matching. We then replace the force-matched bonded interactions with the Boltzmann inverted bonded interactions. Note that there is a minor inconsistency here: the nonbonded interactions are no longer completely consistent since the force-matched bonded interactions have been exchanged for the Boltzmann inverted bonded interactions. To check that our approach is accurate, we again conducted force matching with a new force data set, where the histogram-based bonded interactions were subtracted. The result is a new set of nonbonded interactions that was then used with the histogram-based bonded interactions. We then ran CG MD simulations using this new force field. For both cellobiose and cellotetraose, the results were indistinguishable from the results with the original, inconsistent hybrid force

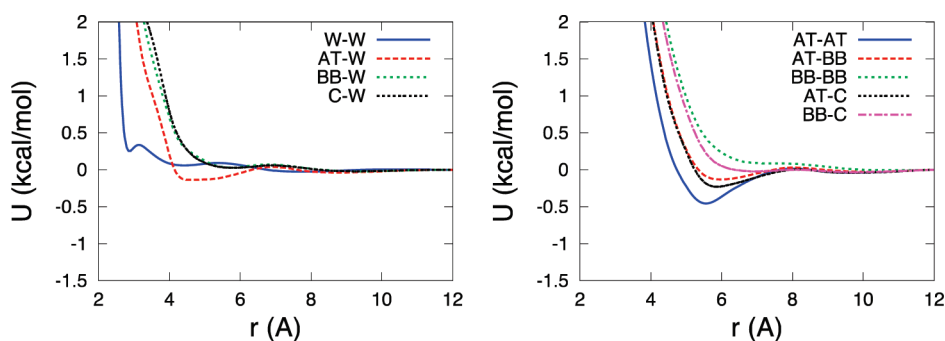


Figure 8. Nonbonded pair-potential between CG water (W) and beads AT, BB, and C, in the CG glucose model.

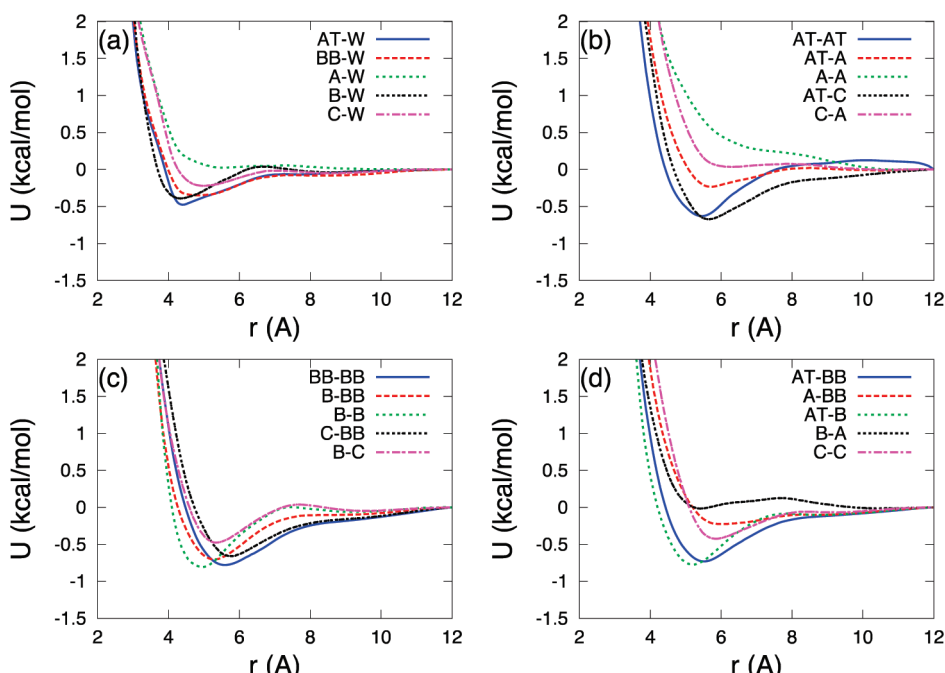


Figure 9. Nonbonded interactions in CG cellobiose model.

field without the subtraction of the histogram-based bonded interactions. Therefore, we decided to use the original hybrid force field.

Figure 2 shows the contour plot for the hybrid method. Comparing the main figure with the atomistic result in the inset, we see that the hybrid method results in cellobiose configurations that closely agree with the atomistic data. In summary, the coarse-graining process for cellobiose and cellotetraose was as follows:

- Step 1 Perform atomistic simulations
- Step 2 Calculate nonbonded CG force field using force matching
- Step 3 Calculate bonded CG force field using Boltzmann inversion
- Step 4 Perform the CG simulations using the force fields from steps 2 and 3

Implementation in CHARMM. To perform the CG MD simulations using CHARMM, we modified CHARMM to use lookup tables for bond length, angular, and dihedral interactions. To prevent CHARMM from assigning nonbonded interactions between beads in the same cellobiose or cellotetraose unit, we added “zero-force” bonded interactions to the model. These are simply bonded force lookup tables that are set to zero at all distances. For cellobiose, we

added AT–A₂, AT–BB, and B₁–BB bonds and AT–BB–C, B₁–BB–C, and C–B₁–BB angles. For cellotetraose, we added AT–A₃, AT–A₄, B₁–B₂, B₁–B₃, B₂–BB, C₁–B₂, C₁–B₃, C₁–B₄, C₃–B₁, A₃–C₁, A₄–C₁, C₁–BB, C₂–BB, C₃–BB, A₁–A₂, A₁–A₃, A₁–CB, B₁–CB, B₂–CB, B₃–CB, C₁–BB, C₂–BB, and C₃–BB bonds and AT–A₂–B₂, AT–A₂–B₃, AT–A₂–BB, AT–A₂–C₂, AT–A₂–C₃, AT–A₂–CB, B₁–B₂–A₃, B₁–B₂–A₄, C₁–AT–C₃, C₁–AT–CB, A₂–AT–B₃, A₂–AT–BB, B₂–AT–A₄, and A₃–AT–CB angles.

In our force-matching program, the nonbonded force field was represented using lookup tables with bin width $\Delta x = 0.05$ Å and low and high cutoffs of 2.3 Å and 12.0 Å, respectively. The bonded, angular, and dihedral force lookup tables had bin widths of $\Delta x = 0.02$ Å, $\Delta\theta = 1.0^\circ$, and $\Delta\varphi = 1.0^\circ$, respectively. The program automatically detects and sets the upper and lower bounds of bonded, angular, and dihedral force tables.

RESULTS AND DISCUSSION

Coarse-Grained Force Fields. Figures 4 and 5 show the bonded pair potentials for glucose, cellobiose, and cellotetraose.

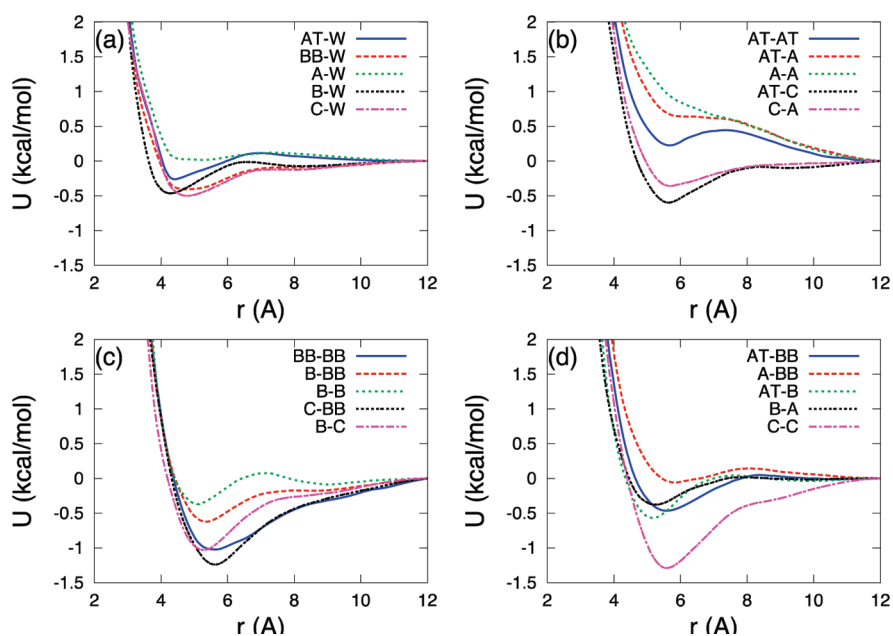


Figure 10. Nonbonded interactions in the CG cellotetraose model.

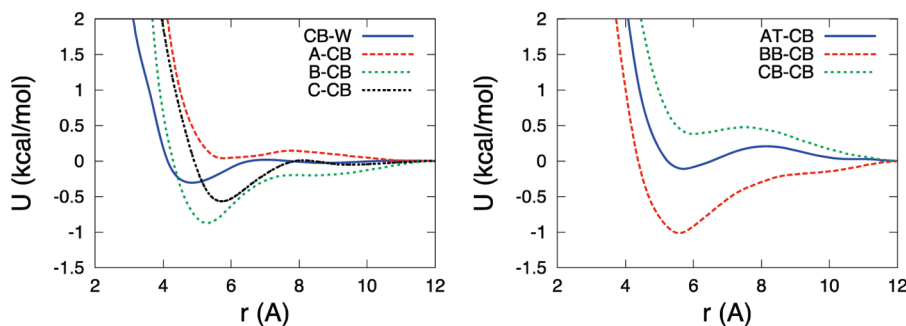


Figure 11. Nonbonded interactions in the CG cellotetraose model (continued from Figure 10).

We are showing the postprocessed potentials where we have added repulsive barriers on the edges of the potentials. In practice, this was done via a linear extrapolation of the force data. This gives rise to quadratic repulsion in the potential energy.

To facilitate a comparison between the pair potentials, the graphs in Figures 4 and 5 are plotted on the same axis scale. The location of the global minimum of the glucose AT–BB potential matches with the cellobiose and cellotetraose A–BB potentials, but does not match with the cellobiose and cellotetraose AT–B or A–B potentials. This can be explained by the difference in the center of mass between the AT and A and the BB and B beads. The AT bead only has an additional hydrogen atom compared to A; hence the center of mass of AT is very similar to that of A. Compared to B, the BB bead has an additional oxygen and hydrogen atoms; hence the center of mass of BB is different from that of B.

The B–A bonded potential is shown Figure 4b and d for cellobiose and cellotetraose, respectively. We see that the B–A potential agrees well between cellobiose and cellotetraose, both having a global minimum at around 3.5 Å, and a local minimum at around 2.6 Å. The C–C bonded pair potential is shown in Figure 5b and c for cellobiose and cellotetraose, respectively. As

can be seen, the C–C potential agrees well between cellobiose and cellotetraose. The minimum in the C–C potential is at 6.3 Å, which compares well with the minimum in the crystalline cellulose where the equilibrium distance is 6.7 Å.

The double well behavior shown in Figure 4a for the BB–C interaction is due to the change in conformation of the primary alcohol group in the atomistic simulations. This behavior is seen in multiple cases for glucose, cellobiose, and cellotetraose in Figure 4b–d and in Figure 5a. The local minimum in Figure 4a for the AT–BB interaction is due to the carbon ring puckering. The same effect of ring puckering is seen in Figure 4b and Figure 5a for cellobiose and cellotetraose. Ring puckering is a relatively rare event occurring approximately once for one of the 64 glucose molecules during a 20 ns simulation. Therefore, the lack of puckering in Figure 4c and d is most likely due to inadequate sampling in these rarely visited regions.

Figure 6a and b show the angular potentials for cellobiose and cellotetraose, respectively. Comparing the figures, we see that the angular potentials for the two models are in good agreement.

Figure 7a shows the dihedral potentials for cellobiose, and Figure 7b,c show the dihedral potentials for cellotetraose. Comparing Figure 7a–c, we see that these dihedral potentials

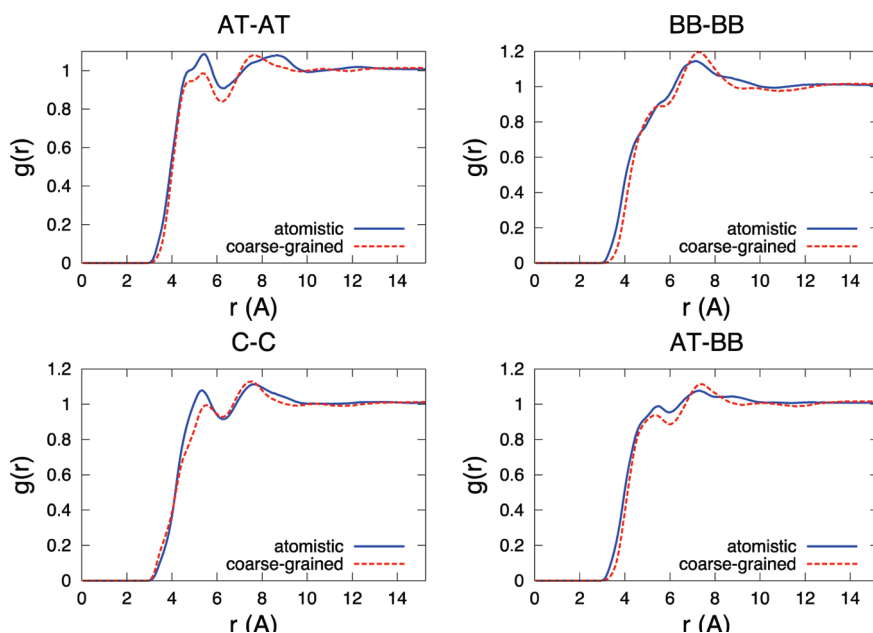


Figure 12. Radial distribution functions from atomistic (solid line) and CG (dashed line) simulations of glucose for AT–AT, BB–BB, C–C, and AT–BB bead pairs.

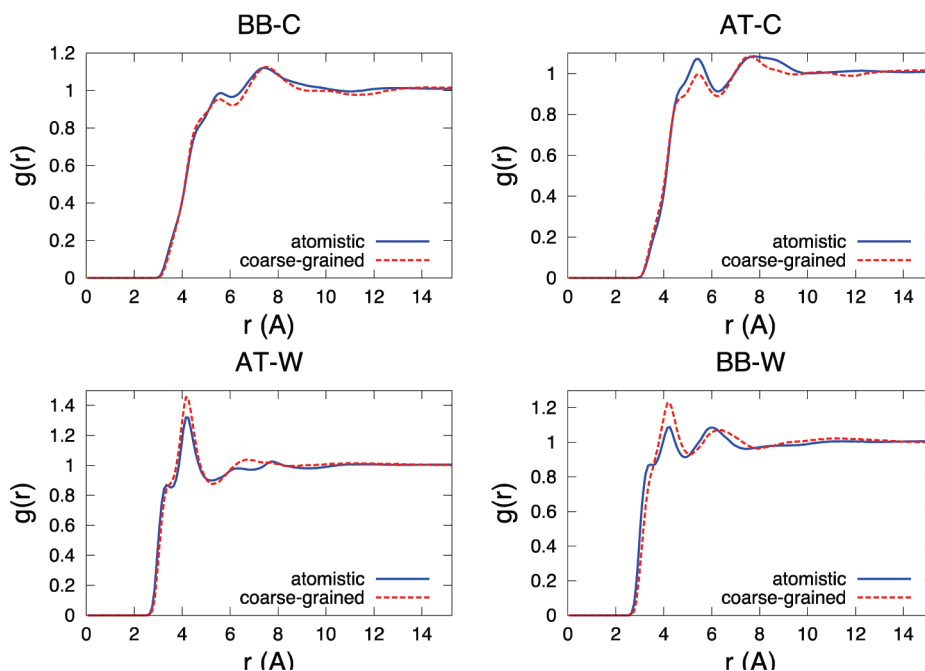


Figure 13. Radial distribution functions from atomistic (solid line) and CG (dashed line) simulations of glucose for BB–C, AT–C, AT–W, and BB–W bead pairs.

agree quite well between cellobiose and cellotetraose. The only major difference is in the A–B–C–A potential shown in Figure 7c, which has a secondary local minimum at around 90° .

It is interesting to note that, as can be seen from Figures 4–7, almost all bonded potentials are nonharmonic and therefore cannot be fitted with an analytical harmonic function such as has been done in some previous work.^{20,21}

Figure 8 shows the nonbonded interaction potentials for the CG glucose model. As shown in Figure 8, the nonbonded

interaction potentials cannot be easily described by a functional potential, such as the Lennard-Jones potential.

Figure 9 shows the nonbonded interactions in the CG cellobiose model. Figures 10 and 11 show the nonbonded interactions for the CG cellotetraose model. For the cellobiose force field to be suitable for use as a force field for CG cellulose, it should have nonbonded interactions that are similar to the nonbonded interactions in the cellotetraose model. It is therefore interesting to compare the nonbonded potentials in the cellobiose

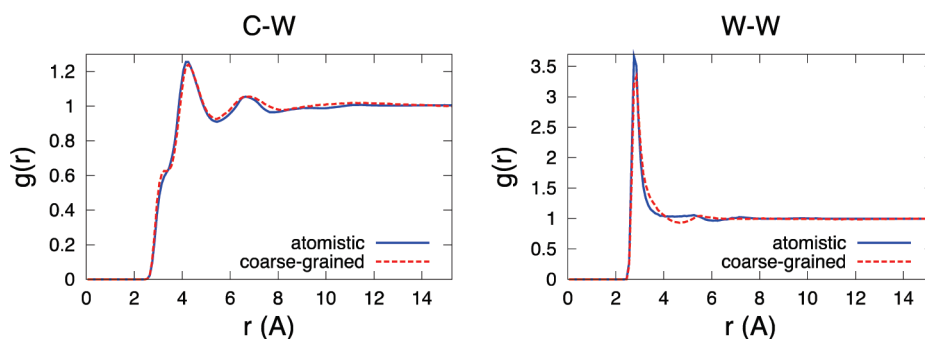


Figure 14. Radial distribution functions from atomistic (solid line) and CG (dashed line) simulations of glucose for C–W and W–W bead pairs.

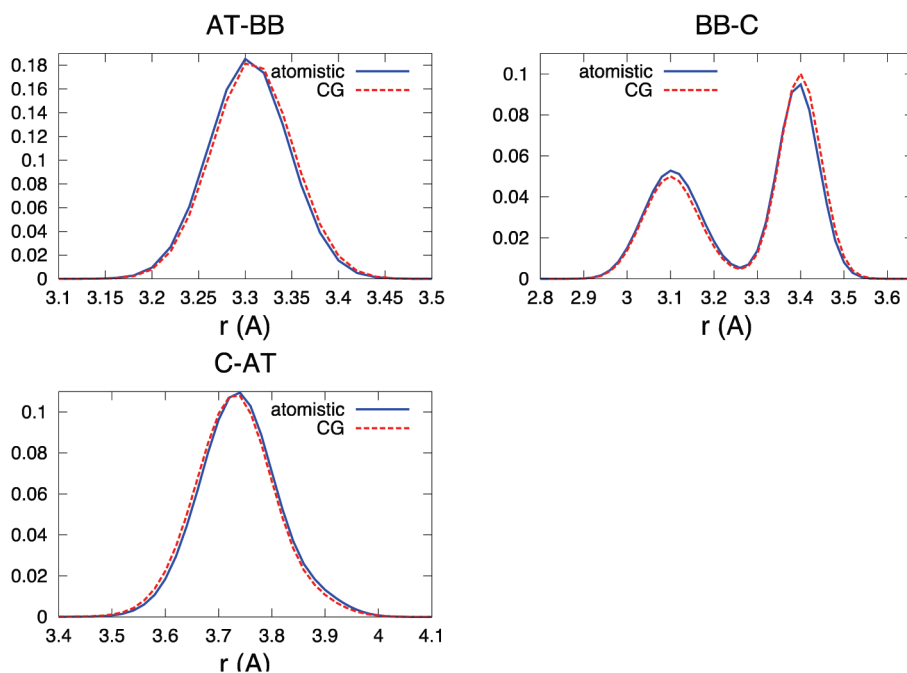


Figure 15. Histograms of bonds from the atomistic (solid line) and CG (dashed line) simulations of glucose.

and cellotetraose models. We make the comparison for the B–A, B–C, and C–A nonbonded potentials. These potentials are shown in Figure 9b–d and Figure 10b–d for cellobiose and cellotetraose, respectively. For example, we compare the B–C interaction potential. As shown, the B–C potential has an attractive well around 5.5 Å for both cellobiose and cellotetraose. However, the depth of the well for cellotetraose is about twice as much as it is for cellobiose. A similar comparison for B–A and C–A nonbonded potentials shows the same trend where the cellotetraose potential is always more attractive than the cellobiose potential.

Comparison of Coarse-Grain and Atomistic Simulations.

To validate the coarse-grain models, we compare the radial distribution functions (RDF) from the atomistic and coarse-grain simulations. The atomistic RDF was calculated by mapping the CG beads onto the atomistic trajectory. To avoid the RDFs being dominated by the bonded interactions, we exclude the contributions from beads in the same glucose, cellobiose, or cellotetraose units.

Figures 12–14 show the comparison for glucose. As we can see, the CG and atomistic results are in good agreement for glucose.

Figure 15 shows the histograms of the bonds AT–BB, BB–C, and C–AT from the atomistic and CG simulations of glucose. As we can see, the agreement between the atomistic and CG simulations is excellent.

Figures 16 and 17 show selected RDFs for cellobiose and cellotetraose, respectively. As we can see, the agreement with the atomistic results is not as good for cellobiose and cellotetraose than for glucose. From Figures 16 and 17, we see that the agreement is very good for the RDF between the C bead and CG water (W). This is true in general for both cellobiose and cellotetraose; i.e., the agreement is very good for all RDFs that are between a bead and CG water (including the water–water RDF). The rest of the RDFs for both cellobiose and cellotetraose are shown in the Supporting Information.

Our eventual goal is to create a CG model for cellulose. To model the cellulose fibril crystal structure correctly, one has to correctly model the intrachain A–B, A–C, and B–C interactions. Figure 17 shows the comparison between the atomistic and CG RDFs for A–B, A–C, and B–C, respectively. As can be seen from the figures, the location of the first peak matches well between the two results. This suggests that our CG force field for

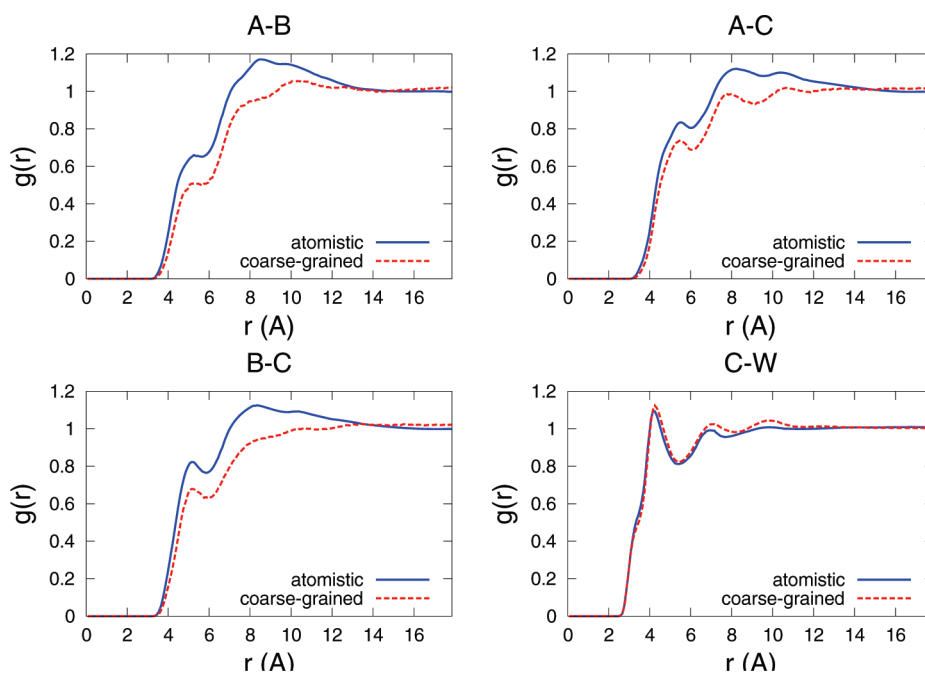


Figure 16. Radial distribution functions from atomistic (solid line) and CG (dashed line) simulations of cellobiose.

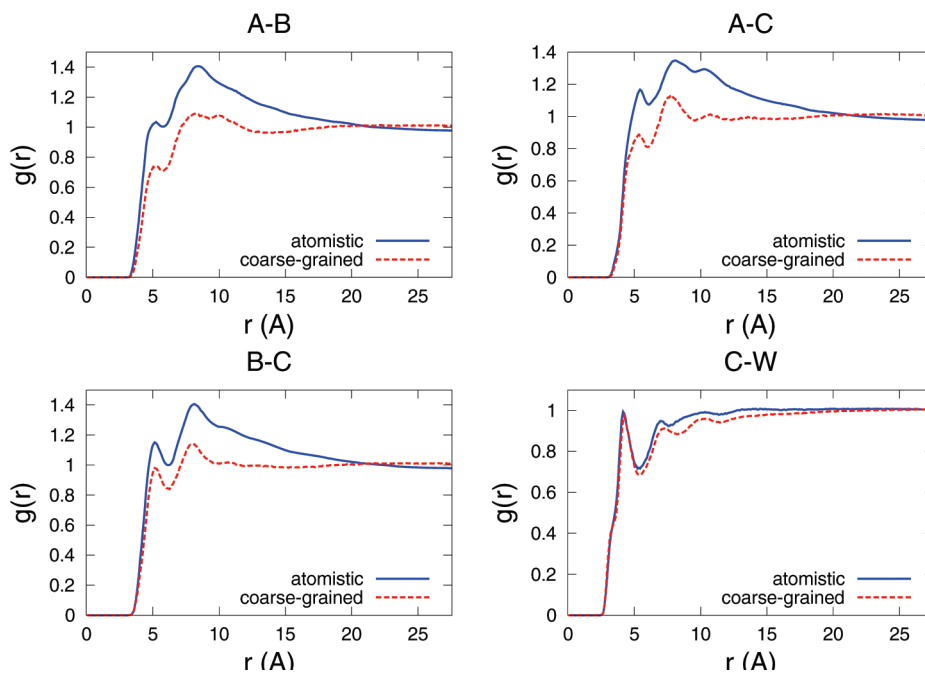


Figure 17. Radial distribution functions from atomistic (solid line) and CG (dashed line) simulations of cellotetraose.

cellotetraose has the potential for correctly reproducing the cellulose fibril crystal structure, which we will examine in a future study.

Figure 18 shows the B–C–C–A dihedral angle probability distribution for cellotetraose from atomistic and CG simulations. As can be seen, the shapes of the distributions match fairly well at the most stable configuration, around -100° . We also see that the finer details of the distribution, such as the metastable configuration at -25° , is not reproduced by the CG force field.

Finally, we make a comparison for the end-to-end distance for a single cellotetraose chain in water. We ran both atomistic and

CG simulations of a single cellotetraose chain and 1470 water molecules in a box with a size of $29.22 \times 38.75 \times 39.10 \text{ \AA}^3$. Figure 19 shows the probability distribution $p(L/n;n) = n \times P(L/n;n)$ of the end-to-end distance L . (Note that this is the same order parameter that was used in the work by Shen et al.⁴) The solid (blue) and dashed (red) lines plot the probability distribution between AT and BB beads from the atomistic and CG simulations, respectively. As can be seen, the location of the maximum of the probability distribution agrees fairly well, although the shape of the probability distribution is broader in the CG results.

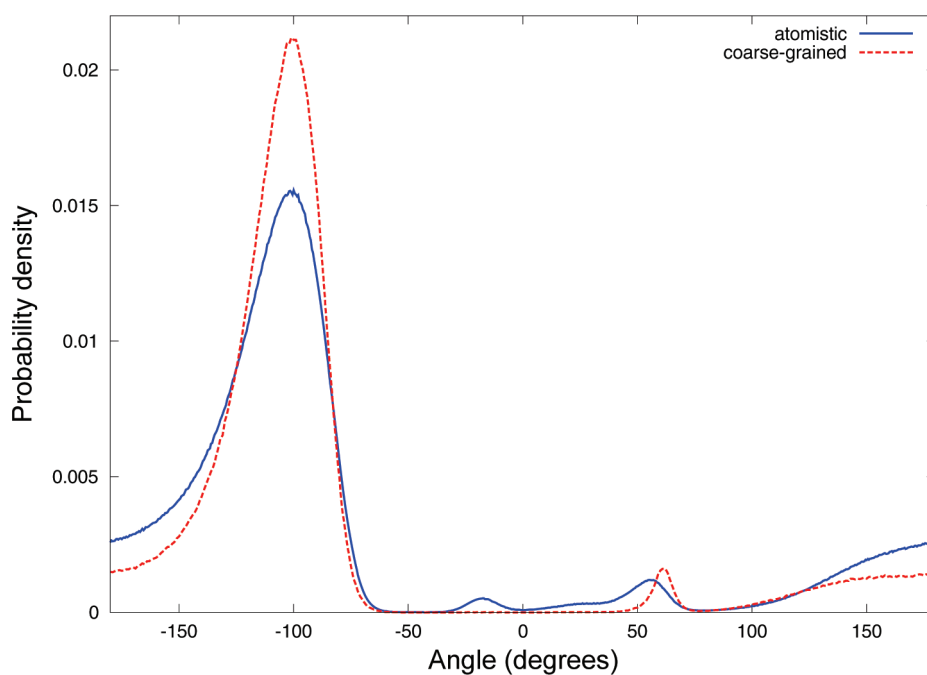


Figure 18. B–C–C–A dihedral angle distribution for cellotetraose from atomistic (solid line) and CG (dashed line) MD simulations.

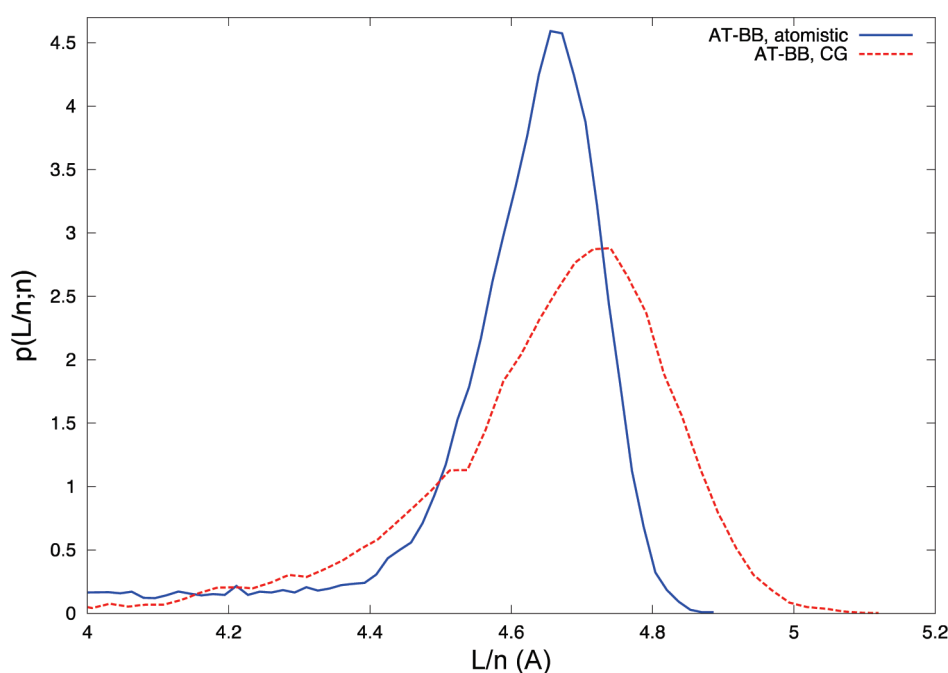


Figure 19. Probability distribution of the end-to-end chain distance for cellotetraose. The solid (blue) and dashed (red) lines show the result for the distance between AT and BB beads.

CONCLUSIONS

We have developed a CG force field for glucose, cellobiose, and cellotetraose using a combination of force matching and Boltzmann inversion. In our CG model, every glucose unit consists of three spherical beads, and water is modeled as a single bead. We note that for all three models, both the bonded and the nonbonded interaction potentials had highly nonanalytical features; i.e., fitting of the CG force field with Lennard-Jones type

and harmonic potentials would likely not be accurate. This model represents the first step to develop an accurate CG model for cellulose crystals in solution, which will make tractable multiple studies on cellulose biosynthesis, interconversion between different cellulose structures, and material properties of cellulose structure. Future research involves extending the CG model for longer cellulose fibrils by extending the cellotetraose model to cellohexaose, then applying the force field to long cellulose fibrils.

■ ASSOCIATED CONTENT

S Supporting Information. Additional radial distribution functions are given, and the bonded and nonbonded interactions force field is provided in an electronic form. This information is available free of charge via the Internet at <http://pubs.acs.org/>

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Mark.Nimlos@nrel.gov.

■ ACKNOWLEDGMENT

This work was supported by the National Renewable Energy Laboratory Directed Research & Development program. Computational resources for this research were supported in part by the Golden Energy Computing Organization and the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory. Computer time was provided in part by the NREL Computational Sciences Center supported by the DOE Office of EERE under contract number DE-AC36-08GO28308. We thank Professor Jih-Wei Chu of UC Berkeley for helpful discussions.

■ REFERENCES

- Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804.
- Matthews, J. F.; Skopec, C. E.; Mason, P. E.; Zuccato, P.; Torget, R. W.; Sugiyama, J.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2006**, *341*, 138.
- Bergenstrahle, M.; Wohler, J.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2010**, *345*, 2060.
- Shen, T.; Langan, P.; French, A. D.; Johnson, G. P.; Gnanakaran, S. *J. Am. Chem. Soc.* **2009**, *131*, 14786.
- Yui, T.; Hayashi, S. *Biomacromolecules* **2007**, *8*, 817.
- Yui, T.; Hayashi, S. *Cellulose* **2009**, *16*, 151.
- Yui, T.; Nishimura, S.; Akiba, S.; Hayashi, S. *Carbohydr. Res.* **2006**, *341*, 2521.
- Yui, T.; Okayama, N.; Hayashi, S. *Cellulose* **2010**, *17*, 679.
- Gross, A. S.; Chu, J.-W. *J. Phys. Chem. B* **2010**, *114*, 13333.
- Bellesia, G.; Asztalos, A.; Shen, T.; Langan, P.; Redondo, A.; Gnanakaran, S. *Acta Crystallogr., Sect. D* **2010**, *66*, 1184.
- Ford, Z. M.; Stevens, E. D.; Johnson, G. P.; French, A. D. *Carbohydr. Res.* **2005**, *340*, 827.
- French, A. D.; Johnson, G. P. *Cellulose* **2009**, *16*, 959.
- Yoneda, Y.; Mereiter, K.; Jaeger, C.; Brecker, L.; Kosma, P.; Rosenau, T.; French, A. D. *J. Am. Chem. Soc.* **2008**, *130*, 16678.
- Zhong, L.; Matthews, J. F.; Crowley, M. F.; Rignall, T.; Talon, C.; Cleary, J. M.; Walker, R. C.; Chukkapalli, G.; McCabe, C.; Nimlos, M. R.; Brooks, C. L.; Himmel, M. E.; Brady, J. W. *Cellulose* **2008**, *15*, 261.
- Zhong, L. H.; Matthews, J. F.; Hansen, P. I.; Crowley, M. F.; Cleary, J. M.; Walker, R. C.; Nimlos, M. R.; Brooks, C. L.; Adney, W. S.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2009**, *344*, 1984.
- Bergenstrahle, M.; Berglund, L. A.; Mazeau, K. *J. Phys. Chem. B* **2007**, *111*, 9138.
- Beckham, G. T.; Matthews, J. F.; Peters, B.; Bomble, Y. J.; Himmel, M. E.; Crowley, M. F. *J. Phys. Chem. B* **2011**, *115*, 4118.
- Payne, C. M.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. *J. Phys. Chem. Lett.* **2011**, *2*, 1546.
- Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Huenenberger, P. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2009**, *5*, 3195.
- Molinero, V.; Goddard, W. *J. Phys. Chem. B* **2004**, *108*, 1414.
- Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566.
- Shen, T.; Gnanakaran, S. *Biophys. J.* **2009**, *96*, 3032.
- Bu, L.; Beckham, G. T.; Crowley, M. F.; Chang, C. H.; Matthews, J. F.; Bomble, Y. J.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R. *J. Phys. Chem. B* **2009**, *113*, 10994.
- Beckham, G. T.; Bomble, Y. J.; Bayer, E. A.; Himmel, M. E.; Crowley, M. F. *Curr. Opin. Biotech.* **2010**, *22*, 231.
- Kroon-Batenburg, L. M. J.; Bouma, B.; Kroon, J. *Macromolecules* **1996**, *29*, 5695.
- Liu, H. B.; Sale, K. L.; Holmes, B. M.; Simmons, B. A.; Singh, S. *J. Phys. Chem. B* **2010**, *114*, 4293.
- Ramakrishnan, S.; Collier, J.; Oyetunji, R.; Stutts, B.; Burnett, R. *Bioresour. Technol.* **2010**, *101*, 4965.
- Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D. *J. Am. Chem. Soc.* **2002**, *124*, 4974.
- Singh, S.; Simmons, B. A.; Vogel, K. P. *Biotechnol. Bioeng.* **2009**, *104*, 68.
- Hu, S.-Q.; Gao, Y.-G.; Tajima, K.; Sunagawa, N.; Zhou, Y.; Kawano, S.; Fujiwara, T.; Yoda, T.; Shimura, D.; Satoh, Y.; Munekata, M.; Tanaka, I.; Yao, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 17957.
- Paredes, A. R.; Somerville, C. R.; Ehrhardt, D. W. *Science* **2006**, *312*, 1491.
- Fontes, C. M. G. A.; Gilbert, H. J. *Annu. Rev. Biochem.* **2010**, *79*, 655.
- Bayer, E. A.; Belaich, J.-P.; Shoham, Y.; Lamed, R. *Annu. Rev. Microbiol.* **2004**, *58*, 521.
- Bomble, Y. J.; Beckham, G. T.; Matthews, J. F.; Nimlos, M. R.; Himmel, M. E.; Crowley, M. F. *J. Biol. Chem.* **2011**, *286*, 5614.
- Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869.
- Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730.
- Reith, D.; Putz, M.; Muller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624.
- Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469.
- Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.
- Ercolessi, F.; Adams, J. B. *Europhys. Lett.* **1994**, *26*, 583.
- Ruhle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theory Comput.* **2009**, *5*, 3211.
- Noid, W. G.; Liu, P.; Wang, Y.; Chu, J. W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 244115.
- Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2009**, *113*, 4443.
- Izvekov, S.; Voth, G. A. *J. Chem. Theory Comput.* **2006**, *2*, 637.
- Shi, Q.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 15045.
- Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 954.
- Krishna, V.; Noid, W. G.; Voth, G. A. *J. Chem. Phys.* **2009**, *131*.
- Das, A.; Andersen, H. C. *J. Chem. Phys.* **2010**, *132*, 164106.
- Larini, L.; Lu, L.; Voth, G. A. *J. Chem. Phys.* **2010**, *132*, 164107.
- Molinero, V.; Goddard, W. A. *Phys. Rev. Lett.* **2005**, *95*, 045701.
- Bu, L.; Himmel, M. E.; Nimlos, M. R. In *Computational Modeling in Lignocellulosic Biofuel Production*; Nimlos, M. R., Crowley, M. F., Eds.; American Chemical Society: Washington, DC, 2010; pp 99–117.
- Im, W.; Feig, M.; Brooks, C. L. I. *Biophys. J.* **2003**, *85*, 2900.
- Im, W.; Lee, M. S.; Brooks, C. L. I. *J. Comput. Chem.* **2003**, *24*, 1691.
- Beckham, G. T.; Matthews, J. F.; Bomble, Y. J.; Bu, L.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R.; Crowley, M. F. *J. Phys. Chem. B* **2010**, *114*, 1447.
- Wohler, J.; Berglund, L. A. *J. Chem. Theory Comput.* **2011**, *7*, 753.
- Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074.
- Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.
- Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

- (59) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926.
- (60) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (61) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- (62) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2008**, *29*, 2543.
- (63) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2009**, *9*, 2353.
- (64) Anderson, E.; Bai, Z.; Bischof, C.; Blackford, S.; Demmel, J.; Dongarra, J.; Croz, J. D.; Greenbaum, A.; Hammarling, S.; McKenney, A.; Sorensen, D. *LAPACK Users' Guide*, 3rd ed.; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1999.

Application of Molecular Dynamics Simulations in Molecular Property Prediction. 1. Density and Heat of Vaporization

Junmei Wang^{*,†} and Tingjun Hou^{*,‡}

[†]Department of Pharmacology, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9050, United States

[‡]Institute of Nano & Soft Materials (FUNSOM) and Jiangsu Key Laboratory for Carbon-Based Functional Materials & Devices, Soochow University, Suzhou, Jiangsu 215123, P. R. China

S Supporting Information

ABSTRACT: Molecular mechanical force field (FF) methods are useful in studying condensed phase properties. They are complementary to experiments and can often go beyond experiments in atomic details. Even if a FF is specific for studying structures, dynamics, and functions of biomolecules, it is still important for the FF to accurately reproduce the experimental liquid properties of small molecules that represent the chemical moieties of biomolecules. Otherwise, the force field may not describe the structures and energies of macromolecules in aqueous solutions properly. In this work, we have carried out a systematic study to evaluate the General AMBER Force Field (GAFF) in studying densities and heats of vaporization for a large set of organic molecules that covers the most common chemical functional groups. The latest techniques, such as the particle mesh Ewald (PME) for calculating electrostatic energies and Langevin dynamics for scaling temperatures, have been applied in the molecular dynamics (MD) simulations. For density, the average percent error (APE) of 71 organic compounds is 4.43% when compared to the experimental values. More encouragingly, the APE drops to 3.43% after the exclusion of two outliers and four other compounds for which the experimental densities have been measured with pressures higher than 1.0 atm. For the heat of vaporization, several protocols have been investigated, and the best one, P4/ntt0, achieves an average unsigned error (AUE) and a root-mean-square error (RMSE) of 0.93 and 1.20 kcal/mol, respectively. How to reduce the prediction errors through proper van der Waals (vdW) parametrization has been discussed. An encouraging finding in vdW parametrization is that both densities and heats of vaporization approach their “ideal” values in a synchronous fashion when vdW parameters are tuned. The following hydration free energy calculation using thermodynamic integration further justifies the vdW refinement. We conclude that simple vdW parametrization can significantly reduce the prediction errors. We believe that GAFF can greatly improve its performance in predicting liquid properties of organic molecules after a systematic vdW parametrization, which will be reported in a separate paper.

1. INTRODUCTION

Molecular dynamics simulation has become increasingly important in studying the structures, dynamics, and functions of biomolecules.¹ Nevertheless, it is also an important tool for predicting a variety of molecular properties.² The power of MD simulations lies in that it can provide atomic details on the dynamics of a simulation system under conditions that are beyond the scope of experiments. Even for some specific force fields, such as CHARMM,^{3,4} AMBER,^{5,6} OPLS,⁷ and GROMOS,^{8,9} which were developed mainly to study the biomolecules, it is often critical for them to reproduce the bulk properties of some small molecules that represent the moieties of biomolecules. With the continually increased demand of “ideal” molecular mechanical models to study ligand–protein and ligand–nucleic acid interactions, it is important to develop a general purpose force field that is totally compatible with one or a set of biomolecular force fields for studying the interactions between biomolecules and organic molecules. The General AMBER force field (GAFF)¹⁰ was developed in this spirit to study biomolecule–ligand interactions in conjunction with the AMBER force

fields.^{5,6}

$$V_{\text{potential}} = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\varphi - \gamma)] + \sum_{i < j} \left\{ \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] + \frac{q_i q_j}{\epsilon R_{ij}} \right\} \quad (1)$$

To maintain the compatibility between GAFF and the biomolecular AMBER force fields, GAFF applies the same harmonic function form used by the AMBER additive force fields (eq 1), and it has been developed using the same force field parametrization strategy. The total potential energy is a sum of the intramolecular bonded terms (the first three terms) and intra- and intermolecular nonbonded terms (the last term). In eq 1, the first three terms represent the contribution to the total energy from bond stretching, bond angle bending, and torsion angle twisting, respectively. The last term is the sum of van der Waals

Received: February 24, 2011

Published: May 30, 2011

Table 1. List of the Experimental Densities (g cm^{-3}) and Heats of Vaporization (kcal/mol) for 71 Compounds (Temperatures are in $^{\circ}\text{C}$)

no.	name	density	T_{density}	ΔH_{vap}	T_{vap}	# res ^a	# atom ^a
1	ethane	0.546	-88.63	3.51 ^c	-88.63	196	8
2	propane	0.581	-42.07	4.49 ^d	-42.07	216	11
3	butane	0.602	-0.5	5.02 ^e	-0.5	196	14
4	isobutane	0.551	25	4.57	25	216	14
5	cyclohexane	0.774	25	7.89 \pm 0.48 ^f	25	180	18
6	propene	0.505 ^b	25	3.40	-47.65	324	9
7	<i>trans</i> -2-butene	0.598	25	5.15	25	216	12
8	1,3-butadiene	0.615 ^b	25	4.99	25	216	10
9	1-butyne	0.678	0	5.58	25	216	10
10	benzene	0.877	20	7.89 \pm 0.48 ^f	25	216	12
11	water	0.997	25	10.51	25	624	3
12	methanol	0.791	20	8.84 \pm 0.48 ^f	25	216	6
13	ethanol	0.789	20	10.04 \pm 0.48 ^f	25	245	9
14	propanol	0.800	25	11.35 \pm 0.10 ^f	25	216	12
15	2-propanol	0.781	25	10.85	25	180	12
16	2-methyl-2-propanol	0.781	25	11.14	25	180	15
17	phenol	1.055	45	13.82	25	180	13
18	<i>m</i> -cresol	1.034	20	14.75	25	180	16
19	methanethiol	0.888	5.96	5.87	5.96	180	6
20	ethanethiol	0.832	25	6.52	25	245	9
21	propanethiol	0.836	25	7.62	25	216	12
22	dimethyl ether	0.735	-24.6	5.14	-24.6	216	9
23	ethyl methyl ether	0.721	7.35	5.91	7.35	196	12
24	diethyl ether	0.714	20	6.48	25	180	15
25	dimethoxymethane	0.854	25	6.90	25	180	13
26	dimethyl sulfide	0.848	20	6.61	25	196	9
27	ethyl methyl sulfide	0.837	25	7.61	25	180	12
28	dimethyl disulfide	1.057	25	9.06 ^c	25	180	10
29	acetaldehyde	0.783	18	6.24 ^g	25	216	7
30	propanal	0.791	25	7.10 \pm 0.05 ^f	25	216	10
31	acetone	0.785	25	7.47 ^c	25	180	10
32	butanone	0.800	25	8.35 ^c	25	216	13
33	acetic acid	1.045	25	12.33 ^c	25	180	8
34	propanoic acid	0.988	25	13.15 ^c	25	256	11
35	methyl formate	0.971	20	6.93 \pm 0.48 ^f	25	196	8
36	methylamine	0.656 ^b	25	5.59	25	245	7
37	propylamine	0.717	20	7.47	25	147	13
38	dimethyl amine	0.680	0	6.08 ^c	25	245	10
39	diethylamine	0.706	20	7.48	25	192	16
40	trimethylamine	0.627 ^b	25	5.18	25	180	13
41	triethylamine	0.728	20.000	8.33	25	216	22
42	aniline	1.022	20	12.91 \pm 0.96 ^f	25	180	14
43	acetonitrile	0.786	20	7.87	25	196	6
44	nitromethane	1.137	20	9.17 ^c	25	216	7
45	N-methylformamide	1.011	19	13.43	25	216	9
46	N,N-dimethyl formamide	0.945	25	11.37 ^h	25	216	12
47	acetamide	0.981	100	13.40	221.15	245	9
48	N-methyl acetamide	0.894	100	13.30	100	196	12
49	N-ethylacetamide	0.942	4			196	15
50	N,N-dimethylacetamide	0.936	25	11.75	25	216	15
51	N-methylpropanamide	0.931	25	15.50	25	180	15
52	dimethyl sulfoxide	1.101	25	10.30	189	196	10
53	methyl phosphate	1.214	20			216	17

Table 1. Continued

no.	name	density	T_{density}	ΔH_{vap}	T_{vap}	# res ^d	# atom ^d
54	triphenyl phosphine oxide	1.212	23			120	34
55	fluorobenzene	1.023	20	8.29 ^c	25	180	12
56	trichloromethane	1.479	25	7.48	25	216	5
57	tetrachloromethane	1.594	20	7.75	25	180	5
58	bromomethane	1.676	20	5.45	25	216	5
59	iodomethane	2.279	20	6.68	25	216	5
60	tetrahydrofuran	0.883	25	7.65	25	216	13
61	1,3-dioxolane	1.060	25	8.48 ⁱ	25	216	11
62	pyrrolidine	0.859	20	8.97 ^c	25	180	14
63	morpholine	1.001	20	8.87	128	216	15
64	furan	0.951	20	6.62 ^c	25	216	9
65	thiophene	1.065	20	8.29	25	180	9
66	pyrrole	0.970	20	10.84 ^c	25	216	10
67	pyridine	0.982	20	9.56 ± 0.72 ^f	25	180	11
68	pyridazine	1.104	23	12.78	25	216	10
69	4-methyl thiazole	1.112	25	10.48	25	216	11
70	quinoline	1.098	15	14.18 ^g	25	180	17
71	dimethyl phosphate	1.323	20			216	14

^a # res, the number of residues (# res) in a simulation box; # atom, the number of atoms in a residue. ^b Densities were measured with pressures larger than 1 atm (propene, 1,3-butadiene, methylamine, trimethylamine). ^c Ref 23. ^d Ref 24. ^e Ref 25. ^f Average value of several pieces of experimental data, specifically, 10 data points of cyclohexane, 7 for benzene, 6 for methanol, 7 for ethanol, 8 for propanol, 8 for propanal, 6 for methyl formate, 7 for aniline, and 7 for pyridine. The individual data points are adopted from <http://webbook.nist.gov/>. ^g Ref 26. ^h Ref 27. ⁱ Ref 28. ^j Ref 29.

and electrostatic energies. A_{ij} and B_{ij} are the Lennard-Jones parameters for repulsion and attraction, which can be expressed in terms of effective van der Waals radii and well depths, R_{ij}^* and ε_{ij} (eqs 2a and 2b), which are further obtained from atomic parameters using mixing rules (eqs 2c and eq 2d). The point charges q_i and q_j are derived to reproduce the *ab initio* electrostatic potential (ESP).

$$A_{ij} = \varepsilon_{ij}(R_{ij}^*)^{12} \quad (2a)$$

$$B_{ij} = 2\varepsilon_{ij}(R_{ij}^*)^6 \quad (2b)$$

$$R_{ij}^* = R_i^* + R_j^* \quad (2c)$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \quad (2d)$$

The most striking difference between AMBER and other force fields is that the atomic partial charges in AMBER are derived to reproduce *ab initio* ESP using the program RESP (Restrained Electrostatic Potential).^{11,12} The quantum-mechanics-based charge model is more general and generalizable than a purely empirical approach which automatically includes inherent polarization of the molecules relative to their gas phase charge distribution in an average way. Fortunately, it was found that the 6-31G* basis set enhances the polarity of liquids by about the same magnitude relative to gas phase moments as does the empirical charge models.^{13,14} There are two types of charges used in GAFF, the HF/6-31G* RESP¹¹ and the AM1-BCC.^{15,16} The former is the same charge method used to develop AMBER parm94 and parm99 force fields; on the other hand, AM1-BCC, a fast charge method that was parametrized to mimic HF/6-31G* RESP charges, is more suitable for the study of a large number of molecules, such as in virtual screenings. The HF/6-31G* RESP

charge model was utilized to develop GAFF in order to maximize its compatibility with AMBER biomolecular force fields.

As van der Waals parameters generally are dominated by the inner closed shell of electrons, they are typically transferable, and only one set of parameters is needed for different atom types of an element, irrespective of its chemical environment (note that hydrogen is a special case, as it has no inner shell of electrons at all).¹³ Therefore, most of the van der Waals parameters in GAFF are adopted from the AMBER biomolecular force fields even for the newly introduced chemical functional groups. For the “hard” force field parameters, the equilibrium bond lengths r_{eq} and bond angles θ_{eq} were obtained through statistical analysis on a large number of optimized structures at the MP2/6-31G* level. The torsional angle potentials were optimized to reproduce both the experimental relative conformational energies and the high-level *ab initio* rotational profiles of model compounds.

GAFF has been extensively tested to predict the structures and energies for a variety of molecules, and an overall satisfactory performance has been achieved. Those tests include a comparison between the GAFF minimized and the crystallographic structures, the interaction energies between nucleic acid base pairs, and the relative energies of a set of conformational pairs. Moreover, GAFF has also been extensively evaluated to calculate the binding free energies by its users.^{17,18} Mobley et al. used GAFF to predict the hydration free energies in an explicit solvent for a large set of molecules.¹⁹

However, GAFF has not been evaluated on predicting the bulk properties of liquids. In this paper series, we set out to systematically study several important molecular properties using GAFF, which include the density, heat of vaporization, isobaric heat capacity, isochoric heat capacity, isothermal compressibility, thermal expansion coefficient, static dielectric constant, free energy of hydration, diffusion constant, rotational correlation time, etc. There are two objectives of this series. First of all, we

hope to establish a set of computational protocols to predict the above molecular properties using GAFF through MD simulations. Second, GAFF will be thoroughly evaluated in this procedure, and the hints of the modification of force field parameters to reduce the prediction errors can guide us to develop a new version of general AMBER force fields.

In this study, we focus on two molecular properties, density and heat of vaporization. Why do we study these two molecular properties? First of all, bulk density and heat of vaporization are almost exclusively associated with the nonbonded interactions.^{7,20} The successful optimization of nonbonded force field parameters (as the atomic partial charges in GAFF are obtained in a deterministic fashion, the nonbonded parameters mainly refer to the van der Waals parameters) to reproduce the two condensed phase data sets lays the groundwork for a molecular mechanical force field including those aimed to study biomolecules.

2. METHODS

Data Sources. In Table 1, the compound names and the experimental mass densities (ρ) and heats of vaporization (ΔH_{vap}) are listed. The 71-molecule data set covers diverse functional groups in organic chemistry, which includes hydrocarbons (aliphatic and aromatic, cyclic and acyclic), alcohols, thiols, phenols, ethers, esters, aldehydes, ketones, carboxylic acids, amines, amides, nitriles, nitro-derivatives, disulfides, thiophenes, sulfides, sulfoxides, sulfones, phosphates, halides, and heterocyclic compounds. The structures of the 71 molecules are shown in Figure S1 of the Supporting Information. The experimental values of density are adopted from the CRC Handbook of Chemistry and Physics (edition 86).²¹

Heat of vaporization (ΔH_{vap}) is defined as the enthalpy change in the conversion of 1 mol of liquid to a gas at constant temperature. In experiments, the heat of vaporization can be measured at the boiling point of the neat liquid through calorimetry or determined from the vapor pressure–temperature (P – T) plot using the Clausius–Clapeyron equation.²⁰ When high-quality P – T data over a wide temperature range are available, it is possible to derive the force field parameters to fit those data.²² All of the experimental data of ΔH_{vap} are adopted from the CRC Handbook of Chemistry and Physics (edition 86) unless explicitly noted. Other experimental data resources include the publications of Majer and Svoboda,²³ Kemp and Egan,²⁴ Pedley et al.,²⁵ Wiberg et al.,²⁶ Geller,²⁷ Pihlaja and Heikkilä,²⁸ and Steele et al.²⁹ When multiple experimental data are available for a compound, the overall guidance is to use the latest one. However, for nine compounds (nos. 5, 10, 12–14, 30, 35, 42, and 67 of Table 1), which have multiple data points (6–10 values) from different sources, the average values are used. For those compounds, the average RMS error is 0.47 kcal/mol and the largest RMS error is 0.96 kcal/mol for aniline.

Similar to QSAR (quantitative structure–activity relationship) studies,³⁰ it is important to use high-quality experimental data in force field development. Mackerell et al.²⁰ provided an excellent example on how to cherry-pick the most reliable experimental data. They found that the wide range of heat of vaporization values of *N*-methylacetamide reported in the literature is due to two basic things: incorrect reporting of the temperatures at which the original values were extracted and limitations in the quality of experimental vapor pressure–temperature data over a wide range of temperatures. We also

noticed that the experimental data of ΔH_{vap} for acetic acid and propanoic acid at 25 °C adopted by the CRC Handbook of Chemistry and Physics (edition 86) are much smaller than from the other resources. For acetic acid, the CRC value of 23.36 kJ/mol is less than half of the values reported by Majer and Svoboda (51.6 kJ/mol)²³ and Konicek and Wadso (51.6 ± 1.5 kJ/mol);³¹ for propanoic acid, the CRC value of 32.14 kJ/mol is also much smaller than the values reported by Majer and Svoboda (55.0 kJ/mol)²³ and Konicek and Wadso (55.0 ± 2.0 kJ/mol).³¹ We adopted the values of Majer and Svoboda in this work, not only because they are relatively newer, but also because the calculated heats of vaporization by both GAFF and OPLS reproduce their experimental values well.

Molecular Mechanical Models. Consistent with the strategy of parametrizing GAFF, the point charges of 71 molecules in the data set were derived by RESP to fit the HF/6-31G* electrostatic potentials which were generated using the Gaussian 03 software package.³² The other force field parameters come from GAFF in AMBER10.³³ The residue topology files were prepared using the Antechamber module³⁴ in AMBER 10.³³ For each molecule, an internal program was used to generate a rectangular parallelepiped box filled with multiple copies of the monomer (N_{res} in Table 1), and then the Leap program in AMBER10 was applied to generate the topologies.

Liquid Phase Molecular Dynamics Simulations. All of the liquid phase MD simulations were performed with the periodic boundary condition to produce isothermal–isobaric ensembles using the Sander program of AMBER10.³³ The Particle Mesh Ewald (PME) method^{35–37} was used to calculate the full electrostatic energy of a unit cell in a macroscopic lattice of repeating images. Except for TIP3P water, which is described with a special “three-point” algorithm, all degrees of freedom were constrained,³⁸ the other molecules had all the degrees of freedom free in MD simulations. The nonbonded cutoff of calculating van der Waals and electrostatic energies was set to 9.0 Å, and a continuum model correction term was added to the van der Waals energies. How nonbonded cutoff affects the MD simulations was discussed by Shirts et al.³⁹

The integration of the equations of motion was conducted at a time step of 1 fs (except TIP3P water, for which a time step of 2 fs was used). Temperature was regulated using the Langevin dynamics⁴⁰ with a collision frequency of 5 ps.^{41–43} Pressure regulation was achieved with isotropic position scaling, and the pressure relaxation time was set to 1.0 ps.

There are three phases in a liquid phase MD simulation, namely, the relaxation phase, the equilibrium phase, and the sampling phase. In the relaxation phase, the main chain atoms were gradually relaxed by applying a series of restraints, and the force constants decreased progressively: from 20 to 10, 5, and 1.0 kcal/mol/Å². For each force constant, the position-restrained MD simulation was run for 20 ps. In the following equilibrium phase, the system was further equilibrated for 2 ns without any restraint and constraint. In the sampling phases, the snapshots as well as the structural and energetic properties were recorded at an interval of 2 ps. In total, 1500 snapshots were saved for post analysis.

Gas Phase MD Simulations. Gas phase MD simulations were performed for single molecules using four different temperature regulation methods, namely, ntt (the temperature scaling keyword in AMBER) = 0, which corresponds to constant total energy classic dynamics; ntt = 1, which uses the weak-coupling algorithm;⁴⁴ ntt = 2, which utilizes the Andersen coupling

scheme;^{45,46} and $\text{ntt} = 3$, which applies the Langevin dynamics with a collision frequency of 5 ps^{-1} to scaling temperatures.⁴⁰ The systems were equilibrated for 1 ns followed by another 1 ns sampling phase, and in total 500 snapshots were recorded for post analysis.

Hydration Free Energy Calculations. It is a standard practice in today's force field development to test if molecular mechanical models can reproduce the hydration free energies of model compounds. Although GAFF has been extensively evaluated by Mobley et al.¹⁹ for this molecular property using a large data set, we need to find out how the newly derived vdW parameters affect the calculation performance. The hydration free energy of a molecule was calculated using thermodynamic integration (TI). In TI calculations, the system evolves according to a mixed potential, $V(\lambda) = (1 - \lambda)^k V_0 + [1 - (1 - \lambda)^k] V_1$, where λ and k are mixing parameters and V , V_0 , and V_1 are the mixed, the unperturbed, and perturbed potentials, respectively. The free energy change, ΔG , is calculated numerically using the following equation: $\Delta G = G_{\lambda=1} - G_{\lambda=0} = \int_0^1 \langle \partial V / \partial \lambda \rangle_\lambda d\lambda = \sum_i w_i \langle \partial V / \partial \lambda \rangle_i$. The free energy of hydration of a molecule was calculated by summing up the free energy changes in four perturbations, i.e., the gas phase and aqueous phase charge disappearing and the gas phase and aqueous phase atom disappearing. For the charge disappearing, a linear mixing rule was applied ($k = 1$); for the atom disappearing, k was set to 6, as suggested by Steinbrecher et al.⁴⁷ In each free energy perturbation, 12 different free energy simulations with the weights (w_i) fitting the Gaussian quadrature formula were performed to numerically estimate the integral.⁴⁸ A similar MD protocol to the above liquid MD simulation was applied to do the free energy simulations, except that the time step of integration was set to 2 fs and the systems were equilibrated for 1 ns followed by production for another nanosecond.

Density Calculations. The average bulk density ρ was computed from the average volume of the simulation box, $\langle V \rangle$ using eq 3, where N_{res} is the number of residues in the simulation box, M is molar mass of the molecule in the study, and N_A is the Avogadro constant. The bulk densities were printed out in the output files of MD simulations by Sander in default.

$$\langle \rho \rangle = \frac{N_{\text{res}} M}{N_A \langle V \rangle} \quad (3)$$

Heat of Vaporization. The heat of vaporization or the enthalpy of vaporization ΔH_{vap} can be calculated with eq 4, where $H(p, T)_{\text{gas}}$ and $H(p, T)_{\text{liquid}}$ are the enthalpies in the gas and liquid phases, respectively. E_{gas} and E_{liquid} are the total energies of the gas and liquid phases, respectively.

$$\begin{aligned} \Delta H_{\text{vap}}(T) &= H(p, T)_{\text{gas}} - H(p, T)_{\text{liquid}} \\ &= E_{\text{gas}}(T) - E_{\text{liquid}}(T) + p(V_{\text{gas}} - V_{\text{liquid}}) \end{aligned} \quad (4)$$

V_{liquid} is negligible compared to V_{gas} . Under the assumption that the gas is ideal so that the kinetic energies of a molecule in the gas and liquid phases are identical, eq 4 becomes eq 5, where $E_{\text{gas}}^{\text{potential}}$ and $E_{\text{liquid}}^{\text{potential}}$ are the potential energies in the gas and liquid phases, respectively; C is the correction term.

$$\Delta H_{\text{vap}}(T) = E_{\text{gas}}^{\text{potential}}(T) - E_{\text{liquid}}^{\text{potential}}(T) + RT + C \quad (5)$$

Both of the potential energy terms include the vibrational energies which are obtained through molecular simulations.

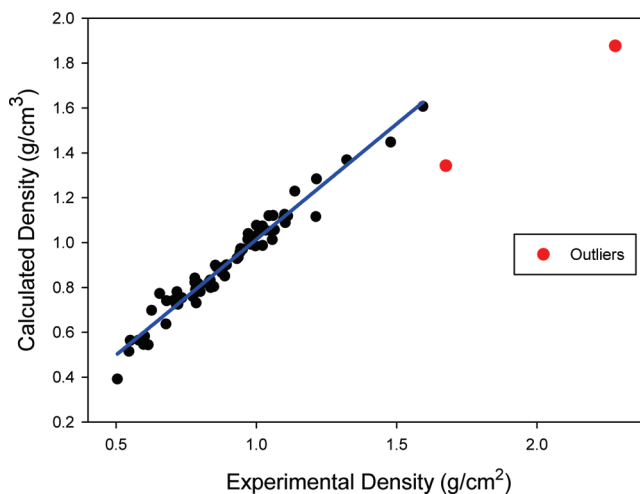


Figure 1. Correlation between experimental and calculated densities for 71 molecules. The two outliers are bromomethane and iodomethane. With the new van der Waals parameters of halides, the experimental densities can be well reproduced: 1.642 versus 1.676 for bromomethane and 2.247 versus 2.279 for iodomethane, respectively.

Note that in MD simulations the actual temperatures T_{MD} may be slightly different from the desirable temperature T . $E_{\text{gas}}^{\text{potential}}$ can also be estimated using eq 6, where $E_{\text{gas}}^{\text{minimized}}$ is the minimized energy and N_{atom} and N_{cons} are the number of atoms in the molecules and the number of the constrained degrees of freedom, respectively. Equation 6 has been used by Caldwell and Kollman¹³ and Fox and Kollman¹⁴ in calculating the heat of vaporization. $E_{\text{liquid}}^{\text{potential}}$ consists of two terms, the intramolecular energy $E_{\text{liquid}}^{\text{intra}}$ and intermolecular energy $E_{\text{liquid}}^{\text{inter}}$.

$$E_{\text{gas}}^{\text{potential}}(T) = E_{\text{gas}}^{\text{minimized}} + \frac{1}{2} RT(3N_{\text{atom}} - 6 - N_{\text{cons}}) \quad (6)$$

Equation 5 is the basic equation for calculating the heat of vaporization.^{7,49–54} The correction term C in eq 5 accounts for the difference in vibration energies calculated quantum mechanically and classically, as well as the polarization and nonideal gas effects. More details on how to calculate the correction term are presented by Horn et al.⁵⁵ In most scenarios, the correction term C is rather small, and it is usually neglected. For example, the quantum correction of the vibrational energies, the major contribution to the correction term C , is estimated to be -0.055 kcal/mol for water.⁵³

If the correction term is neglected, and all degrees of freedoms are frozen in MD simulations or the assumption that the intramolecular energy in the liquid phase is the same as that in the gas phase is applied, eq 5 can be further simplified to eq 7. We used eq 7 to calculate the heat of vaporization of TIP3P water.

$$\Delta H_{\text{vap}}(T) = E_{\text{liquid}}^{\text{inter}}(T) + RT \quad (7)$$

In this work, several computation protocols were applied to calculate the heats of vaporization of 67 molecules in the data set. In the first protocol (P1), eq 7 was used to calculate ΔH_{vap} , assuming the energies of the residue are the same in the gas and the liquid phases. In the second protocol (P2), eqs 5 and 6 were applied to calculate ΔH_{vap} . The third protocol, P3, is similar to P2 except that $E_{\text{gas}}^{\text{minimized}}$ is an average value of minimized energies of multiple conformations sampled in gas phase MD simulations. In the last protocol, P4, ΔH_{vap} was calculated using eq 8, a hybrid

Table 2. Calculation Results of Heats of Vaporization (kcal/mol) for 67 Compounds Using Four Different Basic Protocols (P1–P4)^a

no.	expt	T _{Expt}	T _{MD} ^b	P1	P2	P3		P4	
						NTT=0	NTT=3	NTT=0	NTT=3
1	3.51	-88.63	-88.31	3.31	2.68	2.68	2.91	2.68	3.31
2	4.49	-42.07	-41.64	4.52	3.43	3.43	3.88	3.43	4.47
3	5.02	-0.5	0.04	5.51	4.42	4.42	5.18	3.84	5.38
4	4.57	25	25.42	5.20	3.48	3.48	4.37	3.48	5.20
5	7.89	25	25.50	8.60	6.49	6.49	7.49	6.49	8.56
6	3.40	-47.65	-47.38	3.90	3.06	3.06	3.39	3.06	3.91
7	5.15	25	25.50	4.60	3.08	3.08	3.83	3.08	4.59
8	4.99	25	25.52	3.97	2.69	2.69	3.32	2.69	3.82
9	5.58	25	25.32	5.02	4.20	4.20	4.84	4.20	5.03
10	7.89	25	25.25	7.59	6.38	6.38	7.08	6.38	7.60
11	10.51	25	25.19	10.76	10.76	10.76	10.76	10.76	10.76
12	8.84	25	25.07	10.76	9.72	9.72	10.11	9.72	10.38
13	10.04	25	25.49	11.85	10.15	10.15	10.54	10.45	11.47
14	11.35	25	25.54	13.94	12.31	11.93	12.69	11.94	13.36
15	10.85	25	25.49	14.28	12.44	12.50	13.29	12.50	13.87
16	11.14	25	25.48	15.15	13.23	13.23	14.08	13.23	14.87
17	13.82	25	25.57	13.27	11.98	11.98	12.65	11.98	13.24
18	14.75	25	25.45	17.25	15.17	15.18	15.90	15.37	16.61
19	5.87	5.96	6.11	4.88	4.26	4.26	4.44	4.26	4.89
20	6.52	25	24.99	6.06	4.95	4.95	5.42	5.00	6.04
21	7.62	25	25.55	7.37	5.72	5.72	6.50	5.88	7.30
22	5.14	-24.6	-24.27	5.55	4.64	4.64	4.97	4.64	5.55
23	5.91	7.35	7.79	6.34	4.81	4.81	5.50	4.99	6.42
24	6.48	25	25.50	7.39	5.24	5.24	6.15	5.55	7.41
25	6.90	25	25.40	8.43	9.28	7.19	8.55	6.49	8.14
26	6.61	25	25.43	5.53	4.44	4.44	4.79	4.44	5.52
27	7.61	25	25.30	6.76	5.23	5.45	6.28	5.33	6.77
28	9.06	25	25.40	7.74	6.64	6.64	7.26	6.64	7.77
29	6.24	25	25.49	7.14	6.57	6.57	6.87	6.58	7.13
30	7.10	25	25.38	8.30	7.36	7.36	7.81	7.36	8.31
31	7.47	25	25.45	8.16	7.20	7.20	7.74	7.20	8.11
32	8.35	25	25.45	8.94	7.57	7.68	8.31	7.71	8.91
33	12.33	25	24.83	14.63	13.38	13.38	13.70	13.38	13.94
34	13.15	25	25.57	16.41	14.95	14.45	15.16	14.59	15.57
35	6.93	25	25.32	9.06	8.19	8.19	8.61	8.19	9.01
36	5.59	25	25.60	8.78	7.47	7.47	7.86	7.47	8.42
37	7.47	25	25.53	11.45	9.25	9.25	9.91	9.28	10.95
38	6.08	25	25.38	7.76	6.37	6.37	7.06	6.37	7.66
39	7.48	25	25.45	9.08	6.90	6.90	7.84	7.04	9.11
40	5.18	25	25.72	6.24	4.60	4.60	5.41	4.60	6.19
41	8.33	25	25.62	9.66	8.44	6.78	8.19	6.75	9.60
42	12.91	25	25.62	13.64	12.16	12.16	12.73	12.16	13.50
43	7.87	25	25.50	7.56	7.19	7.19	7.30	7.19	7.57
44	9.17	25	25.13	12.94	12.66	12.67	12.68	12.67	12.82
45	13.43	25	25.43	13.93	13.28	13.15	13.44	13.16	13.85
46	11.37	25	25.73	12.24	11.48	11.44	11.79	11.44	12.22
47	13.40	221.15	222.09	13.14	11.19	11.19	11.98	11.20	12.76
48	13.30	100	100.82	13.82	12.57	12.58	13.26	12.57	13.78
50	11.75	25	25.56	12.47	11.13	11.11	11.92	11.11	12.46
51	15.50	25	25.55	15.45	14.07	14.12	14.59	14.12	15.36
52	10.30	189	189.74	10.50	8.84	8.84	9.71	8.84	10.47

Table 2. Continued

no.	expt	T_{Expt}	T_{MD}^b	P1	P2	P3		P4	
						NTT=0	NTT=3	NTT=0	NTT=3
55	8.29	25	25.35	7.80	6.70	6.70	7.21	6.70	7.77
56	7.48	25	25.31	7.03	6.84	6.84	6.88	6.84	7.01
57	7.75	25	25.18	8.09	8.02	8.02	8.03	8.02	8.07
58	5.45	25	25.24	4.39	3.89	3.89	4.06	3.89	4.36
59	6.68	25	25.17	5.14	4.65	4.65	4.98	4.65	5.13
60	7.65	25	25.61	8.79	7.59	7.58	8.28	7.59	8.81
61	8.48	25	25.37	10.66	9.49	9.49	10.03	9.49	10.56
62	8.97	25	25.46	10.54	9.01	8.75	9.51	8.88	10.46
63	8.87	128	128.57	12.24	9.66	9.66	10.68	9.66	11.87
64	6.62	25	25.44	7.22	6.40	6.40	6.90	6.40	7.25
65	8.29	25	25.52	7.73	6.95	6.95	7.37	6.95	7.73
66	10.84	25	25.19	12.46	11.36	11.36	11.93	11.36	12.42
67	9.56	25	25.37	9.87	8.81	8.81	9.24	8.81	9.86
68	12.78	25	25.07	13.24	12.37	12.37	12.91	12.37	13.25
69	10.48	25	25.48	10.97	10.40	10.19	10.68	10.20	10.94
70	14.18	25	25.28	14.74	13.18	13.18	13.92	13.18	14.72

^a For protocols P3 and P4, only the results of using temperature regulation methods of ntt = 0 and ntt = 3 in gas phase MD are presented. Compounds 49, 53, 54, and 71 are excluded because of a lack of experimental data. ^b T_{MD} is the temperature of MD simulation in the liquid phase (°C)

equation of eqs 5 and 6, where $\Delta T = T_{\text{liquid}}^{\text{MD}} - T_{\text{gas}}^{\text{MD}}$ is the difference between the temperatures of MD simulations in the liquid and gas phases. Note that T is the desired temperature, and $T_{\text{liquid}}^{\text{MD}}$ and $T_{\text{gas}}^{\text{MD}}$ are the actual temperatures in the liquid phase and gas phase MD simulations.

$$\Delta H_{\text{vap}}(T) = E_{\text{gas}}^{\text{potential}}(T) - E_{\text{liquid}}^{\text{potential}}(T) + \frac{1}{2} R \Delta T (3N_{\text{atom}} - 6 - N_{\text{cons}}) + RT \quad (8)$$

With the above-discussed MD protocols, $T_{\text{liquid}}^{\text{MD}}$ is essentially close to T , but $T_{\text{gas}}^{\text{MD}}$ may deviate from T for some temperature regulation protocols. For the last two basic protocols, P3 and P4, each has four subprotocols corresponding to the four temperature regulation methods of the gas phase MD simulations (ntt = 0, 1, 2 and 3). For convenience, we use abbreviations to describe the subprotocols in the following text. For example, the abbreviation “P4/ntt0” describes the computation method of ΔH_{vap} using Protocol P4 in conjunction with ntt = 0 in gas phase MD simulations. Other abbreviations, P3/ntt0, P3/ntt1, P3/ntt2, P3/ntt3, P4/ntt1, P4/ntt2, and P4/ntt3, can be interpreted in a similar way.

Statistical Uncertainty Estimation. The density and most energetic terms in heat of vaporization calculations are ensemble averages. The uncertainty of a term (densities, temperatures, and energies) was estimated by the RMS deviation of a series of accumulated means. For the liquid phase terms, the means were calculated using the first 1000, 1025, 1050, 1075, 1100, ...1500 snapshots; for the gas phase terms, the means were calculated using the first 250, 260, 270, ...500 snapshots.

3. RESULTS AND DISCUSSION

Given the fact that GAFF inherited its van der Waals parameters from the AMBER biomolecular force fields, it is important to systematically assess its performance in reproducing some important bulk properties. Rather than studying many molecular properties for a few compounds,^{7,13,14,56} we adopt a different

presentation style, which studies one or two molecular properties for a large number of compounds. In this work, we focused on density and heat of vaporization, the two most important bulk properties in van der Waals parametrization.

Density Calculations. The calculated densities of 71 molecules are listed in Table S1 of the Supporting Information. The correlation between the experimental and the calculated densities is shown in Figure 1. Bromomethane and iodomethane are recognized as outliers who have the largest prediction errors. This is not a surprise at all since the van der Waals (vdW) parameters of bromide and iodide in GAFF are borrowed from other force fields. The average unsigned error (AUE), root-mean-square error (RMSE), and average percent error (APE) of the whole data set are 0.0436 g/cm⁻³, 0.0756 g/cm⁻³, and 4.5%, respectively. The total assigned density difference $\Sigma(\rho_{\text{expt}} - \rho_{\text{calc}})$ is 0.0169 for 71 molecules, suggesting the symmetrical error in our calculations is minimal.

Exclusion of the two outliers, bromomethane and iodomethane, leads to a significant decrease of AUE, RMSE, and APE, which now are 0.0342, 0.0440, and 4.1%, respectively. It is worth noting that the experimental densities of propene, 1,3-butadiene, methylamine, and trimethylamine were measured with a pressure larger than 1 atm. If we exclude these four molecules, the prediction performance is further improved (AUE = 0.0306, RMSE = 0.0386, APE = 3.430%).

Heat of Vaporization. Four basic protocols and 10 subprotocols (P1, P2, P3/ntt0, P3/ntt1, P3/ntt2, P3/ntt3, P4/ntt0, P4/ntt1, P4/ntt2, and P4/ntt3) have been explored to predict the heats of vaporization of 67 compounds. The calculated heats of vaporization of the 10 subprotocols are listed in Table 2 (P1, P2, P3/ntt0, P3/ntt3, P4/ntt0, and P4/ntt3) and Table S2 (Supporting Information; P3/ntt1, P3/ntt2, P4/ntt1, and P4/ntt2). The individual energetic terms as well as the MD temperatures are listed in Table S3 (Supporting Information).

The performance of calculation is summarized in Table 3. Several interesting conclusions can be drawn from this table. First

Table 3. Performance of Heat of Vaporization Calculations^a

protocol	P1	P2	P3				P4			
			ntt=0	ntt=1	ntt=2	ntt=3	ntt=0	ntt=1	ntt=2	ntt=3
AUE	1.184	1.081	1.061	1.071	1.053	1.069	0.932	2.410	1.019	1.095
RMSE	1.560	1.320	1.285	1.292	1.276	1.288	1.201	3.541	1.303	1.402
R ²	0.864	0.864	0.875	0.878	0.878	0.877	0.880	0.642	0.879	0.877

^aAll energies are in kcal/mol.

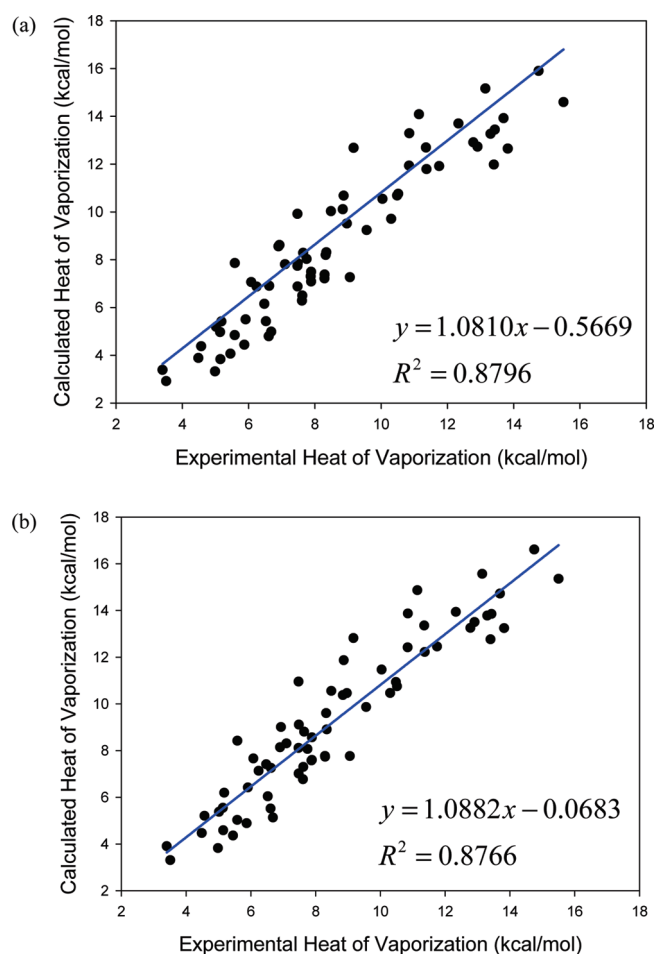


Figure 2. Performance of heats of vaporization calculation using two subprotocols of P4. (a) P4/ntt0, (b) P4/ntt3.

of all, it is a surprise that protocol P4 with the gas phase constant energy classical MD simulations (P4/ntt0) achieves the best performance: the AUE and RMSE are 0.932 and 1.201 kcal/mol, respectively. The total signed ΔH_{vap} difference $\Sigma(\Delta H_{\text{vap}}^{\text{expt}} - \Delta H_{\text{vap}}^{\text{calc}})$ is -8.54 for 67 molecules, suggesting that ΔH_{vap} is overestimated by about 0.13 kcal/mol on average. The correlation between the experimental and the calculated heats of vaporization with this model is shown in Figure 2a. The second-best subprotocol in terms of AUE, P4/ntt2, has an AUE of 1.019 kcal/mol. Second, the newly introduced temperature regulation method, Langevin dynamics (ntt = 3),⁴⁰ performs slightly worse than the Andersen coupling scheme (ntt = 2) in combination with both the P3 and P4 basic protocols. The correlation between the experimental and the calculated heats of

vaporization with P4/ntt3 is shown in Figure 2b. Third, it is not a surprise that all four temperature scaling methods have similar performances in heat of vaporization calculations using protocol 3, since the snapshots sampled by gas MD simulations were further minimized and the means of the minimization energies have much smaller deviations than the MD energies without minimization. Fourth, protocol 4, with the weak coupling algorithm (ntt = 1), performs much worse than the other models, implying that the gas phase energies sampled using ntt = 1 are much different from those from the other temperature regulation methods. Finally, although protocol 1 has the largest approximation in theory, its performance is only marginally worse than the other protocols, except P4/ntt1 (the worst), indicating that the interaction energy makes the largest contribution to the heat of vaporization in most scenarios.

The uncertainty of the heat of vaporization calculations is summarized in Table S4 (Supporting Information). Protocol 1 has the largest average statistical uncertainty of 0.19 kcal/mol for 67 molecules. For the other calculation protocols, the uncertainties are all smaller than 0.085 kcal/mol, indicating that our MD protocols can reliably predict ΔH_{vap} using protocols P2, P3, and P4.

The compounds that have the largest prediction errors via P4/ntt0 are typically those less common compounds in biological systems, such as sulfides (Nos. 26, 27 and 28), nitromethane (No. 44), halides (Nos. 55, 58 and 59), and alkenes (Nos. 7 and 8). This phenomenon suggests that, for those compounds, more effort is needed to tune their van der Waals parameters to reproduce experimental densities and heats of vaporization.

Temperature Dependence of Density and ΔH_{vap} for TIP3P Water. It is important for a molecular mechanical model to accurately predict molecular properties of a broad range of thermodynamic states described by temperature, volume, pressure, etc. Given the fact that water has plenty of experimental data for density and ΔH_{vap} as a function of temperature, we assessed the protocols of MD simulations and ΔH_{vap} prediction using TIP3P water. As shown in Figure 3a, the calculated density of TIP3P decreases more sharply than the experimental value, and the two lines cross around 280 K. At 298 K, the density of TIP3P is 0.985 g/cm³, about 1.2% lower than the experiment (0.997). For the temperature range from 260 to 310 K, the prediction error is smaller than 2% for TIP3P water in our MD simulations. A ρ - T plot with a similar pattern to ours was reported by Price and Brooks.⁵⁴

It is of interest to locate the temperature in the ρ - T plot where TIP3P water has the maximum density. In a pioneer work, Billeter and co-workers investigated the maximum density temperature of the SPC/E water model using energy-volume correlations.⁵⁷ Jorgensen and Jenson also studied the temperatures of maximum density for TIP3P and TIP4P water models.⁵⁸ We estimate that the maximum density temperature is around

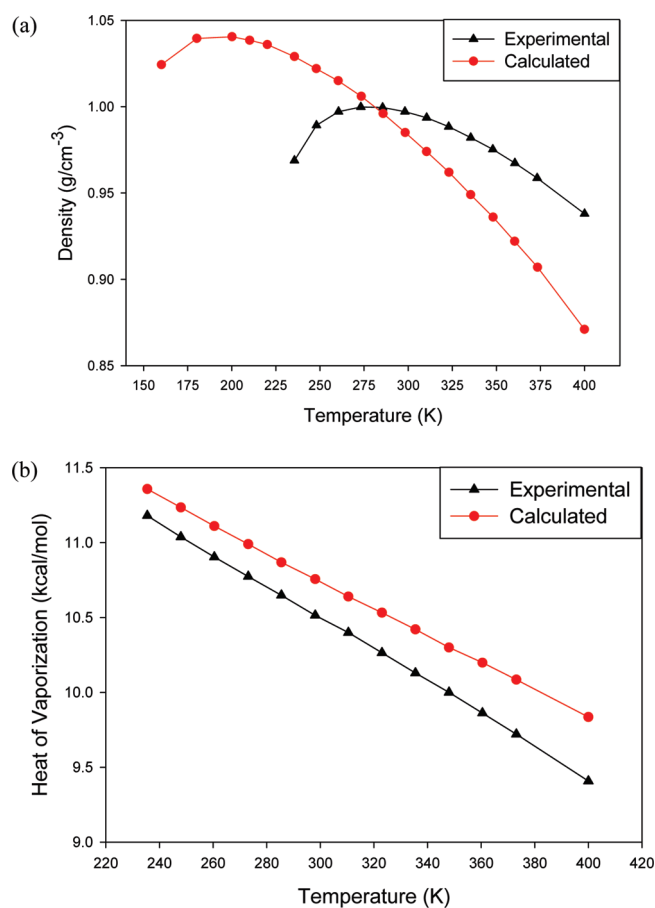


Figure 3. Temperature dependence of density and heat of vaporization of TIP3P water: (a) density, (b) heat of vaporization.

200 K for the TIP3P water. As shown in Figure 3a, when the temperature is above 220 K, the TIP3P model shows monotonically increasing density as the temperature decreases. From 180 to 220 K, there is a flat region, and the maximum density is reached. When the temperature drops below 180 K, the density decreases. This phenomenon is similar to that of TIP4P except TIP4P has its maximum density temperature and flat region around 260 K, as reported by Jorgensen and Jenson.⁵⁸ It is clear that the maximum density temperature of TIP4P is much closer to the experimental value (4 °C). Interestingly, Jorgensen and Jenson did not find the maximum density temperature for TIP3P since the temperatures in their Monte Carlo simulations were limited to -50 to 100 °C.⁵⁸

As a special molecule, TIP3P has all its degrees of freedom frozen; as such, all of the protocols of calculating ΔH_{vap} produce the same value. The temperature dependence of heat of vaporization for TIP3P is shown in Figure 3b. It is clear that the calculated ΔH_{vap} is systematically larger than the experimental value at a temperature range of 230–400 K. At 298 K, the calculated ΔH_{vap} is overestimated by about 2.3% (10.51 vs 10.75 kcal/mol). The experimental and calculated data used for plotting Figure 3 are listed in Table S5 of the Supporting Information.

The Major Factors That Affect Density and ΔH_{vap} Calculations. Given the physical limitations of the current nonpolarizable force fields, it is important to derive force field parameters to reproduce not only the in vacuo ab initio properties but also pure-liquid and hydrated properties. The development of the

OPLS-AA force field perfectly demonstrates this philosophy.⁷ The quality of molecular mechanical models certainly is one of the major factors that determine the performance of density and ΔH_{vap} calculations. However, it is challenging to conduct a critical assessment of different force fields; hidden flaws (such as wrongly using the mixing rules of van der Waals parameters) and some technical differences could lead to discrepancy. In this work, we only make comparison using the published data.

As a pioneer of developing molecular mechanical force fields through liquid simulations, Jorgensen and co-workers have studied a large set of organic compounds that covers most chemical functional groups. The following is a quick summary of the calculation errors for each compound class using the OPLS-AA force field, and the two values in the appended parentheses represent the average percent errors of density and ΔH_{vap} , respectively: hydrocarbon (3%, 2%);⁷ alcohols (1.8%, 2.2%);⁷ sulfur compounds (1.8%, 4%);⁷ ethers (1.5%, 2.8%);⁷ amides (2%, 2.1%);⁷ acids (1.9%, 0.7%);⁷ aldehydes and ketones (1.0%, 3.7%);⁷ pyridine and diazenes (0.8%, 2.7%);⁴⁹ pyrrole, furan, diazole, and oxazoles (2%, 9.9%);⁵⁰ amines (1.4%, 3.5%);⁵² esters (1.4%, 6.2%);⁵¹ nitriles (1.8%, 2.7%);⁵¹ and nitro compounds (1.1%, 2.0).⁵¹ Recently, Price and Brooks further studied a set of 28 mono- and disubstituted benzenes using the OPLS-AA force field.⁵⁹ They found that the agreement between OPLS-AA and experimental results is remarkable, with average errors of 1.8% for densities and 0.69 kcal/mol for ΔH_{vap} .

Unlike the remarkable performance of OPLS-AA, the performance of GAFF, with an APE of 3.43% if not counting outliers for density and an AUE of 0.93 kcal/mol (P4/ntt0) for heat of vaporization, is still satisfactory given the following two facts: the partial charges in OPLS-AA are tunable parameters, while GAFF was parametrized using HF/6-31G* RESP charges, and the van der Waals parameters in OPLS-AA were parametrized to reproduce the bulk properties including density and heat of vaporization, while GAFF inherits its vdW parameters directly from the AMBER force fields without further optimization.

Besides the molecular mechanical models, sampling and the other computational techniques could affect the density and ΔH_{vap} calculation. There has been a dramatic change in electrostatic energy calculations of a unit cell since the vdW parameters of GAFF were developed. Nowadays, particle mesh Ewald (PME) has become a standard method in handling the long-range electrostatic interactions in most molecular simulation packages.^{35–37} Other methods of calculating long-range electrostatic interactions, such as isotropic periodic sum (IPS),^{60–62} begin to emerge. On the other hand, van der Waals parameters implemented in GAFF were developed with short electrostatic cutoffs, and the long-range effect had been embedded in the local interactions in an average sense. It is understandable that van der Waals parameters developed in the old days have some deviations in molecular property calculations using today's simulation techniques (such as PME and Langevin dynamics to regulate temperature).

Exploration of Force Field Parameters. Given the fact that GAFF has not been optimized to reproduce the bulk properties, there should be a lot of room for us to significantly improve the performance of predicting the density and heat of vaporization through proper van der Waals parametrization. Here, we simply carried out an exploring study to prove the concepts, and the systematic van der Waals parametrization for GAFF is the next target in our effort of developing high-quality general purpose force fields.

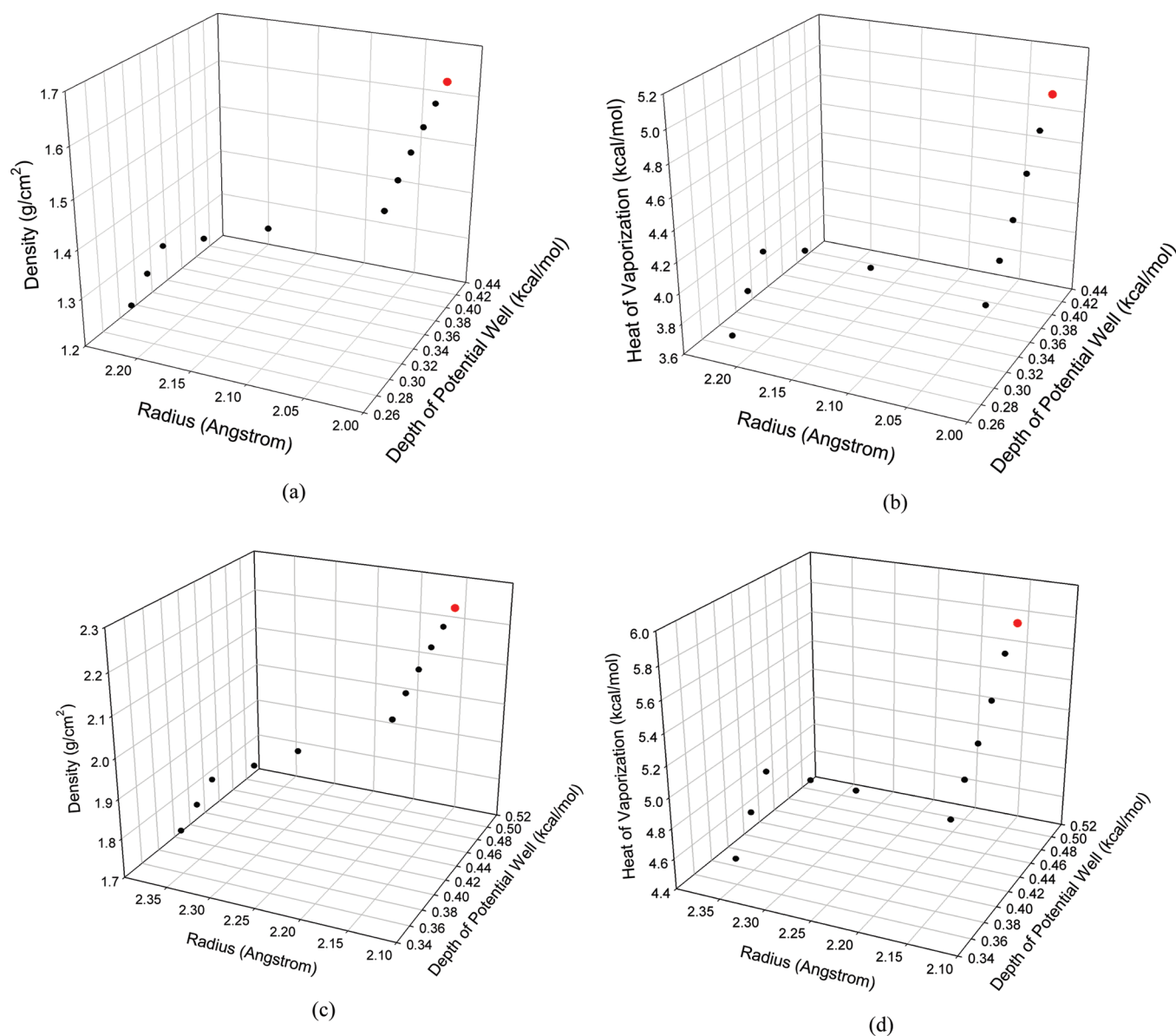


Figure 4. How the calculated density and heat of vaporization change as a function of van der Waals parameters. (a) Density of bromomethane. (b) Heat of vaporization of bromomethane. (c) Density of iodomethane. (d) Heat of vaporization of iodomethane.

First of all, we optimized the vdW parameters of bromide and iodide. There are no parameters for bromide and iodide in the AMBER biomolecular force fields. The vdW parameter of bromide was borrowed from MM3,⁶³ and that of iodine was from Weiner et al.⁶⁴ The ideal parameters of radius R and depth of potential well ϵ were located using systematic searches. How the density and ΔH_{vap} change as a function of R and ϵ is shown in Figure 4. Comparing Figure 4a and b for bromomethane and 4c and d for iodomethane, we can reach the following conclusion: the patterns of density and heat of vaporization change as a function of R and ϵ are essentially similar, and they almost reach the points of experimental values synchronously. In summary, the vdW parameter of bromide changes from $R = 2.22$ and $\epsilon = 0.32$ to $R = 2.02$ and $\epsilon = 0.42$, and the density and ΔH_{vap} of the new parameter are 1.642 g cm^{-3} and 4.952 kcal/mol ($P4/\text{ntt} = 0$), respectively. In contrast, the density and ΔH_{vap} of the old parameter are 1.343 g cm^{-3} and 3.89 kcal/mol , respectively

(the experimental values are 1.676 and 5.45, respectively). As to iodide, the vdW parameter changes from $R = 2.35$ and $\epsilon = 0.40$ to $R = 2.15$ and $\epsilon = 0.50$. Similarly, both density and ΔH_{vap} can much better reproduce the experimental values ($\rho = 2.279 \text{ g cm}^{-3}$ and $\Delta H_{\text{vap}} = 6.68$) using the new parameter ($\rho = 2.247$ and $\Delta H_{\text{vap}} = 5.782$) than the original one ($\rho = 1.877$ and $\Delta H_{\text{vap}} = 4.65$). The prediction errors of densities and ΔH_{vap} are summarized in Table 4.

In the next example, we attempted to improve the density and ΔH_{vap} of aromatic compounds by tuning the vdW parameter of aromatic carbon. Benzene (No. 10) was selected as the model compound for the vdW parametrization. We found that the two properties of benzene are not sensitive to the radius parameter R ; therefore, only the depth of potential well ϵ of aromatic carbon was scanned using the original radius parameter $R = 1.908 \text{ \AA}$. As shown in Figure 5, the density and ΔH_{vap} are proportional to ϵ , and the ideal value of ϵ ranges from 0.090 to 0.095 kcal/mol.

Table 4. Performance of Density and Heat of Vaporization Calculation Using the Original and the Revised GAFF

molecular class	#mols	original GAFF				revised GAFF			
		density (g/cm ³)		ΔH_{vap} (kcal/mol)		density (g/cm ³)		ΔH_{vap} (kcal/mol)	
		AUE	RMSE	AUE	RMSE	AUE	RMSE	AUE	RMSE
bromide	1	0.333	0.333	1.56	1.56	0.034	0.034	0.50	0.50
iodide	1	0.402	0.402	2.03	2.03	0.032	0.032	0.90	0.90
benzene derivatives	5	0.023	0.051	0.72	1.61	0.020	0.045	0.48	1.07
amine	3	0.041	0.044	1.28	1.41	0.026	0.026	0.44	0.70

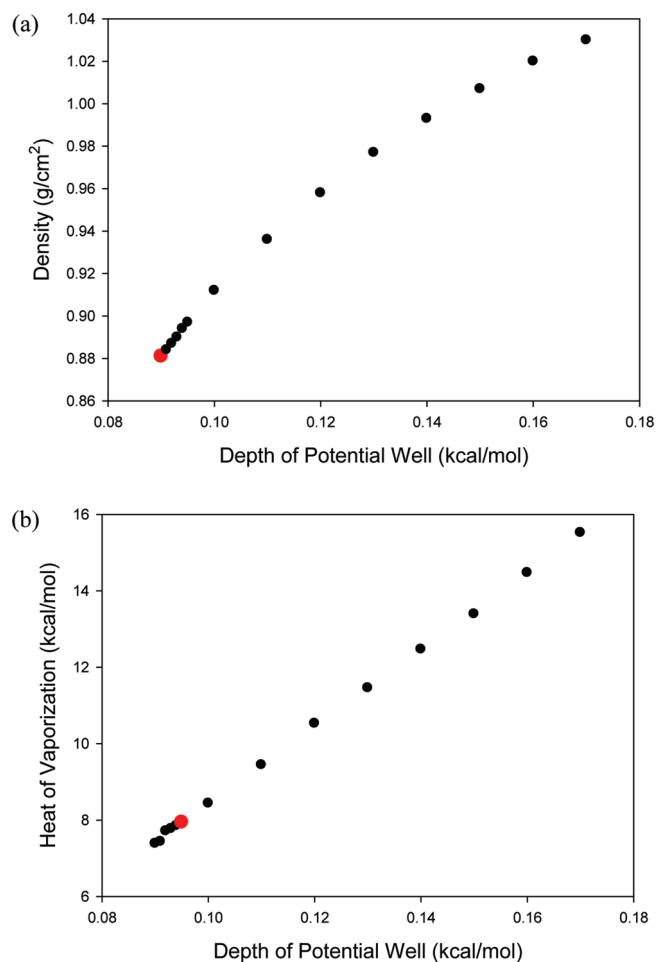


Figure 5. How the density and heat of vaporization of benzene change as a function of the depth of potential well of the aromatic carbon: (a) density, (b) heat of vaporization. The red dots indicate which van der Waals parameter best reproduces the experimental values.

When $\varepsilon = 0.095$, ΔH_{vap} has the best value (7.944 by P4/ntt0) compared to the experimental result (8.086 kcal/mol), while a smaller ε is required to achieve the best density ($\rho = 0.877$ when $\varepsilon = 0.091$). Finally, we set the depth of the potential well of aromatic carbon to 0.092 kcal/mol as a compromise. The calculated densities and ΔH_{vap} for five aromatic compounds (Nos. 10, 17, 18, 42, and 55) using the new vdW parameter are listed in Table S6 (Supporting Information). Although only the depth of potential well ε of aromatic carbon was adjusted, the performance of the density and ΔH_{vap} prediction is significantly

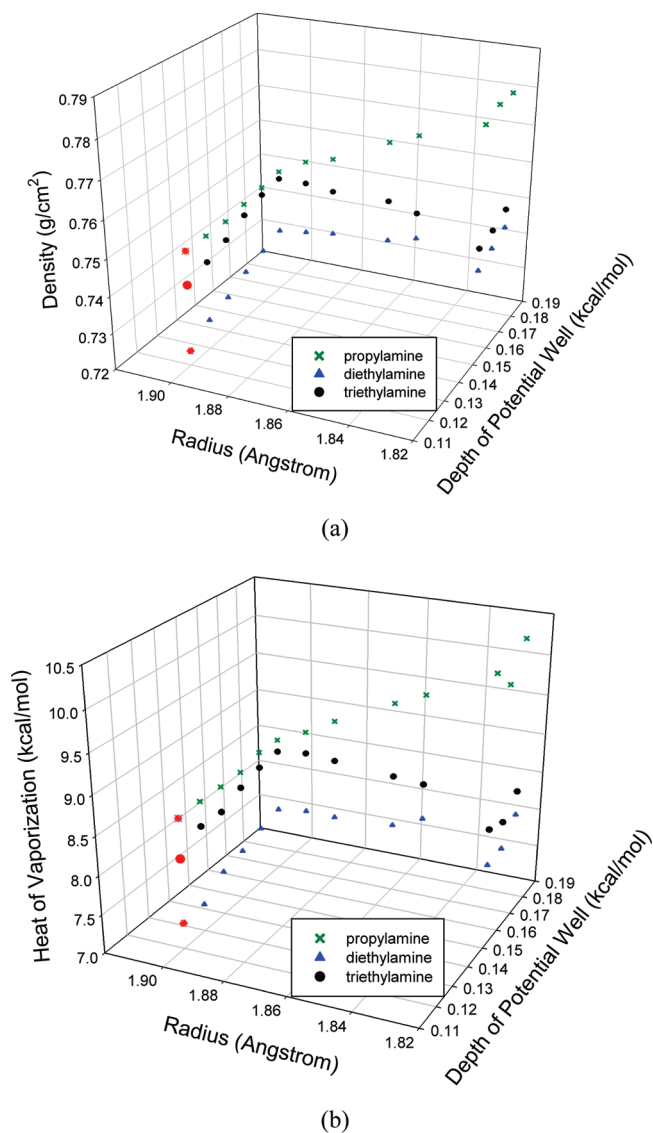


Figure 6. How the density and heat of vaporization of three amines change as a function of the van der Waals parameters of amine nitrogen: (a) density, (b) heat of vaporization. The red symbols indicate which van der Waals parameter best reproduces the experimental values.

improved: the AUE, RMSE, and APE of density calculations using the new parameter are 0.02, 0.045, and 1.83%, respectively, in comparison to AUE = 0.023, RMSE = 0.051, and APE = 2.07% using the original parameter. More significant improvement was

Table 5. List of the Electrostatic Part (Charge Disappearing), the van der Waals Part (Atom Disappearing), and the Calculated Free Energies (kcal/mol) of Hydration for 10 Representative Molecules

no.	name	exptl ^a	elec.	original GAFF			revised GAFF		
				vdW	calcd	elec.	vdW	calcd	
1	bromomethane	−0.82	−1.40	2.10	0.70	−1.55	1.46	−0.10	
2	iodomethane	−0.89	−1.43	1.98	0.55	−1.54	1.38	−0.16	
3	propylamine	−4.39	−7.26	1.84	−5.42	−6.54	2.02	−4.53	
4	diethylamine	−4.07	−3.97	2.33	−1.63	−3.64	2.19	−1.45	
5	triethylamine	−3.22	−0.92	2.56	2.37	−0.79	2.25	2.20	
6	benzene	−0.86	−2.69	1.77	−0.92	−2.70	1.37	−1.33	
7	phenol	−6.61	−5.92	1.48	−4.44	−5.88	1.15	−4.73	
8	<i>m</i> -cresol	−5.49	−7.52	1.47	−6.06	−7.41	1.43	−5.98	
9	aniline	−5.49	−5.92	1.30	−4.62	−5.87	1.11	−4.76	
10	fluobenzene	−0.80	−2.28	2.02	−0.26	−2.25	2.05	−0.19	

^aThe experimental data are from ref 19.

achieved for ΔH_{vap} , and the AUE, RMSE, and APE of the new parameter are 0.477, 1.067, and 4.63%, respectively, in comparison to AUE = 0.720, RMSE = 1.610, and APE = 6.99% of the original parameter of the aromatic carbon. We believe that the prediction performance can be further improved when more atom types are tuned simultaneously. The new depth of potential well of aromatic carbon, 0.092, is much larger than the original value, 0.086. Interestingly, we found that a larger value of depth of potential well of carbon can significantly reduce the prediction errors of interaction energies for a large set of amino acid analog pairs.^{65,66}

In the last example, the vdW parameter of amine carbon was optimized using three different types of amines, namely, propylamine (No. 37), diethylamine (No. 39), and triethylamine (No. 41). How the densities and heats of vaporization change as a function of the radius parameter R and depth of potential well ϵ is shown in Figure 6. Encouragingly, the best prediction values of densities and heats of vaporization for the three compounds are obtained synchronously. The optimized parameter of amine nitrogen ($R = 1.900$ and $\epsilon = 0.140$) has a larger radius and smaller depth of potential well compared to the original one ($R = 1.824$ and $\epsilon = 0.170$). The calculated densities and ΔH_{vap} for three amines using the new vdW parameter are listed in Table S7 (Supporting Information). Again, the new parameter significantly reduces the prediction errors of both molecular properties: for density, the AUE, RMSE, and APE of the new parameter are 0.026, 0.026, and 3.57%, respectively, in comparison to AUE = 0.041, RMSE = 0.044, and APE = 5.67% of the original parameter. As to heat of vaporization, the improvement is even more significant: the AUE, RMSE, and APE of the new parameter are 0.441, 0.699, and 5.89%, respectively, in comparison to AUE = 1.276, RMSE = 1.408, and APE = 16.34% of the original parameter. The performance of the original and revised GAFF in calculating density and ΔH_{vap} is summarized in Table 4.

It is worth noting that tuning the partial charges can also improve the molecular property prediction. For example, in OPLS-AA after turning both the partial charges and Lennard-Jones parameters, Jorgensen and Rizzo could reproduce the densities, the heats of vaporization, and hydration free energies of pure liquid amines very well.⁵² In another example, with the adjustment of several bond charge correction parameters, AM1-BCC can better reproduce the absolute solvation energies.^{15,16} Certainly, it is also possible for us to tune BCC parameters so that

the calculated densities, heats of vaporization, and free energies of hydration agree better with experimental results. However, in this study, the HF/6-31G* RESP charges were used since GAFF was originally parametrized with this charge method.

In summary, the prediction errors of density and heat of vaporization can be significantly reduced by tuning one or two van der Waals parameters. Density and ΔH_{vap} approach their ideal values synchronously in all three cases, although the best parameters for density and ΔH_{vap} may be slightly different. It should be pointed out that although the new vdW parameters were optimized using Protocol P4/ntt0 to calculate ΔH_{vap} , the calculation errors of other protocols (such as P2, P4/ntt3) are also reduced (data not shown). Certainly, the optimum vdW parameter sets are slightly different from one protocol to another.

Hydration Energies Calculations. As the hydration free energy of a molecule is also directly affected by the vdW parameters. It is important for us to check whether the newly developed vdW parameters deteriorate the hydration energy calculations or not. Ten molecules, which are directly affected by the newly developed vdW parameters, were selected to conduct hydration free energies using TI. Encouragingly, the revised GAFF significantly improves the hydration energy calculations, as shown in Table 5: the AUE is reduced from 1.62 to 1.38 kcal/mol and RMSE from 2.21 to 2.05 kcal/mol. It is pointed out that most errors come from three amines, i.e., propylamine, diethylamine, and triethylamine. It is widely known that the hydration energies of amine and amide series are difficult to predict, and the previous studies using free energy perturbation with AMBER force fields strongly overestimate the free energy changes from methylamine to dimethylamine and from dimethylamine to trimethylamine.⁵² If the three amines are excluded, the AUE and RMSE of the revised GAFF become 0.80 and 0.92 kcal/mol, respectively. All three compound classes in Table 5 achieve a better performance of hydration free energy calculation using the revised GAFF. Specifically, for the two halides, both the AUE and RMSE are reduced by 0.75 kcal/mol. For the three amines, the AUE and RMSE are reduced by 0.3 and 0.1 kcal/mol, respectively. For five benzene derivatives, the AUE and RMSE are reduced by 0.01 and 0.11 kcal/mol, respectively. In summary, although the prediction of hydration free energies was not taken into account when we optimized vdW parameters, the revised GAFF achieved better performance in hydration energy calculations using TI. The statistical uncertainties of individual energy

terms of TI calculations are summarized in Table S8 of the Supporting Information.

4. CONCLUSIONS

This is the first paper in a series of predicting molecular properties using the General AMBER Force Field (GAFF). The bulk densities and heats of vaporizations of 71 organic molecules that cover the most common chemical functional groups have been predicted through molecular dynamics simulations. The overall performance of the prediction is satisfactory: the average unsigned error (AUE) of the density of the whole molecular set is 0.0436 g cm^{-3} , and the AUE of the heat of vaporization calculation using Protocol P4/ntt0 is 0.932 kcal/mol . In total, 10 different heat of vaporization subprotocols (P1, P2, P3/ntt=0, P3/ntt=1, P3/ntt=2, P3/ntt=3, P4/ntt=0, P4/ntt=1, P4/ntt=2, P4/ntt=3) have been investigated, and P4/ntt0 outperforms the other subprotocols. van der Waals parametrization has been conducted for three molecular subsets in order to reduce the prediction errors of both density and heat of vaporization. Encouragingly, both density and heat of vaporization can much better reproduce the experimental data after adjusting one or two van der Waals parameters. Moreover, the free energies of hydration for 10 molecules also achieve a much better performance for the revised GAFF. Given the fact that GAFF inherited van der Waals parameters from the AMBER biomolecular force fields without further optimization, it is very likely that GAFF can be significantly improved on predicting molecular properties after a systematic van der Waals parametrization.

■ ASSOCIATED CONTENT

S Supporting Information. The calculated densities of 71 molecules are listed in Table S1. The structures of the 71 molecules are shown in Figure S1. The calculated heats of vaporization using Protocols P3/ntt1, P4/ntt1, P3/ntt2, and P4/ntt2 are summarized in Table S2. The individual energetic terms and statistical uncertainties are listed in Tables S3 and S4, respectively. The data used to prepare Figure 3 are listed in Table S5. In Tables S6–S7, the calculated densities and heats of vaporization using new van der Waals parameters of aromatic carbon (Table S6) and amine nitrogen (Table S7) are listed. The statistical uncertainties of TI calculations are listed in Table S8. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Authors

*Tel.: (214)-645-5966 (J.W.). E-mail: junmei.wang@utsouthwestern.edu, tjhou@suda.edu.cn.

■ ACKNOWLEDGMENT

We are grateful for the research support from NIH (R01GM79383, Y. Duan, P.I.) and Natural Science Foundation of China (No. 20973121, T. Hou, P.I.) and the TeraGrid for the computational time (TG-CHE090098, J. Wang, P.I.).

■ ABBREVIATIONS

GAFF, General AMBER Force Field; MD, Molecular Dynamics; vdW, van der Waals; ρ , density; ΔH_{vap} , heat of vaporization; ϵ ,

depth of potential well; AUE, average unsigned errors; RMSE, root-mean-square errors; APE, average percent errors; R^2 , correlation coefficient square

■ REFERENCES

- (1) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (2) van Gunsteren, W. F.; Mark, A. E. Validation of molecular dynamics simulation. *J. Chem. Phys.* **1998**, *108*, 6109–6116.
- (3) Foloppe, N.; MacKerell, A. D. J. All-atom empirical force field for nucleic acids. 1) Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86–104.
- (4) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (5) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (6) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (7) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (8) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F.; van Gunsteren, W. F. Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* **2005**, *34*, 273–284.
- (9) Soares, T. A.; Hunenberger, P. H.; Kastenholz, M. A.; Krautler, V.; Lenz, T.; Lins, R. D.; Oostenbrink, C.; van Gunsteren, W. F. An improved nucleic acid parameter set for the GROMOS force field. *J. Comput. Chem.* **2005**, *26*, 727–737.
- (10) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (11) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges- the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (12) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers- charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357–1377.
- (13) Caldwell, J. W.; Kollman, P. A. Structure and Properties of Neat Liquids Using Nonadditive Molecular-Dynamics - Water, Methanol, and N-Methylacetamide. *J. Phys. Chem.* **1995**, *99*, 6208–6219.
- (14) Fox, T.; Kollman, P. A. Application of the RESP methodology in the parametrization of organic solvents. *J. Phys. Chem. B* **1998**, *102*, 8070–8079.
- (15) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast efficient generation of high quality atomic charges. AM1-BCC model: I. method. *J. Comput. Chem.* **2000**, *21*, 132–146.
- (16) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
- (17) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* **2005**, *48*, 4040–4048.

- (18) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. I. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69–82.
- (19) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350–358.
- (20) MacKerell, A. D.; Shim, J. H.; Anisimov, V. M. Re-evaluation of the reported experimental values of the heat of vaporization of N-methylacetamide. *J. Chem. Theory Comput.* **2008**, *4*, 1307–1312.
- (21) Physical constants of organic compounds. In *CRC Handbook of Chemistry and Physics*, 86th ed.; Lide, D. R. E., Eds.; CRC Press: Boca Raton, FL, 2005; pp 1–3.
- (22) Siepmann, J. I.; Karaborni, S.; Smit, B. Simulating the Critical-Behavior of Complex Fluids. *Nature* **1993**, *365*, 330–332.
- (23) *Enthalpies of Vaporization of Organic Compounds: A Critical Review and Data Compilation*; Majer, V., Svoboda, V., Eds.; Blackwell Scientific Publications: Oxford, U. K., 1985; pp 1–304.
- (24) Kemp, J. D.; Egan, C. J. Hindered Rotation of the Methyl Groups in Propane. The Heat Capacity, Vapor Pressure, Heats of Fusion and Vaporization of Propane. Entropy and Density of the Gas. *J. Am. Chem. Soc.* **1938**, *60*, 1521–1525.
- (25) Pedley, J. B.; Naylor, R. D.; Kirby, S. P. *Thermochemical Data of Organic Compounds*, 2nd ed.; Chapman and Hall: New York, 1986; pp 1–792.
- (26) Wiberg, K. B.; Crocker, L. S.; Morgan, K. M. Thermochemical Studies of Carbonyl-Compounds 0.5. Enthalpies of Reduction of Carbonyl Groups. *J. Am. Chem. Soc.* **1991**, *113*, 3447–3450.
- (27) Geller, B. E. Properties of the dimethylformamide water system. I. Thermochemical properties. *Russ. J. Phys. Chem.* **1961**, *35*, 542–564.
- (28) Pihlaja, K.; Heikkilä, J. Enthalpies of formation of cyclic acetals. 1,3-dioxolane, 2-methyl-1,3-dioxolane, and 2,4-dimethyl-1,3-dioxolanes. *Acta Chem. Scand.* **1969**, *23*, 1053–1055.
- (29) Steele, W. V.; Archer, D. G.; Chirico, R. D.; Collier, W. B.; Hossenlopp, I. A.; Nguyen, A.; Smith, N. K.; Gammon, B. E. The thermodynamic properties of quinoline and isoquinoline. *J. Chem. Thermodyn.* **1988**, *20*, 1233–1264.
- (30) Wang, J.; Hou, T.; Xu, X. Aqueous solubility prediction based on weighted atom type counts and solvent accessible surface areas. *J. Chem. Inf. Model.* **2009**, *49*, 571–81.
- (31) Konicek, J.; Wadso, I. Enthalpies of Vaporization of Organic Compounds 0.7. Some Carboxylic Acids. *Acta Chem. Scand.* **1970**, *24*, 2612–2626.
- (32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.
- (33) Case, D. A.; Darden, T. A.; Cheatham, I.; Simmerling, C.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER10*; University of California at San Francisco: San Francisco, CA, 2008.
- (34) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Model.* **2006**, *25*, 247–260.
- (35) Darden, T.; Perera, L.; Li, L.; Pedersen, L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **1999**, *7*, R55–60.
- (36) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (37) Sagui, C.; Pedersen, L. G.; Darden, T. A. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *J. Chem. Phys.* **2004**, *120*, 73–87.
- (38) Miyamoto, S.; Kollman, P. A. Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (39) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. Accurate and efficient corrections for missing dispersion interactions in molecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 13052–13063.
- (40) Uberuaga, B. P.; Anghel, M.; Voter, A. F. Synchronization of trajectories in canonical molecular-dynamics simulations: observation, explanation, and exploitation. *J. Chem. Phys.* **2004**, *120*, 6363–6374.
- (41) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **2001**, *114*, 2090–2098.
- (42) Larini, L.; Mannella, R.; Leporini, D. Langevin stabilization of molecular-dynamics simulations of polymers by means of quasisymplectic algorithms. *J. Chem. Phys.* **2007**, *126*, 104101.
- (43) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides - the Frictional Dependence of Isomerization Rates of N-Acetylalanil-N'-Methylamide. *Biopolymers* **1992**, *32*, 523–535.
- (44) Harvey, S. C.; Tan, R. K. Z.; Cheatham, T. E. The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.* **1998**, *19*, 726–740.
- (45) Andersen, H. C. Molecular-Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- (46) Andrea, T. A.; Swope, W. C.; Andersen, H. C. The Role of Long Ranged Forces in Determining the Structure and Properties of Liquid Water. *J. Chem. Phys.* **1983**, *79*, 4576–4584.
- (47) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.* **2007**, *127*, 214108.
- (48) Hummer, G.; Szabo, A. Calculation of free-energy differences from computer simulations of initial and final states. *J. Chem. Phys.* **1996**, *105*, 2004–2010.
- (49) Jorgensen, W. L.; McDonald, N. A. Development of an all-atom force field for heterocycles. Properties of liquid pyridine and diazenes. *THEOCHEM* **1998**, *424*, 145–155.
- (50) McDonald, N. A.; Jorgensen, W. L. Development of an all-atom force field for heterocycles. Properties of liquid pyrrole, furan, diazoles, and oxazoles. *J. Phys. Chem. B* **1998**, *102*, 8049–8059.
- (51) Price, M. L. P.; Ostrovsky, D.; Jorgensen, W. L. Gas-phase and liquid-state properties of esters, nitriles, and nitro compounds with the OPLS-AA force field. *J. Comput. Chem.* **2001**, *22*, 1340–1352.
- (52) Rizzo, R. C.; Jorgensen, W. L. OPLS all-atom model for amines: Resolution of the amine hydration problem. *J. Am. Chem. Soc.* **1999**, *121*, 4827–4836.
- (53) Yu, H. B.; Hansson, T.; van Gunsteren, W. F. Development of a simple, self-consistent polarizable model for liquid water. *J. Chem. Phys.* **2003**, *118*, 221–234.
- (54) Price, D. J.; Brooks, C. L. A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* **2004**, *121*, 10096–10103.
- (55) Horn, H. W.; Swope, W. C.; Pitner, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved

four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(56) Yu, H. B.; Geerke, D. P.; Liu, H. Y.; van Gunsteren, W. E. Molecular dynamics simulations of liquid methanol and methanol-water mixtures with polarizable models. *J. Comput. Chem.* **2006**, *27*, 1494–1504.

(57) Billeter, S. R.; King, P. M.; Vangunsteren, W. F. Can the Density Maximum of Water Be Found by Computer-Simulation. *J. Chem. Phys.* **1994**, *100*, 6692–6699.

(58) Jorgensen, W. L.; Jenson, C. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *J. Comput. Chem.* **1998**, *19*, 1179–1186.

(59) Price, D. J.; Brooks, C. L. Detailed considerations for a balanced and broadly applicable force field: A study of substituted benzenes modeled with OPLS-AA. *J. Comput. Chem.* **2005**, *26*, 1529–1541.

(60) Wu, X.; Brooks, B. R. Isotropic periodic sum: a method for the calculation of long-range interactions. *J. Chem. Phys.* **2005**, *122*, 44107.

(61) Klaua, J. B.; Wu, X.; Pastor, R. W.; Brooks, B. R. Long-range Lennard-Jones and electrostatic interactions in interfaces: application of the isotropic periodic sum method. *J. Phys. Chem. B* **2007**, *111*, 4393–400.

(62) Takahashi, K.; Yasuoka, K.; Narumi, T. Cutoff radius effect of isotropic periodic sum method for transport coefficients of Lennard-Jones liquid. *J. Chem. Phys.* **2007**, *127*, 114511.

(63) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular Mechanics - the MM3 Force-Field for Hydrocarbons. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.

(64) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force-Field for Simulations of Proteins and Nucleic-Acids. *J. Comput. Chem.* **1986**, *7*, 230–252.

(65) Wang, J. M.; Cieplak, P.; Li, J.; Hou, T. J.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations I: Parameterization of Atomic Polarizability. *J. Phys. Chem. B* **2011**, *115*, 3091–3099.

(66) Wang, J. M.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M. J.; Lei, H. X.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations II: Induced Dipole Models Significantly Improve Accuracy of Intermolecular Interaction Energies. *J. Phys. Chem. B* **2011**, *115*, 3100–3111.

NOTE ADDED AFTER ASAP PUBLICATION

This manuscript was originally published on May 30, 2011 with the given name and surname of one of the authors transposed. The corrected manuscript was published on June 22, 2011.

High-Performance Solution of Hierarchical Equations of Motion for Studying Energy Transfer in Light-Harvesting Complexes

Christoph Kreisbeck,[†] Tobias Kramer,^{*,†} Mirta Rodríguez,[‡] and Birgit Hein[†]

[†]Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany

[‡]Instituto de Estructura de la Materia CSIC, C/Serrano 121, 28006 Madrid, Spain

ABSTRACT: Excitonic models of light-harvesting complexes, where the vibrational degrees of freedom are treated as a bath, are commonly used to describe the motion of the electronic excitation through a molecule. Recent experiments point toward the possibility of memory effects in this process and require one to consider time nonlocal propagation techniques. The hierarchical equations of motion (HEOM) were proposed by Ishizaki and Fleming to describe the site-dependent reorganization dynamics of protein environments (*J. Chem. Phys.* **2009**, *130*, 234111), which plays a significant role in photosynthetic electronic energy transfer. HEOM are often used as a reference for other approximate methods but have been implemented only for small systems due to their adverse computational scaling with the system size. Here, we show that HEOM are also solvable for larger systems, since the underlying algorithm is ideally suited for the usage of graphics processing units (GPU). The tremendous reduction in computational time due to the GPU allows us to perform a systematic study of the energy-transfer efficiency in the Fenna–Matthews–Olson (FMO) light-harvesting complex at physiological temperature under full consideration of memory effects. We find that approximative methods differ qualitatively and quantitatively from the HEOM results and discuss the importance of finite temperature to achieving high energy-transfer efficiencies.

1. INTRODUCTION

Light-harvesting complexes (LHCs) are pigment protein complexes that act as the functional units of photosynthetic systems, capable of absorbing the energy of a photon and transferring it toward the reaction center where it is converted into chemical energy usable for the cell. The transfer of energy in such systems is described by electronic exciton dynamics coupled to the vibrations and other mechanical modes of the complex.¹ Laser spectroscopy shows quantum coherent effects in the energy transfer in LHCs at temperatures up to 300 K.^{2–4}

Theoretical studies of model Hamiltonians at different levels of approximation^{5–11} show that the interplay between coherent transport and dissipation leads to high efficiencies in the energy transport in these systems. LHC provide a remarkable example of systems where noise or dissipation aids the transport. Understanding these systems is relevant as it gives insight into the optimal design of artificial systems such as novel nanofabricated structures for quantum transport or optimized solar cells.

The modeling of LHCs is challenging due to the lack of atomistic ab initio methods and requires one to resort to effective descriptions. This is most apparent in the treatment of the vibrational excitations, which are commonly described by a structureless mode distribution. Then, the energy transfer is calculated by the time propagation of a density matrix, which couples the electronic exciton dynamics to the vibrational environment. For LHCs, the rearrangement of the molecular states after the absorption of the photon has to be taken into account and is described by the reorganization energy. The hierarchical equations of motion (HEOMs)^{12–14} for the time evolution of the density matrix were adapted by Ishizaki and Fleming¹⁵ to include the reorganization process in the transport

equations and is exact within the model of exciton dynamics coupled to a bath with a Drude-Lorentz spectral density.

In principle, the HEOM can be extended to other spectral densities by using a superposition of Drude-Lorentz peaks.^{16,17} Previous calculations for the energy-transfer efficiency of the FMO complex did not consider memory effects and used a weak coupling perturbation theory.^{7,8} Other models try to get around these limitations by using the generalized Bloch-Redfield equations⁹ but yield different results compared to the HEOM solution of the same model system. Prolonged coherent dynamics is predicted due to the slow dissipation of reorganization energy to the vibrational environment.¹⁸ Theoretical descriptions must go beyond the rotating-wave approximation and perturbation theory and require a full incorporation of time nonlocal effects and physiological temperature. The HEOMs fulfill all of these premises.

To date, only the exciton population dynamics for the FMO model has been studied within the full hierarchical approach,^{18,19} whereas the calculation of efficiency or 2D absorption spectra have been considered out-of-range for present computational power, since they require stable algorithms to propagate enlarged system matrices over many more time steps. The adverse computational scaling of the HEOM stems from the need to propagate a complete hierarchy of coupled auxiliary equations, which need to be simultaneously accessed in memory and propagated in time. The insufficient computational power and memory-transfer bandwidth of conventional CPU clusters²⁰ has limited the application of the HEOM to study energy-transfer efficiency in small dimer systems, where other methods are

Received: February 22, 2011

Published: May 25, 2011

available for comparison around $T = 0$ K.^{21–24} The advent of high-performance graphics processing units (GPUs) with several hundred stream processors working in parallel and with a high-bandwidth memory has led us to perform the full HEOM approach for the exciton model of LHC. The efficiency calculations for the FMO system in the strong coupling regime require one to propagate 240 000 auxiliary matrices up to 50 ps (corresponding to 20 000 time steps). The full HEOM approach takes only hours of computational time on a single GPU, whereas the corresponding CPU calculation would run several weeks and becomes completely unfeasible for bigger LHCs due to the large communication overhead. We use the GPU algorithmic advances to characterize the exciton energy-transfer efficiency in LHCs for a wide range of reorganization energies under full consideration of the memory effects and at $T = 300$ K. Our calculations reveal several important mechanisms which are not contained within the approximative methods. The GPU–HEOM method opens the window to a widespread utilization of the HEOM, including the calculation of two-dimensional nonlinear spectra of LHCs, as we will discuss elsewhere. Also, the implementation of a scaled version of the HEOM,¹⁹ which reduces the number of auxiliary matrices, could be achieved on a GPU and reduces the computational effort of hierarchical methods further.

For the development of new theoretical chemistry and physics algorithms, GPUs are important devices and considerably enlarge the class of solvable problems if one manages to devise a program code which takes full advantage of the GPU stream-processing architecture. For interacting many-body systems, this cannot be generally achieved by porting an existing program to the GPU but requires one to follow the vector-programming paradigm from the onset.^{25,26}

The manuscript is organized as follows: In section 2, we set up the model for energy transfer to the reaction center in the FMO complex. In section 3, we calculate the key quantities used to characterize the energy flow, namely, the efficiency and the transfer time to the reaction center. We compute them for a wide range of reorganization energies and bath correlation times within the hierarchical approach. This section contains a detailed discussion of the differences of the HEOM results compared to calculations based on approximative methods. We highlight the main mechanism behind the high efficiency, the delicate balance between the requirements of an energy gradient toward the reaction center, and the detuning of the energies, as shown in section 4. In section 5, we discuss how the transport efficiency is optimized with respect to physiological temperature and comment on the thermalization properties of the HEOM. Finally, we summarize our findings in section 6. Throughout the article, we provide detailed information about the computational times and requirements and collect in the appendices additional detailed information about the algorithms used and our GPU implementation.

2. MODEL

The FMO protein is part of the light-harvesting complex that appears in green sulfur bacteria. Its structure has been widely studied both with X-ray and optical spectroscopic techniques.^{27–29} It has a trimer structure, with each of the monomers consisting of seven bacteriochlorophyll (BChl) pigment molecules, which are electronically excited when the energy flows from the antenna to the reaction center. An ab initio calculation of the energy-transfer process within an atomistic model is far beyond present computational capabilities. Instead,

Table 1. Exciton Hamiltonian in the Site Basis in (cm^{-1})^a

	BChl ₁	BChl ₂	BChl ₃	BChl ₄	BChl ₅	BChl ₆	BChl ₇
BChl ₁	12410	−87.7	5.5	−5.9	6.7	−13.7	−9.9
BChl ₂	−87.7	12530	30.8	8.2	0.7	11.8	4.3
BChl ₃	5.5	30.8	12210	−53.5	−2.2	−9.6	6.0
BChl ₄	−5.9	8.2	−53.5	12320	−70.7	−17.0	−63.3
BChl ₅	6.7	0.7	−2.2	−70.7	12480	81.1	−1.3
BChl ₆	−13.7	11.8	−9.6	−17.0	81.1	12630	39.7
BChl ₇	−9.9	4.3	6.0	−63.3	−1.3	39.7	12440

^a Bold font shows the dominant couplings and site energies. Values taken from ref 32.

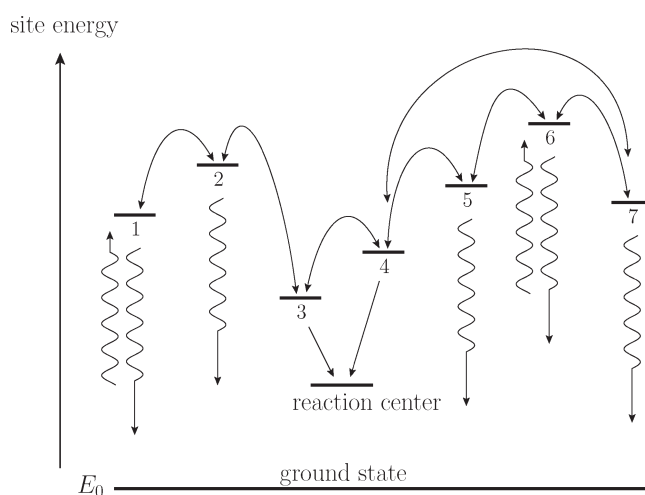


Figure 1. Sketch of the exciton energies of the FMO complex (Table 1), the reaction center, and the ground state. Each site, designated with a number, represents a BChl pigment of the FMO complex. The arrows indicate the dominant intersite couplings. The excitation enters the FMO complex through the chlorosome antenna located close to sites 1 and 6. The incoming excitation, depicted with wavy arrows pointing upward, follows two energy pathways to the reaction center. Wavy arrows pointing downward indicate radiative loss channels leading to the electronic ground state. In addition, each site is coupled to a phonon bath which accounts for the protein environment surrounding the pigments.

one has to develop effective model Hamiltonians such as the widely used excitonic Frenkel–Hamiltonian.^{1,30,31} Within the Frenkel model, which assumes that excitations enter the system one at a time, the seven BChl pigments of the FMO complex are treated as individual sites which are coupled to each other and also to the protein environment. The excitonic Hamiltonian is given by

$$\mathcal{H}_{\text{ex}} = E_0|0\rangle\langle 0| + \sum_{m=1}^N (\varepsilon_m^0 + \lambda_m)|m\rangle\langle m| + \sum_{m>n} J_{mn}(|m\rangle\langle n| + |n\rangle\langle m|) \quad (1)$$

where $N = 7$, $|m\rangle$ corresponds to an electronic excitation of the chromophore BChl_{*m*}, and $|0\rangle$ denotes the electronic ground state of the pigment protein complex where we fix the zero of energy $E_0 = 0$. The site energies $\varepsilon_m = \varepsilon_m^0 + \lambda_m$ of the chromophores consist of the “zero-phonon energies” ε_m^0 and a reorganization energy λ_m , which takes into account the rearrangement of the

complex during excitation due to the phonon bath:¹

$$\mathcal{H}_{\text{reorg}} = \sum_{m=1}^N \lambda_m |m\rangle \langle m| \quad (2)$$

In the following, we will consider identical couplings for all sites, $\lambda_m = \lambda$.

The intersite couplings J_{mn} are obtained by fits to experimentally measured absorption spectra.²⁹ In this contribution, we use the designations and parameters of ref 32, Table 4 (trimer column) and Table 1 (column 4), summarized in Table 1. A sketch of the dominant couplings is shown in Figure 1. The protein environment surrounding the pigments is modeled as identical featureless spectral bath densities coupled to each BChl. For simplicity, we neglect correlations between the baths. The electronic excitations at each site couple linearly with strength d_i to the vibrational phonon modes b_i^\dagger of frequency ω_i . The coupling Hamiltonian is given by

$$\mathcal{H}_{\text{ex-phon}} = \sum_{m=1}^N \left(\sum_i \hbar \omega_i d_i (b_i + b_i^\dagger) \right)_m |m\rangle \langle m| \quad (3)$$

where we assume identical baths at every site. Note that the reorganization energy is related to the coupling by $\lambda = \sum_i \hbar \omega_i d_i^2 / 2$.

We model the losses due to radiative decay from the exciton to the electronic ground state $|0\rangle$ introducing a dipole coupling to an effective radiation photon field a_ν^\dagger

$$\mathcal{H}_{\text{ex-phot}} = \sum_{m=1}^N \sum_\nu (a_\nu + a_\nu^\dagger) \mu_m^\nu (|0\rangle \langle m| + |m\rangle \langle 0|) \quad (4)$$

which results in a finite lifetime for the exciton. The reaction center (RC) is treated as a population-trapping state

$$\mathcal{H}_{\text{trap}} = E_{\text{RC}} |\text{RC}\rangle \langle \text{RC}| \quad (5)$$

and enlarges the system Hamiltonian to a 9×9 matrix. Adolphs and Renger³² suggest that pigments 3 and 4, which have the largest overlap with the energetically lowest exciton-state, couple to the reaction center. Recent experimental evidence shows that pigment 3 is orientated toward the reaction center.³³ In addition, it has been proposed that an eighth pigment may play a role in the initial stages of the energy transfer.³⁴ Here, we include the reaction center by introducing leakage rates from pigments 3 and 4 to the reaction center, which acts as a population-trapping state. Thus, the coupling term to the reaction center reads

$$\mathcal{H}_{\text{ex-RC}} = \sum_{m=3}^4 \sum_{\nu'} (a_{\nu'} + a_{\nu'}^\dagger) \mu_{\text{RC}}^{\nu'} (|\text{RC}\rangle \langle m| + |m\rangle \langle \text{RC}|) \quad (6)$$

where the sum runs over the photon modes at the reaction center. As shown in Appendix A, eqs 20 and 21, the coupling can be expressed in terms of a trapping rate Γ_{RC} , and similarly for the radiative decay in eq 4 with the rate Γ_{phot} . The total Hamiltonian of the system is thus given by

$$\mathcal{H} = \mathcal{H}_{\text{ex}} + \mathcal{H}_{\text{trap}} + \mathcal{H}_{\text{ex-phon}} + \mathcal{H}_{\text{ex-phot}} + \mathcal{H}_{\text{ex-RC}} + \mathcal{H}_{\text{phon}} + \mathcal{H}_{\text{phot}}^0 + \mathcal{H}_{\text{phot}}^{\text{RC}} \quad (7)$$

where $\mathcal{H}_{\text{phon}} = \sum_{i,m} (\hbar \omega_i b_i^\dagger b_i)_{m,m}$, $\mathcal{H}_{\text{phot}}^0 = \sum_{\nu,m} (\hbar \nu a_\nu^\dagger a_\nu)_{m,m}$ and $\mathcal{H}_{\text{phot}}^{\text{RC}} = \sum_{\nu',m=3,4} (\hbar \nu' a_{\nu'}^\dagger a_{\nu'})_{m,m}$. The time evolution of the total density operator $R(t)$ is described by the Liouville equation:

$$\frac{d}{dt} R(t) = -\frac{i}{\hbar} [\mathcal{H}, R(t)] \quad (8)$$

We assume that at initial time $t = 0$ the total density operator factorizes in system and bath components

$$R(t = 0) = \rho(t = 0) \otimes \rho_{\text{phon}} \otimes \rho_{\text{phot}}^0 \otimes \rho_{\text{phot}}^{\text{RC}} \quad (9)$$

while at later times the system and the bath get entangled. Since we are only interested in the exciton dynamics, we trace out the degrees of freedom of the phonon and photon environments $\alpha = \{\text{phon}, \text{phot}^0, \text{phot}^{\text{RC}}\}$ and propagate the reduced 9×9 density matrix in the Schrödinger picture:

$$\begin{aligned} & \rho(t) \\ &= \text{Tr}_\alpha (\exp(-it/\hbar (\mathcal{L}_0 + \mathcal{L}_{\text{ex-phon}} + \mathcal{L}_{\text{ex-phot}} + \mathcal{L}_{\text{ex-RC}} + \mathcal{L}_{\text{bath}})) R(0)) \end{aligned} \quad (10)$$

for the exciton system $\{|m\rangle\}_{m=1,\dots,7}$, the ground electronic state $|0\rangle$, and the reaction center $|\text{RC}\rangle$.

Equation 10 is obtained by formal integration of the Liouville equation, eq 8. The operator $\mathcal{L}_0 = [\mathcal{H}_{\text{ex}} + \mathcal{H}_{\text{trap}}, \bullet]$ represents the coherent dynamics, and $\mathcal{L}_{\text{ex-phon}}$ accounts for dephasing and energy relaxation due to vibrations induced by the interaction with the protein environment, while the recombination and energy trapping are expressed by $\mathcal{L}_{\text{ex-phon}}$ and $\mathcal{L}_{\text{ex-RC}}$, respectively. The parts describing the different baths are summarized in $\mathcal{L}_{\text{bath}} = [\mathcal{H}_{\text{phon}} + \mathcal{H}_{\text{phot}}^0 + \mathcal{H}_{\text{phot}}^{\text{RC}}, \bullet]$. The coupling to the phonon and photon baths can be studied with different degrees of approximation.

We calculate the energy flow within a hybrid formulation which treats the exciton dynamics and the vibrational environment within the HEOM and the trapping to the reaction center and the radiative decay within a Markov model. The Markovian treatment of the photon modes is justified as it occurs on a very different time scale and no backward energy flow to the system is allowed. We abbreviate our model ME-HEOM, see Appendix A. We solve the hierarchical equations using GPUs, which are ideally suited for this task and lead to huge speedups of the algorithm. Details of the computational implementation are collected in Appendix B.

3. TRAPPING TIME FOR DIFFERENT REORGANIZATION ENERGIES

The strong coupling of the excitonic system to the vibrational environment, which is on the same order as the excitonic energy differences (100 cm^{-1}), requires a detailed treatment of the phonon bath over the time-scale of the correlations present in the system. The coupling is quantified by the parameter γ , eq 30, ranging from 35 to 166 fs^{-1} for models of light-harvesting complexes.¹⁸ We calculate the efficiency of the energy transfer from an initially excited site to the reaction center using the hierarchical eqs 39 and 40. The efficiency η is defined as the population of the reaction center at long times:

$$\eta = \langle \text{RC} | \rho(t \rightarrow \infty) | \text{RC} \rangle \quad (11)$$

For the FMO complex, two sites are located near the light-absorbing antenna.³² We consider initial excitations at either site 1 or 6, which give rise to two energy pathways to the reaction center. One pathway starts from site 1 and transfers energy via site 2 to site 3, and the second pathway starts from site 6 with the energy flowing via site 7 or 5 to site 4, see Figure 1.

We fix the upper limit of time propagation at t_{max} defined such that the remaining population in the system, excluding the ground state and reaction center, has dropped from initially 1

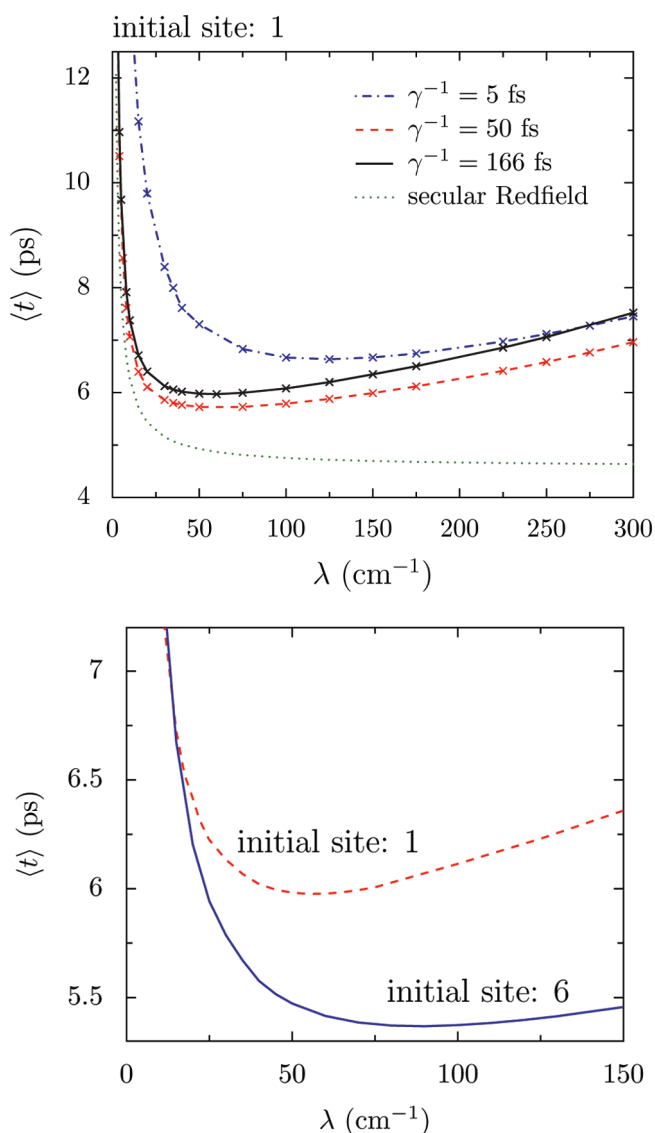


Figure 2. Trapping time from eq 12 as a function of reorganization energy λ at temperature $T = 300$ K. Trapping rate to BChl 3 and 4, $\Gamma_{\text{RC}}^{-1} = 2.5$ ps and $\Gamma_{\text{phot}}^{-1} = 250$ ps. Upper panel: Secular Redfield result with $\gamma^{-1} = 166$ fs and the ME-HEOM results for three different bath correlation times $\gamma^{-1} = 166$ fs, 50 fs, and 5 fs. The excitation enters at site 1. Lower panel: Comparison of the trapping times for the two possible pathways in the FMO when the energy is entering the complex starting from site 1, or at site 6 for a bath correlation time of $\gamma^{-1} = 166$ fs.

to 10^{-5} . To our knowledge, no solid experimental data exist for the coupling strength in eq 6, given in terms of the trapping rate Γ_{RC} of sites 3 or 4 to the reaction center. In the following, we assume values of $\Gamma_{\text{RC}}^{-1} = 2.5$ ps and $\Gamma_{\text{phot}}^{-1} = 250$ ps, which are on the same order of magnitude as in other theoretical studies.^{7,9,11} The actual time scale of the energy trapping is quantified by the trapping time:

$$\langle t \rangle = \int_0^{t_{\text{max}}} dt' t' \left(\frac{d}{dt} \langle \text{RC} | \rho(t) | \text{RC} \rangle \right)_{t=t'} \quad (12)$$

where we replace the upper limit of the integral by t_{max} . The trapping time depends strongly on the reorganization energy, as

shown in Figure 2. For reorganization energies $\lambda < 50$ cm^{-1} , the coupling to the environment assists the transport, and the trapping time decreases when λ increases.

Evaluating the equations of motion (eqs 39 and 40) in the ME-HEOM approach requires one to truncate the hierarchy at N_{max} , which has to be large enough to reach convergence. In Figure 2, we adjust the truncation such that the trapping times for $N_{\text{max}} = N$ and $N_{\text{max}} = N + 1$ differ at most by 0.02 ps. The required truncation increases with the reorganization energy, and for $\lambda = 300$ cm^{-1} , we need $N_{\text{max}} = 16$ where we have to propagate 245 157 auxiliary matrices over 22 000 time steps ($\Delta t = 2.5$ fs), leading to a GPU computation time of 3.7 h. On a standard CPU, the same calculation takes more than one month, and a systematic study of parameters is not feasible.

In the upper panel of Figure 2, we compare the ME-HEOM result with the secular Redfield theory, which employs the time-local Born–Markov approximation in combination with the rotating-wave approximation. For stronger values of the coupling, the hierarchical approach strongly deviates from the plateau obtained within the secular Redfield theory, which assumes a fast decay of the phonon bath. The secular Redfield limit (see Appendix A) reflects, as expected, the qualitative behavior only for small reorganization energies and overestimates the energy transfer to the reaction center for $\lambda > 10$ cm^{-1} .

An interesting question is the existence of an optimal value for the coupling λ and the bath correlation rate γ , for which the trapping time is minimized (and the efficiency maximized). Secular and full Redfield do not yield a local minimum of the trapping time and, thus, no corresponding optimal λ . Introducing the bath correlations and memory effects by the parameter γ in the ME-HEOM gives rise to a local minimum and an optimal value of λ , as shown in Figure 2. In addition, an optimal value of γ emerges around $\gamma^{-1} = 25$ –35 fs. For a small value $\gamma^{-1} = 5$ fs, the theory predicts a rapid loss of efficiency.

The lower panel of Figure 2 details the changes of the trapping time for the two different pathways of the energy flow in the FMO complex as function of the reorganization energy. The optimal reorganization energy for an initial excitation of site 1 is given by $\lambda_{\text{opt}}^1 = 55$ cm^{-1} ($\langle t \rangle_{\text{opt}}^1 = 6.0$ ps), while for an initial excitation of site 6, we obtain $\lambda_{\text{opt}}^6 = 85$ cm^{-1} ($\langle t \rangle_{\text{opt}}^6 = 5.4$ ps).

Optimal values of trapping times have been calculated within the generalized Bloch–Redfield (GBR) approximation.⁹ Using the same parameters, couplings, and Hamiltonian as in ref 9, the ME-HEOM yields qualitative and quantitative differences with a 0.9-ps-longer trapping time for an initial excitation of site 1. For an initial excitation located at site 6, the ME-HEOM and GBR results for the trapping time differ by 0.2 ps.

4. EFFICIENCY FOR REARRANGED ENERGY LEVELS

In this section, we study the relevance of the spacings of the energy levels in the FMO complex to see if the experimentally obtained energy levels (Table 1) are close to an optimal value with respect to transport efficiency at physiological temperature.

The isolated excitonic system shows coherent oscillations of energy between the initially populated site and the delocalized excitonic states. Coupling to the environment gives rise to several mechanisms leading to a nonreversible energy transfer. In the simplest Haken–Strobl model, only dephasing is incorporated,^{7,35} but the temperature is fixed at $T = \infty$. Only by adjusting the dephasing rate can temperature effects be included on a rudimentary level. The ME-HEOM approach enables us to calculate

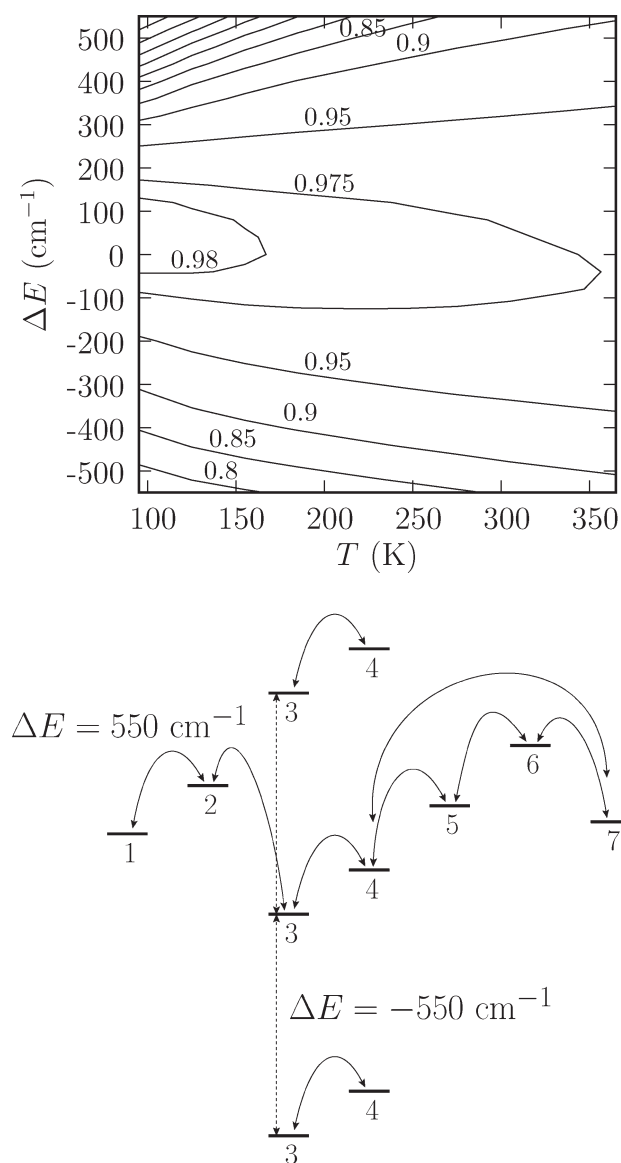


Figure 3. Upper panel: Energy transfer efficiency η in eq 11 as function of temperature and site-energy shifts $\varepsilon_{3/4} \rightarrow \varepsilon_{3/4} + \Delta E$. ME-HEOM parameters: $\lambda = 35 \text{ cm}^{-1}$, $\gamma^{-1} = 166 \text{ fs}$, $\Gamma_{\text{phot}}^{-1} = 250 \text{ ps}$, and $\Gamma_{\text{RC}}^{-1} = 2.5 \text{ ps}$. The hierarchy is truncated at $N_{\text{max}} = 8$. Lower panel: Energy-level shifts considered in the parameter range of the left panel.

the transport at physiological temperature ($T = 300 \text{ K}$) and brings into the picture another crucial mechanism to achieve highly efficient energy transfer, namely, the temperature-dependent stationary site populations. Since the system is in contact with a thermal environment at finite temperature, there is energy dissipation, and the system relaxes to thermal equilibrium. This process guides the excitons to the lowest energy states (for the FMO complex within a few picoseconds) and is not contained in pure dephasing models.

For a small coupling λ and under the assumption that the system and bath degrees of freedom factorize, the thermal state of the system is given by the Gibbs measure

$$\begin{aligned} \rho_{\text{thermal}} &= \exp(-\beta \mathcal{H}_{\text{ex}}) / \text{Tr} \exp(-\beta \mathcal{H}_{\text{ex}}), \\ \beta &= 1/(k_{\text{B}} T) \end{aligned} \quad (13)$$

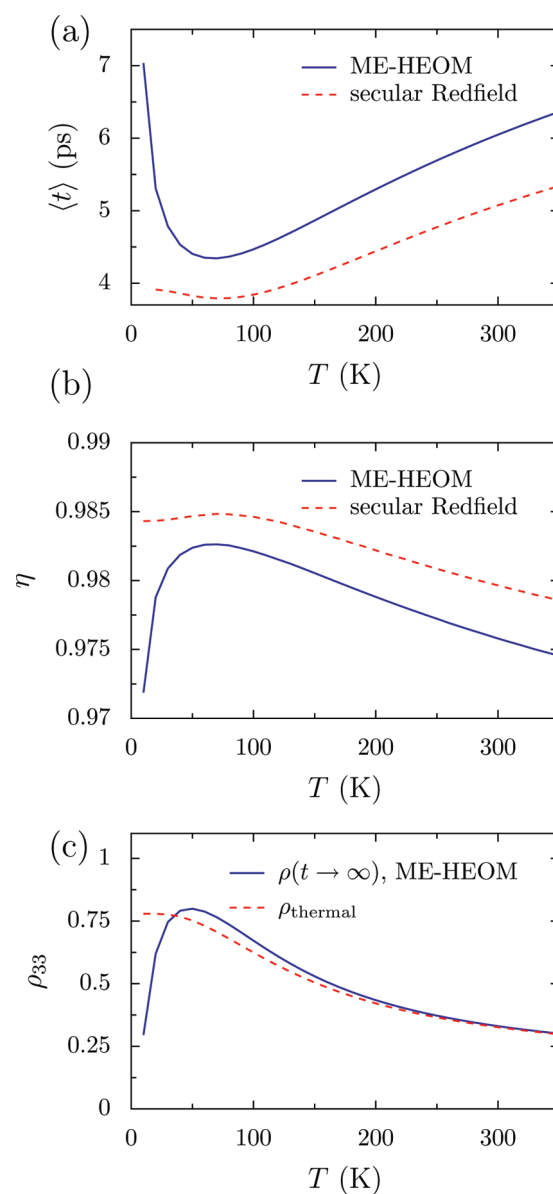


Figure 4. Energy transfer as a function of temperature for the secular Redfield approximation and the exact ME-HEOM calculation with $\gamma^{-1} = 166 \text{ fs}$, $\Gamma_{\text{phot}}^{-1} = 250 \text{ ps}$, $\Gamma_{\text{RC}}^{-1} = 2.5 \text{ ps}$, and truncation $N_{\text{max}} = 8$. Both approaches use a reorganization energy of $\lambda = 35 \text{ cm}^{-1}$ and start with initial population at site 1. (a) Trapping time as a function of temperature. (b) Efficiency as a function of temperature. (c) Population of site 3 for different temperatures in the Boltzmann thermal equilibrium state ρ_{thermal} and for the isolated FMO (decoupled from the reaction center and the radiative decay) using the ME-HEOM, $\rho(t \rightarrow \infty)$. Note that the ME-HEOM needs further corrections at temperatures below 100 K in order to reach the thermal state.

which populates the eigenstates of \mathcal{H}_{ex} according to the Boltzmann statistics. Stronger couplings lead to deviations from the Boltzmann statistics.³⁶ Since the coupling to the reaction center, where the system deposits its excitation, is linked to sites 3 and 4, the efficiency depends strongly on the population and actual site energies 3 and 4. To study this relation, we shift levels $\varepsilon_{3/4} \rightarrow \varepsilon_{3/4} + \Delta E$ and compute the efficiency of the energy transfer. Figure 3 shows the efficiency evaluated with the ME-HEOM. We observe an almost symmetric behavior of the

efficiency for positive and negative energy shifts, with slightly higher efficiencies toward negative energy shifts.

A shift to lower energies increases the energy gradient in the FMO, as the thermal state prefers to populate the low-lying sites. This mechanism improves the transfer efficiency but shifts the two sites out of resonance, and they get decoupled from the other levels of the FMO. Thus, coherent transport becomes more difficult, and the energy transfer to the reaction center is expected to slow down. Similar arguments hold when the energies ε_3 and ε_4 are shifted to higher energies. On the one hand, $\Delta E > 0$ brings sites 3 and 4 closer to resonance and increases the coupling to the remaining sites, thus enhancing coherent transport. On the other hand the thermal state gets delocalized over all sites of the FMO complex, and there is no special preference to populate site 3 and site 4. In such a case, the FMO loses its property to act as an energy funnel, and environment-assisted transport to the reaction center is hindered.

Figure 3 shows how the delicate interplay between coherent delocalization and energy dissipation toward the reaction center gives rise to an optimal arrangement of site energies. We obtain maximal efficiency around $\Delta E = 0$, corresponding to the original parameters in Table 1, and the optimum value is robust against small variations in the site energies.

5. TRAPPING TIME FOR DIFFERENT TEMPERATURES

As discussed in the previous section, the environment assists the transport toward the thermal equilibrium state. In the FMO complex, sites 3 and 4 are coupled to the reaction center and present the lowest exciton energies in the system (see Figure 1); thus the energy dissipation in the phonon environment enhances the population of those sites and hence the efficiency. With increasing temperature, one might expect high transfer efficiencies because thermalization occurs on a faster time scale. However, with increasing temperature, higher energy states have a higher thermal-equilibrium population, and thus the transport efficiency toward the reaction center decreases.

These two competing mechanisms result in an optimal temperature with maximal efficiency. Both mechanisms are already present in the secular Redfield limit, and the optimal energy transfer is obtained around 75 K, see Figure 4a. Our ME-HEOM calculations predict optimal efficiency at a slightly lower temperature of 70 K, but this value is outside the range where our high-temperature implementation is supposed to work (see Appendix A). We obtain a steep increase of the trapping time for low temperatures, shown in Figure 4a, which is also reflected in the efficiencies in Figure 4b. This increase in trapping time and decrease in efficiency is not present in the secular Redfield approach, which saturates for $T \rightarrow 0$. Although we take into account the lowest-order quantum correction to the Boltzmann statistics,¹⁸ at low temperatures, more correction terms are required. One criterion to validate the HEOM is to check if the stationary state $\rho(t \rightarrow \infty)$ of the population dynamics of the isolated FMO, which is decoupled from the reaction center and radiative decay, approaches the thermal state. As is shown in Figure 4c, the HEOM high-temperature implementation fails to approach the thermal state for temperatures below 100 K, where the HEOMs predict an unphysical steep descent of population at low energy site 3, and hence transfer efficiency is underestimated. For temperatures above 100 K, the high temperature limit agrees very well with the thermal state, and the ME-HEOM results are reliable. Comparing our ME-HEOM results above 100 K to the

secular Redfield ones shown in Figure 4a and b, we conclude that the Redfield approach, which is known to be valid in the weak coupling limit only, overestimates the efficiency and underestimates the trapping time.

6. CONCLUSIONS

We have shown that the HEOMs are computationally feasible for calculating the energy transfer for large systems following our GPU implementation. This algorithmic advance allowed us to calculate the efficiency and trapping time of the energy transfer in the FMO complex for a wide range of parameters. The results point to qualitative and quantitative deficiencies of approximative methods and show that an accurate treatment of memory effects and reorganization processes in the system-bath coupling of LHC is needed to evaluate the precise role of temperature, exciton energy differences, coupling strength, and time correlations in the bath. The ME-HEOMs yield longer trapping times and indicate the importance of memory effects and correlations in order to maximize the efficiency in the FMO complex at physiological temperature. Interestingly, the zero-shift energies of the FMO complex provide an almost optimal arrangement for funneling the energy flow to the reaction center at $T = 300$ K. Beyond the results for the FMO complex, our fast computational GPU algorithm for the HEOM provides a robust and scalable way to treat bigger systems and allows us to calculate two-dimensional spectra of LHC, which requires one to enlarge the dimension of the density matrix by taking into account double-excitonic states.

APPENDIX

A. Hybrid Markov–HEOM Approach

A.1. Secular Redfield Approximation. We follow ref 37 and expand the phonon part $\mathcal{L}_{\text{phon}}$ up to second order in the exciton phonon coupling. We use the Born–Markov and secular approximation to obtain

$$-\frac{i}{\hbar} \mathcal{L}_{\text{ex-phon}} = \sum_{m,\omega} iL_m(\omega) [V_m^\dagger(\omega)V_m(\omega), \bullet] + \sum_{m,\omega} i\gamma_m(\omega) \mathcal{D}(V_m(\omega)) \quad (14)$$

where $V_m = |m\rangle\langle m|$ stands for the exciton operators and $\mathcal{D}(V)\rho = V\rho V^\dagger - \frac{1}{2}V^\dagger V\rho - \frac{1}{2}\rho V^\dagger V$. The Lamb shift reads

$$L_m(\omega) = \text{Im} \int_0^\infty dt' \exp(-i\omega t') \langle u_m(t') u_m(0) \rangle_{\text{phon}} \quad (15)$$

and the decoherence rates are given by

$$\gamma_m(\omega) = 2\text{Re} \int_0^\infty dt' \exp(-i\omega t') \langle u_m(t') u_m(0) \rangle_{\text{phon}} \quad (16)$$

with the phonon operators $u_{m,\text{phon}} = (\sum_i \hbar\omega_i d_i (b_i + b_i^\dagger))_m$. The exciton operators

$$V_m(\omega) = \sum_{\omega, M, N} c_m^*(M) c_m(N) |M\rangle \langle N| \delta(\omega - E_M + E_N) \quad (17)$$

are evaluated in the excitonic eigenbasis $|M\rangle = \sum_m c_m(M) |m\rangle$ with $\mathcal{H}_{\text{ex}} |M\rangle = E_M |M\rangle$. For simplicity, we assume that the phonon environments of the individual chromophores are uncorrelated. We additionally neglect the Lamb-type renormalization term.

For the explicit evaluation of the decoherence rates, we quantify the strength of the exciton–phonon coupling and introduce a Drude-Lorentz spectral density

$$J_m(\omega) = 2\lambda_m \frac{\omega\gamma_m}{\omega^2 + \gamma_m^2} \quad (18)$$

The decoherence rates are then given by

$$\gamma(\omega) = \begin{cases} 2\pi J(-\omega)(n(-\omega) + 1), & \text{if } \omega < 0 \\ 2\pi \frac{k_B T}{\hbar} \frac{dJ(\omega)}{d\omega}, & \text{if } \omega = 0 \\ 2\pi J(\omega)n(\omega), & \text{if } \omega > 0 \end{cases} \quad (19)$$

where $n(\omega) = (\exp(\hbar\omega/k_B T) - 1)^{-1}$ corresponds to the phonon statistics. Note that we drop the index m and use the same parameters for all sites.

We describe the radiative decay and trapping to the reaction center using a Lindblad ansatz

$$-\frac{i}{\hbar}\mathcal{L}_{\text{ex-phot}} = \sum_{m=1}^N \Gamma_{\text{phot}} \mathcal{D}(|0\rangle\langle m|) \quad (20)$$

and

$$-\frac{i}{\hbar}\mathcal{L}_{\text{ex-RC}} = \sum_{m=3}^4 \Gamma_{\text{RC}} \mathcal{D}(|\text{RC}\rangle\langle m|) \quad (21)$$

with identical trapping rates for all sites $\Gamma_{\text{phot}} = 2\pi|\mu|^2$ and $\Gamma_{\text{RC}} = 2\pi|\mu_{\text{RC}}|^2$, respectively. Equations 20 and 21 are derived by a master equation approach employing the rotating wave approximation for Hamiltonians $\mathcal{H}_{\text{phot}}^0$ and $\mathcal{H}_{\text{phot}}^{\text{RC}}$, respectively. The strength of the exciton-photon coupling is defined by a constant spectral density $J(\omega) = 2\pi$. We further assume that the photon field cannot create excitations. The same holds for the trapping to the reaction center, as no backward energy flow to the system is allowed. This is equivalent to previous works, where trapping and exciton recombination are included in the Hamiltonian in the form of anti-Hermitian parts.⁷

The equation of motion for the density operator in secular Redfield form reads

$$\frac{d}{dt}\rho(t) = -\frac{i}{\hbar}(\mathcal{L}_0 + \mathcal{L}_{\text{ex-phon}} + \mathcal{L}_{\text{ex-phot}} + \mathcal{L}_{\text{ex-RC}})\rho(t) \quad (22)$$

A.2. Combined Master-Equation–HEOM Approach (ME-HEOM). For the FMO complex, the coupling of the excitons to the phonon environment is large, and the reorganization process occurs on a time scale which is comparable to the system dynamics. The secular Redfield approximation is not valid for the coupling to the phonon bath, and a nonperturbative treatment is required. We will follow the derivation in ref 15 and introduce a set of hierarchically coupled equations. For trapping-time and efficiency calculations, we introduce slight modifications, and in particular, we include the coupling to the reaction center and the radiative decay. We derive a combined ME-HEOM approach which treats the exciton–phonon coupling exactly, whereas the leakage to the reaction center and exciton ground state is described in the Born–Markov limit.

We start with the Liouville equation for the total density operator eq 8 and assume that the total density operator

factorizes eq 9. In the interaction picture with

$$\mathcal{H}_0 = \mathcal{H}_{\text{ex}} + \mathcal{H}_{\text{trap}} + \mathcal{H}_{\text{phon}} + \mathcal{H}_{\text{phot}}^0 + \mathcal{H}_{\text{phot}}^{\text{RC}} \quad (23)$$

where we denote operators with

$$\tilde{\mathcal{O}}(t) = \exp(i\mathcal{H}_0 t/\hbar) \mathcal{O} \exp(-i\mathcal{H}_0 t/\hbar) \quad (24)$$

the Liouville equation reads

$$\begin{aligned} \frac{d}{dt}\tilde{\mathcal{R}}(t) &= -\frac{i}{\hbar}[\tilde{\mathcal{L}}_{\text{ex-phon}} + \tilde{\mathcal{L}}_{\text{ex-phot}} + \tilde{\mathcal{L}}_{\text{ex-RC}}, \tilde{\mathcal{R}}(t)] \\ &= -\frac{i}{\hbar}(\tilde{\mathcal{L}}_{\text{ex-phon}} + \tilde{\mathcal{L}}_{\text{ex-phot}} + \tilde{\mathcal{L}}_{\text{RC}})\tilde{\mathcal{R}}(t) \end{aligned} \quad (25)$$

After formal integration and tracing out the bath degrees of freedom $\alpha = \{\text{phon}, \text{phot}^0, \text{phot}^{\text{RC}}\}$, we get a formal solution for the reduced density operator describing the exciton degrees of freedom

$$\tilde{\rho}(t) = \tilde{\mathcal{U}}(t)\tilde{\rho}(0) \quad (26)$$

with time evolution operator

$$\begin{aligned} \tilde{\mathcal{U}}(t) &= \text{Tr}_\alpha(T_+ \exp(-i/\hbar \int_0^t ds (\tilde{\mathcal{L}}_{\text{ex-phon}}(s) \\ &+ \tilde{\mathcal{L}}_{\text{ex-phot}}(s)) + \tilde{\mathcal{L}}_{\text{ex-RC}}(s)) \times \rho_{\text{phon}} \otimes \rho_{\text{phot}}^0 \otimes \rho_{\text{phot}}^{\text{RC}}) \end{aligned} \quad (27)$$

We make use of the Gaussian nature of the harmonic baths to reduce the bath expectation values to two-time correlation functions. Hence, the influence of the environment is characterized by the symmetrized correlation

$$S_{m,\alpha}(t) = \frac{1}{2}\langle[\tilde{u}_{m,\alpha}(t), \tilde{u}_{m,\alpha}(0)]_+\rangle \quad (28)$$

and the response function

$$\chi_{m,\alpha}(t) = \frac{1}{2}\langle[\tilde{u}_{m,\alpha}(t), \tilde{u}_{m,\alpha}(0)]\rangle \quad (29)$$

where $u_{m,\text{phot}} = \sum_\nu \mu_m^\nu (a_\nu + a_\nu^\dagger)$. We assume that each site is coupled to an independent phonon bath and that there are no correlations between the radiative decay and trapping at different sites. For the exciton–phonon coupling, we employ a Drude-Lorentz spectral density

$$J_m(\omega) = 2\lambda_m \frac{\omega\gamma_m}{\omega^2 + \gamma_m^2} \quad (30)$$

and obtain, in the high temperature limit

$$S_m(t) \simeq \frac{2\lambda_m}{k_B T} \exp(-\gamma_m t) \quad (31)$$

$$\chi_m(t) = 2\lambda_m \gamma_m \exp(-\gamma_m t) \quad (32)$$

The parameter γ_m describes the time scale of correlations in the vibrational environment of the protein. Note that as we consider identical couplings for all sites, the notation is simplified in the main text, and the subindex m is removed from the time correlation scale of the bath γ .

The time evolution operator becomes

$$\begin{aligned} \tilde{U}(t) = & T_+ \prod_{m=1}^N \exp\left(\int_0^t ds \tilde{W}_{m,\text{phon}}(s)\right) \\ & \times \prod_{m=1}^N \exp\left(\int_0^t ds \tilde{W}_{m,\text{phot}^0}(s)\right) \\ & \times \prod_{m=3}^4 \exp\left(\int_0^t ds \tilde{W}_{m,\text{phot}^{\text{RC}}}(s)\right) \end{aligned} \quad (33)$$

with

$$\begin{aligned} \tilde{W}_{m,\alpha} = & -\frac{1}{\hbar^2} \int_0^t ds \tilde{V}_{m,\alpha}(s)^\times \\ & \times \left[S_{m,\alpha}(t-s) \tilde{V}_{m,\alpha}(s)^\times - i\frac{\hbar}{2} \chi_{m,\alpha} \tilde{V}_{m,\alpha}(s)^\circ \right] \end{aligned} \quad (34)$$

We denote the commutation relations by $\mathcal{O}^\times f = [\mathcal{O}, f]$ and $\mathcal{O}^\circ f = [\mathcal{O}, f]_+$. The time evolution of the reduced density matrix is given by

$$\begin{aligned} \frac{d}{dt} \tilde{\rho}(t) = & T_+ \left(\sum_{m=1}^7 \tilde{W}_{m,\text{phon}}(t) + \sum_{m=1}^7 \tilde{W}_{m,\text{phot}^0}(t) \right. \\ & \left. + \sum_{m=3}^4 \tilde{W}_{m,\text{phot}^{\text{RC}}}(t) \right) \tilde{\rho}(t) \end{aligned} \quad (35)$$

Note that the time ordering operator affecting the integration in eqs 34 and 35 is time nonlocal. In the following, we treat the exciton–photon and exciton–reaction center couplings in the Born–Markov limit. That is, the time nonlocal operators $T_+ \sum_{m=1}^7 \tilde{W}_{m,\text{phot}^0}(t)$ and $T_+ \sum_{m=3}^4 \tilde{W}_{m,\text{phot}^{\text{RC}}}(t)$ are replaced by their time-local Born–Markov limits $\mathcal{L}_{\text{ex-phot}}$ and $\mathcal{L}_{\text{ex-RC}}$ defined in eqs 20 and 21, respectively. Equation 35 finally reduces to

$$\begin{aligned} \frac{d}{dt} \tilde{\rho}(t) = & -\frac{i}{\hbar} \mathcal{L}_{\text{phot}} \tilde{\rho}(t) - \frac{i}{\hbar} \mathcal{L}_{\text{RC}} \tilde{\rho}(t) \\ & + T_+ \sum_{m=1}^7 \tilde{W}_{m,\text{phon}}(t) \tilde{\rho}(t) \end{aligned} \quad (36)$$

We define auxiliary operators

$$\begin{aligned} \tilde{\sigma}^{(n_1, \dots, n_7)}(t) = & T_+ \prod_{m,k,l} \left(\int_0^t ds \exp(-\gamma_m(t-s) \tilde{\theta}_m(s)) \right)^{n_m} \\ & \times \exp\left(\int_0^t ds \tilde{W}_{m,\text{phon}}(s)\right) \exp\left(\int_0^t ds \tilde{W}_{k,\text{phot}^0}(s)\right) \\ & \times \exp\left(\int_0^t ds \tilde{W}_{l,\text{phot}^{\text{RC}}}(s)\right) \end{aligned} \quad (37)$$

with

$$\begin{aligned} \tilde{\theta}_m(s) = & i \left(\frac{2\lambda_m}{k_B T \hbar^2} \tilde{V}_{m,\text{phon}}^\times(s) - i\frac{\lambda_m}{\hbar} \gamma_m \tilde{V}_{m,\text{phon}}^\circ(s) \right), \\ \tilde{\sigma}^{(0, \dots, 0)}(t) = & \tilde{\rho}(t) \end{aligned} \quad (38)$$

Table 2. Comparison of CPU and GPU Computation Time of the Population Dynamics of the Isolated FMO Complex^a

N_{max}	# σ matr.	CPU	GPU	speed up	GPU utilization
4	330	120 s	1 s	$\times 120$	22%
6	1716	676 s	3 s	$\times 225$	56%
8	6435	2636 s	7 s	$\times 376$	82%
10	19448	8275 s	19 s	$\times 435$	93%
12	50388	21972 s	48 s	$\times 458$	97%

^a We propagate 1000 time steps. The GPU (NVIDIA C2050) calculations are performed in single precision. Double precision (not required here for converged results) increases the GPU computation time by a factor of 2.

and rewrite the time nonlocal effects into hierarchically coupled equations of motion

$$\begin{aligned} \frac{d}{dt} \rho(t) = & -\frac{i}{\hbar} (\mathcal{L}_{\text{ex}} + \mathcal{L}_{\text{phot}} + \mathcal{L}_{\text{RC}}) \rho(t) \\ & + \sum_m iV_{m,\text{phon}}^\times \sigma^{(n_1, \dots, n_m + 1, \dots, n_7)}(t) \end{aligned} \quad (39)$$

with

$$\begin{aligned} \frac{d}{dt} \sigma^{(n_1, \dots, n_7)}(t) = & \left[-\frac{i}{\hbar} (\mathcal{L}_{\text{ex}} + \mathcal{L}_{\text{phot}} + \mathcal{L}_{\text{RC}}) + \sum_m n_m \gamma_m \right] \sigma^{(n_1, \dots, n_7)}(t) \\ & + \sum_m iV_{m,\text{phon}}^\times \sigma^{(n_1, \dots, n_m + 1, \dots, n_7)}(t) + \sum_m n_m \theta_m \sigma^{(n_1, \dots, n_m - 1, \dots, n_7)}(t) \end{aligned} \quad (40)$$

where we again have used the Born–Markov limit for the trapping and radiative decay. The auxiliary operators keep track of the memory effects of the bath and account for the removal of the reorganization energy. The σ matrices are initially set to zero. For a sufficiently large $N_{\text{max}} = \sum_m n_m$ the diagonal coupling in eq 40 becomes the dominant term, and we can truncate the hierarchy.

B. Algorithm for Implementing the Hierarchical Method on Graphics Processing Units. For the large reorganization energies typically found in LHC, one needs to go beyond the Born–Markov approach and consider non-local temporal effects. We do this by solving the system dynamics within the hierarchical approach shown in the previous section. The method requires considerable memory and computational efforts, and a large number of auxiliary matrices are needed to store the time non-local bath properties. Since all auxiliary matrices have to be accessed to perform the next propagation step, the huge communication overhead renders conventional parallelization schemes, where distributed computing nodes are connected by Ethernet,²⁰ ineffective. GPUs have the two-fold advantage of a fast memory bandwidth and the availability of several hundred stream processors. By assigning one stream processor to each auxiliary matrix, we obtain a speedup of the hierarchical method by the number of processors. The numerical calculations in this manuscript are performed on a NVIDIA Fermi C2050 GPU with 448 processors (1.15 GHz) and 3 gigabytes of ECC-protected on-board memory.

The first step of the algorithm initializes the system of the σ matrices of the hierarchy. With increasing truncation N_{max} ,

the total number of σ matrices grows factorially $N_{\text{tot}} = (N + N_{\text{max}})! / (N! N_{\text{max}}!)$, where N corresponds to the number of sites.¹⁸ As shown in Table 2, the calculation of population dynamics of the FMO complex with $N = 7$ and $N_{\text{max}} = 12$ already requires 50 388 matrices, whereas 330 matrices are sufficient for a truncation at $N_{\text{max}} = 4$. The memory of the σ matrices is allocated on the graphics board and initialized to zero. It is not necessary to transfer the σ matrices to the main-processor memory at any time during the calculation. The only memory transfer between CPU and GPU involves the $N \times N$ entries of the reduced density operator ρ . To advance the propagation one time step in eq 40 requires a large number of matrix multiplications. Each single σ matrix is connected to $2N$ neighbors; these connections are stored in GPU memory in a linked list. The GPU uses 448 cores in parallel with fast GPU memory transfer and thus provides an immense reduction of the computation time up to a factor of 458 for the matrix multiplications. For benchmarking the algorithm, we propagate 1000 time steps using a fourth order Runge-Kutta integrator. For the final output into files, a short memory transfer from the GPU to the CPU is required.

In Table 2, we summarize the computational speedup of the C2050-GPU compared to a standard CPU (Intel 2.40 GHz). The GPU computation is performed using single precision, which yields sufficient accuracy for the problem at hand. For the population dynamics of the FMO complex using $\lambda = 35 \text{ cm}^{-1}$, $\gamma^{-1} = 166 \text{ fs}$, a temperature of 300 K, a propagation time of 10 ps with step size $\Delta t = 10 \text{ fs}$, and truncation $N_{\text{max}} = 12$, the populations are accurate within single precision to six digits $|\rho_{ii}^{\text{single}}(t) - \rho_{ii}^{\text{double}}(t)| < 5 \times 10^{-7}$. This switch from single to double precision increases the computation time approximately by a factor of 2 on the C2050-GPU.

AUTHOR INFORMATION

Corresponding Author

*E-mail: tobias.kramer@mytum.de.

ACKNOWLEDGMENT

This work has been supported by the DAAD project 50240755 and the Spanish MINCINN AI DE2009-0088 (Acciones Integradas Hispano-Alemanas), the Emmy-Noether program of the DFG, KR 2889/2, the Spanish MICINN project FIS2010-18799, and the Ramón y Cajal program.

REFERENCES

- (1) May, V.; Kühn, O. In *Charge and Energy Transfer Dynamics in Molecular Systems*; Wiley-VCH: Weinheim, Germany, 2004; p 62.
- (2) Engel, G. S.; Calhoun, T. R.; Read, E. L.; Ahn, T.-K.; Mancal, T.; Cheng, Y.-C.; Blankenship, R. E.; Fleming, G. R. *Nature* **2007**, *446*, 782.
- (3) Collini, E.; Wong, C. Y.; Wilk, K. E.; Curmi, P. M. G.; Brumer, P.; Scholes, G. D. *Nature* **2010**, *463*, 644.
- (4) Panitchayangkoon, G.; Hayes, D.; Fransted, K.; Caram, J.; Harel, E.; Wen, J.; Blankenship, R.; Engel, G. *PNAS* **2010**, *107*, 12766.
- (5) Gaab, K. M.; Bardeen, C. J. *J. Chem. Phys.* **2004**, *121*, 7813.
- (6) Plenio, M. B.; Huelga, S. F. *New J. Phys.* **2008**, *10*, 113019.
- (7) Rebentrost, P.; Mohseni, M.; Kassar, L.; Lloyd, S.; Aspuru-Guzik, A. *New J. Phys.* **2009**, *11*, 033003.
- (8) Fassiooli, F.; Olaya-Castro, A. *New J. Phys.* **2010**, *12*, 085006.
- (9) Wu, J.; Liu, F.; Shen, Y.; Cao, J.; Silbey, R. J. *New J. Phys.* **2010**, *12*, 105012.
- (10) Hsin, J.; Strümpfer, J.; Sener, M.; Qian, P.; Hunter, C. N.; Schulten, K. *New J. Phys.* **2010**, *12*, 085005.

- (11) Hoyer, S.; Sarovar, M.; Whaley, K. B. *New J. Phys.* **2010**, *12*, 065041.
- (12) Yan, Y.; Yang, F.; Liu, Y.; Shao, J. *Chem. Phys. Lett.* **2004**, *395*, 216.
- (13) Xu, R.-X.; Cui, P.; Li, X.-Q.; Mo, Y.; Yan, Y. *J. Chem. Phys.* **2005**, *122*, 041103.
- (14) Ishizaki, A.; Tanimura, Y. *J. Phys. Soc. Jpn.* **2005**, *74*, 3131.
- (15) Ishizaki, A.; Fleming, G. R. *J. Chem. Phys.* **2009**, *130*, 234111.
- (16) Meier, C.; Tannor, D. J. *J. Chem. Phys.* **1999**, *111*, 3365.
- (17) Kleinekathöfer, U. *J. Chem. Phys.* **2004**, *121*, 2505–14.
- (18) Ishizaki, A.; Fleming, G. R. *PNAS* **2009**, *106*, 17255.
- (19) Zhu, J.; Kais, S.; Rebentrost, P.; Aspuru-Guzik, A. *J. Chem. Phys. B* **2011**, *115*, 1531–1537.
- (20) Strümpfer, J.; Schulten, K. *J. Chem. Phys.* **2009**, *131*, 225101.
- (21) Anders, F. B.; Bulla, R.; Vojta, M. *Phys. Rev. Lett.* **2007**, *98*, 210402.
- (22) Thorwart, M.; Eckel, J.; Reina, J.; Nalbach, P.; Weiss, S. *Chem. Phys. Lett.* **2009**, *478*, 234.
- (23) Roden, J.; Schulz, G.; Eisfeld, A.; Briggs, J. J. *J. Chem. Phys.* **2009**, *131*, 044909.
- (24) Prior, J.; Chin, A. W.; Huelga, S. F.; Plenio, M. B. *Phys. Rev. Lett.* **2010**, *105*, 050404.
- (25) Olivares-Amaya, R.; Watson, M. A.; Edgar, R. G.; Vogt, L.; Shao, Y.; Aspuru-Guzik, A. *J. Chem. Theory Comput.* **2010**, *6*, 135–144.
- (26) Kramer, T.; Krueckl, V.; Heller, E.; Parrott, R. *Phys. Rev. B: Condens. Matter* **2010**, *81*, 205306.
- (27) Olson, J. *Photosynth. Res.* **2004**, *80*, 181–187.
- (28) Brixner, T.; Stenger, J.; Vaswani, H. M.; Cho, M.; Blankenship, R. E.; Fleming, G. R. *Nature* **2005**, *434*, 625–8.
- (29) Milder, M. T. W.; Brüggemann, B.; van Grondelle, R.; Herek, J. L. *Photosynth. Res.* **2010**, *104*, 257.
- (30) Leegwater, J. A. J. *J. Phys. Chem.* **1996**, *100*, 14403.
- (31) Ritz, T.; Park, S.; Schulten, K. *J. Phys. Chem. B* **2001**, *105*, 8259.
- (32) Adolphs, J.; Renger, T. *Biophys. J.* **2006**, *91*, 2778.
- (33) Wen, J.; Zhang, H.; Gross, M. L.; Blankenship, R. E. *PNAS* **2009**, *106*, 6134.
- (34) Schmidt am Busch, M.; Müh, F.; El-Amine Madjet, M.; Renger, T. *J. Phys. Chem. Lett.* **2011**, *2*, 93.
- (35) Chin, A. W.; Datta, A.; Caruso, F.; Huelga, S. F.; Plenio, M. B. *New J. Phys.* **2010**, *12*, 065002.
- (36) Zürcher, U.; Talkner, P. *Phys. Rev. A* **1990**, *42*, 3267–3277.
- (37) Rebentrost, P.; Chakraborty, R.; Aspuru-Guzik, A. *J. Chem. Phys.* **2009**, *131*, 184102.

Calculation of Hyperfine Tensors and Paramagnetic NMR Shifts Using the Relativistic Zeroth-Order Regular Approximation and Density Functional Theory

Jochen Autschbach,^{*,†} Serguei Patchkovskii,[‡] and Ben Pritchard[†]

[†]Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260-3000, United States

[‡]National Research Council of Canada, 100 Sussex Drive Ottawa, Ontario K1A 0R6, Canada

 Supporting Information

ABSTRACT: Density functional theory (DFT) calculations of molecular hyperfine tensors were implemented as a second derivative property within the two-component relativistic zeroth-order regular approximation (ZORA). Hyperfine coupling constants were computed for systems ranging from light atomic radicals to molecules with heavy d and f block elements. For comparison, computations were also performed with a ZORA first-order derivative approach. In each set of computations, Slater-type basis sets have been used. The implementation allows for nonhybrid and hybrid DFT calculations and incorporates a Gaussian finite nucleus model. A comparison of results calculated with the PBE nonhybrid and the PBE0 hybrid functional is provided. Comparisons with differing basis sets and incorporation of finite-nucleus corrections are discussed. The second derivative method is applied to calculations of paramagnetic NMR ligand chemical shifts of three Ru(III) complexes. The results are consistent with those calculated using a first-order derivative method, and the results are consistent for different functionals used. A comparison of two different methods of calculating pseudo-contact shifts, one using the full hyperfine tensor and one assuming a point-charge paramagnetic center, is made for the Ru(III) complexes.

1. INTRODUCTION

Given a molecule with unpaired electrons in an external magnetic field \mathbf{B} , the interaction between the electrons, nuclei, and the external magnetic field can be described through the electron paramagnetic resonance (EPR) spin Hamiltonian:^{1–4}

$$H_{\text{spin}} = -g\beta_e\mathbf{B}\cdot\mathbf{S} - g_N\beta_N\mathbf{B}\cdot\mathbf{I}_N - \mathbf{S}\cdot\mathbf{A}\mathbf{I}_N \quad (1)$$

Here, g is the g -tensor (or g_e for a free electron) and g_N is the nuclear g -factor; \mathbf{S} is the effective spin of the electronic system, and \mathbf{I}_N is the nuclear spin operator. The first two terms in eq 1 are the electronic and nuclear Zeeman terms, describing the interaction between the electron and the external field, and between a nucleus and the external field, respectively. The final term is of interest to this work and describes the magnetic hyperfine interaction between the magnetic moments of the electron and the nuclei. The tensor \mathbf{A} is the hyperfine coupling tensor. In order to compute the hyperfine tensor, it is necessary to consider three contributions: the “first-order” Fermi contact + spin–dipole (FC+SD) terms, and, in relativistic theories, a spin–orbit (SO) coupling cross term with the nuclear paramagnetic spin–orbital (PSO) operator (with the latter also being responsible for the paramagnetic NMR shielding in Ramsey’s theory⁵). Details regarding the theory are provided in section 2, “Theoretical Methods”. In nonrelativistic theory, the isotropic Fermi-contact term can be related to the excess (unpaired) spin density at a nucleus ($\rho^{\alpha-\beta}(0)$)^{1,2,6} (assuming point nuclear charges, in atomic units where $\beta_e = 1/2$ and $\mu_0/(4\pi) = c^{-2}$):

$$A_{\text{iso}}^{\text{nr}} = \left(\frac{4\pi}{6c^2}\right) g_e g_N \beta_N \langle S_z \rangle^{-1} \rho^{\alpha-\beta}(0) \quad (2)$$

Here, $\rho^{\alpha-\beta}(0)$ is the excess spin density at the nucleus in question (the “contact” spin density). From eq 2, it can be seen that the value of A_{iso} may be positive or negative, depending on whether there exists an excess of α or β spin density at a particular nucleus. The sign of the hyperfine coupling constant is taken to be positive where the spins of the electron and nucleus are antiparallel.¹ In relativistic theories, the orbitals and the electron density have weak singularities at point nuclei, and the “contact” operators are modified accordingly in order to sample the electronic structure very close to, but not at, the nuclei.⁷ The derivation in section 2 provides a case in point. The contact or near-contact nature of the relativistically generalized FC mechanism can generally be expected to be subject to large relativistic effects (scalar effects in particular). The anisotropic part of the hyperfine tensor is often not calculated because it does not contribute to the EPR hyperfine coupling for freely rotating molecules (for instance, in gas phase or solution). However, it is an important ingredient for calculations of paramagnetic NMR (pNMR) pseudo-contact chemical shifts.⁸ In pNMR, large A_{iso} further give rise to sizable contact shifts.^{8,9} The spin–orbit contribution to the hyperfine tensor is often neglected for light atomic systems, but it becomes increasingly important with increasing atomic number.¹⁰ Inclusion or neglect of scalar relativistic and spin–orbit (SO) effects can greatly affect the value of magnetic properties, including hyperfine tensors, due to the requirement of describing electrons in close proximity to the nucleus and therefore exhibiting

Received: February 25, 2011

Published: June 06, 2011

relativistic behavior. Arbuznikov et al. have investigated spin-orbit effects on hyperfine tensors and found that the effects can be very significant for a range of molecules and nuclei.¹⁰ In general, it was found that the correlation to experiment improved when SO terms were included in the calculation of the hyperfine tensor.

The treatment of SO effects in calculations of hyperfine coupling tensors may be assigned to one of two types. These two approaches mirror those previously discussed by us in a paper on calculating electronic g -tensors¹¹ and, therefore, are only briefly summarized described here. In the first type of calculation, SO coupling is included variationally in the ground state. The hyperfine tensor is then calculated as a first-order derivative of the energy; therefore, this route may be termed the “first-order” or “expectation value” (EV) approach, and has been implemented in the framework of the two-component relativistic zeroth-order regular approximation (ZORA)¹² by van Lenthe, van der Avoird, and Wormer.¹³ The method has been successfully applied in conjunction with density functional theory (DFT) to transition-metal complexes and small organic molecules.¹³

Alternatively, SO coupling can be introduced as a perturbation on top of nonrelativistic or scalar-relativistic calculations. In this case, the hyperfine tensor is calculated from double-perturbation theory as a second derivative of energy with respect to the nuclear spin angular momentum and the effective electronic spin. The SO contributions to the hyperfine tensor are calculated to lowest order in this approach from solving a set of linear response equations, while the FC+SD terms are still computed from an expectation value integral involving the ground-state orbitals/wave function only. This type of calculation may be termed the “second-order” or “linear response” (LR) approach. One purpose of this article is to present an implementation and first application of the LR approach in conjunction with the ZORA formalism in the calculation of hyperfine tensors and hyperfine coupling constants. The approach presented here closely follows a recent development of a ZORA-based LR method for density functional theory (DFT) calculations of electronic g -tensors.¹¹

DFT is an attractive electronic structure for calculating many molecular properties because of its relatively low computational expense and its often acceptable-to-good accuracy. DFT been used previously for calculations of magnetic properties of open-shell molecules.^{6,10,14,15} Calculations of hyperfine tensors and, in particular, isotropic hyperfine coupling constants, have been previously carried out and benchmarked by various groups using DFT. Hermosilla et al. have analyzed a set of organic and inorganic radicals with the B3LYP hybrid functional and several basis sets, concluding through the use of regression analysis, that such a computational approach is appropriate.^{16,17} Later, Barone et al. calculated hyperfine tensors for 208 free radicals using the B3LYP hybrid functional, again deeming the functional suitable.¹⁴ It should be noted that, in these studies, relativistic effects were not included since they are likely unimportant for the relatively light atoms studied. Compounds with heavier nuclei (in particular, transition-metal complexes) have been studied by a number of authors, in some cases including SO effects^{18–20} and with relativistic effects included using a second-order approach.¹⁰

Herein, we investigate the performance of a LR method of calculating hyperfine tensors using DFT and the two-component relativistic ZORA Hamiltonian. The new implementation allows one to directly compare results calculated with this method to those calculated with the aforementioned EV method reported

by van Lenthe et al.¹³ Differences may be expected for molecules where SO coupling is large, such as radicals with third-row transition metals, lanthanides, or actinides. Both our new LR method and the EV method of van Lenthe et al.¹³ are implemented in the Amsterdam Density Functional (ADF) package,²¹ which is a DFT code that makes use of Slater-type orbital (STO) basis sets. There are several reasons why it is desirable to have a relativistic LR approach available for hyperfine tensor computations. First, in terms of computational efficiency, variational SO DFT computations for large molecules can become very demanding on the computational resources, whereas the LR approach has a computational cost that is comparable to that of scalar relativistic ground-state computations with real orbitals. For systems with light to intermediate strength of SO effects, the analytic LR approach may also be advantageous in terms of numerical accuracy. Another reason pertains the prediction of the absolute and relative signs of the g - and the hyperfine tensors; the sign, which is important for pNMR applications, is straightforwardly determined in LR calculations. In our recent work on g -tensors, we have highlighted an example (NpF₆) where the prediction of the sign of a tensor appears to be problematic. We note that the new LR implementation is capable of handling hybrid functionals and it incorporates finite nucleus effects, which can be significant for hyperfine coupling tensors of heavy atoms.²⁰

After presenting the theoretical background (section 2) and some details regarding the computations (section 3, “Computational Details”), the results of computations using the LR ZORA DFT implementation and other methods are compared in section 4.1, in “Results and Discussion”, using a test set of small molecules with light and heavy atoms. Comparisons are also made with previously calculated values and with experiment, where available. Effects due to basis set, functional (nonhybrid vs hybrid), and finite nucleus corrections are investigated in some detail. We find that the LR method performs well in comparison with the EV method, even for systems with very heavy atoms. For heavy-nucleus hyperfine couplings, finite nucleus effects can be substantial. The sensitivity of hyperfine couplings to features of the electronic structure near the nuclei highlights the need for augmentation of the basis set with high exponent functions (in particular, in relativistic computations). As an additional application, calculated pNMR contact and pseudo-contact shifts are reported for three Ru(III) complexes (see section 4.2). These and several other complexes have been used recently for a computational pNMR benchmark by Rastrelli and Bagno.^{9,22} In ref 22, relativistic effects on the ligand contact shifts were considered using the EV ZORA implementation of ADF, in conjunction with nonhybrid functionals, whereas the influence of the functional has been assessed by a comparison with nonrelativistic B3LYP LR calculations, using Gaussian-type basis sets. We take the opportunity here to revisit a subset of the complexes studied by Rastrelli and Bagno for which the new LR code allows us to investigate the performance of various functionals consistently within the ZORA LR regime using the same STO basis sets. For the pseudo-contact shifts, a comparison is made between using equations for the dipolar interaction, assuming a point paramagnetic center and using the dipolar hyperfine and g -tensors calculated from DFT.

2. THEORETICAL METHODS

2.1. Hyperfine Tensors Calculated As Second Derivatives of the Scalar Relativistic Electronic Energy. What follows is a

brief outline of the theory as it applies to hyperfine tensors. For a related ZORA-based derivation for LR calculations of g -tensors providing additional details and references, we refer the reader to ref 11. If spin-orbit (SO) effects are to be treated as a perturbation, element u, ν of the hyperfine coupling tensor with $u, \nu \in \{x, y, z\}$ can be defined as a second derivative of energy, with respect to electron and nuclear spin:

$$A_{uv} = \frac{\partial^2 E}{\partial I_u \partial S_\nu} = g_N \beta_N \frac{\partial^2 E}{\partial \mu_{N,u} \partial S_\nu} \quad (3)$$

Here, and in the following, derivatives are assumed to be taken at $I_u = 0$, to eliminate higher-order terms from the nuclear spin perturbation. Assuming either a complete basis set or a basis set that is not dependent on the derivative parameters (which is the case for hyperfine coupling), the hyperfine tensor components for nucleus N are within a spin-unrestricted Kohn-Sham DFT, which is given as follows: The “first-order” (EV) FC+SD term reads

$$A_{uv}^{\text{FC} + \text{SD}} = \frac{2g_N \beta_N}{n_\alpha - n_\beta} \sum_{r,s} P_{sr}^{(0)\alpha-\beta} \langle \chi_r | \hat{h}^{(u,\nu)} | \chi_s \rangle \quad (4a)$$

and the second-order (linear response) paramagnetic spin orbital-spin orbit (PSOSO) term is given as

$$A_{uv}^{\text{PSOSO}} = \frac{2g_N \beta_N}{n_\alpha - n_\beta} \sum_{r,s} P_{sr}^{(v)\alpha-\beta} \langle \chi_r | \hat{h}^{(u)} | \chi_s \rangle \quad (4b)$$

Here, n_α and n_β are the numbers of occupied α and β spin orbitals, respectively, expressed in the basis set $\{\chi_r\}$ with MO coefficients $C_{ri}^{(0)}$ (unperturbed) and $C_{ri}^{(v)}$ (perturbed by the ν component of the nuclear spin magnetic moment or by the ν component of the SO operator derivative with respect to the electron spin operator). Furthermore, $P_{rs}^{(0)\alpha-\beta}$ and $P_{rs}^{(v)\alpha-\beta}$ are elements of the scalar relativistic unperturbed and first-order spin density matrices:

$$P_{rs}^{(0)\alpha-\beta} = \sum_i n_i^\alpha C_{ri}^{(0)\alpha} C_{si}^{*(0)\alpha} - \sum_i n_i^\beta C_{ri}^{(0)\beta} C_{si}^{*(0)\beta} \quad (5a)$$

$$P_{rs}^{(v)\alpha-\beta} = \sum_i n_i^\alpha [C_{ri}^{(0)\alpha} C_{si}^{*(v)\alpha} + C_{ri}^{(v)\alpha} C_{si}^{*(0)\alpha}] - \sum_i n_i^\beta [C_{ri}^{(0)\beta} C_{si}^{*(v)\beta} + C_{ri}^{(v)\beta} C_{si}^{*(0)\beta}] \quad (5b)$$

The ZORA¹² one-electron Fock operator in the absence of electromagnetic fields used in DFT computations with a local effective potential V reads (in atomic units)

$$\begin{aligned} \hat{h} &= V + \frac{1}{2}(\vec{\sigma} \cdot \hat{\mathbf{p}}) \mathcal{K} (\vec{\sigma} \cdot \hat{\mathbf{p}}) \\ &= V + \frac{1}{2} \hat{\mathbf{p}} \mathcal{K} \hat{\mathbf{p}} + \frac{1}{2} i \vec{\sigma} \cdot (\hat{\mathbf{p}} \mathcal{K} \times \hat{\mathbf{p}}) \end{aligned} \quad (6)$$

with

$$\mathcal{K} = \frac{2c^2}{2c^2 - V} \quad (7)$$

The potential in \mathcal{K} is typically approximated by a sum of local atomic potentials, which represents an efficient but quite accurate approximation of ZORA used in several electronic structure program packages.^{23–26} In a hybrid DFT scheme, part of the effective exchange potential in eq 6 may be combined with a fraction of

Hartree-Fock exchange. In eq 6, $\vec{\sigma}$ is the 3-vector of 2×2 Pauli spin matrices, with components $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, and $\hat{\mathbf{p}} = -i \nabla$ is the momentum vector operator. Regarding the importance of two-electron SO terms see refs 11 and 27. The scalar relativistic part of eq 6 is taken to be the zeroth-order Fock operator,

$$\hat{h}^{(0)} = V + \frac{1}{2} \hat{\mathbf{p}} \mathcal{K} \hat{\mathbf{p}} \quad (8)$$

while the last term on the right-hand side of eq 6, the ZORA SO operator, is considered to be one of the perturbations. The nuclear spin magnetic perturbation is included in the formalism, following the derivations of refs 11, 13, and 28. The relevant perturbation operators to calculate hyperfine tensors for point nuclei are given by

$$\hat{h}^{(u)} = -\frac{i}{2} [\mathcal{K} (\mathbf{U}_N \times \nabla)_u + (\mathbf{U}_N \times \nabla)_u \mathcal{K}] \quad (9a)$$

$$\hat{h}^{(v)} = \frac{i}{2} (\hat{\mathbf{p}} \mathcal{K} \times \hat{\mathbf{p}})_v \quad (9b)$$

$$\hat{h}^{(u,\nu)} = \frac{1}{2} \{ \delta_{uv} \nabla \cdot (\mathcal{K} \mathbf{U}_N) - \nabla_u (\mathcal{K} \mathbf{U}_{N,\nu}) \} \quad (9c)$$

In the previous equations,

$$\mathbf{U}_N = c^{-2} \left(\frac{\mathbf{r}_N}{r_N^3} \right) \quad (10)$$

assuming a nuclear point magnetic dipole. Here, \mathbf{r}_N is the electron-nucleus distance vector and r_N is its length. (For incorporation of a finite nucleus model, see below.) The operator in eq 9c is the sum of the ZORA analogs of the FC+SD operator derivatives, and eq 9a is the paramagnetic spin-orbit (PSO) operator. Regarding eq 9c, it is important to keep in mind that the derivations leading to the operators as published in refs 28 and 29 implicate that ∇ in this operator only acts inside the operator, not on any function to the right-hand side of the expression. To indicate this, the operator has been enclosed between curly brackets $\{\dots\}$. The operator in eq 9b does not occur explicitly in eqs 4a and 4b. It is used to calculate the perturbed MO coefficients and density matrices (eq 5b) using methods to solve the coupled-perturbed Kohn-Sham (CPKS) equations that have been described elsewhere.^{29–31} We have also implemented the LR part of the computations with the reverse order of perturbation operators, i.e. by solving the CPKS equations for the perturbation eq 9a for each nucleus, instead of using eq 9b, and obtained identical results as required by the interchange theorem of double perturbation theory.

2.2. Nonrelativistic Limit, Pauli Approximation, and Point-Charge Models. The nonrelativistic limit within the ZORA framework is given by $\mathcal{K} \rightarrow 1$. In this case, the hyperfine tensor has only the EV contribution, because the spin derivative of the SO operator ($\hat{h}^{(v)}$) vanishes. If the hyperfine tensor is calculated by using the Pauli approximation of the SO operator, the operator of eq 9b approximates to

$$\hat{h}^{(v)}(\text{Pauli}) = \frac{i}{4c^2} (\hat{\mathbf{p}} V \times \hat{\mathbf{p}})_v \quad (11)$$

corresponding to the substitution $\mathcal{K} \rightarrow V/(2c^2)$. Schreckenbach and Ziegler³² used the operator that is described by eq 11 for LR calculations of molecular g -tensors based on the Pauli approximation. For LR calculations of g -tensors, the SO “spin perturbation” yielding the operator $\hat{h}^{(v)}$ is the same as that used in the calculations of hyperfine tensors.

The nonrelativistic limit of the operator $\hat{h}^{(u)}$ in eq 9a is the well-known PSO operator that is used to calculate nonrelativistic nuclear magnetic shielding and J -coupling tensors. The factor of $1/c^2$ in this operator indicates the smallness of nuclear hyperfine terms and corresponds to $\mu_0/(4\pi)$ in SI units converted to atomic units. Therefore, this factor is not altered when taking a $c \rightarrow \infty$ nonrelativistic limit.

The nonrelativistic limit of the FC+SD operator derivative is obtained with the help of

$$\nabla_u \left(\frac{\mathbf{r}_N}{r_N^3} \right) = \left(\frac{4\pi}{3} \right) \delta_{uv} \delta(\mathbf{r}_N) + \frac{\delta_{uv}}{r_N^3} - 3 \left(\frac{r_{N,u} r_{N,v}}{r_N^5} \right)$$

After letting $\mathcal{K} \rightarrow 1$, taking the derivatives of \mathbf{U}_N in eq 9c, and regrouping terms, the operator reads

$$\hat{h}^{(u,v)}(\text{nrel}) = \frac{1}{2c^2} \left(\left(\frac{8\pi}{3} \right) \delta_{uv} \delta(\mathbf{r}_N) - \left[\frac{\delta_{uv}}{r_N^3} - 3 \left(\frac{r_{N,u} r_{N,v}}{r_N^5} \right) \right] \right) \quad (12)$$

Using this operator in the first term on the rhs of eq 4a yields an expression that is equivalent to the hyperfine tensor contributions given in eqs 2 and 3 of Eriksson,² apart from a factor of $g_e \beta_e$, which equals 1 in atomic units used here, and a factor of c^{-2} that is not present in Eriksson's equation. This factor, again originating from $\mu_0/(4\pi)$ in SI units, indicates the smallness of the nuclear hyperfine terms, relative to other perturbations.

Consider a situation where the hyperfine interaction is caused by one singly occupied molecular orbital (SOMO), with all α orbitals, for the sake of simplicity, assumed to have the same shape as their β counterparts, and $n_\alpha - n_\beta = 1$ in eqs 4a and 4b. In this case, the density matrices in eqs 5a and 5b are both reduced to one term from the SOMO. The assumption that the SOMO is completely localized on one paramagnetic center (for example, a heavy metal) located at the coordinate origin leads to a purely dipolar "through space" hyperfine tensor. Approximating the SOMO spin density as a delta distribution (a point spin density) simplifies the nonrelativistic EV in eq 4a to

$$A_{u,v}^{\text{dip}}(\text{nrel}) \approx \frac{g_N \beta_N}{c^2} \left[3 \left(\frac{r_{N,u} r_{N,v}}{r_N^5} \right) - \frac{\delta_{uv}}{r_N^3} \right] \quad (13)$$

(in atomic units). This simplification is commonly applied in calculations of pseudo-contact (PC) shifts.³³

2.3. Calculation of Operator Matrix Elements, Finite Nucleus Model. The perturbation operators for hyperfine coupling and g -factors, such as other magnetic perturbation operators in ZORA, involve derivatives of \mathcal{K} . In numerical integrations, it is desirable to avoid the calculation of these derivatives. By using the turnover rule for the momentum operator and/or partial integration, the derivatives can be switched over to the basis functions χ_μ and χ_ν instead, assuming that the basis functions vanish at the integration limits.

For the PSO operator (eq 9a) needed for hyperfine coupling, the AO matrix elements are the same as those reported in ref 29, which read as follows (after partial integration):

$$h_{rs}^{(u)} = -\frac{i}{2} \int d^3r \cdot \mathcal{K} (\mathbf{U}_N \times [\chi_r^* (\nabla \chi_s) - (\nabla \chi_r^*) \chi_s])_u \quad (14)$$

To calculate the analogous matrix elements needed for g -tensors, one simply replaces \mathbf{U}_N by $\mathbf{r}/2$, as outlined in ref 11. In some of the calculations, we have adopted a finite nucleus model in the form of Gaussian nuclear charge distributions.^{34,35} In this case,

the nuclear charge density is given as

$$\rho_N^{\text{Gauss}}(\mathbf{R}) = Z_N \left(\frac{\xi_N}{\pi} \right)^{3/2} \exp \left(-\xi_N |\mathbf{R} - \mathbf{R}_N|^2 \right) \quad (15)$$

where \mathbf{R}_N is the charge center of nucleus N and \mathbf{R} is a position in space where the nuclear charge density is evaluated. The exponent ξ_N is readily calculated from the nuclear isotope mass.³⁴ We define $\tilde{r}_N = (\xi_N)^{1/2} r_N$. With the finite nucleus model, the matrix elements of the PSO operator can be straightforwardly calculated from

$$h_{rs}^{(u)} = -\frac{i}{2} \int d^3r \cdot \mathcal{K} P \left(\frac{3}{2}, \tilde{r}_N^2 \right) (\mathbf{U}_N \times [\chi_r^* (\nabla \chi_s) - (\nabla \chi_r^*) \chi_s])_u \quad (16)$$

where $P(a,x)$ is the lower incomplete gamma function ratio:

$$P(a,x) = \frac{1}{\Gamma(a)} \int_0^x dt t^{a-1} \exp(-t) \quad (17)$$

For a point nucleus, the value of $P(3/2, \tilde{r}_N^2)$ in eq 16 is equal to 1 and eq 14 is recovered. The adoption of a Gaussian nucleus model in other hyperfine integrals leads to the same substitution of $\mathcal{K} \mathbf{U}_N$ by $\mathcal{K} P(3/2, \tilde{r}_N^2) \mathbf{U}_N$, which we therefore adopt in the following. Furthermore, the potential energy terms used to determine \mathcal{K} in eq 7 and the electron–nucleus potential used in $\hat{h}^{(0)}$ are calculated based on the Gaussian nucleus model. For further details (presented in the context of calculations of nuclear indirect spin–spin coupling), please see ref 31.

For the bilinear FC+SD operator in eq 9c, the derivatives are only acting within the operator. Thus, a simple partial integration shifts the derivative to the product of the basis functions instead:

$$h_{rs}^{(u,v)} = -\frac{1}{2} \delta_{uv} \int d^3r \mathcal{K} P \left(\frac{3}{2}, \tilde{r}_N^2 \right) \mathbf{U}_N \cdot \nabla (\chi_r^* \chi_s) + \frac{1}{2} \int d^3r \mathcal{K} P \left(\frac{3}{2}, \tilde{r}_N^2 \right) \mathbf{U}_{N,v} \nabla_u (\chi_r^* \chi_s) \quad (18)$$

For the matrix elements of the spin–orbit operator derivative $\hat{h}^{(v)}$, we have adopted the same technique that was used for our recent development for g -tensor computations. The relevant AO integrals are calculated numerically using

$$h_{uv}^{(v)} = \frac{i}{2} \int d^3r (\mathcal{K} - 1) [\{ \nabla \chi_\mu^* \} \times \{ \nabla \chi_\nu \}] \quad (19)$$

As for the other operator matrix elements, with the Gaussian nucleus model, the function \mathcal{K} used in the numerical integration is not the same as when a point nucleus model is adopted.

3. COMPUTATIONAL DETAILS

All computations were carried out with a developers version (pre-2010 release) of the Amsterdam Density Functional (ADF) package.²¹ Geometry optimizations of the molecules in the test set employed the BP86 functional^{36–38} and a triple-zeta polarized STO all-electron basis set with two sets of polarization functions for all atoms (TZ2P from the ADF basis set library), and the scalar ZORA spin-unrestricted formalism. The small molecule test set is a subset of that used in our recent benchmark of ZORA g -tensor calculations.¹¹ Point-group symmetry was not explicitly applied in the computations. Calculations of the

hyperfine tensors were based on the optimized geometries and employed a setting of 7.0 for the numerical integration accuracy parameter to ensure well-converged results and accurate perturbation operator integrals.

The functionals used in the hyperfine-tensor calculations were the Perdew–Burke–Ernzerhof (PBE) nonhybrid functional and, for comparison, the PBE-based PBE0 hybrid functional,^{39,40} which affords 25% Hartree–Fock exchange. In the calculation of hyperfine tensors, a basis set that includes high-exponent functions must be used to accurately calculate the Fermi-contact term.⁴¹ Therefore, a custom STO basis set (JCPL) was used, based on previous work on nuclear spin–spin coupling.^{42,43} This basis set includes functions with high exponents that are required to model the electronic structure close to the nucleus. To determine the effect of the point-nuclei approximation, calculations using a Gaussian finite-nucleus correction were carried out with HgH, HgF, and NpF₆, using the implementation in ref 31, and as outlined above. The comparison between the LR method outlined in Section 2 and the EV approach made use of the ADF implementation by van Lenthe et al.¹³ For comparison, computations with both codes were performed with spin-unrestricted scalar ZORA orbitals, as well as with spin-restricted orbitals populated using configurations with $n_{\alpha} \neq n_{\beta}$.

The structures of Ru(III) complexes chosen for comparison with the work of Rastrelli and Bagno²² were optimized according to the procedure outlined by Rastrelli and Bagno, which is almost identical to that detailed above (BP86/TZ2P, no symmetry, spin-unrestricted scalar ZORA).²² Hyperfine calculations were then carried out as previously described, and also with the Becke88–Perdew86 (BP) nonhybrid functional and with the popular Becke three-parameter hybrid B3LYP.⁴⁴ In order to save some computational expense, the JCPL basis was used solely for the hydrogen atoms for these three molecules; the regular ZORA-optimized TZ2P basis from the ADF basis set library was used for all other atoms. Since the valence basis functions in JCPL are derived from TZ2P, the use of this locally dense basis is not expected to lead to a basis set imbalance for the Ru(III) complexes.

When calculating properties using basis sets as large as JCPL used in the current study, near-linear dependencies may form in the basis set. Overcompleteness of the AO basis can be remedied by removing problematic linear combinations of the basis functions. Such a case occurred with some of the molecules when using the hybrid functionals (PBE0 and B3LYP), in which case a check for linear dependency is automatically switched on in the ADF program. Some molecular orbitals (MOs) were removed accordingly from the self-consistent field procedure and from the MO set used in the calculation of the magnetic properties in order to reduce numerical noise.

4. RESULTS AND DISCUSSION

4.1. Hyperfine Coupling: General Performance of the LR ZORA Approach, Basis Set Effects, and Finite Nucleus Corrections. A summary of the results of the calculation for the suite of test molecules using the various approaches is shown in Table 1. Most results are given with four significant figures; some small numbers are given to within 3 decimal places. Very small values may be affected by the numerical precision of the calculation. All hyperfine couplings are given in SI units of MHz. These can be converted to units of Gauss (G), using the

following formula:³

$$A_{\text{MHz}} = 2.8025 \left(\frac{g_{\text{iso}}}{g_e} \right) A_{\text{Gauss}} \quad (20)$$

where g_{iso} is the isotropic g -factor (LR data consistent with the hyperfine couplings may be taken from ref 11), and g_e is the free-electron g -value. Electron paramagnetic resonance (EPR) experiments cannot straightforwardly deduce the sign of the isotropic hyperfine coupling; therefore, the sign of the experimental hyperfine coupling has been listed in agreement with the sign predicted by the computations.

There is excellent agreement between the method developed in this study and the EV method of van Lenthe et al. for atoms in the first to third rows. For the heavier elements (e.g., Hg), the agreement is still very close, although it can be seen that, percentage-wise, there are significant differences in the PSOSO mechanism for these linear molecules. A breakdown of the perturbational treatment of the SO operator for linear systems echoes that for g -tensors; for a detailed discussion and further references, we refer the reader to refs 11, 54, and 55. The overall good agreement between the LR and EV treatment of hyperfine coupling is not surprising, since the hyperfine tensor has large nonrelativistic and scalar relativistic contributions, which are calculated to be the same with both methods.

In the spin-restricted calculations, we have calculated the FC+SD and the PSOSO terms obtained from the first-order EV approach separately. When comparing these results to those from the spin-restricted LR calculations, it is apparent that for light atomic systems where SO coupling is small, the two mechanisms yield essentially identical results. Continuing with the comparison of the spin-restricted results, starting with SiOH and SiSH, one begins to observe small deviations in the dominating FC+SD term between the two approaches, which is due to SO coupling slightly affecting the shape of the ground-state unperturbed MOs in the EV computations. The PSOSO mechanism also begins to exhibit some differences; we remind the reader that, in the EV approach, this contribution is calculated from an expectation value, just like the FC+SD mechanism, whereas in the LR approach the PSOSO contribution is calculated from the linear response of the orbitals to either the PSO or the SO operator (see section 2).

Larger deviations between calculations and experiment are found for the hyperfine coupling constant of the neptunium nucleus of NpF₆, with the first-order spin-polarized EV treatment underestimating the isotropic coupling and the LR spin-polarized treatment overestimating the isotropic coupling. The spin-restricted calculations are clearly not suitable at all for this coupling constant. Sizable deviations also occur for the fluorine nuclei in HgF, TiF₃, and NpF₆, with both the LR and EV methods underestimating the magnitude of the experimental isotropic hyperfine value. This underestimation is often correlated with the number of lone pairs on the atom of interest and the deficiency of commonly used (LDA and GGA, but also hybrid) functionals in accurately modeling the spin densities with nuclei-containing multiple lone electron pairs.^{10,56,57}

The effect of SO coupling is explicitly (separately) obtained in the set of LR data. While this effect generally small for first- and second-row elements, the need for SO treatment increases with increasing atomic number. Where the SO term is large (in HgH and NpF₆), its inclusion substantially improves the results, compared to a method neglecting this effect. Indeed, for the

Table 1. Summary of Calculated and Experimental Isotropic Hyperfine Coupling Constants (All Values Given in MHz)^a

	Spin-Polarized LR ^b (PBE)			Spin-Polarized LR ^b (PBE0)			Spin-Restricted LR ^b (PBE)			Spin-Restricted EV ^c			exp.	previous calc.	
	FC+SD	PSOSO	total	FC+SD	PSOSO	total	FC+SD	PSOSO	total	FC+SD	PSOSO	total			
CH ₂															
C	208.3	-0.180	208.2	223.7	-0.180	223.5	133.5	-0.153	133.4	209.1					
H	-19.39	0.009	-19.38	-25.51	0.009	-25.50	39.66	0.010	39.67	-19.43					
CH ₃															
C	66.64	-0.152	66.49	82.05	-0.157	81.89	0.071	-0.100	-0.029	67.02	0.070	-0.100	-0.029	108 ^d	
H	-65.47	0.018	-65.45	-74.20	0.017	-74.19	0.002	0.022	0.024	-65.44	0.001	0.022	0.024	-64.46 ^d	
HCO															
C	385.4	-0.478	384.9	393.4	-0.495	392.9	367.0	-0.581	366.4	386.8	367.0	-0.591	366.4	365 ^e	
H	375.6	-0.083	375.5	378.5	-0.084	378.5	334.6	-0.109	334.5	375.2	334.6	-0.119	334.5	354 ^e	343.1 ^k
O	-22.11	0.490	-21.62	-32.21	0.468	-31.74	-13.95	0.538	-13.41	-21.72	-13.95	0.540	-13.41		
HSiO															
Si	-541.8	0.878	-540.9	-557.4	0.859	-556.5	-586.7	0.902	-585.8	-541.2	-586.7	0.903	-585.8	630 ^f	
H	434.0	-0.027	434.0	438.4	-0.027	438.4	359.3	-0.032	359.2	435.4	359.3	-0.032	359.2	450 ^f	397.1 ^k
O	-5.718	0.099	-5.619	-9.740	0.050	-9.689	-4.492	-0.068	-4.560	-5.738	-4.492	-0.060	-4.561		
HSiS															
Si	-509.0	1.648	-507.3	-517.8	1.725	-516.1	-536.1	1.782	-534.3	-506.5	-536.0	1.792	-534.2		
H	328.0	-0.066	327.9	328.8	-0.070	328.8	269.1	-0.077	269.0	328.6	269.1	-0.078	269.0		304.9 ^g
S	-0.105	0.048	-0.057	0.268	0.186	0.454	4.993	0.409	5.402	0.078	4.991	0.399	5.406		
SiOH															
O	-22.45	0.724	-21.73	-22.27	0.736	-21.53	-24.19	3.602	-20.59	-21.75	-23.97	3.616	-20.48		
Si	30.19	6.546	36.74	15.63	7.030	22.66	-2.347	27.41	25.06	37.87	-1.849	26.37	24.48		
H	54.72	-0.151	54.56	52.13	-0.149	51.98	53.17	-0.698	52.47	55.37	52.79	-0.877	52.11		51.3 ^g
SiSH															
S	26.01	-0.164	25.85	21.34	-0.160	21.18	34.59	-0.256	34.34	26.03	34.57	-0.345	34.30		
Si	39.75	4.666	44.42	24.44	5.013	29.45	-0.616	7.816	7.200	44.21	-0.572	7.816	7.202		
H	131.8	-0.059	131.7	122.3	-0.058	122.2	114.5	-0.098	114.4	129.7	114.4	-0.099	114.3		109.6 ^g
HgH															
Hg	7259	-189.7	7070	7708	-192.3	7516	9725	-221.8	9503	6989	9746	-333.5	9413	7002 ^g	6357 ⁱ
H	758.8	-1.800	757.0	742.0	-1.624	740.4	566.3	-1.909	564.4	728.4	543.5	-1.928	541.6	710 ^g	
HgF															
Hg	18480	-22.76	18460	19870	-35.94	19830	20170	-25.55	20140	18420	20070	-63.06	20010	22163 ^h	19292 ⁱ
F	248.9	-144.8	104.2	299.3	-102.9	196.5	196.8	-183.9	12.91	87.57	188.0	-187.8	0.209	670 ^h	
TiF ₃															
Ti	-235.8	2.085	-233.7	-202.6	2.796	-199.8	-270.0	3.417	-266.6	-234.1	-269.9	3.435	-266.4	-184.8 ⁱ	-260.8 ^m
F	1.381	0.535	1.916	-18.53	0.597	-17.93	22.59	1.872	24.46	4.484	22.56	1.889	24.45	23.6 ⁱ	26.1 ^m

Table 1. Continued

	Spin-Polarized LR ^b (PBE)			Spin-Polarized LR ^b (PBE0)			Spin-Restricted LR ^b (PBE)			Spin-Restricted EV ^c			previous calc.		
	FC+SD	PSOSO	total	FC+SD	PSOSO	total	FC+SD	PSOSO	total	EV	FC+SD	PSOSO		total	exp.
NpF ₆ ^o															
Np	-293.1	-2337	-2630	-19.20	-2541	-2560	5.948	-3398	-3392	-1437	47.94	479.2	527.1	-1994 ^d	-2020 ⁿ
F	-36.81	-10.33	-47.13	-50.71	-29.54	-80.25	0.000	-2.218	-2.218	-30.20	-1.770	8.723	22.90	-72.67 ^f	-63 ⁿ

^aAll calculations used a modified TZ2P basis set augmented by high-exponent functions added to all atoms (JCPL⁴³). The point-nuclei approximation was used throughout. Where applicable, calculated coupling constants for symmetry-equivalent atoms have been averaged. ^bEV = expectation value approach ("first-order"). ^cLR = linear response approach ("second-order"). ^dData taken from ref 45. ^eData taken from ref 47. ^fData taken from ref 49. ^gData taken from ref 51. ^hData taken from ref 53. ⁱData taken from ref 46. ^kB3LYP/6-311++G**. ^lData taken from ref 48. ^mDKS2-RL. ⁿData taken from ref 19. ^oSpin-restricted EV. ^pDirac scattered-wave, first-order. ^qData taken from ref 52. ^rSix basis function combinations were removed because of linear dependency.

heaviest atom in our test set, neptunium, the isotropic coupling is predominantly caused by the PSOSO mechanism (~85% in the LR calculations). The results for NpF₆ somewhat resemble those that we obtained recently for the Δg -tensor of this complex,¹¹ in the sense that SO coupling treated at the LR perturbational level appears to be suitable for predicting the huge magnitude and the sign of its EPR parameters.

For most test compounds, calculations using the hybrid functional PBE0 yield results that are close to those obtained using the nonhybrid PBE functional. Again, larger deviations are confined to fluorine nuclei in the heavier compounds. For F in NpF₆, the hybrid DFT result agrees best with the experiment. For Np in NpF₆, the two functionals agree, in terms of the total isotropic coupling, but there are some differences regarding relative contributions of the FC+SD and PSOSO mechanisms. The calculations with PBE predict the FC+SD term to be ~10% of the total coupling, whereas in the calculation with the PBE0 functional, this mechanism becomes insignificant for Np.

The coupling constant calculated with the spin-restricted methods have a much wider variability than those calculated with other methods. While this approach seems fair for the medium atomic weight molecules in our test suite, it yields a vanishing hyperfine coupling for the simplest singlet radical that was tested with this method (CH₃). The planar geometry of CH₃, combined with the nature of the SOMO, is a typical case where spin polarization is essential to obtain the correct result. If the spin-restricted SOMO is a pure C 2p orbital perpendicular to the CH₃ plane, there cannot be any spin density at the carbon nucleus or at the hydrogen nuclei. As a result, the all-important contact term vanishes. The situation resembles that for the phenyl radical as discussed by Rieger.³

It is important to note that, when using spin-restricted orbitals, the sign of the isotropic coupling is fixed. That is, starting from a restricted set of orbitals and assigning the SOMO as one of the α orbitals always results in the $\rho^{\alpha-\beta}(0)$ term in eq 2 being positive. Some of the hyperfine couplings listed in Table 1 contain negative FC+SD contributions because of the negative magnetic moments for oxygen, silicon, and titanium nuclei. The spin-polarized calculations for CH₃ show that spin-polarization effects can be large enough to change the sign of the contact density at a nucleus away from the atom upon which an α spin SOMO is centered.

The calculations discussed up to this point used point charges for the nuclei. This approximation is not always suitable, especially when hyperfine properties for very heavy nuclei are considered. Indeed, Malkin et al. have reported sizable finite nucleus size effects calculated for hyperfine coupling constants of a set of Group 12 compounds including Hg, and for Group 11 atoms.²⁰ Spin-free relativistic DFT calculations using the Douglas–Kroll–Hess Hamiltonian and a Gaussian nuclear charge distribution yielded finite-nucleus effects that reduced the magnitude of Hg hyperfine coupling constants by as much as 20% relative to calculations using a point nucleus, with finite nucleus effects improving the agreement with experiment. Computations with a four-component relativistic method also showed reductions in calculated hyperfine coupling constants, in particular for mercury-containing compounds, although for other molecules the reduction was not as severe.⁵⁸

Similar effects have also been investigated in ZORA computations of NMR spin–spin coupling (J -coupling), which involves perturbation operators that have the same r_N dependence as those that dominate the hyperfine coupling.³¹ Finite nucleus

Table 2. Effect of the Finite-Nucleus Approximation on Calculated Hyperfine Coupling (Isotropic Couplings Given in MHz)^a

	FC+SD		PSOSO		Total		experimental
	point	finite	point	finite	point	finite	
Nonhybrid (PBE)							
HgH							
Hg	7259	6677	−189.7	−190.5	7070	6487	7002 ^b
H	758.8	758.9	−1.800	−1.793	757.0	757.1	710 ^b
HgF							
Hg	18480	17180	−22.76	−21.32	18460	17160	22163 ^c
F	248.9	250.6	−144.8	−146.8	104.2	103.8	670 ^c
NpF ₆ ^e							
Np	−293.1	−216.8	−2337	−2338	−2630	−2555	−1994 ^d
F	−36.81	−47.20	−10.33	−10.35	−47.13	−47.20	−72.67 ^d
Hybrid (PBE0)							
HgH							
Hg	7708	7084	−192.3	−192.9	7516	6891	7002 ^b
H	742.0	741.7	−1.624	−1.617	740.4	740.0	710 ^b
HgF							
Hg	19870	18440	−35.94	−35.01	19830	18410	22163 ^c
F	299.3	300.2	−102.9	−103.8	196.5	196.4	670 ^c
NpF ₆ ^e							
Np	−19.20	13.40	−2541	−2543	−2560	−2530	−1994 ^d
F	−50.71	−50.78	−29.54	−29.61	−80.25	−80.39	−72.67 ^d

^a A Gaussian nuclear charge distribution was used as described in ref 59 and section 2; otherwise, the computational protocol remained the same as previously described. ^b Data taken from ref 50. ^c Data taken from ref 51. ^d Data taken from ref 46. ^e Six basis function combinations removed due to linear dependency.

effects reduced the magnitude of one-bond Hg–ligand coupling constants typically by up to $\sim 10\%$. The correction terms arising from the finite nuclear volume have different origins: one is the modification of the electronic structure due to the weakened electrostatic electron–nucleus potential in the vicinity of the nuclear radius; the other one arises from the finite range of the nuclear current density in the perturbation operators, which is responsible for the nuclear magnetic moment. The level of treatment by a spherical Gaussian distribution may be considered relatively crude, yet the bulk of effects on isotropic hyperfine couplings (and NMR spin–spin coupling) is obtained with this straightforward-to-implement model and should be sufficient for computations of EPR parameters using DFT.

In order to investigate the magnitude of finite nucleus effects in the present LR ZORA hyperfine coupling calculations, we make a comparison between point nucleus and Gaussian nucleus calculations for HgH, HgF, and NpF₆. The results of these calculations are shown in Table 2. The finite nucleus correction decreases the magnitude of the isotropic hyperfine value for all molecules, but strongly so only for the heavy nuclei. In the PBE computations, as well as the PBE0 computations, the effect for Hg is an 8%–9% reduction of the isotropic coupling. The reduction is less pronounced than that which has been reported previously, but this is largely a consequence of the JCP basis set not being able to reach a converged point-nucleus hyperfine coupling constant with respect to the augmentation of the basis with high exponent functions. This point has already been made in ref 31; therefore, we forego additional computations with basis sets that are more compatible with point nucleus computations. In previous work by Malkin et al. using Gaussian basis sets and the second-order

Douglas–Kroll–Hess relativistic two-component Hamiltonian, the point nucleus Hg hyperfine couplings were larger than those in the experiment; therefore, finite nucleus effects improved the agreement with the experiment.²⁰ In our calculations, the trends for the finite nucleus corrections are the same, reducing the Hg hyperfine couplings. Because of the overall slightly lower magnitudes, the agreement with experiment deteriorates slightly when changing from point nuclei to finite nuclei. Ultimately, however, the performance of the computations should be assessed using the most realistic computational model which, for hyperfine coupling, should afford finite nuclei. Table 2 demonstrates that the finite nucleus model primarily affects the FC+SD mechanism, which is to be expected, given the “contact” nature of the FC operator. As a consequence, the hyperfine coupling for neptunium in NpF₆ is hardly affected by finite nucleus corrections because the hyperfine coupling is dominated by the PSOSO mechanism. In the PBE computations, where the FC+SD mechanism is not negligible for Np, a strong reduction (by $\sim 34\%$) of this mechanism due to finite nucleus effects is found. The JCP basis used for Hg set represents an economic choice in terms of balancing accuracy and required computational resources. However, additional high-exponent functions may be needed to fully converge the finite nucleus results for Hg,³¹ with respect to augmentation by high-exponent basis functions on Hg.

Additional computations were performed for HgF with basis sets of different flexibility. The results of varying basis set size with HgF are collected in Table 3. The series of basis sets is not designed to guarantee monotonous convergence but, instead, is used here to illustrate the variability of the results when using small, computationally efficient, basis sets (DZ, DZP), compared

Table 3. Effect of Basis Set Flexibility on the Calculated Isotropic Hyperfine Coupling Constants of HgF (Isotropic Couplings Given in MHz)^a

	DZ	TZP	TZ2P	QZ4P	JCPL	TZ2P3 ^b	experiment ^c
Non-Hybrid (PBE)							
Hg							
FC+SD	14930	15390	15410	1704	17180	17510	
PSOSO	-15.01	-15.53	-13.57	-21.16	-21.32	-18.7	
total	14920	15370	15400	17020	17160	17490	22163
F							
FC+SD	-14.94	54.22	62.52	230.9	250.6	256.0	
PSOSO	-150.0	-146.8	-141.7	-146.1	-146.8	-143.7	
total	-165.0	-92.58	-79.21	84.87	103.8	112.3	670
Hybrid (PBE0)							
Hg							
FC+SD	16070	16550	16620	18300	18440	18760	
PSOSO	-27.77	-30.30	-28.04	-34.77	-35.01	-31.80	
total	16040	16520	16600	18260	18410	18730	22163
F							
FC+SD	-0.650	101.4	108.8	273.5	300.2	307.5	
PSOSO	-104.4	-103.1	-100.3	-104.6	-103.8	-102.5	
total	-105.1	-1.686	8.467	168.9	196.4	205.0	670

^a Computations with a Gaussian nuclear model. ^b TZ2P3 on Hg (see ref 31 and text for details); JCPL for F. ^c Experimental data taken from ref 51.

to basis sets that one may consider as being of good to very good quality (QZ4P and, to some extent, TZ2P). Regarding the convergence for high exponent augmentation, please see ref 31. Both JCPL and QZ4P have several high-exponent functions for Hg but only JCPL has exponents that significantly exceed the nuclear charge for F. Therefore, the effect from the high-exponent augmentation is more strongly seen in the FC+SD contributions of the F hyperfine coupling. For Hg, the JCPL and QZ4P results are more similar, relative to the magnitude of the FC+SD term. Across the table, there is a readily apparent trend for the FC+SD mechanism: increasing the number of basis functions increases the FC+SD contribution, better matching experiment. The sensitivity of the FC mechanism and its relativistic generalizations in nuclear spin–spin coupling is well-documented^{7,60,61} and also is seen here to affect the hyperfine coupling significantly. Table 3 affords additional results with a Hg basis of higher augmentation (TZ2P3), affording Slater exponents of up to 4×10^4 but otherwise comparable to JCPL, which was previously used for a *J*-coupling benchmark.³¹ Compared to the JCPL basis, which is limited to exponents up to 1×10^4 , there is a slight increase in the Hg hyperfine couplings, but not sufficient to obtain quantitative agreement with the experiment. We observe an effect on the F hyperfine couplings when comparing JCPL and TZ2P3, showing that an improved description of the Hg valence orbitals core tails is coupled to changes in the outer region of these orbitals as well. The coupling for fluorine in HgF, although showing some improvement with the larger basis sets, remains conspicuously underestimated, compared to experiment. According to Table 1, with the PBE functional, the comparison of the LR approach with the EV approach reveals some shortcomings of the LR approach for the linear HgF molecule, but not to an extent that would indicate a severe breakdown of the LR calculations. There is a strong cancellation between the FC+SD and the PSOSO mechanisms for fluorine. Such a balance of opposing terms tends to expose

deficiencies in the computational model. We tentatively attribute the discrepancy between the calculated and the experimental fluorine hyperfine coupling for HgF to approximations in the density functionals preventing an accurate description of the spin density distribution and its linear response in this system.

4.2. Paramagnetic NMR Effects in Some Ru(III) Complexes.

Knowledge of the hyperfine tensor can be used in the prediction and analysis of paramagnetic NMR shifts.^{4,8,62,63} The contact shift due to the hyperfine interaction is, to a first approximation, given by^{64,65}

$$\delta^{\text{FC}} = \frac{g_{\text{iso}}\beta_e}{g_N\beta_N} \frac{S(S+1)}{3kT} A_{\text{iso}} \quad (21)$$

where g_{iso} is the isotropic average of the *g*-tensor. Because of this dependence on the *g*-tensor, prediction of pNMR shifts is not straightforward, as the dependence of the pNMR shift on excess α or β electron spin density at a nucleus is to be combined with the sign and magnitude of g_{iso} . For a given calculated sign of A_{iso} , it is possible for the contribution to pNMR shift to be of the opposite sign if the value of g_{iso} is negative. This situation is not typically encountered for organic radicals, but it may well be the case for complexes that contain heavy elements.

In a previous study, Rastrelli and Bagno investigated the effect of relativistic effects in calculations of ligand pNMR shifts of several Ru(III) compounds of potential pharmaceutical interest, with particular emphasis placed on the contribution of contact shifts to the overall pNMR shifts.²² As a first application of our new LR ZORA method for computations of hyperfine tensors, calculations were carried out for three complexes selected from the test set of Rastrelli and Bagno: NAMI, Ru(acac)₃, and Ru(tfac)₃, whose structures are shown in Figure 1. Because the isotropic *g*-factor plays a role (as seen in eq 21), *g*-tensors were calculated using a compatible ZORA method (i.e., first-order EV *g*-tensors were combined with EV hyperfine tensor calculations,

and LR g -tensors were combined with LR hyperfine tensor calculations).

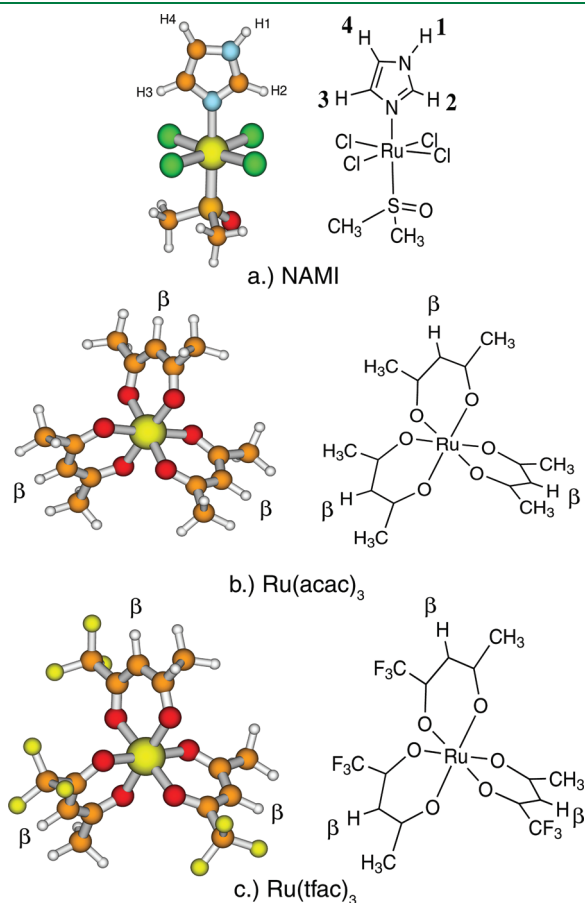


Figure 1. Structures selected from the Rastrelli and Bagno ruthenium complex benchmark set. See ref 22.

Table 4. Isotropic Hyperfine Couplings and g -Factors Calculated for the Present Work and Comparison with Data from Rastrelli and Bagno^{22 a}

	Rastrelli and Bagno ²²		EV ^d		LR ^e		
	Non-Rel. ^b	Rel. ^c	(PBE)	(BP86)	(PBE)	(BP86)	
NAMI							
g_{iso}	1.858	2.194	2.193	2.193	2.203	2.203	2.233
CH ₃	-0.20	-0.32	-0.316	-0.307	-0.327	-0.323	-0.336
H2	0.08	0.094	0.123	0.128	-0.034	-0.019	-0.049
H3	-0.03	-0.11	-0.067	-0.067	-0.188	-0.234	-0.186
H4	-0.10	-0.15	-0.142	-0.145	-0.142	-0.143	-0.163
NH1	-0.06	-0.072	-0.063	-0.062	-0.070	-0.074	-0.102
Ru(acac)₃							
g_{iso}	2.683	2.138	2.103	2.102	2.123	2.122	2.237
CH ₃	-0.54	-0.59	-0.676	-0.709	-0.669	-0.711	-0.726
H- β	-0.68	-0.87	-1.382	-1.557	-1.436	-1.595	-1.286
Ru(tfac)₃							
g_{iso}	2.587	2.084	2.082	2.082	2.103	2.102	2.209
CH ₃	-0.81	-0.89	-0.693	-0.792	-0.702	-0.744	-0.828
H- β	-1.50	-1.87	-1.617	-1.846	-1.730	-1.899	-1.586

^a Isotropic g -factors were calculated using a method equivalent to that used in the calculation of the hyperfine tensor. ^b Hyperfine couplings in MHz. Gaussian09/B3LYP with cc-pVTZ on H,C,N,O,S,F,Cl and DZVP on Ru. ^c ADF/EV/BP86/TZ2P. ^d EV = expectation value approach ("first-order"). ^e LR = linear response approach ("second-order").

In Table 4, calculated isotropic hyperfine couplings and isotropic g -factors are collected and compared with data previously reported by Rastrelli and Bagno. All calculations that employ a relativistic Hamiltonian are seen to yield quite comparable results, with very little deviation among the g -tensors and only slight variations in the isotropic hyperfine couplings. The differences in the hyperfine couplings are likely a consequence of the use of the JCP basis for the protons, as opposed to the universal use of the TZ2P basis by Rastrelli and Bagno. Comparison between the nonhybrid functionals BP86 and PBE shows almost negligible differences with regard to both EV and LR approaches. At the relativistic level, switching to a hybrid functional shows a tendency to increase both the hyperfine coupling magnitudes (with some notable exceptions) as well as the isotropic g -factors. The effects from switching to the hybrid are noticeable but not dramatic for the three Ru(III) complexes.

Table 5 lists the pNMR contact shifts predicted from eq 21, based on the calculated isotropic hyperfine and g -tensor data of Table 4. There is generally good agreement in both magnitude and sign among all relativistic methods, with the main exception being the H2 proton of NAMI. Overall, the comparison with the experiment was not quantitative in ref 22, but the overall trends for the set of complexes were reproduced. Because of the smaller benchmark set used here to test the new implementation, we forego a discussion of the chemical shifts and refer the reader to ref 22. In agreement with Rastrelli and Bagno, the calculated contact shifts do not seem to be negative enough for the H2–H4 and NH protons of NAMI to bring the sign of the overall calculated and experimental chemical shifts for these protons to agreement. The calculated contact shifts for Ru(acac)₃ and Ru(tfac)₃ are large enough such that the sign of the overall calculated chemical shifts match the experiment, but the magnitude of the contact shifts are too large for the methyl protons.

Knowledge of the hyperfine tensor, along with the g -tensor, can also be used in the prediction of pNMR pseudo-contact (PC)

Table 5. Comparison of pNMR Contact Shifts for the Ru(III) Complexes of Figure 1 Estimated from eq 21, Using the Data from Table 4^a

	Rastrelli and Bagno ²²		EV ^d		LR ^e		
	Non-Rel. ^b	Rel. ^c	PBE	BP86	PBE	BP86	B3LYP
NAMI^f							
CH ₃	-6.25	-9.41	-9.166	-8.905	-9.513	-9.412	-9.914
H2	2.45	2.73	3.580	3.726	-0.991	-0.554	-1.447
H3	-0.98	-3.21	-1.954	-1.955	-5.478	-6.818	-5.494
H4	-3.29	-4.29	-4.107	-4.219	-4.138	-4.167	-4.814
NH1	-1.99	-2.11	-1.820	-1.793	-2.040	-2.156	-3.013
Ru(acac)₃^g							
CH ₃	-19.3	-16.8	-18.36	-19.27	-18.35	-19.50	-21.00
H-β	-24.2	-24.6	-37.56	-42.32	-39.39	-43.75	-37.19
Ru(tfac)₃^h							
CH ₃	-25.4	-24.7	-18.85	-21.53	-19.26	-20.43	-23.86
H-β	-46.7	-51.5	-43.96	-50.17	-47.50	-52.12	-45.72

^a Physical constants used in eq 21 are those reported in ref 66. ^b Gaussian09/B3LYP with cc-pVTZ on H,C,N,O,S,F,Cl and DZVP on Ru. ^c ADF/EV/BP86/TZ2P. ^d EV = expectation value approach ("first-order"). ^e LR = linear response approach ("second-order"). ^f Data obtained at 25 °C. ^g Data obtained at 32 °C. ^h Data obtained at 29 °C. See ref 22.

shifts. An approach for calculating the isotropic PC contribution was previously derived by McConnell and Robertson,⁶⁷ expanded upon by Kurland and McGarvey,⁶⁵ and discussed in detail by Bertini et al.³³ This approach approximates the paramagnetic center as a point similar to eq 13. The overall PC shift resulting from the point approximation may take many forms, including³³

$$\delta^{\text{pcs}} = \frac{1}{12\pi r^5} \text{Tr}\{3\mathbf{r} \otimes (\mathbf{r} \cdot \boldsymbol{\chi}) - r^2 \boldsymbol{\chi}\} \quad (22a)$$

$$\delta^{\text{pcs}} = \frac{1}{4\pi r^3} \left[(\chi_{zz} - \bar{\chi}) \frac{2z^2 - x^2 - y^2}{2r^2} + (\chi_{xx} - \chi_{yy}) \frac{x^2 - y^2}{2r^2} + \chi_{xy} \frac{2xy}{r^2} + \chi_{xz} \frac{2xz}{r^2} + \chi_{yz} \frac{2yz}{r^2} \right] \quad (22b)$$

$$\delta^{\text{pcs}} = \frac{1}{4\pi r^3} \left[(\chi_{zz} - \bar{\chi}) \frac{3 \cos^2 \theta - 1}{2} + (\chi_{xx} - \chi_{yy}) \frac{\sin^2 \theta \cos 2\phi}{2} + \chi_{xy} \sin^2 \theta \sin 2\phi + \chi_{xz} \sin 2\theta \cos \phi + \chi_{yz} \sin 2\theta \sin \phi \right] \quad (22c)$$

where \mathbf{r} is the vector from the paramagnetic center to the NMR nucleus of interest, r is the distance $|\mathbf{r}|$, and θ and ϕ are polar angles with respect to the principal axes. The principal components of the magnetic susceptibility tensor $\boldsymbol{\chi}$ can be obtained using the principal elements of the g -tensor:³³

$$\chi_{ii} = \mu_0 \beta_e^2 g_{ii}^2 \frac{S(S+1)}{3kT} \quad (23)$$

An alternate approach to calculating PC shifts that arises from a rigorous derivation of paramagnetic NMR shielding tensors previously carried out by Moon and Patchkovskii,⁸ and shown by Hrobárik et al. to be applicable to metal complexes,⁶² does not make particular assumptions about the spatial distribution of the spin

density from which the dipolar hyperfine tensor is calculated. The PC shift can be calculated in such a way by

$$\delta^{\text{pcs}} = \frac{\beta_e}{\beta_N g_N} \frac{S(S+1)}{9kT} \text{Tr}[g \mathbf{A}_{\text{dip}}^T] \quad (24)$$

with all constants given in SI units, and the traceless dipolar hyperfine tensor $\mathbf{A}_{\text{dip}} = \mathbf{A} - A_{\text{iso}} \mathbf{1}$ with A_{iso} being the isotropic value of the hyperfine tensor, and $\mathbf{1}$ being a 3×3 identity matrix. The superscript T denotes a transpose. Because of the symmetry of the g -tensor, eq 24 can be written in several equivalent forms, differing in the order of $g \mathbf{A}_{\text{dip}}$ matrix multiplication and whether or not the transpose of \mathbf{A}_{dip} is used. Hrobárik et al. obtained an equivalent expression for the PC shift.⁶² We adopt the coordinate frame used by Moon and Patchkovskii,⁸ where the g -tensor is diagonal and the \mathbf{A} -tensor is transformed to the principal axis system (PAS) of the g -tensor. The \mathbf{A}_{dip} tensor (given in SI units of Joules) can be approximated in the same way as in the derivation leading up to eq 13, or alternatively, from the Hamiltonian expression for interacting magnetic dipoles:

$$\hat{H} = -\frac{\mu_0}{4\pi} \left[\frac{3(\boldsymbol{\mu}_N \cdot \mathbf{r})(\boldsymbol{\mu}_e \cdot \mathbf{r})}{r^5} - \frac{\boldsymbol{\mu}_N \cdot \boldsymbol{\mu}_e}{r^3} \right] \quad (25)$$

$$= \frac{\mu_0}{4\pi} g_e \beta_e g_N \beta_N \left[\frac{3(\mathbf{I} \cdot \mathbf{r})(\mathbf{I} \cdot \mathbf{r})}{r^5} - \frac{\mathbf{I} \cdot \mathbf{S}}{r^3} \right] \quad (26)$$

where $\boldsymbol{\mu}_N$ is the nuclear magnetic moment operator ($\boldsymbol{\mu}_N = \gamma_N \mathbf{I} = g_N \beta_N \mathbf{I}$) and $\boldsymbol{\mu}_e$ is the magnetic moment for the electron. In eq 26, the magnetic moment of a free electron ($\boldsymbol{\mu}_e = -g_e \beta_e \mathbf{S}$) was used, and, under this approximation, the resulting equation for the traceless dipolar hyperfine tensor is identical to eq 13 (given here in SI units):

$$\mathbf{A}_{u,v}^{\text{dip}} \approx \frac{\mu_0}{4\pi} g_e \beta_e g_N \beta_N \left[3 \frac{r_{N,u} r_{N,v}}{r_N^5} - \frac{\delta_{uv}}{r_N^3} \right] \quad (27)$$

In molecules that contain heavy elements, particularly transition metals, such an approximation is likely not valid and a better

Table 6. Comparison of eq 22a and eq 24 in Calculating Pseudo-contact Shifts in Selected Ru(III) Complexes^a

	Chemical Shift (ppm)	
	eq 22a ³³	eq 24
NAMI^b		
CH ₃	-2.31	-1.41
H2	-3.52	-2.09
H3	-4.51	-2.30
H4	-2.19	-0.98
NH1	-2.23	-0.99
Ru(acac)₃^c		
CH ₃	-0.09	-0.04
H-β	-1.05	-1.25
Ru(tfac)₃^d		
CH ₃	-0.09	-0.04
H-β	-0.97	-1.17

^aPBE computations. ^bData obtained at 25 °C. ^cData obtained at 32 °C. ^dData obtained at 29 °C.

expression for the electron spin magnetic moment is $\mu_e = -\beta_e g S$, where g is the full g -tensor.¹ By substituting this into eq 25, one can obtain an expression for the dipolar hyperfine tensor:

$$A_{\text{dip}} \approx \left(\frac{\mu_0 \beta_e g_N \beta_N}{4\pi} \right) \mathbb{T} \cdot g \quad (28)$$

where $\mathbb{T} = r^{-5}[3r_{rr} - r^2\mathbb{1}]$ is the geometric factor seen in eqs 13 and 27; this is consistent with McConnell and Robertson's derivation.⁶⁷ Substituting eq 28 into eq 24 results in the equations given by Bertini et al. (eqs 22a–22c),³³ assuming diagonal g and magnetic susceptibility tensors.

To compare the two approaches, the isotropic PC shifts were calculated for the Ru(III) complexes of the Rastrelli and Bagno study discussed above. The results are collected in Table 6. The calculated g -tensor was diagonalized and both the approximate hyperfine tensor calculated using eq 24 and the tensor from the ZORA DFT calculations were transformed to the PAS of g accordingly. The conversion factor used for the calculated ZORA hyperfine tensor uses g_e . Therefore, in order to directly compare the two approaches, the hyperfine tensor from ADF was first divided by the value of g_e , and then multiplied by the calculated g -tensor in order to arrive at eq 28.

The agreement between the two methods reveals some significant differences. However, there is universal agreement in sign. (See the Supporting Information for more detailed data.) The discrepancies between the two methods must be attributed to the point-dipole approximation underlying eq 22a, which is not used in the PC shifts calculated from eq 24. Visualization of the SOMOs for the Ru(III) complexes (see the Supporting Information) demonstrates that these orbitals have a significant spatial extension; thus, an approximation that neglects the spatial distribution of the spin density can be expected to break down to some extent. The PC shifts calculated from the full ZORA DFT hyperfine tensors (eq 24) are in most cases smaller in magnitude than those calculated with the point-dipole approximation. Qualitatively, the situation is similar to where one considers the difference of the electrostatic energy between two point charges and the interaction between a point charge and a continuous

spherical charge distribution, with the latter being smaller in magnitude. The most noticeable discrepancies are found with the H-β protons on Ru(acac)₃ and Ru(tfac)₃. Two of these protons on each complex are bonded to carbon atoms that contain significant electron density from the SOMO. Considering the spatial distribution of the spin density in the hyperfine tensor computations results in a sizable increase of calculated PC shift magnitudes for Ru(acac)₃, and a slightly smaller (but still significant) increase with the corresponding protons in Ru(tfac)₃. The other H-β protons, which lie along the principal axis of their respective complexes, do not afford SOMO density on the adjacent carbons; therefore, the effect is not as large.

Therefore, the trend is that the spatial extension of the SOMO, compared to a point spin density, will yield a smaller PC shift, potentially excluding cases where the SOMO is significantly delocalized to regions near the nucleus for which the PC shift is being calculated.

Isosurface plots of the SOMOs and tables with individual proton contact and PC shifts are available in the Supporting Information.

5. SUMMARY AND OUTLOOK

A density functional theory (DFT)-based method for calculating the electron paramagnetic resonance hyperfine coupling tensors, using second-order perturbation theory and the relativistic zeroth-order regular approximation (ZORA) Hamiltonian, has been developed and tested for radicals with few atoms and for three Ru(III) complexes. The implementation (1) makes use of Slater-type orbital (STO) basis sets, (2) is capable of both nonhybrid and hybrid DFT computations, and (3) supports a Gaussian finite nucleus model. The new procedure performs well, compared to hyperfine couplings calculated using an expectation value (first-order) approach developed previously by van Lenthe et al.¹³ Using a hybrid functional shows a tendency to increase the magnitude of the calculated hyperfine couplings. The use of a finite nucleus model may significantly improve the agreement with the experiment for very heavy atoms. The role of basis set was also investigated, with the requirement for high-exponent basis functions to describe the core electronic structure along with flexibility in the valence region being illustrated in computations on HgF. Preliminary test calculations of contact and pseudo-contact paramagnetic NMR chemical shifts based on calculated hyperfine and g -tensors for a set of Ru(III) complexes demonstrate that the method is computationally efficient.

■ ASSOCIATED CONTENT

S Supporting Information. Supporting Information is available containing information about individual proton shifts and SOMO distribution for the Ru(III) complexes. This information is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jochena@buffalo.edu.

■ ACKNOWLEDGMENT

The authors acknowledge support of this research from the Center of Computational Research at SUNY Buffalo, and

financial support from the U.S. Department of Energy (Grant No. DE-SC0001136) (BES Heavy Element Chemistry Program).

REFERENCES

- (1) Atherton, N. M. *Ellis Horwood Series in Physical Chemistry*; Ellis Horwood, PTR Prentice Hall: New York, 1993; pp 46–50, 210.
- (2) Eriksson, L. A. ESR Hyperfine Calculations. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; Wiley: Chichester, U.K., 1998; pp 952–958.
- (3) Rieger, P. H. *Electron Spin Resonance. Analysis and Interpretation*; The Royal Society of Chemistry: Cambridge, U.K., 2007; p 3.
- (4) Abragam, A.; Bleaney, B. *Electron Paramagnetic Resonance of Transition Ions*; Clarendon Press: Oxford, U.K., 1970; pp 133–216.
- (5) Pyykkö, P. *Theor. Chem. Acc.* **2000**, *103*, 214–216.
- (6) Kossmann, S.; Kirchner, B.; Neese, F. *Mol. Phys.* **2007**, *105*, 2049–2071.
- (7) Autschbach, J.; Zheng, S. *Annu. Rep. NMR Spectrosc.* **2009**, *67*, 1–95.
- (8) Moon, S.; Patchkovskii, S. First-principles calculations of paramagnetic NMR shifts. In *Calculation of NMR and EPR Parameters. Theory and Applications*; Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley–VCH: Weinheim, Germany, 2004.
- (9) Rastrelli, F.; Bagno, A. *Chem.—Eur. J.* **2009**, *15*, 7990–8004.
- (10) Arbuznikov, A. V.; Vaara, J.; Kaupp, M. *J. Chem. Phys.* **2004**, *120*, 2127–2139.
- (11) Autschbach, J.; Pritchard, B. *Theor. Chem. Acc.* **2011**, *129*, 453–466.
- (12) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1993**, *99*, 4597–4610.
- (13) van Lenthe, E.; van der Avoird, A.; Wormer, P. E. S. *J. Chem. Phys.* **1998**, *108*, 4783–4796.
- (14) Barone, V.; Cimino, P.; Stendardo, E. *J. Chem. Theor. Comput.* **2008**, *4*, 751–764.
- (15) Neese, F. *Coord. Chem. Rev.* **2009**, *253*, 526–563.
- (16) Hermosilla, L.; Calle, P.; Garca de la Vega, J. M.; Sieiro, C. *J. Phys. Chem. A* **2005**, *109*, 1114–1124.
- (17) Hermosilla, L.; Calle, P.; Garca de la Vega, J. M.; Sieiro, C. *J. Phys. Chem. A* **2005**, *109*, 7626–7635.
- (18) Remenyi, C.; Reviakine, R.; Arbuznikov, A. V.; Vaara, J.; Kaupp, M. *J. Phys. Chem. A* **2004**, *108*, 5026–5033.
- (19) Komorovský, S.; Repiský, M.; Malkina, O. L.; Malkin, V. G.; Malkin, I.; Kaupp, M. *J. Chem. Phys.* **2006**, *124*, 084108.
- (20) Malkin, E.; Malkin, I.; Malkina, O. L.; Malkin, V. G.; Kaupp, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4079–4085.
- (21) Baerends, E. J.; Ziegler, T.; Autschbach, J.; Bashford, D.; Bérces, A.; Bickelhaupt, F. M.; Bo, C.; Boerrigter, P. M.; Cavallo, L.; Chong, D. P.; Deng, L.; Dickson, R. M.; Ellis, D. E.; van Faassen, M.; Fan, L.; Fischer, T. H.; Fonseca Guerra, C.; Ghysels, A.; Giammona, A.; van Gisbergen, S. J. A.; Götz, A. W.; Groeneveld, J. A.; Gritsenko, O. V.; Grüning, M.; Gusarov, S.; Harris, F. E.; van den Hoek, P.; Jacob, C. R.; Jacobsen, H.; Jensen, L.; Kaminski, J. W.; van Kessel, G.; Kootstra, F.; Kovalenko, A.; Krykunov, M. V.; van Lenthe, E.; McCormack, D. A.; Michalak, A.; Mitoraj, M.; Neugebauer, J.; Nicu, V. P.; Noodleman, L.; Osinga, V. P.; Patchkovskii, S.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Rodríguez, J. I.; Ros, P.; Schipper, P. R. T.; Schreckenbach, G.; Seldenthuis, J. S.; Seth, M.; Snijders, J. G.; Solà, M.; Swart, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visscher, L.; Visser, O.; Wang, F.; Wesolowski, T. A.; van Wezenbeek, E. M.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; Yakovlev, A. L. Amsterdam Density Functional; SCM, Theoretical Chemistry, Scientific Computing & Modelling (SCM), Theoretical Chemistry, Vrije Universiteit: Amsterdam, The Netherlands. (URL: <http://www.scm.com>.)
- (22) Rastrelli, F.; Bagno, A. *Magn. Reson. Chem.* **2010**, *48* (S1), S132–S141.
- (23) Philipsen, P. H. T.; van Lenthe, E.; Snijders, J. G.; Baerends, E. J. *Phys. Rev. B* **1997**, *56*, 13556–13562.
- (24) Nichols, P.; Govind, N.; Bylaska, E. J.; de Jong, W. A. *J. Chem. Theor. Comput.* **2009**, *5*, 491–499.
- (25) Aquino, F.; Govind, N.; Autschbach, J. *J. Chem. Theor. Comput.* **2010**, *6*, 2669–2686.
- (26) van Wüllen, C. *J. Chem. Phys.* **1998**, *109*, 392–399.
- (27) Patchkovskii, S.; Strong, R. T.; Pickard, C. J.; Un, S. *J. Chem. Phys.* **2005**, *122*, 214101.
- (28) Autschbach, J.; Ziegler, T. *J. Chem. Phys.* **2000**, *113*, 9410–9418.
- (29) Autschbach, J.; Ziegler, T. *J. Chem. Phys.* **2000**, *113*, 936–947.
- (30) Autschbach, J. *J. Chem. Phys.* **2008**, *129*, 094105 (and erratum, *J. Chem. Phys.* **2009**, *130*, 209901).
- (31) Autschbach, J. *ChemPhysChem* **2009**, *10*, 2274–2283.
- (32) Schreckenbach, G.; Ziegler, T. *J. Phys. Chem. A* **1997**, *101*, 3388–3399.
- (33) Bertini, I.; Luchinat, C.; Parigi, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *40*, 249–273.
- (34) Visscher, L.; Dylla, K. *At. Data Nucl. Data Tables* **1997**, *67*, 207–224.
- (35) Andrae, D. *Phys. Rep.* **2000**, *336*, 413–527.
- (36) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (37) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.
- (38) Perdew, J. P. *Phys. Rev. B* **1986**, *34*, 7406.
- (39) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029–5036.
- (40) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (41) Watson, M. A.; Handy, N. C.; Cohen, A. J.; Helgaker, T. *J. Chem. Phys.* **2004**, *120*, 7252–7261.
- (42) Bryce, D.; Autschbach, J. *Can. J. Chem.* **2009**, *87*, 927–941.
- (43) Moncho, S.; Autschbach, J. *J. Chem. Theor. Comput.* **2010**, *6*, 223–234.
- (44) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (45) Weltner, Jr., W. *Magnetic Atoms and Molecules*; Dover Publications, Inc.: New York, 1983.
- (46) Butler, J. E.; Hutchison, C. A., Jr. *J. Chem. Phys.* **1981**, *74*, 3102–3119.
- (47) Holmberg, R. W. *J. Chem. Phys.* **1969**, *51*, 3255–3260.
- (48) Grein, F. *J. Chem. Phys.* **2004**, *296*, 71–78.
- (49) Van Zee, R. J.; Ferrante, R. F.; Weltner, W., Jr. *J. Chem. Phys.* **1985**, *83*, 6181–6187.
- (50) Knight, L. B., Jr.; Weltner, W., Jr. *J. Chem. Phys.* **1971**, *55*, 2061–2070.
- (51) Knight, L. B., Jr.; Fisher, T. A.; Wise, M. B. *J. Chem. Phys.* **1981**, *74*, 6009–6013.
- (52) Case, D. A. *J. Chem. Phys.* **1985**, *83*, 5792–5796.
- (53) De Vore, T. C.; Weltner, W. *J. Am. Chem. Soc.* **1977**, *99*, 4700–4703.
- (54) Belanzoni, P.; van Lenthe, E.; Baerends, E. J. *J. Chem. Phys.* **2001**, *114*, 4421–4433.
- (55) Patchkovskii, S.; Schreckenbach, G. Calculation of EPR g-tensors with density functional theory. In *Calculation of NMR and EPR Parameters. Theory and Applications*; Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley–VCH: Weinheim, Germany, 2004.
- (56) Malkina, O. L.; Salahub, D. R.; Malkin, V. G. *J. Chem. Phys.* **1996**, *105*, 8793–8800.
- (57) Malkin, V. G.; Malkina, O. L.; Salahub, D. R. *Chem. Phys. Lett.* **1994**, *221*, 91–99.
- (58) Malkin, E.; Repiský, M.; Komorovský, S.; Mach, P.; Malkina, O. L.; Malkin, V. G. *J. Chem. Phys.* **2011**, *134*, 044111.
- (59) Hansen, J.; Autschbach, J.; Davies, H. *J. Org. Chem.* **2009**, *74*, 6555–6563.
- (60) Kowalewski, J. *Annu. Rep. NMR Spectrosc.* **1982**, *12*, 81–176.
- (61) Helgaker, T.; Jaszuński, M.; Ruud, K. *Chem. Rev.* **1999**, *99*, 293–352.
- (62) Hrobárik, P.; Reviakine, R.; Arbuznikov, A. V.; Malkina, O. L.; Malkin, V. G.; Köhler, F. H.; Kaupp, M. *J. Chem. Phys.* **2007**, *126*, 024107.
- (63) Kaupp, M.; Köhler, F. H. *Coord. Chem. Rev.* **2009**, *253*, 2376–2386.
- (64) Vega, A. J.; Fiat, D. *Pure Appl. Chem.* **1972**, *32*, 307–315.

- (65) Kurland, R. J.; McGarvey, B. R. *J. Magn. Reson.* **1970**, *2*, 286–301.
- (66) Mohr, P. J.; Taylor, B. N.; Newell, D. B. *Rev. Mod. Phys.* **2008**, *80*, 633–730.
- (67) McConnell, H. M.; Robertson, R. E. *J. Chem. Phys.* **1958**, *29*, 1361–1365.

Surface Hopping Excited-State Dynamics Study of the Photoisomerization of a Light-Driven Fluorene Molecular Rotary Motor

Andranik Kazaryan,[†] Zhenggang Lan,[‡] Lars V. Schäfer,[§] Walter Thiel,^{*,‡} and Michael Filatov^{*,||}

[†]Division of Theoretical Chemistry, Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

[‡]Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470, Mülheim an der Ruhr, Germany

[§]Molecular Dynamics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

^{||}University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

S Supporting Information

ABSTRACT: We report a theoretical study of the photoisomerization step in the operating cycle of a prototypical fluorene-based molecular rotary motor (**1**). The potential energy surfaces of the ground electronic state (S_0) and the first singlet excited state (S_1) are explored by semiempirical quantum-chemical calculations using the orthogonalization-corrected OM2 Hamiltonian in combination with a multireference configuration interaction (MRCI) treatment. The OM2/MRCI results for the S_0 and S_1 minima of the relevant **1-P** and **1-M** isomers and for the corresponding S_0 transition state are in good agreement with higher-level calculations, both with regard to geometries and energetics. The S_1 surface is characterized at the OM2/MRCI level by locating two S_0 – S_1 minimum-energy conical intersections and nearby points on the intersection seam and by computing energy profiles for pathways toward these MECIs. Semiclassical Tully-type trajectory surface hopping (TSH) simulations with on-the-fly OM2/MRCI calculations are carried out to study the excited-state dynamics after photoexcitation to the S_1 state. Fast relaxation to the ground state is observed through the conical intersection regions, predominantly through the lowest-energy one with a strongly twisted central C=C double bond and pyramidalized central carbon atom. The excited-state lifetimes for the direct and inverse photoisomerization reactions (1.40 and 1.79 ps) and the photostationary state ratio (2.7:1) from the TSH-OM2 simulations are in good agreement with the available experimental data (ca. 1.7 ps and 3:1). Excited-state lifetimes, photostationary state ratio, and dynamical details of the TSH-OM2 simulations also agree with classical molecular dynamics simulations using a reparametrized optimized potentials for liquid simulations (OPLS) all-atom force field with ad-hoc surface hops at predefined conical intersection points. The latter approach allows for a more extensive statistical sampling.

INTRODUCTION

A design of molecular machines, such as rotary molecular motors and photoswitches, represents a real challenge for future technology.^{1–10} A wide variety of molecular machines ranging from single molecule switches and motors^{5–7,10–22} to photo-mobile polymer materials^{23–25} has been synthesized. The desired functionality, in these machines, is achieved through light-driven changes in molecular structure. This principle is used in light-driven molecular rotary motors derived from chiral overcrowded alkenes^{5–7,12–22} in which periodic repetition of photoisomerization and thermal relaxation steps leads to a unidirectional rotation of one part of the molecule with respect to another (see e.g., Scheme 1). In these motors, clear design principles were formulated for the thermal helix inversion step, and a considerable increase in the rotation speed (ca. 10^8 times) has been achieved by lowering its barrier.^{13,14,16,17} The photoisomerization step however still remained poorly understood and less amenable to judicious chemical modification. To be able to improve the design of light-driven molecular rotary motors, one needs to better understand the underlying mechanism and the effect of various factors, such as substituents, heteroatoms, and environment, on the dynamics of photorearrangement. Such

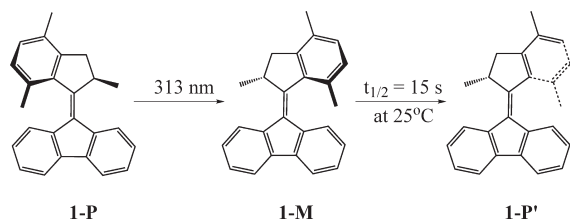
knowledge can be used not only to improve the design of synthetic molecular motors but also to gain deeper insight into the mode of action of biological molecular machines.²⁶

These goals motivated a recent molecular dynamics (MD) study²⁷ of the rotation cycle of a particular molecular motor, 9-(2,4,7-trimethyl-2,3-dihydro-1H-inden-1-ylidene)-9H-fluorene (**1**), see Scheme 1. In this study, the potential energy surfaces (PESs) of the lowest singlet excited state (S_1) and of the electronic ground state (S_0) were represented by a classical force field [optimized potentials for liquid simulations (OPLS) all-atom], with a special parametrization for the excited state.²⁷ The MD-OPLS simulations started in the S_1 state and returned to the S_0 state via conical intersections^{28–30} that play a crucial role in the photoisomerization of **1**. In a general sense, the mechanism of photoisomerization in **1** is basically the same as in the prototypical ethylene molecule. To reach the conical intersection, **1** needs to undergo a substantial geometric distortion that involves a twist about the central double bond and a strong pyramidalization of the central carbon atom of the fluorene moiety. Although

Received: March 23, 2011

Published: May 24, 2011

Scheme 1. Isomerization Cycle of Molecular Motor 1



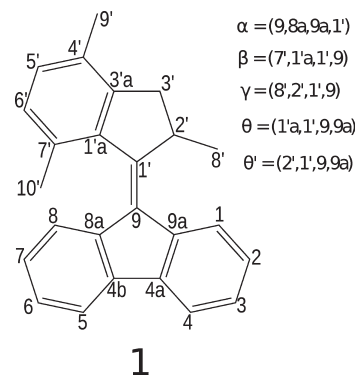
the results of ref 27 for the excited-state lifetimes, the photostationary state ratio, and the thermochemical parameters of the helix inversion step were in good agreement with experiments,¹⁶ the underlying theoretical approach is based on a number of assumptions that could not be thoroughly tested. In particular, it was assumed that the population transfer occurs whenever the molecular conformation reaches a predefined geometry that corresponds to a minimum-energy conical intersection.²⁷ This assumption may artificially exaggerate the role of minimum-energy conical intersections and thus needs to be justified carefully.

A more balanced description of excited-state decay processes can be achieved with the use of semiclassical trajectory surface hopping (TSH) simulations in which the parameters governing population transfer are determined on-the-fly from quantum mechanical (QM) calculations. It is the primary purpose of the present work to carry out such TSH-QM simulations of the photoisomerization cycle of the molecular motor **1** and to obtain detailed insight into its mechanism. The results of this work should not only corroborate those of the previous study²⁷ but also provide an unbiased assessment of the role of conical intersections and avoided crossing regions for the photodynamics of **1**. The sheer size of **1** precludes the use of accurate QM methods in on-the-fly TSH-QM simulations. Therefore we have chosen an efficient semiempirical QM approach for this purpose, the OM2/graphical unitary group approach multireference configuration interaction (GUGA-MRCI) method,^{31–33} which offers a realistic description of the ground and excited states of large molecules^{34,35} and has recently been applied successfully in excited-state dynamics studies.^{36–38} In the present TSH-OM2 simulations, we compute the ground- and excited-state PESs of **1** and the required nonadiabatic coupling parameters by OM2/GUGA-MRCI on-the-fly during the dynamics runs. For comparison, we report classical MD-OPLS simulations of the type of ref 27 using the conical intersection seam structures optimized with the OM2/GUGA-MRCI method. We analyze the results of the two types of simulations in the context of the experimentally observed excited-state lifetimes³⁹ and photostationary state ratios¹⁶ and discuss the implications of these results for the photochemistry of sterically overcrowded alkenes.

METHOD OF CALCULATION

To investigate the potential energy surfaces and the dynamics of **1**, semiempirical QM calculations were performed using the development version of the MNDO program.⁴⁰ The orthogonalization-corrected OM2 semiempirical Hamiltonian^{31,32} and the GUGA-MRCI approach³³ were employed to calculate the required energies, gradients, and nonadiabatic couplings. GUGA-MRCI denotes the graphical unitary group approach to multireference configuration interaction. Three reference configurations

Scheme 2. Definition of Key Geometric Parameters of 1



[closed-shell and single and double highest occupied–lowest unoccupied molecular orbital (HOMO–LUMO) excitations] were used to build the MRCI expansion. The self-consistent field (SCF) calculations were done in the restricted open-shell Hartree–Fock (ROHF) formalism, which provides a better description of the electronic wave function of singlet excited states than the usual restricted closed-shell Hartree–Fock approach.

The geometries of the ground- and excited-state minima were optimized for all the relevant isomers using the OM2/GUGA-MRCI method. The ground-state transition state (TS) for the first step in Scheme 1 was also optimized at the OM2/GUGA-MRCI level. The geometries of the minimum-energy S_0 – S_1 conical intersections (MECIs) were located by using the Lagrange–Newton method with analytically calculated gradients and nonadiabatic coupling vectors.^{41,42} The conical intersection seam in the vicinity of the located MECIs was studied by performing relaxed scans along dihedral angles corresponding to twisting about the C_9 – $C_{1'}$ double bond and pyramidalization of the C_9 atom (see Scheme 2).

The photoinduced nonadiabatic dynamics was investigated with Tully's surface hopping method (see refs 43–46 for further details). All relevant energies, gradients, and nonadiabatic coupling vectors were calculated on-the-fly using OM2/GUGA-MRCI. Nonadiabatic transitions at conical intersections were treated by the fewest switches algorithm,^{43,44,47} with velocity adjustment along the nonadiabatic coupling vector after a successful hop.⁴⁷ The empirical decoherence correction proposed by Truhlar et al.⁴⁸ and Granucci et al.⁴⁹ was applied to improve the internal consistency of the fewest switches scheme, using the suggested value of 0.1 hartree for the empirical constant in this correction. Initial structures and velocities were obtained by Wigner sampling.⁵⁰ The nuclear trajectories were propagated for 2 ps with a time step of 0.1 fs. A step size of 0.001 fs was used for the propagation of the electronic motion.⁴⁷

The active space in the OM2/GUGA-MRCI calculations comprised 12 electrons in 11 orbitals (π and π^*), which was found suitable for stable dynamics simulations and for geometry optimizations along the isomerization path including the conical intersection points. To always retain the π orbitals in the active space, we employed a recently developed method for identifying and tracking the π character of orbitals.⁵¹ To be included in the active space, a molecular orbital (MO) was required to have a π -type population exceeding 0.4 for optimizations and 0.35 for nonadiabatic dynamics (see ref 51 for further details of this approach). The chosen (12,11) active space contained the two

Table 1. Geometry Parameters in Å and deg of the 1-P and 1-M Conformations, the Transition State, and the Conical Intersections Obtained with OM2/GUGA-MRCI (this work) and with the RE-B3LYP/6-31G* and SA-RE-BH&HLYP//RE-B3LYP/6-31G* Methods (from ref 27)^a

structure(state)	method	C _{1'} -C ₉	α	β	γ	θ	pyr ^b
P(S ₀)	OM2/GUGA-MRCI	1.359	0.7	-39.6	102.2	168.2	1.1
	RE-B3LYP	1.368	1.7	-43.1	105.5	169.2	2.8
M(S ₀)	OM2/GUGA-MRCI	1.364	1.3	27.0	41.2	27.2	2.1
	RE-B3LYP	1.376	1.7	30.0	32.4	31.7	2.8
P(S ₁)	OM2/GUGA-MRCI	1.400	16.4	1.5	59.4	110.2	26.7
M(S ₁)	OM2/GUGA-MRCI	1.392	4.3	0.7	58.6	85.3	6.9
TS(S ₀)	OM2/GUGA-MRCI	1.438	0.2	0.2	59.1	90.0	0.3
	RE-B3LYP	1.463	0.0	5.6	66.0	90.0	0.0
CI ₁	OM2/GUGA-MRCI	1.420	35.5	5.7	67.2	133.5	58.0
	SA-RE-BH&HLYP//RE-B3LYP	1.490	34.0	3.8	36.8	120.0	52.7
	CASSCF(8,8)	1.423	35.5	4.9	64.9	131.0	58.1
CI ₂	OM2/GUGA-MRCI	1.405	-32.6	3.2	59.6	56.1	-53.6
	SA-RE-BH&HLYP//RE-B3LYP	1.472	-30.0	4.9	45.5	70.0	-47.9

^aSee Scheme 2 for definitions of geometry parameters for 1-P and 1-M. ^bPyramidalization angle: angle between C_{1'}-C₉ bond and (8a-9-9a) plane.

Table 2. Energies, in kcal/mol, of the Species in Table 1 Obtained from OM2/GUGA-MRCI (this work) and RE-B3LYP/6-31G* or SA-RE-BH&HLYP/6-31G* (from ref 27) Calculations^a

state	method	P	M	TS	P(S ₁) ^b	M(S ₁) ^c	CI ₁	CI ₂
S ₀	OM2/GUGA-MRCI	0.0	2.1	35.7	40.8	37.8	62.7	66.2
	RE-B3LYP	0.0	3.5	32.1			67.6 ^d	70.3 ^d
S ₁	OM2/GUGA-MRCI	90.6	86.5	62.0	59.1	59.0	62.7	66.2
	SA-RE-BH&HLYP	87.6	81.7	68.0			71.2	74.0

^aAll values are given relative to the ground-state energy of the P conformer. ^bEnergy minimum on the S₁ PES structurally closest to the 1-P conformer. ^cEnergy minimum on the S₁ PES structurally closest to the 1-M conformer. ^dEnergy obtained from SA-RE-BH&HLYP//RE-B3LYP single-point calculation.

singly occupied MOs from the ROHF treatment, the 5 highest doubly occupied π MOs, and the 4 lowest unoccupied π MOs, which was considered sufficient for capturing the major correlation effects in the OM2/GUGA-MRCI framework. For this choice, the energy gaps between the active and inactive π MOs were generally rather large, consistent with the fact that the (12,11) active space turned out to be quite robust during dynamics and optimizations.

To verify the molecular geometries obtained at the OM2/GUGA-MRCI level, complete active space SCF (CASSCF) calculations were undertaken for a number of the structures, using an active space with 8 electrons in 8 orbitals and the 6-31G** basis set.⁵² The CASSCF calculations were performed with the MOLPRO2008.1 package.⁵³

The classical force field MD simulations were carried out using the Gromacs (v4.0.5) program package.⁵⁴ The ground- and excited-state potential energy surfaces were described by the OPLS all-atom force field,⁵⁵ as reparametrized in ref 27. Further details on the setup and the results of the classical MD simulations can be found in the Supporting Information. To distinguish in the following between the two types of excited-state dynamics described above, we shall use the label TSH-OM2 for the semiclassical trajectory surface hopping method with on-the-fly OM2/GUGA-MRCI calculations and the label MD-OPLS for the classical MD simulations with the reparametrized OPLS all-atom force field.

RESULTS AND DISCUSSION

OM2/GUGA-MRCI Calculations. The operation cycle of the motor **1** consists of four steps (strokes) of which one photoisomerization step (power stroke) and one thermal helix inversion step are shown in Scheme 1. The geometries of the stable 1-P and the metastable 1-M conformations of **1** in the ground (S₀) electronic state were optimized using the OM2/GUGA-MRCI semiempirical method. The atomic numbering scheme and the definition of the key geometric parameters of **1** are given in Scheme 2, and the results of the calculations are collected in Tables 1 and 2. Due to the increased steric repulsion, the metastable conformation 1-M is destabilized by 2.1 kcal/mol as compared to 1-P, and the central C_{1'}-C₉ bond is elongated by 0.005 Å. These values are in a good agreement with the results of RE-B3LYP/6-31G* calculations ($\Delta E = 3.5$ kcal/mol, $\Delta l = 0.008$ Å) for the ground-state species (see Tables 1 and 2).²⁷ The rearrangement from 1-P to 1-M occurs on the S₀ PES via a transition state that lies 35.7 kcal/mol (OM2/GUGA-MRCI) above the 1-P minimum, see Table 2; the corresponding RE-B3LYP/6-31G* barrier is 32.1 kcal/mol.²⁷ The OM2/GUGA-MRCI transition state possesses an imaginary frequency of 310i cm⁻¹ associated with a twist about the C_{1'}-C₉ double bond. Previously, the activation energy of the thermal helix inversion step (second step in Scheme 1) was found^{16,27} to be 18.7 kcal/mol, which guarantees that the helix inversion is the

dominant pathway of relaxation for the metastable conformation **1-M** on the ground-state S_0 PES.

The vertical excitation energies of the stable **1-P** and the metastable **1-M** conformations from OM2/GUGA-MRCI calculations are compared in Table 3 with the results of SA-RE-BH&HLYP//RE-B3LYP and TD-BH&HLYP//B3LYP calculations²⁷ and with the maxima of the absorption bands of the two conformers measured at -40 °C in hexane solution.¹⁶ The lowest singlet S_1 excited state corresponds to the excitation of one electron from the π -bonding to the π -antibonding orbital of the central $C_{1'}$ – C_9 double bond. This state is more polar than the ground state, as indicated by the electric dipole moments of the **1-P** and **1-M** conformers in the S_1 and S_0 state, respectively (2.53 vs 1.30 D in **1-P** and 2.25 vs 1.20 D in **1-M**). Due to large values of the transition electric dipole moment, 7.46 D for the **1-P** conformer and 7.43 D for the **1-M** conformer, the S_1 state is

Table 3. Excitation Energies, in eV and in nm (in parentheses), of the Stable Conformers **1-P** and **1-M** Calculated with the OM2/GUGA-MRCI, SA-RE-BH&HLYP, and TD-BH&HLYP Methods

basis set	method	1-P	1-M
6-31G*	OM2/GUGA-MRCI	3.93 (315)	3.75 (331)
	SA-RE-BH&HLYP	3.80 (327)	3.40 (365)
	TD-BH&HLYP	3.78 (328)	3.39 (366)
6-311+G**	SA-RE-BH&HLYP	3.71 (334)	3.31 (375)
	TD-BH&HLYP	3.68 (337)	3.28 (378)
	expt ^a	3.44 (360)	3.22 (385)

^a Maxima of the absorption bands measured in ref 16 in hexane solution at -40 °C.

optically accessible, and the **1-P** \rightarrow **1-M** photoisomerization occurs in this state.²⁷ By contrast, the second lowest singlet S_2 excited state is essentially dark. It lies about 0.4 eV above the S_1 state: For the **1-P** conformer, OM2/GUGA-MRCI predicts the S_2 state at 4.31 eV (288 nm) with an oscillator strength of 0.02, compared to the S_1 state at 3.93 eV (315 nm) with an oscillator strength of 0.83.

For illustration purposes, the potential energy surfaces of the S_0 and S_1 states were obtained from relaxed scans on a two-dimensional grid of the dihedral angles θ and α . The positions of the geometric structures listed in Table 1 are shown in Figure 1 by yellow dots located at the respective values of θ and α . The S_1 PES of **1** features an elongated and shallow low-energy region where two nearly degenerate stationary points can be identified. The two structures are denoted in Figure 1 and in Tables 1 and 2 as $P(S_1)$ and $M(S_1)$ based on their geometric proximity to one of the ground-state minima, **1-P** or **1-M**. The $P(S_1)$ minimum lies 31.5 kcal/mol below the Franck–Condon (FC) point corresponding to the **1-P** conformation, and the $M(S_1)$ minimum is 27.5 kcal/mol below the **1-M** FC point. These energy differences suggest that, in the S_1 state, there is a substantial driving force for rotating the upper part of **1** toward twisting angles about the central $C_{1'}$ – C_9 bond of approximately 90° .

The MECI points were located at the OM2/GUGA-MRCI level using the Lagrange–Newton method with analytic non-adiabatic coupling vectors and gradient difference vectors.⁴² The key geometry parameters of these MECI structures and their relative energies are reported in Tables 1 and 2, respectively. In Figure 1, the dihedral angles α and θ corresponding to the optimized MECI structures are shown. Note however that the energies of the optimized MECI points do not exactly match the PES contour plots in this figure (due to differences in the

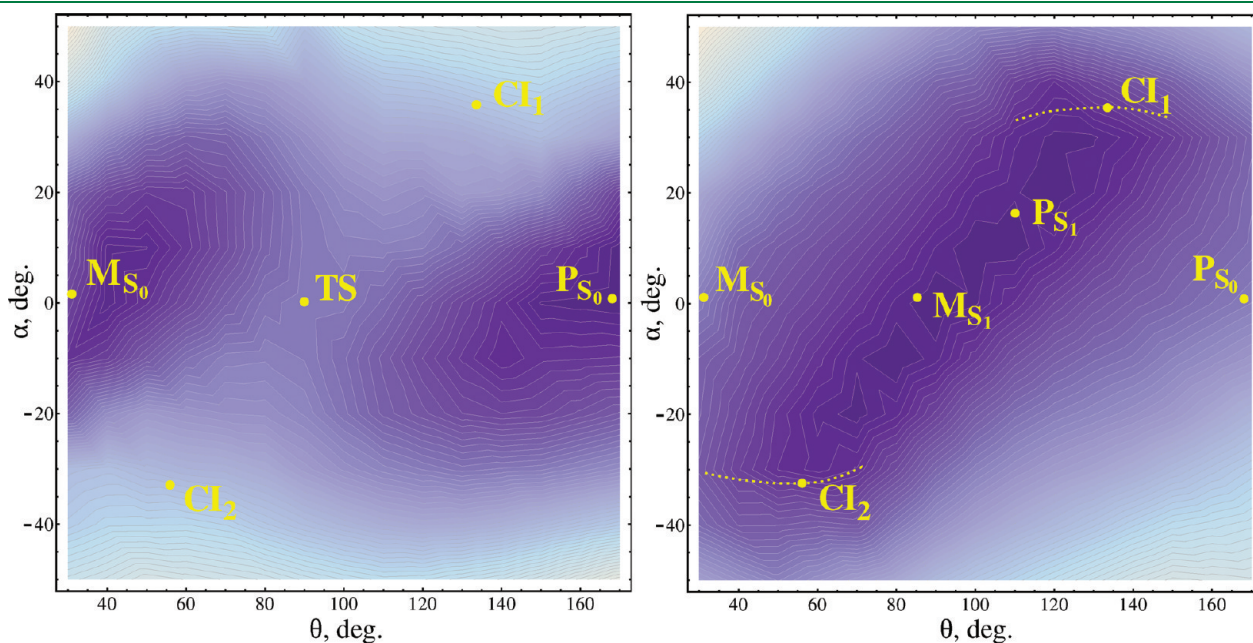


Figure 1. Contour plots of the S_0 and S_1 (right and left panels, respectively) PESs of **1** obtained in OM2/GUGA-MRCI calculations (see text for details). Special points marked on the plots in yellow: P_{S_0} , minimum corresponding to the **1-P** conformation on the S_0 PES; M_{S_0} , minimum corresponding to the **1-M** conformation on the S_0 PES; P_{S_1} , minimum on the S_1 PES geometrically closest to the **1-P** conformation; M_{S_1} , minimum on the S_1 PES geometrically closest to the **1-M** conformation; CI_1 and CI_2 , positions of the two minimum-energy conical intersection points; and TS, transition state on the S_0 PES. Dotted lines on the right panel indicate the positions of the conical intersection seam in the vicinity of MECIs as obtained with the OM2/GUGA-MRCI method.

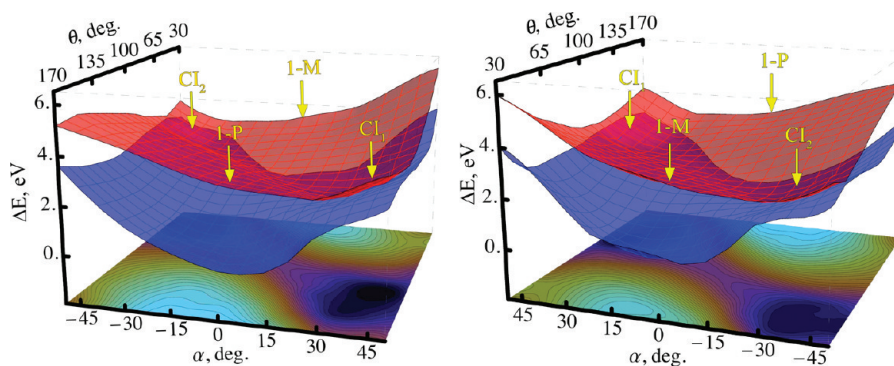


Figure 2. Profiles of the S_0 (blue) and S_1 (red) PESs of **1** obtained from OM2/GUGA-MRCI calculations (see text for details). Left panel: view from the side of the 1-P conformation. Right panel: view from the side of the 1-M conformation. The positions of conical intersection points and Franck–Condon points are shown with yellow arrows. The contour plots below show the difference between the S_1 and S_0 energies.

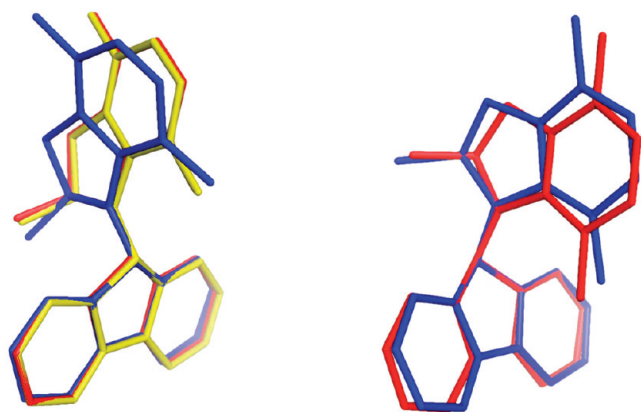


Figure 3. CI_1 and CI_2 (left and right panels, respectively) structures as obtained from OM2/GUGA-MRCI (red), RE-BH&HLYP/6-31G* (blue) and CASSCF(8,8)/6-31G* (yellow) calculations (see text for details).

other geometrical parameters). The two MECIs denoted as CI_1 and CI_2 feature pronounced pyramidalization of the C_9 atom and lie on the opposite sides of the avoided crossing region, i.e., the region in the vicinity of the TS on the S_0 PES, see Figures 1 and 2. Similar conical intersection geometries were obtained previously by Kazaryan et al.²⁷ using the SA-RE-BH&HLYP/6-31G* method. Note however that there was no search to precisely locate the MECI points in ref 27. Nevertheless, the geometries of the CI_1 and CI_2 points from the present work and from ref 27 closely coincide, as can be seen in Figure 3 where molecular drawings of the CI_1 and CI_2 geometries are superimposed. For comparison, a geometry optimized for the CI_1 point using the CASSCF(8,8) method is also shown in Figure 3.

Energetically CI_1 and CI_2 lie 3–5 kcal/mol above the P_{S_1} and M_{S_1} minima on the S_1 PES of **1**. The CI_1 point is easily accessible from the respective FC points of the 1-P and 1-M conformations (see Figure 4), whereas there is a small barrier (ca. 3.8 kcal/mol) en route from 1-M to CI_1 . This is illustrated in Figure 4 where the PES profiles are shown along pathways connecting the following points on the S_0 and S_1 surfaces: 1-P, P_{S_1} , CI_1 , and 1-M; 1-P, P_{S_1} , CI_2 , and 1-M; 1-P, CI_1 , M_{S_1} , and 1-M; and 1-P, CI_2 , M_{S_1} , and 1-M. These PES profiles were obtained by performing relaxed scans in these directions. During the scans, constrained geometry optimizations were carried out in the S_1 state for the intervals

shown in Figure 4 with solid red lines, for fixed values of the dihedral angles θ and α that were obtained by linear interpolation between the respective structures; the S_0 energies (dashed blue lines) were then computed at these optimized S_1 geometries. After reaching a conical intersection, it was assumed that the system switches to the S_0 ground state. Therefore constrained ground-state geometry optimizations were carried out for the intervals shown in Figure 4 with solid blue lines; the S_1 energies (dashed red lines) were obtained from single-point calculations at the S_0 geometries. The resulting PES profiles are presented in Figure 4.

It is well-known that conical intersections play a crucial role in the radiationless relaxation of excited states.^{28–30} The computed PES profiles of **1** suggest that the photoisomerization reaction can proceed via a barrierless pathway, and the expected excited-state lifetime should thus be quite short, of the order of 1 ps.³⁰ Specifically, for 1-P, the energy profiles for the S_1 state indicate that the relaxation should predominantly occur via CI_1 (see Figure 4), which is more easily accessible and also lower in energy than CI_2 (see Table 2). We note that barrierless pathways toward conical intersection points were recently also reported for a molecular switch structurally similar to **1**.^{21,22} However, even if such favorable pathways exist, the actual trajectories on the S_1 PES may take somewhat different routes due to dynamical effects. To obtain more realistic information on the dynamical behavior of the system after photoexcitation and on experimental observables, such as the quantum yields of photoisomerization, one needs to carry out excited-state dynamics simulations starting from the respective FC regions.

Molecular Dynamics Simulations. In our semiempirical TSH-OM2 simulations, 120 trajectories were initiated for each of the 1-P \rightarrow 1-M and 1-M \rightarrow 1-P photoisomerization reactions. The simulations were started in the S_1 state at geometries in the vicinity of the respective FC point, and the nuclear trajectories were propagated for 2 ps. The populations of the S_0 and S_1 states averaged over all trajectories are shown in Figure 5 for the two reactions. The S_1 populations were fitted by an exponential function, which gave S_1 lifetimes of 1.40 and 1.79 ps for the 1-P \rightarrow 1-M and 1-M \rightarrow 1-P reactions, respectively. These values are in a good agreement with the lifetimes of 1.40 ± 0.10 and 1.77 ± 0.13 ps, respectively, reported previously by Kazaryan et al. in their first classical MD-OPLS simulations.²⁷

To gain more detailed structural insight into the radiationless relaxation from S_1 to S_0 , we inspect some characteristic angles in

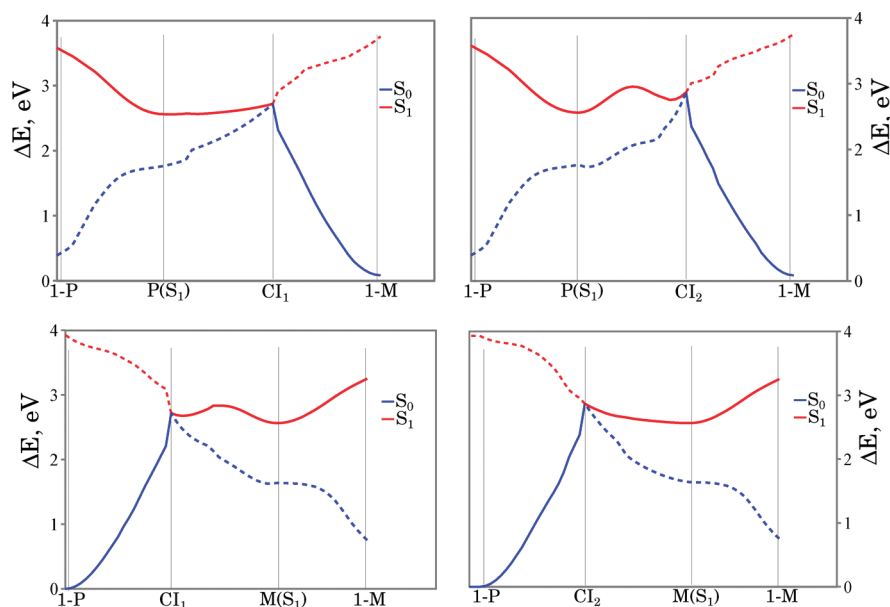


Figure 4. PES profiles connecting the 1-P, $P(S_1)$, CI_1 , and 1-M structures (upper left panel), the 1-P, $P(S_1)$, CI_2 , and 1-M structures (upper right panel), the 1-P, CI_1 , $M(S_1)$, and 1-M structures (lower left panel), and the 1-P, CI_2 , $M(S_1)$, and 1-M structures (lower right panel) on the S_1 (red) and S_0 (blue) PES. The solid lines represent energies obtained from constrained geometry optimization in the respective electronic state, and the dashed lines refer to energies obtained in single-point calculations (see text for further details).

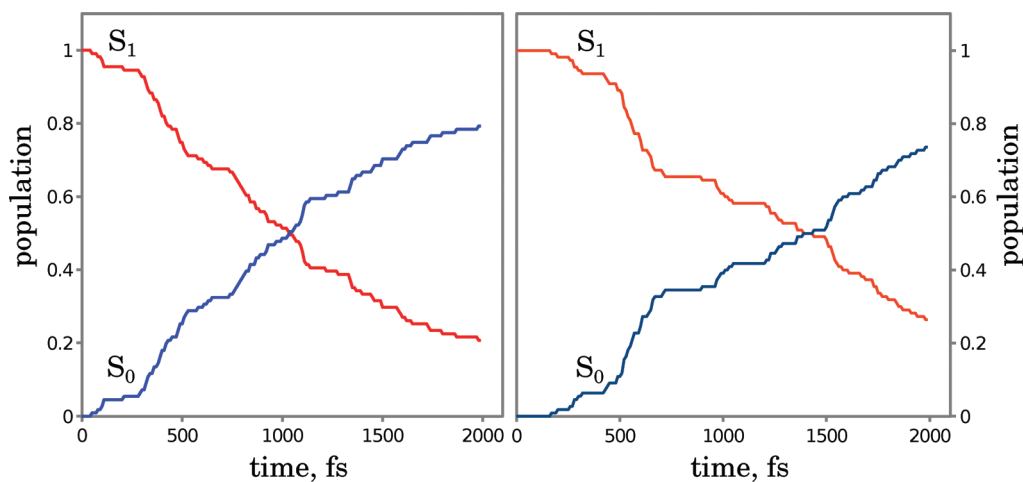


Figure 5. Populations of the S_0 (blue) and S_1 (red) states as a function of time (in fs) for the 1-P \rightarrow 1-M (left panel) and 1-M \rightarrow 1-P (right panel) photoisomerization reactions as obtained from the TSH-OM2 simulations.

the geometries at which surface hops occur during the TSH-OM2 simulations. The histograms in Figure 6 show the probability of surface hops as a function of the absolute values of the twisting angle θ and the pyramidalization angles at the C_9 atom and the $C_{1'}$ atom, for all trajectories generated for the two photoisomerization reactions. Obviously, the surface hops take place predominantly at sufficiently large values of θ (ca. 110° on average) and $\text{pyr-}C_9$ (ca. 35° on average). The $C_{1'}$ atom remains in a practically planar environment. It can therefore be conjectured that the surface hops occur in the vicinity of one of the MECIs found in the preceding OM2/GUGA-MRCI geometry optimizations. Indeed, both MECIs feature a rather strongly pyramidalized C_9 atom and a substantial twist about the $C_9-C_{1'}$ bond. The large average value of the dihedral angle θ also

suggests that, for both photoisomerization reactions, the hops occur predominantly close to the CI_1 geometry (see Table 1). These findings are consistent with the assumptions made in ref 27 when setting up the classical MD simulations.

More detailed information on the populations of the various molecular structures in the S_0 and S_1 states is presented in Figure 7, where the following distinctions are made in the analysis of the TSH-OM2 trajectories: The $P(S_0)$ structure corresponds to all molecular geometries with $\theta \geq 120^\circ$ encountered on the S_0 PES. The $P(S_1)$ structure corresponds to all geometries with $\theta \geq 120^\circ$ encountered on the S_1 PES. Analogous definitions hold for the $M(S_0)$ and $M(S_1)$ structures with $\theta \leq 70^\circ$. The $X(S_0)$ and $X(S_1)$ structures correspond to all intermediate geometries in the S_0 and S_1 states. These plots

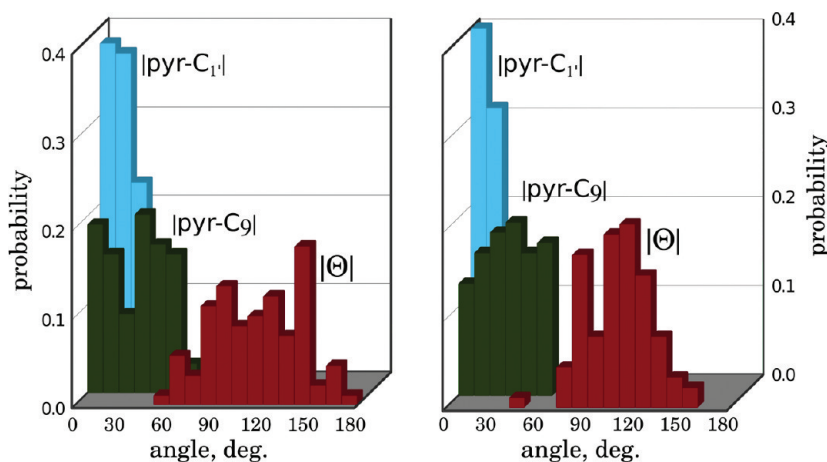


Figure 6. Probability of surface hops as a function of the absolute value of the angles θ (red), pyr-C₉ (green), and pyr-C_{1'} (light blue) in the TSH-OM2 trajectories for the 1-P \rightarrow 1-M (left panel) and 1-M \rightarrow 1-P (right panel) photoisomerization reactions. See Scheme 2 for the definition of θ and atom labels; pyr-C₉ is the angle between C₉-C_{1'} bond and (8a-9-9a) plane; and pyr-C_{1'} is the angle between C₉-C_{1'} bond and (1'a-1'-2') plane.

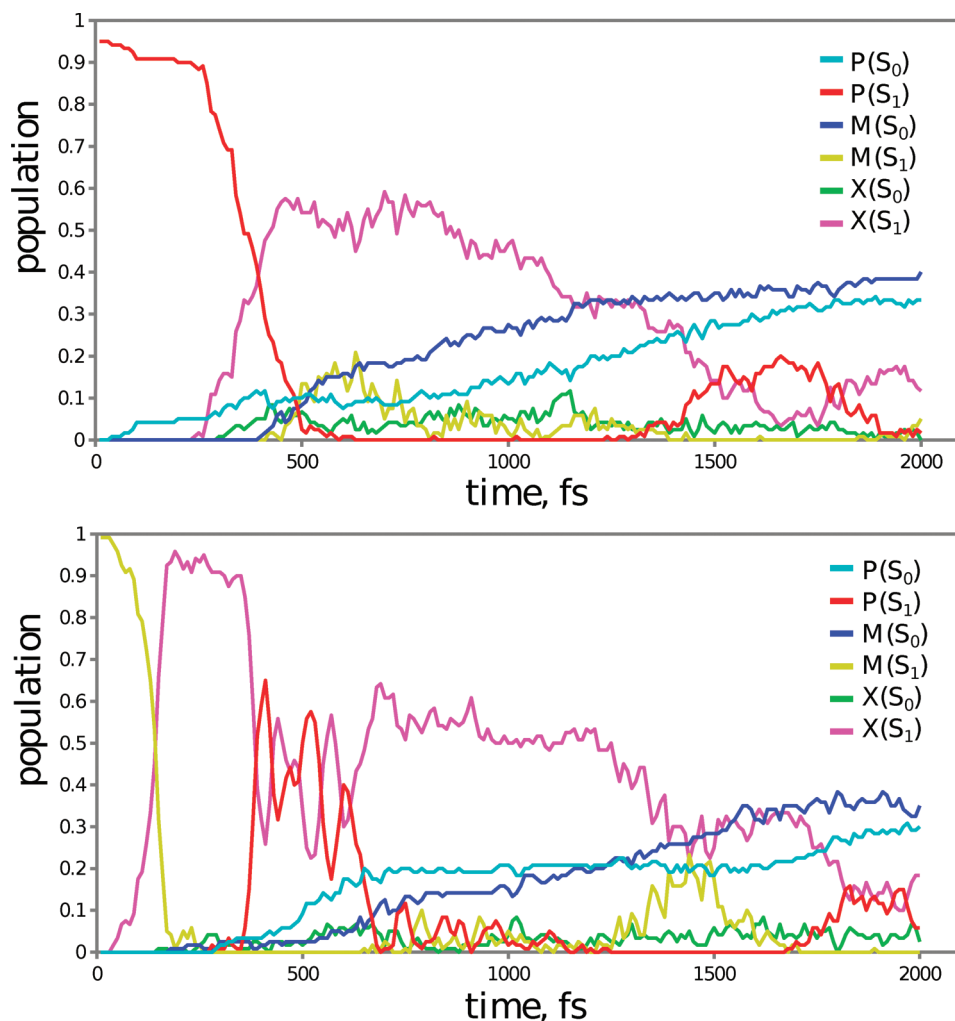


Figure 7. Time evolution of the populations of 1-P and 1-M structures in the ground S_0 and excited S_1 states as obtained from the TSH-OM2 simulations of the 1-P \rightarrow 1-M (upper panel) and 1-M \rightarrow 1-P (lower panel) photoisomerization reactions. Notation: P(S_0), 1-P isomer in the ground state; P(S_1), 1-P isomer in the excited state; M(S_0), 1-M isomer in the ground state; M(S_1), 1-M isomer in the excited state; X(S_0), intermediate molecular structure in the ground state; and X(S_1), intermediate molecular structure in the excited state (see text for details).

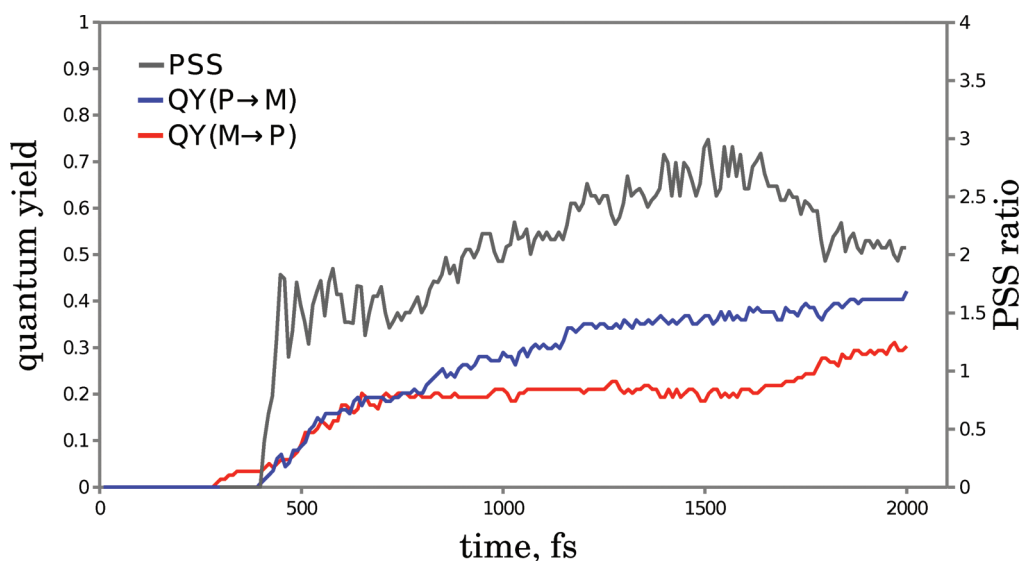


Figure 8. Quantum yields of the $1\text{-P} \rightarrow 1\text{-M}$ (blue) and $1\text{-M} \rightarrow 1\text{-P}$ (red) photoisomerization reactions and photostationary state ratio (gray) of the $1\text{-P} \rightleftharpoons 1\text{-M}$ photoisomerization reaction as obtained from the TSH-OM2 simulations.

indicate that the $1\text{-P} \rightarrow 1\text{-M}$ photoisomerization reaction is practically finished after ca. 1.5 ps, with less than 20% of the trajectories remaining in the excited state near the 1-P geometry. In the case of the $1\text{-M} \rightarrow 1\text{-P}$ reaction, ca. 30% of the trajectories remain in the excited state after ca. 1.6 ps, with the geometries intermediate between the 1-P and 1-M structures. This is as expected from the lifetimes evaluated for the two reactions from the data in Figure 5. We refrain from analyzing the data beyond 1.5 ps because of technical difficulties. In TSH-OM2 runs, the identification of active orbitals and the tracking of the active space may become problematic at longer simulation times, which may result in discontinuous PESs and unphysically large energy gradients.⁵¹ While these problems may be partly overcome by the techniques proposed previously⁵¹ and applied presently, they became unacceptably large for some of the trajectories after propagation times of more than 1.6 ps so that we prefer to focus on the TSH-OM2 results up to 1.5 ps.

In spite of these technical difficulties, we may still estimate the quantum yields of the two photoisomerization reactions and the photostationary state (PSS) ratio from Figure 7. The isomerization quantum yield of $1\text{-P} \rightarrow 1\text{-M}$ can be obtained as the ratio of the population of the $M(S_0)$ structures to the overall population of all structures at the start of the trajectories. The quantum yield of $1\text{-M} \rightarrow 1\text{-P}$ can be defined analogously using the $P(S_0)$ population. The PSS ratio of the $1\text{-P} \rightleftharpoons 1\text{-M}$ photoisomerization reaction is then defined as the ratio of the two quantum yields. To make the calculated PSS ratio comparable to the experimentally measured one, the ratio of quantum yields was multiplied by the ratio of the optical absorption cross sections of the two conformers, 1-P and 1-M ,²⁷ which amounts to 1.5:1.¹⁶ The so-obtained quantum yields and the PSS ratio are shown in Figure 8 as a function of simulation time. It is seen that, between 1.4 and 1.6 ps, the PSS ratio runs into a plateau and then suddenly decreases. Because this decrease is due to technical issues (see above), the final value of the PSS ratio was evaluated as the average value over the interval between 1.4 and 1.6 ps, which yields a value of 2.7:1 in a good agreement with the experimentally measured value of ca. 3:1.¹⁶ Please note that the computed value refers to the gas phase, while the experimental value was determined in perdeuterated toluene.¹⁶

The simulation times that can be achieved in the TSH-OM2 simulations are rather short. To investigate whether the equilibrium parameters derived from the TSH-OM2 simulations are reliable, we used classical two-state MD-OPLS simulations, as outlined in ref 27. We initiated 1000 MD simulations in the 1-P and 1-M states, respectively, and the simulations were run for up to 10 ps, which allowed us to determine converged equilibrium parameters.

A detailed account of the setup and the results of the MD-OPLS simulations is given in the Supporting Information. In the present work, we extended our MD-OPLS approach of ref 27 toward a more detailed description of the CI seam, which is represented by a number of conical intersection points around the two MECIs optimized with OM2/GUGA-MRCI (see Figure 1 in this article and Table 1 in the Supporting Information). In spite of the quite different approaches, the time evolution of the total population of the S_0 and the S_1 states as well as the populations of individual species on the excited- and ground-state PESs of the $1\text{-P} \rightarrow 1\text{-M}$ and $1\text{-M} \rightarrow 1\text{-P}$ photoisomerization reactions obtained in the classical MD-OPLS simulations (see Figures 1 and 2 of the Supporting Information) closely match those from the TSH-OM2 simulations (see Figure 5). The equilibrium values of isomerization quantum yields are reached after ca. 2.5 ps (Figure 3 of Supporting Information). The value of the PSS ratio at ca. 1.5 ps is not very different from the converged value, thus justifying the approach used to calculate the PSS ratio from the results of the TSH-OM2 simulations (see Figure 8).

In summary, the dynamics of the photoisomerization obtained in the TSH-OM2 simulations is consistent with previous theoretical results²⁷ and with the available experimental data.¹⁶ Comparing results of the present work with theoretical work on other photoactive molecules, the character of the lowest excited states and the mechanism of photoisomerization are qualitatively similar to what has been reported for stilbene,^{56–58} which is not too surprising since the molecular rotor **1** contains the stilbene substructure and may be regarded as some kind of stiff stilbene.⁵⁸ Like in ethylene and stilbene, the relevant conical intersection is characterized by a twist about the central double bond and a strong pyramidalization at one of its two carbon

atoms. The radiationless relaxation to the ground state is an ultrafast subpicosecond process in ethylene and stilbene; it is still fast in our case, with an excited-state lifetime of 1.4 ps for **1-P**, albeit clearly slower than in ethylene and stilbene. Qualitatively, this may be due to the stiffness of **1** (arising from the annelated rings at both sides of the central double bond), which should make it more difficult to reach the conical intersection region with one strongly pyramidalized carbon atom. Finally, we note that other types of conical intersections have been reported to be involved in the photoisomerization of longer polyenes^{59,60} and of protonated or alkylated Schiff bases that serve as retinal models⁶¹ or molecular switches.²¹

CONCLUSIONS

In this work, we report the results of semiclassical TSH-OM2 and classical MD-OPLS simulations of the photoisomerization step in the rotational cycle of the fluorene-based molecular rotary motor **1**. In such motors based on overcrowded alkenes, unidirectional rotation is achieved due to periodic repetition of photoisomerization and thermal relaxation steps. Hitherto, clear design principles were formulated for the thermally activated helix inversion step,^{13,14,16} whereas the photoisomerization step remained poorly understood and less amenable to rational modification based on mechanistic understanding. A first attempt to reach such an understanding was made in ref 27 using classical two-state MD simulations for **1** with a reparametrized all-atom force field. Although the results obtained were in good agreement with experiment, the chosen ad-hoc approach was based on a number of assumptions, the validity of which could not be thoroughly tested.

The present work extends the previous study²⁷ in several ways. First, the use of the OM2/GUGA-MRCI method allows a comprehensive investigation of the excited-state dynamics. A thorough characterization of the excited-state PES of **1** has been achieved for the first time, which provides information on the topology of the conical intersection seam, the location of the energy minima, and the potential energy barriers. On the relaxed PES, the conical intersection seam is found to be easily accessible, without the need to cross potential energy barriers. This observation suggests a reinterpretation of the results of transient absorption experiments in hexane,³⁹ in which a short excited-state decay time (ca. 1.7 ps) was interpreted as relaxation time due to crossing a potential barrier. The S_1 decay times obtained in the present gas-phase work from the TSH-OM2 simulations ($\tau_{S_1} \approx 1.40$ – 1.79 ps), as well as those from ref 27 ($\tau_{S_1} \approx 1.40$ – 1.77 ps), indicate that this relaxation time should be interpreted as excited-state lifetime. The other experimentally observed decay time, ca. 12 ps, which was associated with the transfer of population from the excited to the ground state,³⁹ may actually correspond to vibrational cooling of the “hot” ground-state species.

The semiclassical TSH-OM2 simulations clearly show that the radiationless relaxation of the excited electronic state of **1** proceeds via the conical intersection seam and that other parts of the excited state PES do not play a significant role in the mechanism of photoisomerization. Previously, this was only guessed.²⁷ The computed excited-state lifetimes for the direct, **1-P** \rightarrow **1-M**, and inverse, **1-M** \rightarrow **1-P**, photoisomerization reactions and the photostationary state ratio from the TSH-OM2 simulations are in good agreement with previous work²⁷ and with the experimental measurements.^{16,39}

The detailed knowledge of the excited-state PES profiles and of the geometries on the conical intersection seam obtained from OM2/GUGA-MRCI calculations allowed us to amend the approach used in ref 27 and to run classical MD-OPLS simulations using an extended conical intersection seam. The excited-state lifetimes, the detailed dynamics of various conformations, and the photostationary state ratio are in good agreement with the TSH-OM2 results. After proper validation, such excited-state classical MD simulations, with predefined conical intersection structures for hopping to the ground state, can thus be useful for running very long simulations (of the order of 10 ps or more) to reach converged results or for quickly exploring the excited-state dynamics of very large molecular species. A possible application would be to study the effect of chemical modifications on the photoisomerization cycle of rotary molecular motors.

ASSOCIATED CONTENT

S Supporting Information. The setup and the results of the classical MD simulations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: thiel@mpi-muelheim.mpg.de; m.filatov@rug.nl.

ACKNOWLEDGMENT

Z.L. and W.T. are grateful for financial support from Volkswagenstiftung. L.V.S. thanks The Netherlands Organization for Scientific Research (NWO) for a Veni fellowship (700.57.404).

REFERENCES

- (1) Stoddart, J. F. Molecular machines. *Acc. Chem. Res.* **2001**, *34*, 410–411.
- (2) Kottas, G. S.; Clarke, L. I.; Horinek, D.; Michl, J. Artificial Molecular Rotors. *Chem. Rev.* **2005**, *105*, 1281–1376.
- (3) Balzani, V.; Venturi, M.; Credi, A. *Molecular devices and machines, a journey into the nanoworld*; Wiley-VCH: Weinheim, Germany, 2003.
- (4) *Molecular Switches*; Feringa, B. L., Ed.; Wiley-VCH: Weinheim, Germany, 2001.
- (5) Feringa, B. L. In Control of Motion: From Molecular Switches to Molecular Motors. *Acc. Chem. Res.* **2001**, *34*, 504–513.
- (6) Browne, W. R.; Feringa, B. L. Making Molecular Machines Work. *Nat. Nanotechnol.* **2006**, *1*, 25–35.
- (7) Feringa, B. L. The Art of Building Small: From Molecular Switches to Molecular Motors. *J. Org. Chem.* **2007**, *72*, 6635–6652.
- (8) Kay, E. R.; Leigh, D. A.; Zerbetto, F. Synthetic Molecular Motors and Mechanical Machines. *Angew. Chem., Int. Ed.* **2007**, *46*, 72–191.
- (9) Irie, M. Photochromism and molecular mechanical devices. *Bull. Chem. Soc. Jpn.* **2008**, *81*, 917–926.
- (10) Balzani, V.; Credi, A.; Venturi, M. Light powered molecular machines. *Chem. Soc. Rev.* **2009**, *38*, 1542–1550.
- (11) Comstock, M. J.; Levy, N.; Kirakosian, A.; Cho, J.; Lauterwasser, F.; Harvey, J. H.; Strubbe, D. A.; Fréchet, J. M. J.; Trauner, D.; Louie, S. G.; Crommie, M. F. Reversible Photomechanical Switching of Individual Engineered Molecules at a Metallic Surface. *Phys. Rev. Lett.* **2007**, *99*, 038301+.
- (12) Koumura, N.; Zijlstra, R. W. J.; van Delden, R. A.; Harada, N.; Feringa, B. L. Light-driven monodirectional molecular rotor. *Nature* **1999**, *401*, 152–155.
- (13) Koumura, N.; Geertsema, E. M.; van Gelder, M. B.; Meetsma, A.; Feringa, B. L. Second Generation Light-Driven Molecular Motors.

Unidirectional Rotation Controlled by a Single Stereogenic Center with Near-Perfect Photoequilibria and Acceleration of the Speed of Rotation by Structural Modification. *J. Am. Chem. Soc.* **2002**, *124*, 5037–5051.

(14) Pijper, D.; van Delden, R. A.; Meetsma, A.; Feringa, B. L. Acceleration of a Nanomotor: Electronic Control of the Rotary Speed of a Light-Driven Molecular Rotor. *J. Am. Chem. Soc.* **2005**, *127*, 17612–17613.

(15) van Delden, R. A.; ter Wiel, M. K. J.; Pollard, J. M. M.; Vicario; Koumura, N.; Feringa, B. L. Unidirectional molecular motor on a gold surface. *Nature* **2005**, *437*, 1337–1340.

(16) Pollard, M. M.; Meetsma, A.; Feringa, B. L. A redesign of light-driven rotary molecular motor. *Org. Biomol. Chem.* **2008**, *6*, 507–512.

(17) Klok, M.; Boyle, N.; Pryce, A. M. T.; Meetsma; Browne, W. R.; Feringa, B. L. MHz unidirectional rotation of molecular rotary motors. *J. Am. Chem. Soc.* **2008**, *130*, 10484–10485.

(18) London, G.; Carroll, G. T.; Fernández Landaluce, T.; Pollard, M. M.; Rudolf, P.; Feringa, B. L. Light-driven altitudinal molecular motors on surfaces. *Chem. Commun.* **2009**, 1712–1714.

(19) Browne, W. R.; Feringa, B. L. Light Switching of Molecules on Surfaces. *Annu. Rev. Phys. Chem.* **2009**, *60*, 407–428.

(20) Geertsma, E. M.; van der Molen, S. J.; Martens, M.; Feringa, B. L. MHz unidirectional rotation of molecular rotary motors. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 16919–16924.

(21) Sinicropi, A.; et al. An artificial molecular switch that mimics the visual pigment and completes its photocycle in picoseconds. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17642–17647.

(22) Rivado-Casas, L.; Sampedro, D.; Campos, P. J.; Fusi, S.; Zanirato, V.; Olivucci, M. Fluorenylidene-Pyrroline Biomimetic Light-Driven Molecular Switches. *J. Org. Chem.* **2009**, *74*, 4666–4674.

(23) Yu, Y.; Nakano, M.; Ikeda, T. Directed Bending of A Polymer Film by Light. *Nature* **2003**, *425*, 145.

(24) Yamada, M.; Kondo, M.; Mamiya, J.-I.; Yu, Y.; Kinoshita, M.; Barrett, C.-J.; Ikeda, T. Photomobile Polymer Materials: Towards Light-Driven Plastic Motors. *Angew. Chem., Int. Ed.* **2008**, *47*, 4986–4988.

(25) Naka, Y.; Mamiya, J.-I.; Shishido, A.; Washio, M.; Ikeda, T. Direct fabrication of photomobile polymer materials with an adhesive-free bilayer structure by electron-beam irradiation. *J. Mater. Chem.* **2011**, *21*, 1681–1683.

(26) Strambi, A.; Durbeeja, B.; Ferré, N.; Olivucci, M. Anabaena sensory rhodopsin is a light-driven unidirectional rotor. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 21322–21326.

(27) Kazaryan, A.; Kistemaker, J. C. M.; Schäfer, L. V.; Browne, W. R.; Feringa, B. L.; Filatov, M. Understanding the Dynamics Behind the Photoisomerization of a Light-Driven Fluorene Molecular Rotary Motor. *J. Phys. Chem. A* **2010**, *114*, 5058–5067.

(28) Yarkony, D. R. Diabolical Conical Intersections. *Rev. Mod. Phys.* **1996**, *68*, 985–1013.

(29) Bernardi, F.; Olivucci, M.; Robb, M. A. Potential energy surface crossings in organic photochemistry. *Chem. Soc. Rev.* **1996**, *25*, 321–328.

(30) Levine, B. G.; Martinez, T. J. Isomerization Through Conical Intersections. *Annu. Rev. Phys. Chem.* **2007**, *58*, 613–634.

(31) Weber, W. Ph.D. Thesis; Universität Zürich: Zürich, Switzerland, 1996.

(32) Weber, W.; Thiel, W. Orthogonalization corrections for semiempirical methods. *Theor. Chem. Acc.* **2000**, *103*, 495–506.

(33) Koslowski, A.; Beck, M. E.; Thiel, W. Implementation of a general multireference configuration interaction procedure with analytic gradients in a semiempirical context using the graphical unitary group approach. *J. Comput. Chem.* **2003**, *24*, 714–726.

(34) Otte, N.; Scholten, M.; Thiel, W. Looking at Self-Consistent-Charge Density Functional Tight Binding from a Semiempirical Perspective. *J. Phys. Chem. A* **2007**, *111*, 5751–5755.

(35) Silva-Junior, M. R.; Thiel, W. Benchmark of Electronically Excited States for Semiempirical Methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S and INDO/S2. *J. Chem. Theory Comput.* **2010**, *6*, 1546–1564.

(36) Fabiano, E.; Thiel, W. Nonradiative deexcitation dynamics of 9H-adenine: An OM2 surface hopping study. *J. Phys. Chem. A* **2008**, *112*, 6859–6863.

(37) Lan, Z.; Fabiano, E.; Thiel, W. Photoinduced Nonadiabatic Dynamics of 9H-Guanine. *ChemPhysChem* **2009**, *10*, 1225–1229.

(38) Lan, Z.; Fabiano, E.; Thiel, W. Photoinduced Nonadiabatic Dynamics of Pyrimidine Nucleobases: On-the-Fly Surface-Hopping Study with Semiempirical Methods. *J. Phys. Chem. B* **2009**, *11*, 3548–3555.

(39) Augulis, R.; Klok, M.; Feringa, B. L.; van Loosdrecht, P. H. M. Light-Driven Rotary Molecular Motors: an Ultrafast Optical Study. *Phys. Status Solidi C* **2009**, *6*, 181–184.

(40) Thiel, W. MNDO program, version 6.1; Max-Planck-Institut für Kohlenforschung: Mülheim, 2007.

(41) Domcke, W.; Yarkony, D. R.; Köppel, H. *Conical Intersections: Electronic Structure, Dynamics and Spectroscopy*; World Scientific: Singapore, 2004.

(42) Keal, T. W.; Koslowski, A.; Thiel, W. Comparison of algorithms for conical intersection optimization using semiempirical methods. *Theor. Chem. Acc.* **2007**, *118*, 837–844.

(43) Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **1990**, *93*, 1061–1071.

(44) Hammes-Schiffer, S.; Tully, J. C. Proton transfer in solution: Molecular dynamics with quantum transitions. *J. Chem. Phys.* **1994**, *101*, 4657–4667.

(45) Barbatti, M.; Granucci, G.; Persico, M.; Ruckebauer, M.; Vazdar, M.; Eckert-Maksić, M.; Lischka, H. *J. Photochem. Photobiol. A* **2007**, *190*, 228–240.

(46) Fabiano, E.; Groenhof, G.; Thiel, W. Approximate switching algorithms for trajectory surface hopping. *Chem. Phys.* **2008**, *351*, 111–116.

(47) Fabiano, E.; Keal, T. W.; Thiel, W. Implementation of surface hopping molecular dynamics using semiempirical methods. *Chem. Phys.* **2008**, *349*, 334–347.

(48) Zhu, C.; Jasper, A. W.; Truhlar, D. G. Non-Born-Oppenheimer Liouville-von Neumann Dynamics. Evolution of a Subsystem Controlled by Linear and Population-Driven Decay of Mixing with Decoherent and Coherent Switching. *J. Chem. Theory Comput.* **2005**, *1*, 527–540.

(49) Granucci, G.; Persico, M.; Zocante, A. Including quantum decoherence in surface hopping. *J. Chem. Phys.* **2010**, *133*, 134111+.

(50) Wigner, E. On the Quantum Correction For Thermodynamic Equilibrium. *Phys. Rev.* **1932**, *40*, 749–759.

(51) Keal, T. W.; Wanko, M.; Thiel, W. Assessment of semiempirical methods for the photoisomerisation of a protonated Schiff base. *Theor. Chem. Acc.* **2009**, *123*, 145–156.

(52) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650–654.

(53) Werner, H.-J.; Knowles, P. J. et al. MOLPRO, version 2008.1, a package of ab initio programs; University College Cardiff Consultants Limited : Cardiff, U.K., 2008; <http://www.molpro.net>.

(54) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GRO-MACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(55) Jorgensen, W. L.; Tirado-Rives, J. The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

(56) Gagliardi, L.; Orlandi, G.; Molina, V.; Malmqvist, P.-A.; Roos, B. Theoretical Study of the Lowest 1BU States of trans-Stilbene. *J. Phys. Chem. A* **2002**, *106*, 7355–7361.

(57) Quenneville, J.; Martinez, T. J. Ab initio study of cis-trans photoisomerization in stilbene and ethylene. *J. Phys. Chem. A* **2003**, *107*, 829–837.

(58) Improta, R.; Santoro, F. Excited-State Behavior of trans and cis Isomers of Stilbene and Stiff Stilbene: A TD-DFT Study. *J. Phys. Chem. A* **2005**, *109*, 10058–10067.

(59) Celani, P.; Garavelli, M.; Ottani, S.; Bernardi, F.; Robb, M. A.; Olivucci, M. Molecular “Trigger” for the Radiationless Deactivation of

Photoexcited Conjugated Hydrocarbons. *J. Am. Chem. Soc.* **1995**, *117*, 11584–11585.

(60) Garavelli, M.; Celani, P.; Bernardi, F.; Robb, M. A.; Olivucci, M. Force Fields for Ultrafast Photochemistry: The Reaction Path for all-trans-Hexa-1,3,5-triene. *J. Am. Chem. Soc.* **1997**, *119*, 11487–11494.

(61) Gonzalez-Luque, R.; Garavelli, M.; Bernardi, F.; Merchan, M.; Robb, M. A.; Olivucci, M. Computational evidence in favor of a two-state, two-mode model of the retinal chromophore photoisomerization. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9379–9384.

Role of Many-Body Effects in Describing Low-Lying Excited States of π -Conjugated Chromophores: High-Level Equation-of-Motion Coupled-Cluster Studies of Fused Porphyrin Systems

K. Kowalski*

William R. Wiley Environmental Molecular Sciences Laboratory, Battelle, Pacific Northwest National Laboratory, K8-91, P.O. Box 999, Richland, Washington 99352, United States

R. M. Olson

Cray, Incorporated, 380 Jackson St. Suite 210, St. Paul, Minnesota 55101, United States

S. Krishnamoorthy

High Performance Computing, Pacific Northwest National Laboratory, Richland, Washington 99352, United States

V. Tipparaju and E. Aprà*

Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States

ABSTRACT: The unusual photophysical properties of the π -conjugated chromophores make them potential building blocks of various molecular devices. In particular, significant narrowing of the HOMO–LUMO gaps can be observed as an effect of functionalization chromophores with polycyclic aromatic hydrocarbons (PAHs). In this paper we present equation-of-motion coupled cluster (EOMCC) calculations for vertical excitation energies of several functionalized forms of porphyrins. The results for free-base porphyrin (FBP) clearly demonstrate significant differences between functionalization of FBP with one- (anthracene) and two-dimensional (coronene) structures. We also compare the EOMCC results with the experimentally available results for anthracene fused zinc–porphyrin. The impact of various types of correlation effects is illustrated on several benchmark models, where the comparison with the experiment is possible. In particular, we demonstrate that for all excited states considered in this paper, all of them being dominated by single excitations, the inclusion of triply excited configurations is crucial for attaining qualitative agreement with experiment. We also demonstrate the parallel performance of the most computationally intensive part of the completely renormalized EOMCCSD(T) approach (CR-EOMCCSD(T)) across 120 000 cores.

INTRODUCTION

The remarkable photophysical properties of the π -conjugated chromophores, emanating from very small HOMO–LUMO gaps, has recently stimulated a lot of interest. Of special interest are the π -conjugated porphyrin arrays, where the existence of low-lying excited states in the near-infrared absorption spectra makes these systems potential building blocks for molecular wires, optical devices, and light-harvesting systems (for a review of this broad area, see refs 1–10 and references therein). Over the last 2 decades it has been experimentally demonstrated that tuning the electronic structure of these systems is especially efficient when porphyrins are functionalized with the polycyclic aromatic hydrocarbons (PAHs). Theoretical studies of excited states can be very helpful in understanding the impact of various types of functionalization and can predict shifts in absorption spectra. For this purpose, various ab initio methodologies can be used. In addition to providing accurate and predictive results that can be compared with experiment, these wave-function-based ab initio approaches are also important to calibrate and parametrize low-order Hamiltonians.^{11–16} On the other hand, due to the size of

these systems and steep scaling of the high-level excited-state methods, examples of high-level calculations for these molecules are very scarce. For example, the current literature does not show any systematic study of the role played by higher-order effects (for example, collective three-body excitations) for functionalized forms of the porphyrins.

Another important question concerns possible ways of incorporating important excited-state correlation methods. Among many methods, the equation-of-motion of coupled cluster method (EOMCC)^{17–20} (for related approaches see refs 21–30) has evolved into a widely used formalism for calculating vertical excitation energies (VEEs). The well-established family of approximations including higher excitations starting from the rudimentary EOMCCSD model (EOMCC with singles and doubles)^{18–20} to more expensive models with triples (EOMCCSDT)³¹ and quadruples (EOMCCSDTQ),³² etc., have enabled one to reach the full configuration interaction (FCI) limit when all possible

Received: March 30, 2011

Published: May 27, 2011

excitations are included. The EOMCCSD can be used to describe singly excited states, whereas more complex excited states require the expensive EOMCCSDT approximation. In order to find a trade-off between cost and accuracy, in this paper we will resort to the completely renormalized EOMCCSD(T) approach (CR-EOMCCSD(T)),³³ which accounts for the effect of triply excited configurations in a noniterative manner (see also ref 34) and at the same time is characterized by the N^7 scaling (N represents the system size). For systems of the size of the functionalized porphyrins, the applicability of high-order methods faces significant challenges, mostly associated with the very steep numerical scaling. Therefore, in order to make these methods applicable to π -conjugated chromophores, their efficient parallel implementations need to be used.

In this paper, we want to address the following problems: (1) What is the role of various-rank correlation effects in describing the vertical excitation energies in fused porphyrins, and what level of approximation and basis set are required to obtain an agreement of calculated VEEs with the experimentally inferred values? In particular, we will analyze the effect of including triply excited amplitudes. We will illustrate the accuracies of the EOMCCSD and CR-EOMCCSD(T) methods on the example of available experimental data for the anthracene fused zinc-porphyrin.⁷ (2) What is the effect of functionalizing free-base porphyrin by adding one- and two-dimensional structures, anthracene and coronene? (3) How can efficient computer implementations exploit massively parallel computer architectures?

THEORY

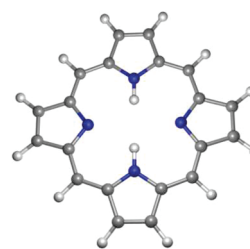
The EOMCC formalism has evolved into one of the most widely used tools in excited-state calculations of low-lying excited states,¹⁷ where various correlation effects can be controlled by the rank of excitations included in the cluster (T) and correlation (R_K) operators used to parametrize K th excited state

$$|\Psi_K\rangle = R_K e^T |\Phi\rangle \quad (1)$$

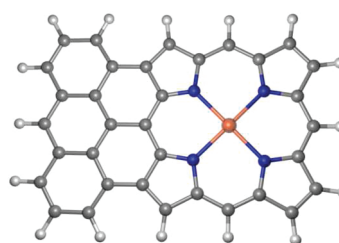
where $|\Phi\rangle$ is the so-called reference function usually chosen as a Hartree–Fock determinant. Various approximation schemes range from the basic EOMCCSD approximation, where the cluster and correlation operators are represented as sums of scalar ($R_{K,0}$; for excitation operator only), single ($T_1, R_{K,1}$), and double ($T_2, R_{K,2}$) excitations,

$$|\Psi_K^{\text{EOMCCSD}}\rangle = (R_{K,0} + R_{K,1} + R_{K,2})e^{T_1 + T_2} |\Phi\rangle \quad (2)$$

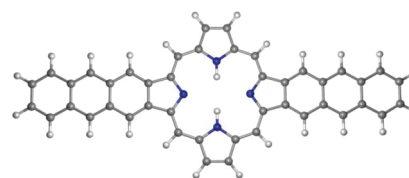
to the more accurate EOMCCSDT approach. Over the last 2 decades it was demonstrated that the progression of methods: EOMCCSD \rightarrow EOMCCSDT \rightarrow EOMCCSDTQ... leads to more accurate estimates of the excited-state energies, which in the limit of N_e (N_e stands for the number of correlated electrons) excitations converge to the FCI energies. However, the rapid growth in the numerical complexity [N^6 (EOMCCSD), N^8 (EOMCCSDT), N^{10} (EOMCCSDTQ)] of the EOMCC methods makes accurate calculations with the EOMCCSDT and EOMCCSDTQ methods prohibitively expensive, even for relatively small systems. Unfortunately, the EOMCCSD method is capable of providing reliable results only for singly excited states. However, as it has recently been demonstrated,³⁵ conspicuous errors in the range of 0.25–0.30 eV with respect to the experimental VEEs persist with increasing system size. In order to narrow the gap between the EOMCCSD and EOMCCSDT VEEs, several noniterative N^7 -scaling methods that mimic the



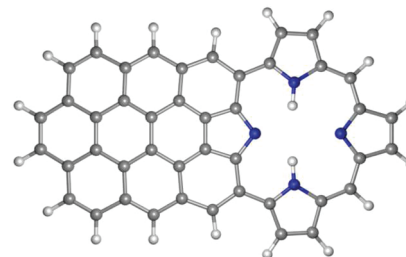
Free-base porphyrin



Zinc porphyrin-f-anthracene



Free-base porphyrin-(f-anthracene)₂



Free-base porphyrin-f-coronene

Figure 1. Benchmark set of molecules considered in this study.

effect of triples in a perturbative fashion have been proposed in the past.^{36–41} The completely renormalized EOMCCSD(T) approach (CR-EOMCCSD(T))³³ falls into this category (see also refs 42 and 43 for the most recent developments). In this approach, the energy correction $\delta_K^{\text{CR-EOMCCSD(T)}}$ is added to the EOMCCSD VEE ($\omega_K^{\text{EOMCCSD}}$)

$$\omega_K^{\text{CR-EOMCCSD(T)}} = \omega_K^{\text{EOMCCSD}} + \delta_K^{\text{CR-EOMCCSD(T)}} \quad (3)$$

where the $\delta_K^{\text{CR-EOMCCSD(T)}}$ is expressed through the trial wave function $\langle \Psi_K |$ and the triply excited EOMCCSD moment operator $M_{K,3}^{\text{EOMCCSD}}$ (see ref 33 for details)

$$\delta_K^{\text{CR-EOMCCSD(T)}} = \frac{\langle \Psi_K | M_{K,3}^{\text{EOMCCSD}} | \Phi \rangle}{\langle \Psi_K | (R_{K,0} + R_{K,1} + R_{K,2}) e^{T_1 + T_2} | \Phi \rangle} \quad (4)$$

The N^7 scaling of the CR-EOMCCSD(T) method stems from the $M_{K,3}^{\text{EOMCCSD}}|\Phi\rangle$ part. Although the CR-EOMCCSD(T) method has the same N^7 scaling as the ground-state CCSD(T) method,⁴⁴ there is a numerical prefactor that makes our CR-EOMCCSD(T) implementation 3 times more expensive for the states of the same symmetry as the ground state and 2 times more expensive for states of different symmetry than the ground state. Moreover three (two) large two-body intermediates of the size proportional to the $n_o n_u^3$ (n_o and n_u refer to the number of occupied and unoccupied spin-orbitals) have to be formed in the calculation of excited states of the same (different) symmetry as the ground state. Therefore, the storage of these quantities requires substantial memory resources. These requirements can be significantly reduced in the active-space CR-EOMCCSD(T) methods.³⁵

Table 1. A Comparison of the EOMCCSD and CR-EOMCCSD(T) VEEs (in eV) for the Two Lowest Singlet States of the Anthracene Molecule in Various Basis Sets

basis set	EOMCCSD	CR-EOMCCSD(T)
	¹ L _a State, Expt 3.60 eV ⁶²	
AVTZ	4.48	4.22
cc-pVDZ	4.16	3.87
POL1	4.00	3.69
	¹ L _b State, Expt 3.64 eV ⁶²	
AVTZ	4.00	3.76
cc-pVDZ	3.96	3.68
POL1	3.90	3.59

Table 2. EOMCCSD and CR-EOMCCSD(T) VEEs (in eV) for the Coronene Molecule in Various Basis Sets

basis set	EOMCCSD	CR-EOMCCSD(T)
	¹ L _b State	
AVTZ	3.46	3.22
cc-pVDZ	3.43	3.16
POL1	3.38	3.10
cc-pVTZ	3.40	3.08
	¹ L _a State, Expt 3.72 eV ⁶⁸	
AVTZ	4.48	4.25
cc-pVDZ	4.17	3.92
POL1	4.08	3.81
cc-pVTZ	4.07	3.77

COMPUTATIONAL DETAILS

All calculations were performed with a development version of NWChem suite of codes.⁴⁵ The EOMCC calculations were carried out with the TCE⁴⁶ implementations of the EOMCCSD and CR-EOMCCSD(T) methods. In all calculations we used the development variant of the CR-EOMCCSD(T) code, which significantly reduces the local memory requirements. In particular, the new implementation offers the possibility of partitioning six-dimensional tensors across the first two dimensions, which enables one to use large-size tiles (or basic building block defining the partitioning of spin-orbital domain and block structure of all tensors), which significantly improves the performance of the iterative part (CCSD and EOMCCSD). Iterative CCSD and EOMCCSD have also undergone significant changes. A major problem in the older implementation was “serial” execution of parallel procedures characterized by task pools of various size. This “vertical” structure has been replaced by a horizontal structure where all procedures are grouped into several classes, each class containing procedures that can be executed independently of each other. All subroutines from a given class contribute to a global task pool characterizing given class, which to a large extent alleviates the problems associated with load balancing. All tensors corresponding to cluster/excitation amplitudes, recursive intermediates, and electron integrals were stored on global arrays, which provide a portable shared-memory interface.

In our calculations, we considered several systems (some of them shown in Figure 1): (1) anthracene, (2) coronene, (3) free-base-porphyrin (FBP), (4) fused anthracene–Zn–porphyrin (ZnP-f-anthracene), (5) double anthracene fused free-base porphyrin [FBP-(f-anthracene)₂] (oligoporphyrin **46** of ref 10), and (6) coronene fused free-base porphyrin (FBP-f-coronene) (oligoporphyrin **65** of ref 10). Except for the anthracene and FBP, geometries of the remaining systems were optimized with the B3LYP approach⁴⁷ using the cc-pVTZ basis set.^{48,49} The geometries of anthracene and FBP are the same as used in refs 50 and 51, respectively. In all correlated calculations the core electrons were kept frozen. The calculations for anthracene were performed using Ahlrichs’ VTZ (AVTZ),⁵³ cc-pVDZ, and POL1⁵² basis sets. For coronene, AVTZ, cc-pVDZ, POL1, and cc-pVTZ basis sets were employed. For larger systems, we used AVTZ (FBP, ZnP-f-anthracene, FBP-f-coronene) and cc-pVDZ (FBP, ZnP-f-anthracene, FBP-(f-anthracene)₂) basis sets. Our calculations were performed with D_{2h} [anthracene, coronene, FBP, FBP-(f-anthracene)₂] and C_{2v} (ZnP-f-anthracene, FBP-f-coronene) symmetries. For larger systems we used relaxed convergence criteria for iterative approaches. Previous studies with the

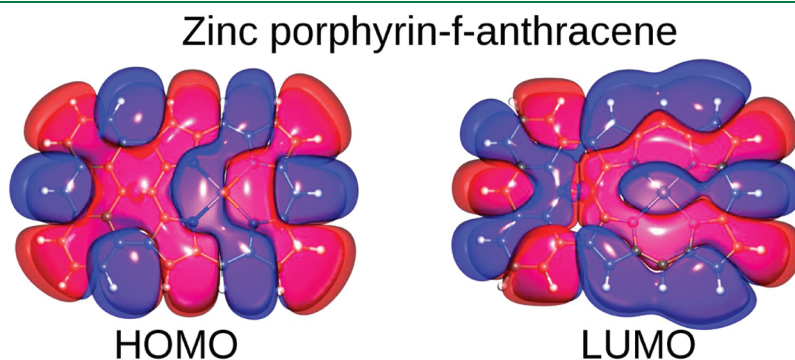


Figure 2. Graphical representation of the RHF HOMO and LUMO orbitals for the ZnP-f-anthracene system obtained with the cc-pVDZ basis set.

EOMCC methods³⁵ suggest that the cc-pVDZ basis set can provide reliable results for valence excited states for systems as large as an oligoporphyrin dimer. The issue of basis set choice will be discussed in the next section.

DISCUSSION

The ZnP and FBP systems have been studied intensively by employing the EOMCCSD, CR-EOMCCSD(T), and active-space CR-EOMCCSD(T) methods. In studies of the effect of the functionalization of the porphyrins, we will refer to the results available in the literature.^{35,54} We start our discussion by presenting the VEEs of the two lowest-lying singlet–singlet $\pi \rightarrow \pi^*$ transitions for components of our target systems: anthracene and coronene. We will employ the Platt nomenclature⁵⁵ for these

Table 3. EOMCCSD and CR-EOMCCSD(T) VEEs (in eV) for the ZnP-f-anthracene System in AVTZ and cc-pVDZ Basis Sets

basis set	EOMCCSD	CR-EOMCCSD(T)
	¹ L _a State, Expt. 1.71 eV ⁷	
AVTZ	2.03	1.79
cc-pVDZ	2.03	1.77
dominant excitation ^a	H→L	

^a See the text for details.

Table 4. A Comparison of the EOMCC VEEs (in eV) of the FBP-(f-anthracene)₂ and FBP-f-coronene with the VEEs (in eV) of the Free-Base Porphyrin Using cc-pVDZ and AVTZ Basis Sets, Respectively

	FBP-(f-anthracene) ₂ (cc-pVDZ)		FBP(cc-pVDZ) ^a	
	¹ L _a	¹ L _b	¹ B _{3u}	¹ B _{2u}
EOMCCSD	1.99	2.57	2.15	2.61
CR-EOMCCSD(T)	1.78	2.37	1.86	2.32
dominant excitation ^b	H→L	H-1→L H→L+1	H-1→L H→L+1H-1→L+1	H→L

	FBP-f-coronene (AVTZ)		FBP(AVTZ)	
	¹ L _a	¹ L _b	¹ B _{3u}	¹ B _{2u}
EOMCCSD	1.63	—	2.15	2.72
CR-EOMCCSD(T)	1.36	—	1.88	2.45
dominant excit. ^b	H→L		H-1→L H→L+1	H→L

^a From ref 35. ^b See the text for details.

excited states, ¹L_a and ¹L_b, where these states are dominated by HOMO→LUMO excitation (L_a) and a combination of HOMO-1→LUMO and HOMO→LUMO+1. Since the excited states of porphyrin-anthracene and porphyrin-coronene complexes studied in this paper have similar configurational structure we will adapt the Platt convention to characterize these states. Using NWChem (autosym) convention, the ¹L_a state of the ZnP-f-anthracene corresponds to the 2¹A₁ state, the ¹L_a and ¹L_b states of FBP-(f-anthracene)₂ correspond to ¹B_{3u} and ¹B_{2u} states, respectively, whereas the ¹L_a state of FBP-f-coronene corresponds to the ¹B₂ state. However, it should be noted that other codes can employ different symmetry conventions, and therefore, the classification of excited states based on specifying leading excitations provides an unambiguous method of their characterization.

Before we start the discussion of the EOMCC results for the anthracene molecule, it is worth mentioning that the VEEs cannot be measured experimentally and the results of theoretical calculations can only be interpreted in terms of approximate positions of the maxima in the experimental UV absorption spectrum (for the pertinent discussion, see ref 56). For example, it was shown that even for small molecules (benzene and furan) the uncertainties can be on the order of 0.1–0.2 eV.^{57–59} Similar conclusions have been drawn for the naphthalene and anthracene molecules, where the 0–0 transition energies were estimated using the CIS method and compared to the accurate gas-phase experiments (see, for example, ref 61), showing that the CIS VEEs are 0.2–0.6 eV higher in energy than the maximum intensity peaks.⁶⁰ Other factors contributing to experiment–theory discrepancies can be attributed to the basis set quality and level of theory used to describe correlation effects. In order to make adequate comparisons, in our studies for anthracene we will use the solvent-corrected experimental VEEs of ref 62.

In previous studies various levels of theory have been applied to these systems, from TDDFT methods to the coupled cluster CC2 formalism²⁵ (see refs 62–64 and references therein). In Table 1 we show the VEEs for anthracene obtained with the AVTZ, cc-pVDZ, and POL1 basis sets. For all basis sets employed, one can observe the significant effects of the CR-EOMCCSD(T) corrections, which lower the EOMCCSD ¹L_a and ¹L_b VEEs by 0.24–0.31 eV. The best predictions of the VEEs corresponding to these states are obtained with the POL1 basis set. In both cases, the CR-EOMCCSD(T) errors are reduced to within 0.1 eV. One should also notice that all EOMCC calculations shown in Table 1 reverse the ordering of the ¹L_a and ¹L_b states. Instead of 0.04 eV above, the ¹L_b state is predicted to be 0.1 eV below the ¹L_a state. This is somewhat puzzling, since the CC2 calculations of ref 62 predict ordering consistent with the experiment, although the ¹L_a and ¹L_b separation of 0.2 eV obtained with the CC2 method is a bit exaggerated.⁶² Similar reverse ordering has been reported in the context of

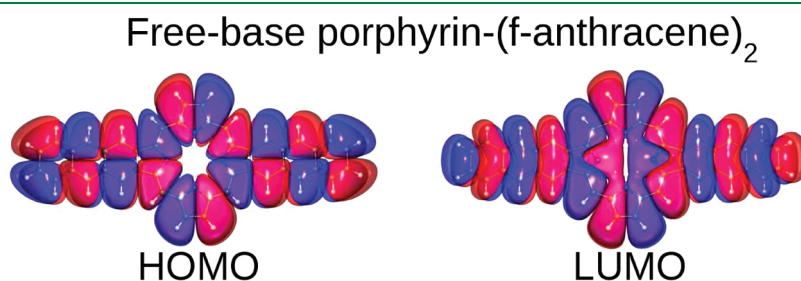


Figure 3. Graphical representation of the RHF HOMO and LUMO orbitals for the FBP-(f-anthracene)₂ system obtained with the cc-pVDZ basis set.

multireference Møller–Plesset (MRPT) theory^{66,67} calculations for low-lying excited states of anthracene (see ref 65). In the case of the MRPT approach, the 0.17 eV separation between 1L_b and 1L_a states is slightly larger than 0.1 eV obtained with the CR-EOMCCSD(T) method for the POL1 basis set. The discrepancies between the CR-EOMCCSD(T) and experimental vertical energies of ref 62 can be attributed to many factors including (1) basis set quality [For the 1L_a state one can observe a strong effect due to the basis set quality. The difference between AVTZ and POL1 CREOMCCSD(T) VEEs for this state is as big as 0.53 eV. This difference, however, is much smaller for the 1L_b state (0.17 eV) predicted to be the lowest singlet excited state by all presented EOMCC calculations.], (2) insufficient treatment of correlation effects, and (3) uncertainties of the TDDFT-based procedure of defining the experimental VEEs described in ref 62.

For the coronene molecule (Table 2), we used AVTZ, cc-pVDZ, POL1, and cc-pVTZ basis sets. The experimental value is available only for the 1L_a state.⁶⁸ As can be seen from Table 2, the best estimate for the experimental VEE of the 1L_a state is obtained with the CR-EOMCCSD(T) method in

cc-pVTZ basis set, where the error is reduced to 0.05 eV. The analogous error for the POL1 basis set amounts to 0.09 eV. It should be stressed that for the 1L_a state a strong dependency on the basis set is observed. For example, the CR-EOMCCSD(T) predictions for the AVTZ and cc-pVTZ basis sets differ by 0.48 eV. However, for the first singlet excited state, the AVTZ and cc-pVTZ CR-EOMCCSD(T) difference is much smaller, 0.14 eV. The same is true for the cc-pVDZ basis set, although for this basis set these differences are reduced to 0.15 eV (1L_a state) and 0.08 eV (1L_b state), respectively. A similar trend can be observed for the lowest states predicted by the EOMCC methods for the anthracene molecule.

The spectral properties of anthracene fused porphyrins have recently been studied experimentally. In ref 7 it was demonstrated that the fusion to the anthracene can significantly change the electronic structure of the porphyrin, shifting the absorption maximum to the near-infrared region. Our ZnP-f-anthracene model can be used to study this effect. The ZnP-f-anthracene molecule can be considered as a model for the β ,meso, β triply fused porphyrin (system **1c** using the nomenclature of ref 7), which shows an absorption maximum at 725 nm (1.71 eV). Our particular goal is to establish the quality of the EOMCC methods in describing this absorption shift. In Table 3 we show the results of the EOMCCSD and CR-EOMCCSD(T) calculations with the AVTZ and cc-pVDZ basis sets. The results obtained with both basis sets show that the lowest excited state is dominated by HOMO→LUMO excitation; i.e., the lowest state can be categorized as the 1L_a state. The absolute values of dominant R_K amplitudes in the cc-pVDZ basis set are as follows: $|R_{L\alpha}^{H\alpha}| = |R_{L\beta}^{H\beta}| \approx 0.57$, $|R_{L+1\alpha}^{H-1\alpha}| = |R_{L+1\beta}^{H-1\beta}| \approx 0.30$. Similar amplitude values were obtained for the AVTZ basis set. The Hartree–Fock (HF) HOMO and LUMO orbitals in the cc-pVDZ basis set are plotted in Figure 2. Moreover, both basis sets give virtually identical EOMCCSD results (2.03 eV), which are 0.32 eV above the experimental 1.71 eV value. By adding CR-EOMCCSD(T) corrections we significantly reduce the EOMCCSD errors to 0.08 and 0.06 eV for AVTZ and cc-pVDZ basis sets, respectively. The 1.77 eV CR-EOMCCSD(T) VEE for the 1L_a state is in good agreement with the experimental 1.71 eV value. The CR-EOMCCSD(T)/cc-pVDZ 1.77 eV value should be compared with the VEEs 2.25, 2.25, and 3.76 eV of lowest-lying singlet excited states obtained with the CR-EOMCCSD(T) method for the zinc–porphyrin (ZnP) system.⁵⁴ It should be stressed that for the ZnP-f-anthracene model the discrepancy between CR-EOMCCSD(T) VEEs obtained with the AVTZ and cc-pVDZ basis sets are smaller than in the case of anthracene and coronene.

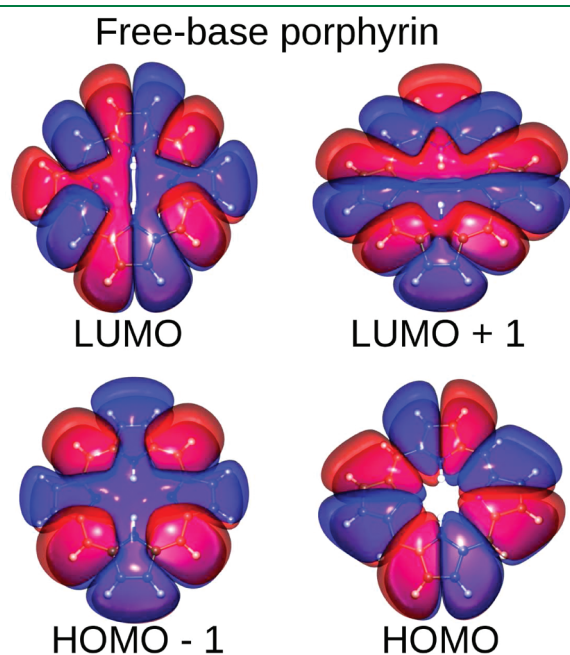


Figure 4. Graphical representation of the RHF HOMO–1, HOMO, LUMO, and LUMO+1 orbitals for the FBP molecule obtained with the cc-pVDZ basis set.

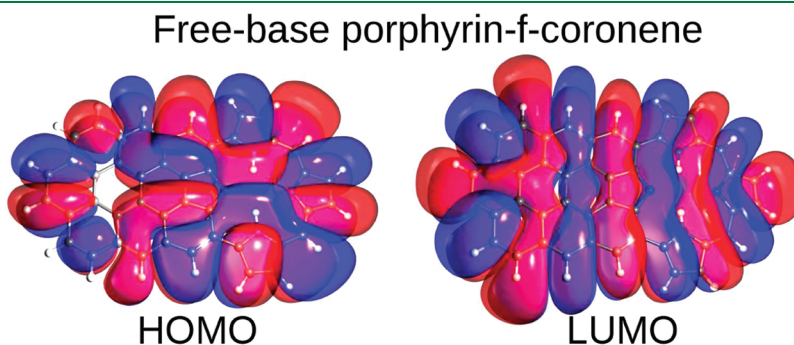


Figure 5. Graphical representation of the RHF HOMO and LUMO orbitals for the FBP-f-coronene system obtained with the AVTZ basis set.

The results discussed above indicate that for larger systems the lowest excited state can be satisfactorily described by the AVTZ or cc-pVDZ basis sets. Using these basis sets we will evaluate the effect of functionalizing free base porphyrin with one- and two-dimensional structures. These problems will be studied with our FBP-(f-anthracene)₂ and FBP-f-coronene models. The results of our calculations are shown in Table 4. For the FBP-(f-anthracene)₂ system we employed the cc-pVDZ basis set, which for ¹L_a and ¹L_b states gives EOMCCSD VEEs equal to 1.99 and 2.57 eV, respectively. As in the ZnP-f-anthracene model, the inclusion of the CR-EOMCCSD(T) corrections lowers the EOMCCSD VEEs to 1.78 and 2.37 eV, respectively. This should be compared with the CR-EOMCCSD(T) energies of 1.86 and 2.32 eV corresponding to the ¹B_{3u} and ¹B_{2u} states of FBP.³⁵ In order to understand the nature of the ¹L_a and ¹L_b states, we plotted the cc-pVDZ HF HOMO and LUMO orbitals of FBP-(f-anthracene)₂ and HOMO-1, HOMO, LUMO, LUMO+1 orbitals of FBP; see Figures 3 and 4, respectively. There are two notable features of the FBP-(f-anthracene)₂'s HOMO and LUMO orbitals. First, both orbitals are delocalized over the whole system. Second, the parts of the FBP-(f-anthracene)₂ HOMO and LUMO orbitals localized on FBP bear great resemblance to the FBP's HOMO and LUMO orbitals. Given that the largest contribution to this state comes from the HOMO-LUMO excitation ($|R_{L\alpha}^{H\alpha}| = |R_{L\beta}^{H\beta}| \approx 0.54$ for the cc-pVDZ basis set) and the second largest contribution to the ¹L_a state corresponds to the HOMO-1→LUMO+1 excitation ($|R_{L+1\alpha}^{H-1\alpha}| = |R_{L+1\beta}^{H-1\beta}| \approx 0.36$ for the cc-pVDZ basis), the ¹L_a should be attributed to the ¹B_{2u} state of FBP ($|R_{L\alpha}^{H\alpha}| = |R_{L\beta}^{H\beta}| \approx 0.48$ and $|R_{L+1\alpha}^{H-1\alpha}| = |R_{L+1\beta}^{H-1\beta}| \approx 0.45$ in the cc-pVDZ basis set), which corresponds to 0.54 eV red-shift using CR-EOMCCSD(T) results. It is also interesting to notice that the CR-EOMCCSD(T) method yields almost identical estimates for the ¹L_a states of the FBP-(f-anthracene)₂ and ZnP-f-anthracene systems. The ¹L_b state of FBP-(f-anthracene)₂ is dominated by the HOMO-1→LUMO and HOMO→LUMO+1 excitations ($|R_{L+1\alpha}^{H\alpha}| = |R_{L+1\beta}^{H\beta}| \approx 0.52$, $|R_{L\alpha}^{H-1\alpha}| = |R_{L\beta}^{H-1\beta}| \approx 0.40$ in the cc-pVDZ basis set) and bears a resemblance to the ¹B_{3u} state of FBP ($|R_{L\alpha}^{H-1\alpha}| = |R_{L\beta}^{H-1\beta}| \approx 0.48$, $|R_{L+1\alpha}^{H\alpha}| = |R_{L+1\beta}^{H\beta}| \approx 0.43$ in the cc-pVDZ basis set).

Due to its size, the calculations for the FBP-f-coronene model were performed only for the ¹L_a state in the AVTZ basis set. From Table 4 it is evident that the effect of functionalization of FBP with coronene has much stronger effects than when anthracene is used. The EOMCCSD VEE corresponding to the ¹L_a state is equal 1.63 eV, while the CR-EOMCCSD(T) yields 1.36 eV, which is more than 0.4 eV below the ¹L_a VEEs for the FBP-(f-anthracene)₂ and ZnP-f-anthracene systems. The HOMO and LUMO orbitals are shown in Figure 5. The ¹L_a is almost entirely dominated by the HOMO→LUMO excitation ($|R_{L\alpha}^{H\alpha}| = |R_{L\beta}^{H\beta}| \approx 0.64$ in the AVTZ basis set). It is interesting to note that the HOMO orbital is to a large extent localized on the FBP component and on the buffer region between FBP and coronene. The opposite tendency can be observed for the LUMO orbital, which is mostly localized on the coronene component.

PARALLEL PERFORMANCE

In all CC/EOMCC calculations reported in this paper, in-core algorithms were used. All cluster/excitation amplitudes, electron integrals, and recursive intermediates were stored in global arrays (GA).⁶⁹ GA forms an abstraction layer that alleviates the task of the developer by isolating most of the complexities involved in parallelizing

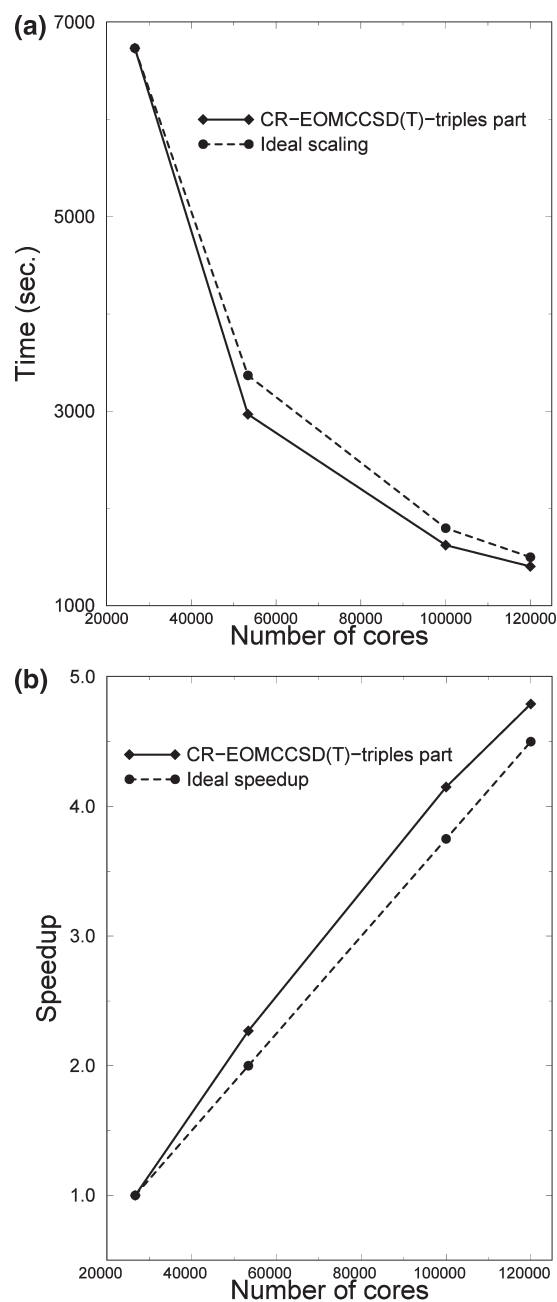


Figure 6. Scalability of the triples part of the CR-EOMCCSD(T) approach for the FBP-f-coronene system in the AVTZ basis set: (a) the time to solution as a function of the number of cores and (b) the corresponding speedups. Timings were determined from calculations on the Jaguar Cray XT5 computer system at NCCS (ORNL).

software that make use of dense matrices. It relies on several components: a message passing library, the ARMCI one-sided communication library, and a memory allocator (MA library). Most of the NWChem modules (including the ones used for this work) make very little use of MPI, since the majority of the communication is managed by the ARMCI one-sided communication library.

In some cases we also employed the CCSD/EOMCCSD implementations utilizing a new task scheduling approach based on the global task pool. Typical timings characterizing our calculations can be summarized as follows.

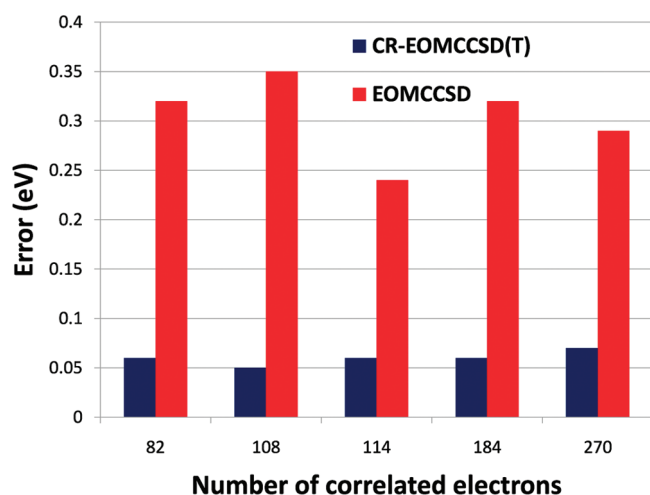


Figure 7. The EOMCCSD and CR-EOMCCSD(T) errors with respect to the experimentally available data for the green fluorescent protein chromophore, coronene, FBP, and oligoporphyrin dimer molecules (see the text for details).

- The coronene molecule in the cc-pVTZ basis set (888 basis set functions, 1020 cores, Chinook HP at EMSL): CCSD iteration, 5.5 min; EOMCCSD iteration, 5.9 min; recursive intermediates for the triples part of the CR-EOMCCSD(T) approach, 28.5 min; triples part of the CR-EOMCCSD(T) method, 2.7 h.
- The FBP-(f-anthracene)₂ system in the cc-pVDZ basis set (802 basis set functions, 1020 cores, Chinook HP at EMSL): CCSD iteration, 21.5 min; EOMCCSD iteration, 21.5 min; recursive intermediates for the triples part of the CR-EOMCCSD(T) approach, 1.48 h; triples part of the CR-EOMCCSD(T) method, 9.6 h.
- The FBP-f-coronene system in the AVTZ basis set (780 basis set functions, 4096 cores, Jaguar Cray XT5 at ORNL): CCSD iteration, 21.3 min; EOMCCSD iteration, 22.8 min; CR-EOMCCSD(T) recursive intermediates, 2.1 h. All these calculations were performed using eight cores per 12 core node. It means that effectively 6144 cores have been used. The scalability of the triples part of the CR-EOMCCSD(T), which is characterized by the N^7 complexity, is shown in Figures 6a,b. In our test calculations this part of code was characterized by a very good parallel performance across 120 000 cores. For example, the 6733 s required by the triples part on 26 672 cores is reduced to 2968, 1623, and 1404 s when 53 326, 100 016, and 120 000 cores are used, respectively. Again, in these tests eight cores per 12 core node have been used, which means that 180 000 cores have been allocated in the largest calculation. Since future parallel architectures may be characterized by smaller memory per core, we performed one calculation for 100 008 cores using all (12) cores per node. The obtained timing of 1787 s is only slightly worse than the 1623 s obtained with 100 016 cores when eight cores per node are used. For eight cores per node calculations we used 600 and 600 MB partitioning for local and global memory, respectively, while in the 12 core per node run we used 600 and 300 MB allocations for the two types of memory. In order to properly address the local memory bottleneck, we have developed the new version of the CR-EOMCCSD(T) approach where all six-dimensional

tensors can be dynamically “sliced” across first two dimensions in order to match available local memory.

CONCLUSIONS

The functionalization of the porphyrins with two-dimensional structures (in the current studies coronene was used) has stronger effects than the functionalization with one-dimensional structures (anthracene). While for the two types of anthracene functionalization considered here the resulting VEEs for lowest excited states of functionalized porphyrins have similar values, the functionalization of the FBP with coronene significantly shifts the VEE of the lowest excited state into the infrared region. We also demonstrated that the functionalization of the porphyrins with the anthracene or coronene significantly alters the electronic structure of the low-lying excited states. Instead of the multiconfigurational character of the lowest excited states of the FBP and ZnP systems, we showed that the lowest states of the ZnP-f-anthracene, FBP-(f-anthracene)₂, and FBP-f-coronene systems are dominated by the HOMO–LUMO excitations. For the ZnP-f-anthracene, the CR-EOMCCSD(T) approach, in the modest AVTZ and cc-pVDZ basis sets, is capable of reproducing the experimental excitation energies to within 0.1 eV. The remaining errors can be attributed to basis set effects, missing correlation effects, and simplified representation of the true molecular system used in the experiment. At the same time, the results obtained for anthracene and coronene suggest that for the VEEs corresponding to higher excited states, basis set effects play an important role. We have also demonstrated that the N^7 part of the CR-EOMCCSD(T) calculations can take advantage of a large number of cores and the scalability across 120 000 cores can be achieved.

Our last conclusion concerns the effect of higher-order correlation effects in describing the VEEs for large systems. It was demonstrated in the previous studies that while for singly excited states the EOMCCSD approach provides a good description of corresponding VEEs, for the proper description of doubly excited states including the effect of triply excited configurations is indispensable for the qualitative accuracy of the EOMCC calculations.³³ Since these conclusions were inferred from calculations for small molecular systems, it is important to answer the same question about the role of particular correlation effects for large systems. The results of this paper in conjunction with previous studies³⁵ will help us to understand these problems at least for singly excited states. Figure 7 contains the deviations of the EOMCCSD and CR-EOMCCSD(T) results from the available experimental values for several systems of different size: Green Fluorescent Protein Chromophore (82 correlated electrons, POL1 basis set), coronene (108 correlated electrons, cc-pVTZ basis set), FBP (114 correlated electrons, POL1 basis set), ZnP-f-anthracene (184 correlated electrons, cc-pVDZ basis set), and oligoporphyrin dimer P₂TA^{10,70} (270 correlated electrons, cc-pVDZ basis set). Although various quality of basis sets have been used, the general trend can be clearly seen from the Figure 7. One can notice that noniterative CR-EOMCCSD(T) corrections significantly reduce the absolute values of the 0.24–0.35 eV EOMCCSD errors to 0.05–0.07 eV (see Figure 7), which clearly demonstrate the increasingly more important role played by triple excitations for singly excited states of larger molecular systems. Similar studies for doubly excited states of large systems are facing problems with locating doubly excited states, which are shifted upward in the energy spectrum by the EOMCCSD method. For this purpose, in

the coming studies we will use the sequence of the EOMCCD, EOMCCSD, and CR-EOMCCSD(T) calculations. The role of the EOMCCD approximation will be to identify the lowest lying doubly excited states, which can be used as a starting vectors for the EOMCCSD approach.

Although the comparison of the EOMCC results with the TDDFT ones is beyond the scope of this paper, it is worth mentioning that while for the anthracene, coronene, ZnP-f-anthracene, and FBP-(f-anthracene)₂ the CR-EOMCCSD(T) and CAM-B3LYP¹² results are in a good agreement, for the FBP-f-coronene one can observe significant discrepancy between the CR-EOMCCSD(T) and CAM-B3LYP excitation energies corresponding to the lowest singlet excited state⁷¹ obtained for the same AVTZ basis set.

AUTHOR INFORMATION

Corresponding Author

*E-mail: karol.kowalski@pnl.gov (K.K.), aprae@ornl.gov (E.A.).

ACKNOWLEDGMENT

The work related to the development of the scalable EOMCCSD and CR-EOMCCSD(T) approaches (K.K.) and development of new parallel tools (S.K.) was supported by the Extreme Scale Computing Initiative, a Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory. Most of the calculations have been performed using EMSL, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory. The Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by the Battelle Memorial Institute under Contract DE-AC06-76RLO-1830. The scalability tests of the CR-EOMCCSD(T) implementation of NWChem have been performed on the Jaguar Cray-XT5 computer system of the National Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

REFERENCES

- (1) Tsuda, A.; Furuta, H.; Osuka, A. *Angew. Chem., Int. Ed.* **2000**, *39*, 2549–2552.
- (2) Tsuda, A.; Furuta, H.; Osuka, A. *J. Am. Chem. Soc.* **2001**, *123*, 10304–10321.
- (3) Tsuda, A.; Osuka, A. *Science* **2001**, *293*, 79–82.
- (4) Anthony, J. E. *Chem. Rev.* **2006**, *106*, 5028–5048.
- (5) Wu, J.; Pisula, W.; Müllen, K. *Chem. Rev.* **2007**, *107*, 718–747.
- (6) Müllen, K.; Rabe, J. P. *Acc. Chem. Res.* **2008**, *41*, 511–520.
- (7) Davis, N. K. S.; Pawlicki, M.; Anderson, H. L. *Org. Lett.* **2008**, *10*, 3945–3947.
- (8) Brédas, L. L.; Norton, J. E.; Cornil, J.; Coropceanu, V. *Acc. Chem. Res.* **2009**, *42*, 1691–1699.
- (9) Imahori, H.; Umeyama, T.; Ito, S. *Acc. Chem. Res.* **2009**, *42*, 1809–1818.
- (10) Reimers, J. R.; Hall, L. E.; Crossley, M. J.; Hush, N. S. *J. Phys. Chem. A* **1999**, *103*, 4385–4397.
- (11) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 35403544.
- (12) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (13) Jensen, L.; Govind, N. *J. Phys. Chem. A* **2009**, *113*, 9761–9765.
- (14) Peach, M. J. G.; Helgaker, T.; Salek, P.; Keal, T. W.; Lutnaes, O. B.; Tozer, D. J.; Handy, N. C. *Phys. Chem. Chem. Phys.* **2006**, *8*, 558–562.
- (15) Tokura, S.; Yagi, K.; Tsuneda, T.; Hirao, K. *Chem. Phys. Lett.* **2007**, *436*, 30–35.
- (16) Glaesemann, K. R.; Govind, N.; Krishnamoorthy, S.; Kowalski, K. *J. Phys. Chem. A* **2010**, *114*, 8764–8771.
- (17) Bartlett, R. J.; Musiał, M. *Rev. Mod. Phys.* **2007**, *79*, 291–352.
- (18) Geersten, J.; Rittby, M.; Bartlett, R. J. *Chem. Phys. Lett.* **1989**, *164*, 57–62.
- (19) Comeau, D. C.; Bartlett, R. J. *Chem. Phys. Lett.* **1993**, *207*, 414–423.
- (20) Stanton, J. F.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 7029–7039.
- (21) Monkhorst, H. J. *Int. J. Quantum Chem.* **1977**, *S11*, 421–432.
- (22) Dalggaard, E.; Monkhorst, H. J. *Phys. Rev. A* **1983**, *28*, 1217–1222.
- (23) Koch, H.; Jørgensen, P. *J. Chem. Phys.* **1990**, *93*, 3333–3344.
- (24) Nakatsuji, H.; Hirao, H. *J. Chem. Phys.* **1978**, *69*, 4535–4547.
- (25) Christiansen, O.; Koch, H.; Jørgensen, P. *Chem. Phys. Lett.* **1995**, *243*, 409–418.
- (26) Christiansen, O.; Koch, H.; Jørgensen, P. *J. Chem. Phys.* **1995**, *103*, 7429–7441.
- (27) Christiansen, O.; Koch, H.; Jørgensen, P.; Olsen, J. *Chem. Phys. Lett.* **1996**, *256*, 185–194.
- (28) Nooijen, M. *J. Chem. Phys.* **1996**, *104*, 2638–2651.
- (29) Krylov, A. I. *Chem. Phys. Lett.* **2001**, *338*, 375–384.
- (30) Musiał, M.; Bartlett, R. J. *J. Chem. Phys.* **2011**, *134*, 034106.
- (31) Kowalski, K.; Piecuch, P. *J. Chem. Phys.* **2001**, *115*, 643–651.
- (32) Hirata, S. *J. Chem. Phys.* **2004**, *121*, 51–59.
- (33) Kowalski, K.; Piecuch, P. *J. Chem. Phys.* **2004**, *120*, 1715–1738.
- (34) Kowalski, K.; Piecuch, P. *J. Chem. Phys.* **2001**, *115*, 2966–2978.
- (35) Kowalski, K.; Krishnamoorthy, S.; Villa, O.; Hammond, J. R.; Govind, N. *J. Chem. Phys.* **2010**, *132*, 154103.
- (36) Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1995**, *233*, 81–87.
- (37) Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1996**, *258*, 581–588.
- (38) Christiansen, O.; Koch, H.; Jørgensen, P. *J. Chem. Phys.* **1996**, *105*, 1451–1459.
- (39) Hirata, S.; Nooijen, M.; Grabowski, I.; Bartlett, R. J. *J. Chem. Phys.* **2001**, *114*, 3919–3928. **2001**, *115*, 3967–3968.
- (40) Shiozaki, T.; Hirao, K.; Hirata, S. *J. Chem. Phys.* **2007**, *126*, 244106.
- (41) Manohar, P. U.; Krylov, A. I. *J. Chem. Phys.* **2008**, *129*, 194105.
- (42) Włoch, M.; Lodriguito, M. D.; Piecuch, P.; Gour, J. R. *Mol. Phys.* **2006**, *104*, 2149–2172. Piecuch, P.; Gour, J. R.; Włoch, M. *Int. J. Quantum Chem.* **2009**, *109*, 3268–3304.
- (43) Kowalski, K. *J. Chem. Phys.* **2009**, *130*, 194110.
- (44) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- (45) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Aprà, E.; Windus, T. L.; de Jong, W. A. *Comput. Phys. Commun.* **2010**, *181*, 1477–1489.
- (46) Hirata, S. *J. Phys. Chem. A* **2003**, *107*, 9887–9897.
- (47) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (48) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (49) Balabanov, N. B.; Peterson, K. A. *J. Chem. Phys.* **2005**, *123*, 064107.
- (50) Hammond, J. R.; de Jong, W. A.; Kowalski, K. *J. Chem. Phys.* **2007**, *127*, 144105.
- (51) Sekino, H.; Kobayashi, H. *J. Chem. Phys.* **1981**, *75*, 3477–3484.
- (52) Sadlej, A. J. *Collect. Czech. Chem. Commun.* **1988**, *53*, 1995–2016.
- (53) Schafer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.
- (54) Fan, P. D.; Valiev, M.; Kowalski, K. *Chem. Phys. Lett.* **2008**, *458*, 205–209.
- (55) Platt, J. R. *J. Chem. Phys.* **1949**, *17*, 484–495.
- (56) Bak, K. L.; Koch, H.; Oddershede, J.; Christiansen, J.; Sauer, S. P. A. *J. Chem. Phys.* **2000**, *112*, 4173–4185.
- (57) Christiansen, O.; Jørgensen, P. *J. Am. Chem. Soc.* **1998**, *120*, 3423–3430.

- (58) Christiansen, O.; Koch, H.; Jørgensen, P.; Helgaker, T.; de Merás, A. S. *J. Chem. Phys.* **1996**, *105*, 6921–6939.
- (59) Christiansen, O.; Stanton, J. F.; Gauss, J. *J. Chem. Phys.* **1998**, *108*, 3987–4001.
- (60) Jas, G. S.; Kuczera, K. *Chem. Phys.* **1997**, *214*, 229–241.
- (61) Lambert, W. R.; Felker, P. M.; Syage, J. A.; Zewail, A. H. *J. Chem. Phys.* **1984**, *81*, 2195–2207.
- (62) Grimme, S.; Parac, M. *ChemPhysChem* **2003**, *4*, 292–295.
- (63) Parac, M.; Grimme, S. *Chem. Phys.* **2003**, *292*, 11–21.
- (64) Wong, B. M.; Hsieh, T. H. *J. Chem. Theory Comput.* **2010**, *6*, 3704–3712.
- (65) Kawashima, Y.; Hashimoto, T.; Nakano, H.; Hirao, K. *Theor. Chem. Acc.* **1998**, *102*, 49–64.
- (66) Hirao, K. *Chem. Phys. Lett.* **1992**, *190*, 374–380.
- (67) Hirao, K. *Chem. Phys. Lett.* **1992**, *196*, 397–403.
- (68) Birks, J. B. *Photophysics of Aromatic Molecules*; Wiley: New York, 1970.
- (69) Nieplocha, J.; Harrison, R. J.; Littlefield, R. J. Global Arrays: A Portable Shared-Memory Programming Model for Distributed Memory Computers. In *Proceedings of Supercomputing 94*; IEEE CS Press: Washington, DC, 1994; 340349. Nieplocha, J.; Harrison, R. J.; Littlefield, R. J. *J. Supercomput.* **1996**, *10*, 169–189.
- (70) Sendt, K.; Johnston, L. A.; Hough, W. A.; Crossley, M. J.; Hush, N. S.; Reimers, J. R. *J. Am. Chem. Soc.* **2002**, *124*, 9299–9309.
- (71) Aprà E.; Lopata, K.; Govind, N.; Kowalski, K. Manuscript in preparation.

Solvation Effects on Electronic Transitions: Exploring the Performance of Advanced Solvent Potentials in Polarizable Embedding Calculations

Tobias Schwabe,^{*,†} Jógvan Magnus Haugaard Olsen,[†] Kristian Sneskov,^{†,§} Jacob Kongsted,[†] and Ove Christiansen^{†,§}

[†]Center for Oxygen Microscopy and Imaging, Department of Chemistry, Aarhus University, Langelandsgade 140, DK-8000 Aarhus C, Denmark

[‡]Department of Physics and Chemistry, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

[§]The Lundbeck Foundation Center for Theoretical Chemistry, Department of Chemistry, Aarhus University, Langelandsgade 140, DK-8000 Aarhus C, Denmark

S Supporting Information

ABSTRACT: The polarizable embedding (PE) approach, which combines quantum mechanics (QM) and molecular mechanics (MM), is applied to predict solvatochromic effects on excitation energies of several representative molecules in aqueous, methanol, acetonitrile, and carbon tetrachloride solutions. Good agreement with experimental results for excitation energies and for solvatochromic shifts is demonstrated on the basis of either density functional theory or coupled cluster methods. Solvent-dependent trends are fully reproduced in this diverse set of solvents. Furthermore, it is shown that the inclusion of higher order multipole moments and anisotropic polarizabilities in the electrostatic embedding potentials leads to a faster convergence with respect to a full QM treatment (within about 0.1 eV of estimated full QM treatments). It is thereby illustrated that the use of advanced solvent potentials can provide higher accuracy compared to various simpler approaches for the prediction of solvent shifts and do so in a computationally competitive manner.

1. INTRODUCTION

Quantum chemistry has developed a thorough knowledge about isolated molecules in vacuum at 0 K. Over the years, many theoretical methods have been developed and benchmarked for performance in different contexts. As a result, methods with known accuracy and computational cost exist for attacking a variety of chemical problems. These methods are primarily grounded in density functional theory (DFT) and post-Hartree–Fock wave function approaches, and benchmarking is an active field in (vacuum) quantum chemistry.^{1–4} However, this picture changes when we turn to condensed phases in general and to solvation in particular. Here, the computational demands increase due to a number of factors: the system size increases rapidly because of the long-range Coulombic interactions, and dynamical effects need to be taken into account due to nonzero temperature effects. Therefore, more often than not, solvation studies have been addressed within the framework of molecular mechanics. Indeed, good results have been reported in the simulation of macroscopic properties.⁵ However, when properties directly related to the electronic structure are the focus, it is necessary to go beyond a completely classical description. One very obvious case for this is the calculation of electronic excitations due to light absorption.

To be able to treat these processes, one may rely on hybrid approaches like combined quantum mechanics and molecular mechanics (QM/MM).⁶ Here, the system is divided into a quantum mechanical subsystem (QM), where the excitation

takes place, and the surrounding environment (MM), which is normally described by the electrostatic potentials of discrete solvent molecules. Alternatively, the bulk solvent effects can be described with an implicit solvation model like the conductor-like solvation model (COSMO)^{7,8} or the polarizable continuum model (PCM).^{9–11} Recently, a combination of QM/MM and PCM (the QM/MM/PCM model) has also been put forward for the calculation of electronic excitation energies in condensed phases.¹² While the gain in computation cost with an implicit solvation model is considerable, solvent-specific effects are not covered well.¹³ In particular, the effect of hydrogen bonding is not well reproduced. A further drawback of an implicit approach is that it is not straightforward to extend to other similar molecular systems where QM/MM approaches, on the other hand, have proven beneficial, e.g., for the study of large biomolecular systems.¹⁴ However, applying explicit solvent approaches for the description of solvation effects on excitation energies is a complex task. Due to temperature effects, it is necessary to sample the configurational space of the solute–solvent system. For this, a molecular dynamics (MD) or Monte Carlo (MC) simulation approach is needed. This may be carried out in a pure MM or a QM/MM framework. Furthermore, within the past decade, full ab initio MD simulations in the framework of Car–Parrinello MD (CPMD) have become feasible.¹⁵ Recently,

Received: April 15, 2011

Published: May 23, 2011

other schemes and combinations thereof have been developed and applied.^{16–19} The MD simulation yields many molecular configurations (often more than 100), all of which have to be subjected to a QM/MM calculation of the electronic excitation energy. Therefore, the quality of the solvent description relies on the quality of the underlying MD simulation, the QM method, the potentials used in the MM part, and the coupling of the QM and MM subsystems. In fact, a further practical complication arises because, normally, depending on the chosen combination for QM/MM, there are no integrated programs carrying out the full protocol.

Overall, there is a need for more accurate QM/MM methods. This requires an improvement of all relevant components in order to increase the overall accuracy. While a lot of effort has been spent to make more accurate QM methods available for QM/MM treatments, the description of the MM subsystem has received less attention. To address the former, a coupled cluster (CC) as well as a DFT QM/MM approach with polarizable force fields have been introduced.^{20–22} Furthermore, redesigned implementations of both approaches have recently been presented using a polarizable embedding (PE) scheme,^{23,24} adding, among other things, the possibility to use multipole moments up to octopoles as well as anisotropic polarizabilities to describe the electrostatic embedding potential.

We emphasize that the PE approach to QM/MM studied here uses a fully self-consistent description of the polarization between the QM (be it DFT or CC) and MM subsystems, for both the ground and excited states. The need for a balanced description of the polarization when investigating solvation effects was recognized 15 years ago by Thompson, who introduced a QM/MM scheme using polarizable force fields and a semiempirical QM method.²⁵ The pioneering work for QM/MM by Warshel and Levitt⁶ also considered mutual polarization of both subsystems. Other approaches in that direction include an alternative implementation of DFT/MM using polarizable force fields,^{26,27} frozen-density embedded density functional theory,^{28,29} the effective fragment potential approach,^{30–32} CASPT2/MM using polarizable force fields,³³ and the many-body expansion scheme.³⁴ The PE scheme differs from other embedding schemes in that it is derived from a long-range perturbational analysis of interacting subsystems. The most important of these effects are then described combining a full quantum description of one part of the full system, with an essentially electrostatic description of the remaining parts and their interaction with the electronic QM system. Therefore, in its present form, the PE scheme does not take into account short-range and dispersion effects on the electronic wave functions and energies, which are of a pure quantum mechanical nature. For further comparison between different embedding schemes, we refer to the literature.³⁵

An important reason for using sophisticated electrostatic potentials is that, in principle, this will allow accurate mimicking of the effects of hydrogen bonding to the QM system. Hence, one avoids the need of including the closer lying solvent molecules in the QM region. Especially for CC methods, the gain in computational efficiency is formidable and may be regarded as a prerequisite for applying CC methods routinely for QM/MM studies. Thorough treatment of hydrogen bonding is also mandatory for QM/MM studies of biomolecular systems where the border between the QM and MM region is often across hydrogen bonds. Advanced electrostatic potentials could provide higher accuracy compared to simpler approaches since, aiming for a

given accuracy, the inclusion of selected parts of the environment into the QM calculations should be far less important. However, these considerations are rather general and at least partially lack numerical evidence. In this paper, we present benchmark calculations that demonstrate how accurate solvatochromic shifts for electronic excitation energies can be obtained for a range of different solvents by applying the PE model in conjunction with electrostatic potentials derived from ab initio calculations. It is shown that the accuracy can be improved systematically by inclusion of higher order multipoles and polarizabilities in the electrostatic potential. We illustrate how these enhanced electrostatic potentials indeed significantly reduce the importance of including the nearest solvent molecules in the QM region while still yielding accurate results.

2. THEORY

In the following, we restrict ourselves to a brief introduction of the electronic QM/MM interface in the PE approach. A detailed derivation of ground state energy and response function expressions for DFT and CC can be found elsewhere.^{23,24} The general idea of the PE model is an expansion of the surrounding charge density in classical terms at discrete expansion points. The atomic sites of the MM region are an obvious choice for the expansion points, but bond midpoints can also be included. Each expansion point may be described by localized multipole moments (up to octopoles) and an anisotropic (or isotropic) polarizability which allows for dipole–dipole polarization. From this, we obtain an electrostatic ($E_{\text{PE}}^{\text{es}}$) and polarization (or induction; $E_{\text{PE}}^{\text{pol}}$) energy contribution. The former is given by the interaction between the permanent multipole moments in the MM region and the electrons and nuclei in the QM region:

$$E_{\text{PE}}^{\text{es}} = \sum_{s=1}^S \sum_{k=0}^K \frac{(-1)^k}{k!} \left(\sum_{m=1}^M Z_m \mathbf{T}_{ms}^{(k)} - \sum_{i=1}^N \mathbf{T}_{is}^{(k)} \right) \mathbf{Q}_s^{(k)} \quad (1)$$

Here, S is the total number of expansion points, M is the number of QM nuclei m , with nuclear charges Z_m , and N is the number of electrons i . The interaction tensors $\mathbf{T}_{st}^{(k)}$ are defined as

$$\mathbf{T}_{st}^{(k)} = \nabla_s^k \frac{1}{(|\mathbf{r}_s - \mathbf{r}_t|)} \quad (2)$$

where ∇_s is a derivative operator with respect to coordinates r_s . $\mathbf{Q}_s^{(k)}$ is a multipole moment of k th order located at expansion point s ; i.e., for $k = 0$, it is a point charge; for $k = 1$, a dipole moment; and so on.

The polarization energy due to the mutual polarization between the environment and QM region is given by

$$E_{\text{PE}}^{\text{pol}} = -\frac{1}{2} \mu^{\text{ind}} (\mathbf{F}^{\text{nuc}} + \mathbf{F}^{\text{elec}} + \mathbf{F}^{\text{mul}}) = -\frac{1}{2} \mu^{\text{ind}} \mathbf{F} \quad (3)$$

where the vector μ^{ind} of length $3S$ contains the complete set of induced dipoles. \mathbf{F} is the electric field vector containing the fields from the nuclei, electrons, and multipole moments in the MM region. To obtain the induced dipoles, one has to include the mutual interaction with the electric fields from all other induced dipoles and from the static electric field. This can be cast into a matrix equation

$$\mu^{\text{ind}} = \mathbf{R} (\mathbf{F}^{\text{nuc}} + \mathbf{F}^{\text{elec}} + \mathbf{F}^{\text{mul}}) = \mathbf{R} \mathbf{F} \quad (4)$$

where \mathbf{R} is a symmetric classical response matrix of dimension $3S \times 3S$. It is defined as

$$\mathbf{R} = \begin{pmatrix} \alpha_1^{-1} & \mathbf{T}_{12}^{(2)} & \cdots & \mathbf{T}_{1S}^{(2)} \\ \mathbf{T}_{21}^{(2)} & \alpha_2^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{T}_{(S-1)S}^{(2)} \\ \mathbf{T}_{S1}^{(2)} & \cdots & \mathbf{T}_{S(S-1)}^{(2)} & \alpha_S^{-1} \end{pmatrix}^{-1} \quad (5)$$

The inverse of the polarizability tensors α_s are placed along the diagonal, while the dipole–dipole interaction tensors are found as off-diagonal elements. An effective operator which couples the QM system and the environment can be derived from the presented energy expressions and further generalized to the calculation of, e.g., vertical electronic excitation energies within the framework of linear response theory. For details concerning the DFT (Hartree–Fock) or CC implementation of the PE model, we refer to the literature.^{23,24}

3. COMPUTATIONAL DETAILS

In the following, we describe the procedure used to compute the solvatochromic effects. The general concept follows the sequential Monte Carlo and QM (S-MC/QM) approach introduced by Coutinho and Canuto.³⁶ A classical MD simulation for the solute in solvation is carried out. From this simulation, we obtain a set of molecular configurations which are treated at a higher level of theory to calculate excitation properties of the system. The configurations are obtained from a thermally coupled simulation, and therefore they already inherit a Boltzmann distribution weighting. Thus, the properties of interest can be calculated directly as the average over all configurations. In principle, further information can be gained on the basis of the distribution of the property values.

Unless otherwise noted, the properties in solution are obtained by averaging over 120 snapshots of an MD simulation (see below). To save computational resources, in some cases, only 30 configurations have been used. To indicate the quality of the averaged results, we report the standard error of the mean (SEM). The effect of using only one-fourth of the configurations does not influence the results significantly. The difference between values averaged from 30 or 120 configurations is about 0.01 eV and therefore well below other effects like basis set incompleteness, the level of theory, or the approximations of the electrostatic environment.

3.1. Potential Determination. The multipole moments and polarizabilities are all obtained in the same way for solutes as well as solvents. First, the molecules are optimized at the B3LYP^{37,38}/aug-cc-pVTZ^{39,40}/PCM^{9,10} level of theory, where we apply the standard PCM parameters as provided in Gaussian 09.⁴¹ Next, we determine the multipole moments up to octopoles as well as the anisotropic polarizabilities by applying the LoProp⁴² approach as implemented in MOLCAS.⁴³ The properties are calculated at the B3LYP/a-aug-cc-pVTZ^{39,40} level for the isolated molecules in a vacuum using the geometry obtained from the PCM optimization. The a-aug-cc-pVTZ basis set is a recontraction of aug-cc-pVTZ to an ANO-type basis, which is necessary for the LoProp procedure.

Finally, we also obtain atomic point charges with the CHelpG procedure⁴⁴ fitted against the electrostatic potential from a B3LYP/aug-cc-pVTZ calculation again at the PCM optimized geometries. As an additional constraint, the reproduction of the

molecular dipole moment is imposed on the fitting. These point charges implicitly contain higher order multipole effects and therefore cannot be combined with a rigorous multipole moment expansion. Thus, whenever we only use point charges, we use those obtained from the CHelpG fitting (except for the comparison of the potentials with their ab initio counterpart), and otherwise the LoProp point charges are used. It should be mentioned that the sole purpose of the potentials is to reproduce as accurate as possible the exact electrostatic environmental potential and its (direct) response to a change in the solute electron density. In the following, we identify different approximations to the electrostatic potential as follows: Mm labels the applied multipole moments, with $m = 0$, point charges; $m = 1$, up to dipoles; $m = 2$, up to quadrupoles; and $m = 3$, up to octopoles, and Pp labels the polarizability of the MM centers, with $p = 0$, no polarization; $p = 1$, isotropic polarizabilities; and $p = 2$, anisotropic polarizabilities. To identify fitted point charges, we use the label M*. All potentials which have been obtained for this study are listed in the Supporting Information.

3.2. Molecular Dynamic Simulations. For all simulations, an equilibration run of 0.4 ns in time steps of 2 fs followed by a production run of 1.2 ns is carried out. Configurations are sampled every 10 ps. This time step is considered sufficient to obtain statistically uncorrelated configurations. The NVT ensemble is applied with $T = 298$ K. All molecules are treated as rigid bodies. Thus, only intermolecular potentials are considered. These are based on a 12-6-Lennard-Jones potential and our M*P1 potentials. The construction of the M*P1 potentials is described in the previous section. The Lennard-Jones parameters are taken from the OPLS-AA force field⁴⁵ as shipped with TINKER.⁴⁶

To model aqueous solutions, we used a slightly different potential, which consists of point charges and a molecular isotropic polarizability located at the oxygen atom, called the SPCpol water model.⁴⁷ This potential was used in all MDs of aqueous solutions. It was chosen to maintain comparability to previous studies. Note that we do not expect significant changes if the M*P1 potential for water had been applied instead. All simulations have been carried out with the MOLSIM program package.⁴⁸

3.3. PE-QM Calculations. The PE-QM calculations are performed using either CAM-B3LYP⁴⁹/aug-cc-pVDZ^{39,40} or CCSD⁵⁰/aug-cc-pVDZ for the QM part. Inputs are generated with the WHIRLPOOL program.⁵¹ For each configuration, the solute geometry is extracted, and a cutoff radius of 12 Å from the center of mass of the solute is applied. All solvent molecules within the cutoff are included in the electrostatic embedding potential. The actual PE-QM computations are done with a development version of the DALTON program.⁵² Excitation energies for which the PCM model has been applied are computed with Gaussian 09.⁴¹ All PCM calculations are performed using default settings: atomic radii are taken from universal force field and scaled by 1.1, and the static/dynamic dielectric constants are 78.355/1.778 (water), 32.613/1.766 (methanol), 35.688/1.807 (acetonitrile), and 2.228/2.132 (carbon tetrachloride).

3.4. Solvent Potential Analysis. To analyze the quality of the solvent potentials used in this work, the electrostatic potentials due to the multipole moments are compared to a QM reference. The QM calculations are performed at the same level of theory and geometry as the corresponding calculation of the potential parameters, i.e., at the B3LYP/aug-cc-pVTZ level using the

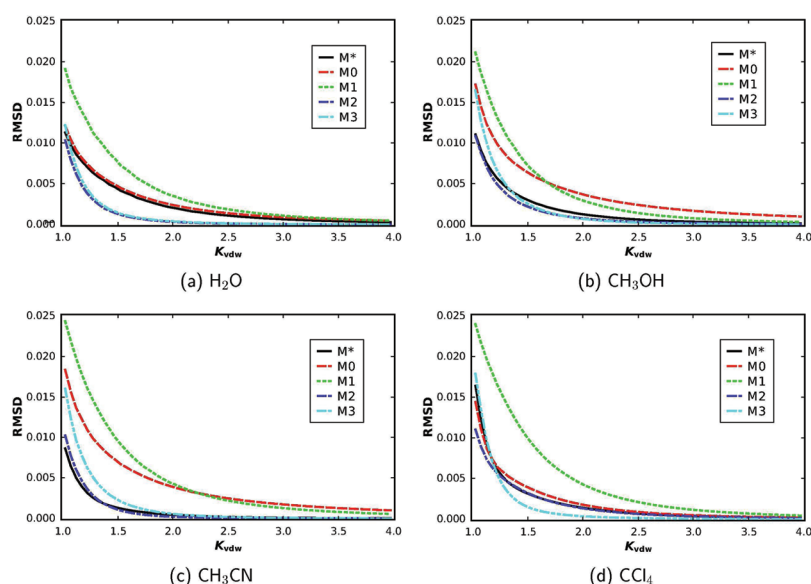


Figure 1. Plots of the RMSD in au between the electrostatic potentials created by permanent multipole moments and a QM reference with respect to distance from the molecular surface defined by interlocking atomic van der Waals spheres.

B3LYP/aug-cc-pVTZ/PCM geometry. For the analysis of the electrostatic potential due to induced dipole moments, we apply a static electric field in the x , y , and z directions with a magnitude of 0.01 au. The electrostatic potential is sampled at a number of points surrounding a single solvent molecule. The sampling volume consists of an inner boundary defined by interlocking atom-centered van der Waals spheres. The outer boundary is similarly defined using van der Waals spheres which have a radius of 4 times the van der Waals radii. The hydrogen van der Waals radius is from ref 53, and all other are from ref 54. Inside the volume, a number of points are uniformly distributed with a distance of 0.1 Å in the x , y , and z directions. This leads to 978 261, 1 343 269, 1 546 866, and 2 435 010 sampling points for water, methanol, acetonitrile, and carbon tetrachloride, respectively. In order to inspect the quality of the electrostatic potentials with respect to the distance from the molecule, we divide the entire volume into smaller subvolumes defined by an inner and outer boundary which is defined by van der Waals radius scaling factors, K_{vdw} and calculate the root-mean-square deviation (RMSD) over all points inside each subvolume. Each subvolume has a thickness of $0.05K_{\text{vdw}}$. The root-mean-square deviation (RMSD) is calculated as

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_a (\varphi_a^{\text{es}} - \varphi_a^{\text{QM}})^2} \quad (6)$$

where a is a sampling point, N is the total number of sampling points, and φ_a^{es} and φ_a^{QM} are the electrostatic potentials at point a due to the solvent potential and the QM reference, respectively. In the case of induced dipoles, the QM reference potential is obtained by subtracting the unperturbed electrostatic potential from the potential obtained in the external field.

The grid generation and calculation of electrostatic potentials due to the multipoles is handled by the Whirlpool program. Whirlpool also creates the input necessary for the QM calculations which are carried out using the DALTON program. The

subsequent comparisons between the electrostatic potentials are also performed using the Whirlpool program.

4. RESULTS AND DISCUSSION

4.1. Analysis of the Solvent Potentials. The quality of the solvent potentials, derived from the LoProp and CHelpG schemes, is evaluated by a comparison between the electrostatic potentials created by the multipole moments and the corresponding QM reference. The RMSDs between the electrostatic potentials with respect to the distance from the molecule are presented in Figure 1 in the case of permanent multipoles and Figure 2 in the case of induced dipoles.

The general observations for the permanent multipole moments derived from the LoProp procedure for all the considered solvents is that the electrostatic potential is reasonably well reproduced far away from the molecule, even at the lowest level, i.e., the M0 potential. However, as we approach midrange distances, the electrostatic potential due to M0 and M1 worsens, while the M2 and M3 potentials still retain good accuracy due to the higher order multipole moments. At short and especially very short range, the deviation from the QM reference increases rapidly for all potentials.

The M1 potential gives the poorest description in all cases, except at mid and long ranges for methanol and acetonitrile, where it performs slightly better than the M0 potential. We also observe that the general short-range performance of the M2 potential is better than or equal to the M3 potential. However, in the case of carbon tetrachloride, the M3 potential performs very well except at very short distances. Here, it is also worth noting that the M0 potential gives similar results to those of the M2 potential. The most noteworthy result is that the M* potential generates electrostatic potentials of almost the same quality as the M2 potential except for a water molecule. This is expected because the charges have been fitted against the QM electrostatic potential. The less accurate description for water can be explained by the fact that there are only three charges, which does not give enough flexibility

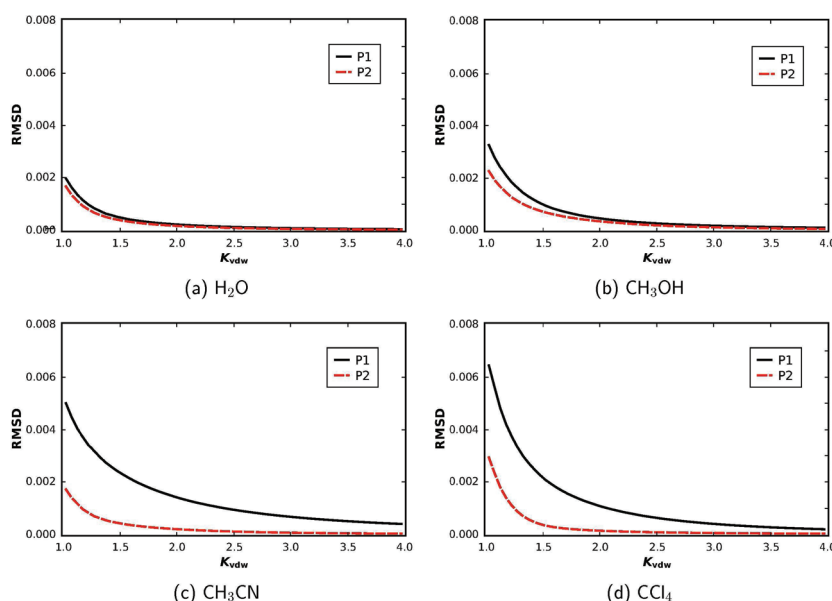


Figure 2. Plots of the RMSD in au between the induced electrostatic potentials created by induced dipole moments and a QM reference with respect to distance from the molecular surface defined by interlocking atomic van der Waals spheres.

Table 1. Excitation Energies (E^{exc}) and Shifts (Δ) for Acrolein in Water with PE-CAM-B3LYP/aug-cc-pVDZ and Different Numbers of Solvent Molecules (#QM-solv) Next to the Double Bond (C=C) or the Oxygen Atom (O) Included in the QM Region^a

force field	#QM-solv (C=C)	#QM-solv (O)	E^{exc}		E^{exc}	
			($n-\pi^*$)	Δ^b	($\pi-\pi^*$)	Δ^b
TIP3P ^c	0	0	4.04	0.26	6.21	-0.20
M2P2	0	0	4.11	0.33	6.05	-0.36
M2P2	2	0	4.11	0.33	5.96	-0.45
M2P2	4	0	4.11	0.33	5.93	-0.48
M2P2	6	0	4.10	0.32	5.92	-0.49
M2P2	0	2	4.07	0.29	6.05	-0.36
M2P2	2	2	4.08	0.30	5.97	-0.44
M2P2	4	2	4.07	0.29	5.94	-0.47
M2P2	8	6	4.04	0.26	5.95	-0.46
exp ^c			3.94	0.25	5.90	-0.52

^a Values are given in eV. All averaged excitation energies have a SEM of ± 0.01 . ^b Gas phase: CAM-B3LYP/aug-cc-pVDZ//B3LYP/aug-cc-pVTZ. ^c Taken from ref 55.

to reproduce higher order multipoles. Thus, only a marginally better description than with the M0 potential is achieved.

The quality of the polarizabilities is tested through the induced electrostatic potential created by induced dipoles generated by a static electric field. In general, we observe that the anisotropic polarizabilities, i.e., P2, give a better reproduction of the QM electrostatic potential than the isotropic polarizabilities, i.e., P1. However, for water and partly also methanol, the difference is less pronounced due to the overall lower magnitude of the polarizabilities of these molecules. The opposite is true in the case of acetonitrile and carbon tetrachloride, where we observe that the anisotropic polarizabilities perform much better—especially at short distances.

4.2. Acrolein in Water. A previous QM/MM study on acrolein solvated in water⁵⁵ revealed that even a point charge-and-polarizable potential (equivalent to M*P1 in our notation) was not sufficient to describe the solvatochromism of the two lowest excited states. Especially the $\pi-\pi^*$ -transition suffered from an inadequate description of the environment, and the inclusion of 14 water molecules into the QM region was needed to achieve converged results. It has already been shown that the new PE-DFT approach remedies these shortcomings if higher multipoles and polarizabilities are taken into account (e.g., a M2P2 potential), and very good agreement with experimental results was obtained.²³ Here, we repeat the investigation of including solvent molecules into the QM region, now with the more advanced M2P2 potential, to investigate the convergence with respect to a full QM treatment. We use the same 120 configurations as in ref 55. Results are given in Table 1.

From this analysis, we first note that we approach the same limit for the excitation energies with an increasing number of QM water: 4.05 eV vs 4.04 eV (this study) for the $n-\pi^*$ state and 5.97 eV vs 5.95 eV (this study) for the $\pi-\pi^*$ state, although the previous study used the TIP3P water potential⁵⁶ for the MM region. This shows that we indeed approach the limit of a full QM treatment. Of course, 14 water molecules are not sufficient to describe bulk effects (and therefore the MM region cannot be neglected), but the numerical experiment shows that the long-range Coulomb effects can be modeled using simpler solvent potentials. This is consistent with the analysis of the solvent potentials in the previous section. Furthermore, we observe that the PE(M2P2)-CAM-B3LYP result without any QM treated solvent molecules is already within 0.1 eV of this limit, which halves the deviation for the $\pi-\pi^*$ excitation in comparison with the point charge only embedding. It should also be pointed out that, with the M2P2 potential, only two additional water molecules in the QM region are necessary to reproduce the limit of 14 water molecules. Depending on the excited state, these molecules have to be those closest to the oxygen atom or to the double bond. No significant additional improvement is obtained

Table 2. Excitation Energies for Acetone in Water with Respect to Different Force Fields for PE-QM Calculations with CAM-B3LYP (CB3) and CCSD (CC) and Different Numbers of Solvent Molecules (#QM-solv) Included in the QM Region^a

solv. model	#QM-solv	CB3		CC	
		$E^{\text{exc}}(n-\pi^*)$	Δ^b	$E^{\text{exc}}(n-\pi^*)$	Δ^c
PCM	0	4.59	0.08		
M*P0	0	4.60 ± 0.01	0.09	4.66 ± 0.01 ^d	0.08
M2P0	0	4.63 ± 0.02 ^d	0.12	4.69 ± 0.02 ^d	0.11
M*P2	0	4.69 ± 0.02 ^d	0.18	4.75 ± 0.02 ^d	0.17
M2P2	0	4.75 ± 0.02	0.24	4.80 ± 0.03 ^d	0.22
M3P2	0	4.74 ± 0.01	0.23		
M2P2	1	4.70 ± 0.01	0.19		
M2P2	2	4.68 ± 0.01	0.17		
exp ^e		4.68	0.22		

^a The aug-cc-pVDZ basis set was used. Excitation energies ($E^{\text{exc}}(n-\pi^*)$) and shifts (Δ) are given in eV and are averaged over 120 configurations, except where it is noted otherwise. ^b Gas phase: CAM-B3LYP/aug-cc-pVDZ//B3LYP/aug-cc-pVTZ. ^c Gas phase: CCSD/aug-cc-pVDZ//B3LYP/aug-cc-pVTZ. ^d Based on 30 configurations. ^e Taken from ref 57.

Table 3. Excitation Energies for Acetone in Methanol with Respect to Different Force Fields for PE-QM Calculations with CAM-B3LYP (CB3) and CCSD (CC) and Different Numbers of Solvent Molecules (#QM-solv) Included in the QM Region^a

solv. model	#QM-solv	CB3		CC	
		$E^{\text{exc}}(n-\pi^*)$	Δ^b	$E^{\text{exc}}(n-\pi^*)$	Δ^c
PCM	0	4.59	0.08		
M*P0	0	4.55 ± 0.00	0.04	4.60 ± 0.01 ^d	0.02
M2P0	0	4.54 ± 0.01 ^d	0.03	4.60 ± 0.01 ^d	0.02
M*P2	0	4.61 ± 0.02 ^d	0.10	4.66 ± 0.02 ^d	0.08
M2P2	0	4.64 ± 0.01	0.13	4.67 ± 0.02 ^d	0.09
M3P2	0	4.63 ± 0.01	0.12		
M2P2	1	4.60 ± 0.01	0.09		
M2P2	2	4.60 ± 0.01	0.09		
exp ^e		4.58	0.12		

^a The aug-cc-pVDZ basis set was used. Excitation energies ($E^{\text{exc}}(n-\pi^*)$) and shifts (Δ) are given in eV and are averaged over 120 configurations, except where it is noted otherwise. ^b Gas phase: CAM-B3LYP/aug-cc-pVDZ//B3LYP/aug-cc-pVTZ. ^c Gas phase: CCSD/aug-cc-pVDZ//B3LYP/aug-cc-pVTZ. ^d Based on 30 configurations. ^e Taken from ref 57.

by the successive inclusion of more molecules. This is in contrast to the TIP3P case where a monotonic convergence with an increasing QM region was observed for the $\pi-\pi^*$ state.⁵⁵ In passing, we note the good agreement of PE-CAM-B3LYP results with experimental results for the excitation energies as well as for the solvatochromic shifts.

4.3. Solvatochromism in Acetone. Encouraged by the good results of the PE model for water as the solvent, we investigated its applicability to other solvents. For this purpose, we examined the lowest excitation of acetone in water, methanol, acetonitrile, and carbon tetrachloride following the same procedure as outlined before. The data are extended by the

Table 4. Excitation Energies for Acetone in Acetonitrile with Respect to Different Force Fields for PE-QM Calculations with CAM-B3LYP (CB3) and CCSD (CC) and Different Numbers of Solvent Molecules (#QM-solv) Included in the QM Region^a

solv. model	#QM-solv	CB3		CC	
		$E^{\text{exc}}(n-\pi^*)$	Δ^b	$E^{\text{exc}}(n-\pi^*)$	Δ^c
PCM	0	4.59	0.08		
M*P0	0	4.52 ± 0.00	0.01	4.59 ± 0.01 ^d	0.01
M2P0	0	4.52 ± 0.01 ^d	0.01	4.58 ± 0.01 ^d	0.00
M*P2	0	4.55 ± 0.01 ^d	0.04	4.61 ± 0.01 ^d	0.03
M2P2	0	4.55 ± 0.00	0.04	4.60 ± 0.01 ^d	0.02
M3P2	0	4.54 ± 0.00	0.03		
M2P2	1	4.54 ± 0.00	0.03		
M2P2	2	4.53 ± 0.00	0.02		
exp ^e		4.52	0.06		

^a The aug-cc-pVDZ basis set was used. Excitation energies ($E^{\text{exc}}(n-\pi^*)$) and shifts (Δ) are given in eV and are averaged over 120 configurations, except where it is noted otherwise. ^b Gas phase: CAM-B3LYP/aug-cc-pVDZ//B3LYP/aug-cc-pVTZ. ^c Gas phase: CCSD/aug-cc-pVDZ//B3LYP/aug-cc-pVTZ. ^d Based on 30 configurations. ^e Taken from ref 57.

Table 5. Excitation Energies for Acetone in Carbon Tetrachloride with Respect to Different Force Fields for PE-QM Calculations with CAM-B3LYP (CB3) and CCSD (CC) and Different Numbers of Solvent Molecules (#QM-solv) Included in the QM Region^a

solv. model	#QM-solv	CB3		CC	
		$E^{\text{exc}}(n-\pi^*)$	Δ^b	$E^{\text{exc}}(n-\pi^*)$	Δ^c
PCM	0	4.52	0.01		
M*P0	0	4.46 ± 0.00	-0.05	4.53 ± 0.00 ^d	-0.05
M2P0	0	4.45 ± 0.00 ^d	-0.06	4.52 ± 0.00 ^d	-0.06
M*P2	0	4.48 ± 0.00 ^d	-0.03	4.55 ± 0.00 ^d	-0.04
M2P2	0	4.47 ± 0.00	-0.04	4.54 ± 0.00 ^d	-0.05
M3P2	0	4.48 ± 0.00	-0.03		
M2P2	1	4.47 ± 0.00	-0.04		
M2P2	2	4.46 ± 0.00	-0.05		
exp ^e		4.43	-0.03		

^a The aug-cc-pVDZ basis set was used. Excitation energies ($E^{\text{exc}}(n-\pi^*)$) and shifts (Δ) are given in eV and are averaged over 120 configurations, except where it is noted otherwise. ^b Gas phase: CAM-B3LYP/aug-cc-pVDZ//B3LYP/aug-cc-pVTZ. ^c Gas phase: CCSD/aug-cc-pVDZ//B3LYP/aug-cc-pVTZ. ^d Based on 30 configurations. ^e Taken from ref 57.

inclusion of octopole moments in the solvent potentials and the use of PE-CCSD.

The results are given in Tables 2–5. Recent experimental⁵⁷ and calculated PCM-DFT values are also listed for comparison. The general trends are solvent-independent. Good results with respect to experimental results are achieved using PE-CAM-B3LYP at the M2P2 level with deviations for excitation energies below 0.1 eV and for solvatochromic shifts of only 0.02 eV or less. Thereby, the largest blue shift found in water (experimental, 0.22 eV; M2P2, 0.24 eV) and the red shift in CCl₄ (experimental, -0.03 eV; M2P2, -0.04 eV) are predicted correctly. The application of multipoles of higher order than quadrupole

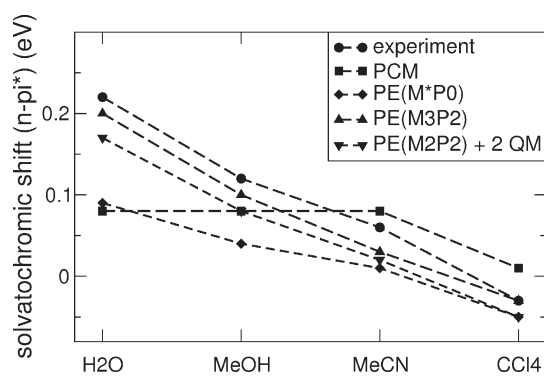


Figure 3. Solvatochromic shifts for the first $n-\pi^*$ electronic excitation energy of acetone. Theoretical results are based on CAM-B3LYP/aug-cc-pVDZ computations with the specified environment model (PCM or PE).

moments in the electrostatic potential has no significant effect. The inclusion of solvent molecules in the QM region does not increase the quality of the $n-\pi^*$ state description. The PE-CAM-B3LYP results are in close correspondence with the PE-CCSD results, which yield slightly higher excitation energies (about 0.05 eV), but the shifts are almost identical.

Finally, a remark about the application of PCM-TDDFT is needed. One can argue that the results for the excitation energies are in an acceptable agreement with experimental results (all within 0.1 eV). However, relative shifts due to the solvent are as important as absolute excitation energies when solvatochromism is considered. In Figure 3, the shifts in different solvents obtained from the experiment as well as from some theoretical methods are shown. Clearly, PCM-TDDFT cannot reproduce this trend and is therefore not sufficient to describe relative shifts between different solvents. As expected, the deviations are largest for water and methanol where hydrogen bonds play a crucial role. These effects are not described at a sufficient level of theory by the PCM treatment. PCM also does not predict a red shift in CCl_4 . For a more extensive discussion about the short-comings of (standard) continuum solvation models to describe the relative solvation shifts in acetone, see ref 13.

The trend of the solvatochromic shifts is improved by using a discrete solvation description, as can be seen from the PE(M*P0) results, but still deviates from what is found in the experiment. When polarization and higher multipoles are taken into account, very good agreement with experimental findings is achieved.

4.4. Solvatochromism in Pyridazine. To present an alternative case, we also calculated the solvatochromic shift of pyridazine in water, methanol, and acetonitrile. We limited the study to the M*P0, M*P1, and M2P2 potentials. For the M2P2 potential, up to two solvent molecules have been included in the QM region. The shifts are shown in Figure 4. Additionally, all excitation energies are listed in the Supporting Information.

The findings are very similar to those from the study of acetone. Convergence of the results is reached at the M2P2 level, and the computed shifts are in good agreement with experimental measurements (within 0.07 eV), while the deviation at the M*P0 level can be as high as 0.17 eV. The absolute excitation energies are about 0.1 eV too high compared to experimental results (see the Supporting Information), but some level of disagreement of the absolute excitation energy is to be expected given the after-all limited accuracy of the QM method, i.e., the CAM-B3LYP functional, as well as the neglect of nuclear

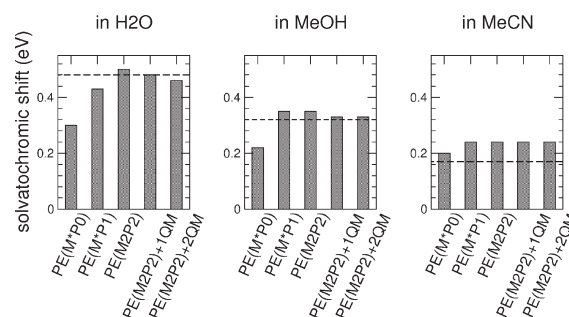


Figure 4. Solvatochromic shifts for the first $n-\pi^*$ electronic excitation energy of pyridazine in different solvents. Theoretical results are based on CAM-B3LYP/aug-cc-pVDZ computations and the indicated polarizable embedding potentials. Experimental results (taken from ref 58) are given as a dashed line.

relaxation and vibrational effects on the spectral profile. For acetonitrile, the difference between the M*P0 and M2P2 levels is least pronounced. Note that the agreement between experimental results and PE-CAM-B3LYP using the M*P0 potential is marginally better than that with the M2P2 potential, but the opposite is true with respect to the inclusion of solvent molecules in the QM region. However, effects are small, and the correct relative trend compared to water and methanol is obtained. In a previous study by one of us regarding the solvent shift in pyridazine in aqueous solution based on the SPCpol potential, which is similar to M*P1, it was found that the QM treatment of the two nearest water molecules is necessary to achieve a satisfying description of the solvent effects.⁵⁹ However, according to our observations, this is not necessary when using the M2P2 potential. Again, this finding is transferable to the other solvent potentials.

5. SUMMARY AND CONCLUSION

The polarizable embedding approach to the QM/MM description of solvent effects has been used for the first time to systematically explore the solvatochromic effects from solvents other than water, namely, methanol, acetonitrile, and carbon tetrachloride. For all solvents, good agreement with experimental results is observed for the solvatochromic shifts. Experimental excitation energies are reproduced with deviations of less than 0.1 eV, and shifts with respect to vacuum are in even better agreement, as they fully reproduce the solvent-dependent trends. The PE method continues to be an accurate and efficient tool also for cases where continuum models are likely to fail, e.g., when hydrogen bonding is important. The level of agreement over such diverse solvents is truly nontrivial and even better than we expected. For fairness, in relation to the much better results of the polarizable embedding QM/MM type of approach compared to the continuum approaches, it should be recalled that the continuum approach requires only a single calculation, while the QM/MM approach requires many calculations and an MD simulation for generating the structures for such calculations.

In this paper, we have shown that the inclusion of higher order multipole moments in the solvent potentials leads to faster convergence with respect to a full QM treatment. Most often, up to quadrupole moments and anisotropic polarizabilities are sufficient to cover environmental effects. The successful application of our approach to new solvents is also encouraging in relation to biomolecular studies. Most likely, the potentials can

be retrieved from ab initio calculations as demonstrated here for several solvents. First results confirm this conclusion.^{24,33,60} In addition, we have shown the great potential of PE-CC to benchmark PE models on the basis of a more simplified QM description.

A good quality potential and explicit consideration of polarization effects are necessities for good results, as has been demonstrated. To improve the PE approach further, other issues of the procedure need to be addressed. The treatment of flexible molecules must be taken into account when larger (and less rigid) molecules in solution are to be considered. Additionally, the QM/MM interface might benefit from a description of Pauli repulsion and of dispersion effects at the electronic level. The documented accuracy and possibility to combine the PE model with CC approaches allows not only a systematic improvement of the QM description but that of the MM region as well. Therefore, the PE method qualifies as a reliable benchmark reference against which other methods can be assessed.

ASSOCIATED CONTENT

S Supporting Information. All PE potentials generated for this study are listed along with the theoretical excitation energies for pyridanzine in different solvents based on different levels of theory. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: tschwabe@chem.au.dk

ACKNOWLEDGMENT

T.S. acknowledges support through a Feodor-Lynen fellowship from the Alexander von Humboldt Foundation. O.C. and J.K. acknowledge support from the Danish Center for Scientific Computing (DCSC), the Danish national research foundation, and the Lundbeck Foundation. O.C. thanks EUROHORCs for a EURYI award.

REFERENCES

- (1) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123*, 124107.
- (2) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397–3406.
- (3) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2010**, *6*, 107–126.
- (4) Zhao, Y.; Truhlar, D. G. *Chem. Phys. Lett.* **2011**, *502*, 1–3.
- (5) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, CA, 1996; Chapter 7, pp 151–176.
- (6) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (7) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799–805.
- (8) Klamt, A.; Jonas, V. *J. Chem. Phys.* **1996**, *105*, 9972–9981.
- (9) Cancés, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032–3031.
- (10) Mennucci, B.; Cancés, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506–10517.
- (11) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3094.
- (12) Steindal, A. H.; Ruud, K.; Frediani, L.; Aidas, K.; Kongsted, J. *J. Phys. Chem. B* **2011**, *115*, 3027–3037.
- (13) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 2829–2844.
- (14) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.
- (15) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- (16) Ohn, A.; Karlstrom, G. *Theor. Chem. Acc.* **2007**, *117*, 441–449.
- (17) Parac, M.; Doerr, M.; Marian, C. M.; Thiel, W. *J. Comput. Chem.* **2010**, *31*, 90–106.
- (18) Murugan, N. A.; Jha, P. C.; Rinkevicius, Z.; Ruud, K.; Ågren, H. *J. Chem. Phys.* **2010**, *132*, 234508.
- (19) Murugan, N. A.; Kongsted, J.; Rinkevicius, Z.; Ågren, H. *Phys. Chem. Chem. Phys.* **2011**, *13*, 1290–1292.
- (20) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *J. Phys. Chem. A* **2003**, *107*, 2578–2588.
- (21) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *J. Chem. Phys.* **2003**, *118*, 1620–1633.
- (22) Nielsen, C. B.; Christiansen, O.; Mikkelsen, K. V.; Kongsted, J. *J. Chem. Phys.* **2007**, *126*, 154112.
- (23) Olsen, J. M.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.
- (24) Sneskov, K.; Schwabe, T.; Kongsted, J.; Christiansen, O. *J. Chem. Phys.* **2011**, *134*, 104108.
- (25) Thompson, M. A. *J. Chem. Phys.* **1996**, *100*, 14492–14507.
- (26) Jensen, L.; van Duijnen, P. T.; Snijders, J. G. *J. Chem. Phys.* **2003**, *118*, 514–521.
- (27) Jensen, L.; van Duijnen, P. T.; Snijders, J. G. *J. Chem. Phys.* **2003**, *119*, 3800–3809.
- (28) Wesolowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 8050–8053.
- (29) Neugebauer, J. *J. Chem. Phys.* **2007**, *126*, 134166.
- (30) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968–1986.
- (31) Gordon, M. S.; Jensen, J. H.; Freitag, M. A.; Bandyopadhyay, P.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293–307.
- (32) Arorar, P.; Slipchenko, L. V.; Webb, S. P.; Gordon, A. D. M. S. *J. Phys. Chem. A* **2010**, *114*, 6742–6750.
- (33) Söderhjelm, P.; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U. *J. Chem. Theory Comput.* **2008**, *5*, 649–658.
- (34) Mata, R. A. *Mol. Phys.* **2010**, *108*, 381–392.
- (35) Jacob, C. R.; Neugebauer, J.; Jensen, L.; Visscher, L. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2349–2359.
- (36) Coutinho, K.; Canuto, S. *Adv. Quantum Chem.* **1997**, *28*, 89–105.
- (37) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (38) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (39) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (40) Kendall, R. A.; T. H. Dunning, J.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (41) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. *J. Gaussian 09*, Revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
- (42) Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **2004**, *121*, 4494–4500.

(43) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.; Neogrády, P.; Bondo Pedersen, T. B.; Pitoňák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Velyazov, V.; Lindh, R. *J. Comput. Chem.* **2010**, *31*, 224–247.

(44) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361–373.

(45) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(46) Ponder, J. W. *TINKER: Software Tools for Molecular Design*, version 5.1; Washington University School of Medicine: Saint Louis, MO, 2010.

(47) Ahlström, P.; Wallqvist, A.; Engström, S.; Jönsson, B. *Mol. Phys.* **1989**, *68*, 563–581.

(48) Linse, P. *MOLSIM 3.3.0*; Lund University: Lund, Sweden, 2001.

(49) Yannai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51–57.

(50) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; J. Wiley: New York, 2000; Chapter 13, pp 648.723.

(51) Aidas, K. *Whirlpool*, version 1.0; University of Copenhagen: Copenhagen, 2010.

(52) DALTON, Release 2.0. See <http://www.kjemi.uio.no/software/dalton/dalton.html> (accessed Jan 1, 2011).

(53) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451.

(54) Batsanov, S. S. *Inorg. Mater.* **2001**, *37*, 871–885.

(55) Aidas, K.; Møgelhøj, A.; Nilsson, E. J. K.; Johnson, M. S.; Mikkelsen, K. V.; Christiansen, O.; Söderhjelm, P.; Kongsted, J. *J. Chem. Phys.* **2008**, *128*, 194503.

(56) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(57) Renge, I. *J. Phys. Chem. A* **2009**, *113*, 10678–10686.

(58) Baba, H.; Goodman, L.; Valenti, P. C. *J. Am. Chem. Soc.* **1966**, *88*, 5410–5415.

(59) Kongsted, J.; Mennucci, B. *J. Chem. Phys.* **2007**, *111*, 9890–9900.

(60) Rocha-Rinza, T.; Sneskov, K.; Christiansen, O.; Ryde, U.; Kongsted, J. *J. Phys. Chem. Chem. Phys.* **2011**, *13*, 1585–1589.

Calibration of the DFT/GGA+U Method for Determination of Reduction Energies for Transition and Rare Earth Metal Oxides of Ti, V, Mo, and Ce

Suzanne Lutfalla, Vladimir Shapovalov, and Alexis T. Bell*

Department of Chemical and Biomolecular Engineering, University of California-Berkeley, Berkeley, California 94720-1462, United States

ABSTRACT: GGA+U calculation were performed for oxides of Ti, V, Mo, and Ce with the objective of establishing the best value of the parameter U_{eff} to use in order to match the calculated reduction and oxidation energies of each oxide with experimental values. In each case, the reaction involved the hydrogen reduction of an oxide to its next lower oxide and the formation of water. Our calculations show that the optimal value of U_{eff} required to match calculated and experimental values of the reaction energy are significantly different from those reported in the literature based on matching lattice parameters or electronic properties and that the use of these values of U_{eff} can result in errors in the calculated redox energies of over 100 kJ/mol. We also found that, when an element exhibits more than two oxidation states, the energy of redox reactions between different pairs of these states are described by slightly different values of U_{eff} .

INTRODUCTION

The oxides of transition and rare earth metals, such as Ti, V, Cu, La, and Ce, are often used as catalysts for industrially important reactions.^{1–6} Consequently, quantum chemical calculations for such elements are of much interest, and density functional theory (DFT) is one of the tools commonly applied to such systems. The catalytic properties of these materials are attributed to their reducibility,^{7–9} where lower oxidation states correspond to occupied d and f orbitals. At the same time, d and f electrons also present difficulties for DFT calculations, because this method tends to delocalize d and f electrons excessively.^{10,11} These difficulties affect a broad spectrum of oxide properties, including crystal lattice parameters, conductivity, and energies of oxide reduction and oxidation.

The difficulty in obtaining accurate property predictions for transition metal oxides from DFT calculations has been recognized for some time. Despite attempts to use hybrid functionals and dynamical mean-field theory to treat the problem, DFT with generalized gradient approximation (GGA) functionals remains an economical choice, and therefore, corrections directed at specific drawbacks of the method are introduced.

It is generally understood that the main source of error in DFT for d and f electrons is their correlated nature. A commonly used ad hoc method for improving the description of d and f electrons is the DFT+U method, in which an “on-site” potential is added to introduce intra-atomic interactions between the strongly correlated electrons. Most recent articles have used the potential proposed by Dudarev et al.,¹² which has the form

$$E = \frac{(U - J)}{2} \sum_{\sigma} (n_{m,\sigma} - n_{m,\sigma}^2) \quad (1)$$

where U and J are the effective Coulomb and exchange parameters, respectively, and n is the occupation number of a d orbital number m with spin σ . U and J can, in principle, be computed

from first principles. In reality, however, the theoretical values of U and J give poor results, and therefore, these parameters are adjusted by fitting to experimental data, such as the oxide band gap or the lattice parameters. Because eq 1 depends on only the difference, $U - J$, can be replaced with one variable $U_{\text{eff}} = U - J$ for the sake of brevity.

The value of U_{eff} is element-specific, and at least one study has suggested that it is transferable between different oxidation states of a given element.¹³ U_{eff} is usually determined empirically, to fit some specific physical property, most often the crystal lattice parameters or the band gap between the occupied and unoccupied states.^{7,14,15} The principal problem with this approach is that no two properties are described well by the same value of U_{eff} , and therefore, a value is picked that minimizes the average error in several properties.

Application of the on-site interaction term to transition metal oxides has been recognized as necessary because of its strong influence on the orbital energies of the occupied d and f states and, as a consequence, on the formation energy of oxygen vacancies formed during reactions that proceed through a Mars–van Krevelen mechanism. For the purposes of catalytic chemistry, we are interested in values of U_{eff} that accurately describe redox reactions. Several authors have shown that the oxygen vacancy formation energy depends strongly on the value of U_{eff} .^{10,16} A notable problem with this approach is that experimental formation energies of oxygen vacancies are difficult to evaluate, and consequently, different authors have reported different values of U_{eff} for the same element.^{7,8,10,13,17–20} Chemisorption energies for probe molecules can serve as a more reliable test of the value of U_{eff} . For example, CO can react with an oxide to form CO₂, which remains adsorbed as a surface carbonate. The net effect of this reaction is that the metal oxide is

Received: March 24, 2011

Published: June 07, 2011

Table 1. Symmetry and Magnetic Properties of the Oxides

oxide	symmetry	magnetic arrangement at STP ^a
TiO ₂	<i>P4₂/mmn</i>	diamagnetic
Ti ₂ O ₃	<i>R3̄</i>	diamagnetic
V ₂ O ₅	<i>Pmmn</i>	diamagnetic
VO ₂	<i>P2₁/c</i>	diamagnetic
V ₂ O ₃	<i>R3̄</i>	paramagnetic
MoO ₃	<i>Pbnm</i>	diamagnetic
MoO ₂	<i>P2₁/c</i>	paramagnetic
CeO ₂	<i>Fm3̄m</i>	diamagnetic
Ce ₂ O ₃	<i>P3m1</i>	antiferromagnetic

^a STP = standard temperature and pressure.

reduced. As the d band of the oxide becomes partially filled, the value of U_{eff} will have an effect on the computed adsorption energy. Huang and Fabris²¹ analyzed the energetics of CO adsorption on CeO₂ as a function of U_{eff} , providing evidence that the value of this parameter presently used in the literature ($U_{\text{eff}} > 4$ eV) can lead to severe overestimation of the binding energy of CO to ceria when surface reduction is involved, whereas the values $U_{\text{eff}} = 2-3$ eV are more appropriate. Still, experimental adsorption energies tend to have a large range, depending on the condition of the surface and the experimental technique used to measure them. Therefore, the effect of the value of U_{eff} on the redox reaction energies has not been settled.

In this work, we consider the redox pairs TiO₂/Ti₂O₃, V₂O₅/VO₂/V₂O₃, MoO₃/MoO₂, and CeO₂/Ce₂O₃. We used bulk oxides in order to avoid the need to determine surface structures. To avoid the difficulties associated with the description of O₂ by DFT,¹³ we chose H₂ as the reducing agent and gas-phase H₂O as the oxidizing agent. The reduction energy of each oxide was compared with values determined from the experimentally obtained energies of formation. We show that values of U_{eff} obtained by fitting the lattice parameters or the band gap can lead to significant errors in the reduction energy of the transition or rare earth oxides.

COMPUTATIONAL APPROACH

All calculations were performed using the VASP 4.6 package.^{22,23} We used the revised Perdew–Burke–Ernzerhof (PBE) functional²⁴ and the projector-augmented wave (PAW) potentials.^{25,26} The plane-wave cutoff was set to 500 eV. For integration over the Brillouin zone, Γ -centered sets of k points were tested to achieve convergence better than 1 meV/atom. The resulting sets are $17 \times 17 \times 7$ for TiO₂, $5 \times 5 \times 5$ for Ti₂O₃, $3 \times 9 \times 9$ for V₂O₅, $5 \times 5 \times 11$ for VO₂, $7 \times 7 \times 7$ for V₂O₃, $6 \times 1 \times 6$ for MoO₃, $13 \times 13 \times 13$ for MoO₂, $11 \times 11 \times 7$ for Ce₂O₃, and $7 \times 7 \times 7$ for CeO₂. Integration was performed using the tetrahedron method with Blöchl corrections.²⁷ The crystal symmetry and magnetic properties of each oxide are listed in Table 1. The corundum structures of Ti₂O₃ and V₂O₃ have two choices of the unit cell in use: the primitive rhombohedral unit cell with compositions V₄O₆ and Ti₄O₆ and the hexagonal unit cell with compositions V₁₂O₁₈ and Ti₁₂O₁₈. The transformation of the lattice vectors and coordinates between the rhombohedral and hexagonal unit cells was described by Cousins.²⁸ We used the rhombohedral cell for our calculations, but the lattice parameters discussed in the text correspond to the hexagonal lattice, as it is the one more commonly discussed. The oxide structures were fully optimized for each value of U_{eff} tested. All calculations were

Table 2. Experimental Enthalpy of Formation and Enthalpy Change between 0 and 298.15 K

	$\Delta_f H_{298.15}^{\circ}$ (kJ/mol)	$H_{298.15}^{\circ} - H_0^{\circ}$ (kJ/mol)
H ₂	0	8 ³⁵
H ₂ O	-242 ³⁶	13 ³⁶
TiO ₂	-944 ³⁷	9 ³⁷
Ti ₂ O ₃	-1521 ³⁸	14 ³⁸
V ₂ O ₅	-1551 ³⁹	21 ³⁹
VO ₂	-714 ³⁷	N/A
V ₂ O ₃	-1217 ³⁸	17 ³⁸
MoO ₃	-745 ⁴⁰	13 ⁴⁰
MoO ₂	-589 ⁴⁰	8 ⁴⁰
CeO ₂	-1090 ⁴¹	10 ⁴¹
Ce ₂ O ₃	-1800 ⁴¹	21 ⁴¹

Table 3. Dependence of Lattice Parameters (in Å) on U_{eff}

	experiment	U_{eff} (eV)			
		0.0	2.0	5.0	8.0
TiO ₂ ⁴²					
<i>a</i>	4.594	4.687	4.701	4.721	4.742
<i>c</i>	2.959	2.981	3.011	3.055	3.096
Ti ₂ O ₃ ⁴³					
<i>a</i>	5.16	5.20	5.24	5.31	5.37
<i>c</i>	13.61	13.78	13.88	14.06	14.24
V ₂ O ₅ ⁴⁴					
<i>a</i>	11.512	11.588	11.583	11.917	11.978
<i>b</i>	3.564	3.597	3.624	4.526	4.549
<i>c</i>	4.368	5.339	5.382	3.685	3.704
VO ₂ ^{a,45}					
<i>a</i>	5.743	5.861	5.696	5.707	5.761
<i>b</i>	4.517	4.609	4.665	4.682	4.701
<i>c</i>	5.375	5.485	5.506	5.516	5.382
V ₂ O ₃ ⁴⁶					
<i>a</i>	4.95	5.03	5.13	5.16	5.29
<i>c</i>	14.00	14.30	14.20	14.58	14.29
MoO ₃ ⁴⁷					
<i>a</i>	3.963	4.047	4.020	3.967	3.904
<i>b</i>	13.855	17.184	17.180	17.140	17.203
<i>c</i>	3.696	3.682	3.704	3.747	3.793
MoO ₂ ⁴⁸					
<i>a</i>	5.611	5.657	5.671	5.699	5.730
<i>b</i>	4.856	4.896	4.908	4.933	4.959
<i>c</i>	5.623	5.675	5.689	5.717	5.747
CeO ₂ ⁴⁹					
<i>a</i>	5.411	5.499	5.504 ^b	5.522	5.536
Ce ₂ O ₃ ⁵⁰					
<i>a</i>	3.891	3.88	3.92	3.96	3.97
<i>c</i>	6.059	6.04	6.11	6.16	6.19

^a $\theta = 122.61^{\circ}$, 122.61° , 121.961° , 121.808° , and 121.841° for experiment and $U_{\text{eff}} = 0.0, 2.0, 5.0,$ and 8.0 eV, respectively. ^b Value computed for $U_{\text{eff}} = 1.0$ eV.

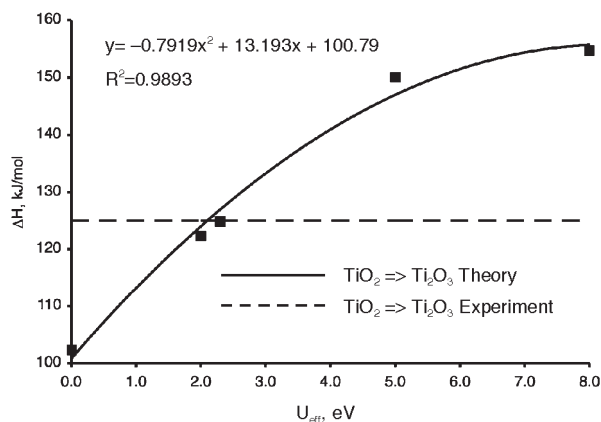


Figure 1. Enthalpy of reduction (kJ/mol) of TiO_2 to Ti_2O_3 versus on-site Coulomb repulsion, U_{eff} (eV).

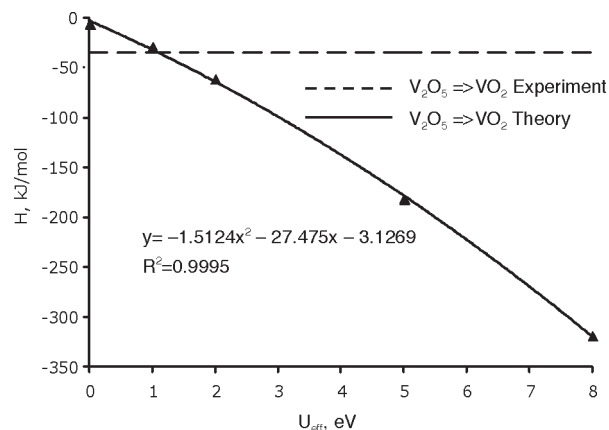


Figure 3. Enthalpy of reduction (kJ/mol) of VO_2 to V_2O_5 versus on-site Coulomb repulsion, U_{eff} (eV).

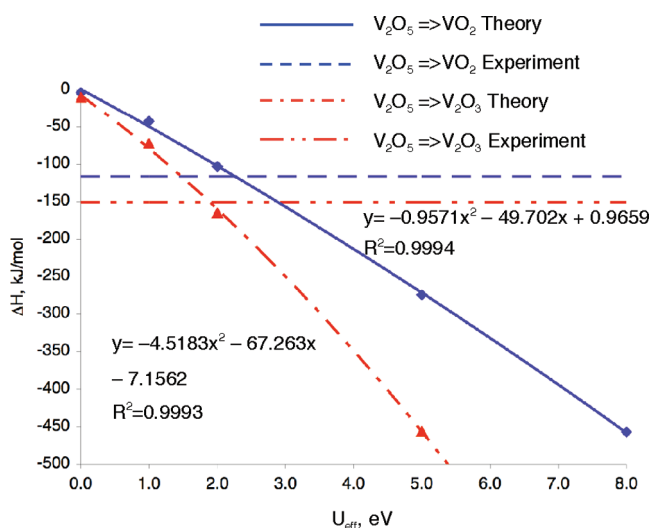


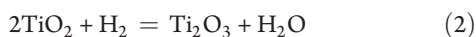
Figure 2. Enthalpy of reduction (kJ/mol) of V_2O_5 to VO_2 and V_2O_3 versus on-site Coulomb repulsion, U_{eff} (eV).

initiated with the experimentally known values of the lattice parameters (see Table 3 below). The effect of U_{eff} on the oxide lattice parameters was determined, as well as the redox energy for each of the oxide pairs of interest.

Experimental enthalpies of formation used in this work are listed in Table 2. Because the energy changes determined by our calculations correspond to 0 K, to compare energies of reactions, it is necessary to subtract the enthalpy difference between 0 and 298.15 K. Unfortunately, we were not able to find the change in entropy of formation for VO_2 between 0 and 298.15 K. We note, however, that, for the overall redox reactions, this correction is below 6 kJ/mol, which is comparable to the error inherent in our theoretical method. Therefore, we chose not to make the correction in the enthalpy change with temperature.

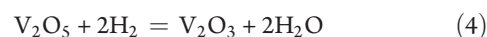
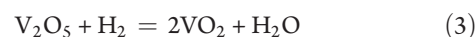
RESULTS AND DISCUSSION

Oxidation Energy. The effect of U_{eff} on the energy of reduction of TiO_2 to Ti_2O_3 according to the reaction

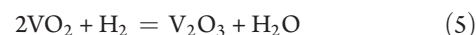


is shown in Figure 1. The experimental energy, 125 kJ/mol, is matched for $U_{\text{eff}} = 2.3$ eV. This value is significantly smaller than that suggested by Morgan and Watson, $U_{\text{eff}} = 4.2$ eV, which was obtained by optimizing the position of the oxygen vacancy states in the electronic spectra of rutile.¹⁴ Our finding is similar to the conclusion of Hu and Metiu in a recent publication.²⁹ They recommended using a U_{eff} value between 2 and 3 eV. The authors used this value of U_{eff} to compare the concentration of oxygen vacancies in rutile and anatase.³⁰ We note, however, that the value of U_{eff} determined by Morgan and Watson overestimates the enthalpy change for reaction 2 by 17 kJ/mol.

Vanadium is the only element for which we considered more than two oxides. Therefore, it is an instructive example of the limitations of the DFT+U methodology. Figures 2 and 3 summarize the dependence on U of the reactions



The two reactions also can be combined to obtain



The experimental reaction enthalpies are -119 kJ/mol for reaction 3, -150 kJ/mol for reaction 4, and -31 kJ/mol for reaction 5. From Figures 2 and 3, one can see that, although the optimal values of the on-site repulsion term are similar for the three reactions, they are not identical. The optimal values of U_{eff} for reactions 3 and 4 are 2.3 and 1.8 eV, respectively. The difference between these two values is in line with that reported by Wang et al.¹³ The optimal value for reaction 5 is $U_{\text{eff}} = 1.1$ eV. Therefore, although the equilibrium between V_2O_5 and V_2O_3 seems to be described reasonably well, VO_2 is artificially destabilized by the same choice of the parameter. By contrast, Scanlon et al. suggested $U_{\text{eff}} = 4.0$ eV based on a comparison of the computed electronic spectra with the experimental ultraviolet photoelectron spectroscopy (UPS) and X-ray photoelectron spectroscopy (XPS) data.⁷ This value results in enthalpies that are lower than those observed experimentally by about 100 kJ/mol for each of the reactions. Figures 2 and 3 also show that the relative energies of the reactions diverge, rather than shifting by a

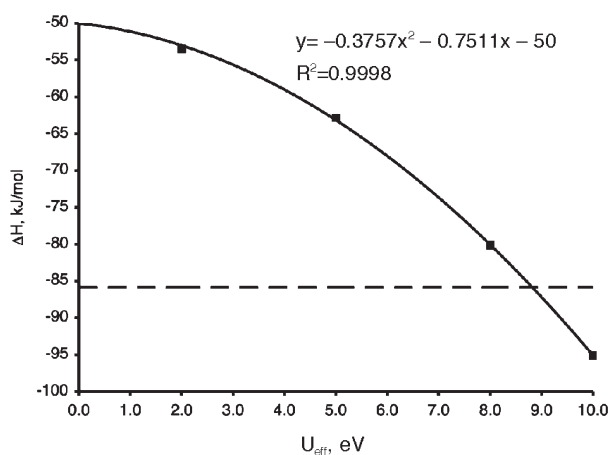
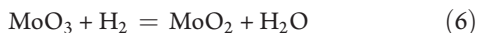


Figure 4. Enthalpy of reduction (kJ/mol) of MoO_3 to MoO_2 versus on-site Coulomb repulsion, U_{eff} (eV).

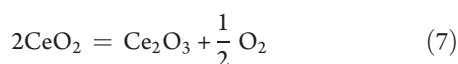
constant value, which makes calculations for metals with multiple oxidation states particularly sensitive to the choice of U_{eff} .

The energy of reduction for the reaction



is 86 kJ/mol and can be matched using a value of $U_{\text{eff}} = 8.6$ eV, as can be seen from Figure 4. Once again, this value differs from those previously reported. Coquet and Willock⁸ used the DFT+U method with the PBE functional to study the formation of oxygen defects on the (010) surface of α - MoO_3 . From comparison of periodic plane-wave and cluster calculations, they arrived at the value of $U_{\text{eff}} = 6.3$ eV. On the other hand, in a study of magnetism in MoO_2 , Wang et al.¹⁷ used a range of values for U_{eff} between 1 and -1 eV, citing weak correlation in Mo.

Cerium dioxide (CeO_2) is used as an oxygen-storage material for the three-way control of automotive emissions. Consequently, a number of authors have examined the issue of parametrization of the DFT+U method. Fabris et al.¹⁰ suggested that the optimal value of U_{eff} required to match the energy for the reaction



is around 1 eV for GGA+U with atomic-like orbital projectors. Jiang et al.¹⁸ used GGA+U to compute the effects of oxygen partial pressure on the CeO_2 surfaces. They calibrated U_{eff} to the electronic spectra and arrived at a value of 6.3 eV. Skorodumova et al.¹⁶ have published several articles on cerium oxides, and in a recent publication, they dealt with the choice of the Coulomb repulsion parameter for reaction 7. The raw data suggested an optimal value of U_{eff} for the GGA functional of about 2 eV; however, the authors noted that the binding energy of O_2 used in their calculations was overestimated, and hence, they recommended the value of $U_{\text{eff}} = 5$ eV. Nolan et al.¹⁹ obtained a value of U_{eff} in their study of ceria surfaces based on the degree of delocalization of the *f* electrons. They observed that, for $U_{\text{eff}} < 5$, the electrons were partly delocalized, but at $U_{\text{eff}} = 5$ eV, localization became complete, leading them to conclude that the appropriate value was $U_{\text{eff}} = 5$ eV. This value of U_{eff} was later used for a study of NO_2 adsorption on ceria.³¹ Castleton et al. optimized the value of U_{eff} for electronic spectra and structure and recommended $U_{\text{eff}} = 5.5$ eV for GGA+U, conceding it as a compromise between several properties. Da Silva et al.³² compared the performances

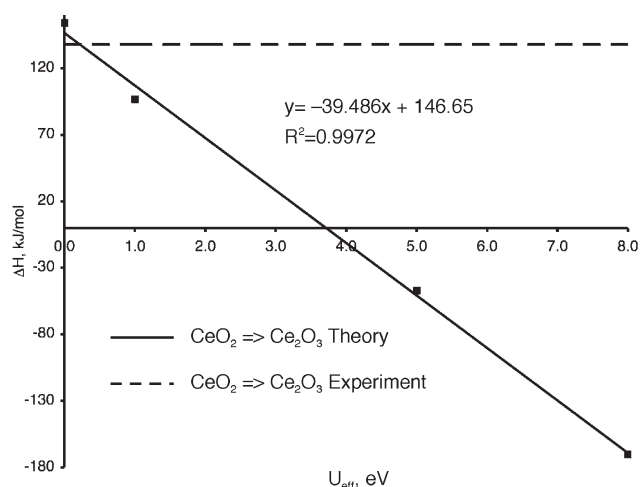
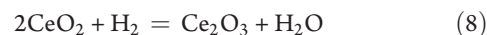


Figure 5. Enthalpy of reduction (kJ/mol) of CeO_3 to Ce_2O_3 versus on-site Coulomb repulsion, U_{eff} (eV).

of the hybrid and GGA functionals, including GGA+U. They noted the large deviation of the oxidation energy with a typical estimation of $U_{\text{eff}} = 4.5$ eV from the experimental values and suggested that the value of 2 eV be used instead. This conclusion coincides with the results of Loschen et al.³³ Huang and Fabris²¹ also suggested a value for U_{eff} of 2–3 eV based on their calculations of the energy for CO adsorption.

The reaction considered in the present study is



As seen in Figure 5, varying the value of U_{eff} from 0 to 6 eV changes the enthalpy of reaction 8 by more than 200 kJ/mol, as well as its sign. The experimental value of 138 kJ/mol is matched with $U_{\text{eff}} = 0.2$ eV. However, if one uses the commonly recommended value of 4.5 eV, the reaction enthalpy becomes -31 kJ/mol, which is almost 170 kJ/mol lower than the experimental value.

Lattice Parameters. The GGA in general produces inaccurate lattice constants, and we would not suggest using these parameters to fit the value of U_{eff} . However, we include a discussion of lattice parameters for two reasons. First, the data are already available from the present calculations. Second, because lattice parameters are sometimes used as one of several parameters to justify the selection of the value of U_{eff} ,^{16,33} it is important to discuss what sort of errors should be expected from a given choice of U_{eff} .

The effect of U_{eff} on the lattice parameters of the oxides investigated in this study is shown in Table 3. In all cases, the lattice parameters are overestimated using the GGA functional, and the extent of overestimation increases as the Coulomb repulsion term grows. Therefore, GGA+U worsens the description of the crystal lattice compared to that obtained with GGA. By contrast, the local density approximation (LDA) functional typically underestimates lattice constants, so the use of LDA+U is a viable means for determining values of U_{eff} required to match calculated and observed lattice parameters.

The single exception to the trend in lattice parameters with U_{eff} is MoO_3 , for which the lattice parameters *a* and *b* generally decrease with increasing values of U_{eff} and the parameter *c* increases. This structure is characterized by well-defined layers perpendicular to the *b* axis. The interactions between the layers are weak and presumably largely of van der Waals in character. Studies by

Coquet and Willock⁸ and Scanlon et al.³⁴ indicated that the effect of such weak binding was the absence of a minimum in the plot of energy versus the length of the *b* lattice parameter. These two studies worked around the problem by freezing the *b* vector at its experimental value and then relaxing the *a* and *c* parameters so as to minimize the total energy. In our study, we were able to locate such a minimum, by frequent updates to the plane wave basis set. The optimized *b* vector was overestimated by a much larger amount than is typical of GGA. Nevertheless, we decided against manually freezing the lattice parameters. We consider that full optimization is more appropriate because, during calculations of reaction or adsorption energies, as a rule, the oxide atoms are fully relaxed. This relaxation releases some of the energy that would be stored in the atomic bonds that would be strained because of the frozen lattice parameters. However, it is not clear how to separate the adsorption or reaction energy from the contribution due to relaxation of the lattice. We believe that, for consistent treatment of MoO₃, all atoms should be allowed to fully relax, even though the lattice constants turn out to be different from those observed experimentally.

CONCLUSIONS

It has been shown that variations in U_{eff} result in significant deviations of the reaction enthalpies from the experimental values. We tried to put this interdependence into the context of a catalytically relevant energy scale. The magnitude of the variation differs from about 5 kJ/mol of enthalpy per 1 eV of U_{eff} for the MoO₃/MoO₂ pair, to about 100 kJ/mol of enthalpy per 1 eV of U_{eff} for the V₂O₅/V₂O₃ pair. The optimal values of U_{eff} corresponding to different pairs of the oxidation states for the same element are close to one another, but not exactly the same. Therefore, to use the DFT+U method, the value of U_{eff} must be adjusted for each reaction. For calculations of the energy of oxygen-atom transfer carried out using the GGA functional, we recommend the following values: $U_{\text{eff}} = 2.3$ eV for Ti, 2.0 eV for V, 8.6 eV for Mo, and 0.2 eV for Ce. Application of the same value of U_{eff} for calculations of different physical properties is questionable. We conclude that values of U_{eff} determined by matching the band gaps or lattice parameters cannot be used to obtain energies for oxide reduction and oxidation that match experimentally observed values.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bell@cchem.berkeley.edu.

ACKNOWLEDGMENT

This work was supported by the Methane Conversion Cooperative funded by BP. Partial support was also provided by the Molecular Graphics and Computation Facility at the College of Chemistry, University of California-Berkeley, under Grant CHE-0233882 from the National Science Foundation. S.L. acknowledges the support of the French Ministry of Research through the master's program at ENS Lyon. We also thank Andrew Getsoian at University of California-Berkeley for his contribution to calculations for MoO₂.

REFERENCES

- (1) Kullgren, J.; Castleton, C. W. M.; Mueller, C.; Ramo, D. M.; Hermansson, K. *J. Chem. Phys.* **2010**, *132*, 054110.
- (2) Hilaire, S.; Wang, X.; Luo, T.; Gorte, R. J.; Wagner, J. *Appl. Catal. A* **2004**, *258*, 271–276.
- (3) Bollinger, M. A.; Vannice, M. A. *Appl. Catal. B* **1996**, *8*, 417–443.
- (4) Lintz, H.-G.; Turek, T. *Appl. Catal. A* **1992**, *85*, 13–25.
- (5) Avgouropoulos, G.; Papavasiliou, J.; Ioannides, T. *Chem. Eng. J.* **2009**, *154*, 274–280.
- (6) Weckhuysen, B. M.; Keller, D. E. *Catal. Today* **2003**, *78*, 25–46.
- (7) Scanlon, D. O.; Walsh, A.; Morgan, B. J.; Watson, G. W. *J. Phys. Chem. C* **2008**, *112*, 9903–9911.
- (8) Coquet, R.; Willock, D. J. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3819–3828.
- (9) Castleton, C. W. M.; Kullgren, J.; Hermansson, K. *J. Chem. Phys.* **2007**, *127*, 244704.
- (10) Fabris, S.; de Gironcoli, S.; Baroni, S.; Vicario, G.; Balducci, G. *Phys. Rev. B* **2005**, *72*, 237102.
- (11) Fabris, S.; de Gironcoli, S.; Baroni, S.; Vicario, G.; Balducci, G. *Phys. Rev. B* **2005**, *71*, 041102.
- (12) Dudarev, S. L.; Botton, G. A.; Savrasov, S. Y.; Humphreys, C. J.; Sutton, A. P. *Phys. Rev. B* **1998**, *57*, 1505–1509.
- (13) Wang, L.; Maxisch, T.; Ceder, G. *Phys. Rev. B* **2006**, *73*, 195107.
- (14) Morgan, B. J.; Watson, G. W. *Surf. Sci.* **2007**, *601*, 5034–5041.
- (15) Laubach, S.; Schmidt, P. C.; Thissen, A.; Fernandez-Madrigal, F. J.; Wu, Q.-H.; Jaegermann, W.; Klemm, M.; Horn, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2564–2576.
- (16) Andersson, D. A.; Simak, S. I.; Johansson, B.; Abrikosov, I. A.; Skorodumova, N. V. *Phys. Rev. B* **2007**, *75*, 035109.
- (17) Wang, F.; Pang, Z.; Lin, L.; Fang, S.; Dai, Y.; Han, S. *Phys. Rev. B* **2010**, *81*, 134407.
- (18) Jiang, Y.; Adams, J. B.; van Schilfgaarde, M. J. *Chem. Phys.* **2005**, *123*, 064701.
- (19) Nolan, M.; Grigoleit, S.; Sayle, D. C.; Parker, S. C.; Watson, G. W. *Surf. Sci.* **2005**, *576*, 217–229.
- (20) Skorodumova, N. V.; Ahuja, R.; Simak, S. I.; Abrikosov, I. A.; Johansson, B.; Lundqvist, B. I. *Phys. Rev. B* **2001**, *64*, 115108.
- (21) Huang, M.; Fabris, S. J. *Phys. Chem. C* **2008**, *112*, 8643–8648.
- (22) Kresse, G.; Hafner, J. *Phys. Rev. B* **1994**, *49*, 14251–14269.
- (23) Kresse, G.; Furthmüller, J. *Comput. Mater. Sci.* **1996**, *6*, 15–50.
- (24) Hammer, B.; Hansen, L. B.; Nørskov, J. K. *Phys. Rev. B* **1999**, *59*, 7413–7421.
- (25) Blöchl, P. E. *Phys. Rev. B* **1994**, *50*, 17953–17979.
- (26) Kresse, G.; Joubert, D. *Phys. Rev. B* **1999**, *59*, 1758–1774.
- (27) Blöchl, P. E.; Jepsen, O.; Andersen, O. K. *Phys. Rev. B* **1994**, *49*, 16223–16233.
- (28) Cousins, C. S. G. *J. Phys. C: Solid State Phys.* **1981**, *14*, 1585–1602.
- (29) Hu, Z.; Metiu, H. J. *Phys. Chem. C* **2011**, *115*, 5841–5845.
- (30) Morgan, B. J.; Watson, G. W. *J. Phys. Chem. C* **2010**, *114*, 2321–2328.
- (31) Galea, N. M.; Scanlon, D. O.; Morgan, B. J.; Watson, G. W. *Mol. Simul.* **2009**, *35*, 577–583.
- (32) Da Silva, J. L. F.; Ganduglia-Pirovano, M. V.; Sauer, J.; Bayer, V.; Kresse, G. *Phys. Rev. B* **2007**, *75*, 045121.
- (33) Loschen, C.; Carrasco, J.; Neyman, K. M.; Illas, F. *Phys. Rev. B* **2007**, *75*, 035115.
- (34) Scanlon, D. O.; Watson, G. W.; Payne, D. J.; Atkinson, G. R.; Egdel, R. G.; Law, D. S. L. *J. Phys. Chem. C* **2010**, *114*, 4636–4645.
- (35) Scientific Group Thermodata Europe (SGTE). SpringerMaterials: The Landolt–Börnstein Database. Thermodynamic Properties of Elements, H to Ni. http://dx.doi.org/10.1007/10652891_6 (accessed Feb 7, 2011).
- (36) Scientific Group Thermodata Europe (SGTE). SpringerMaterials: The Landolt–Börnstein Database. Thermodynamic Properties of Compounds, SH to SiH₂. http://dx.doi.org/10.1007/10688868_4 (accessed Feb 7, 2011).
- (37) Scientific Group Thermodata Europe (SGTE). SpringerMaterials: The Landolt–Börnstein Database. Thermodynamic Properties of Compounds, SbO₂ to Rh₂O₃. http://dx.doi.org/10.1007/10688868_15 (accessed Feb 7, 2011).

(38) Scientific Group Thermodata Europe (SGTE). SpringerMaterials: The Landolt–Börnstein Database. Thermodynamic Properties of Compounds, RuO₃ to Ti₃O₅. http://dx.doi.org/10.1007/10688868_16 (accessed Feb 7, 2011).

(39) Scientific Group Thermodata Europe (SGTE). SpringerMaterials: The Landolt–Börnstein Database. Thermodynamic Properties of Compounds, V₂O₅ to ThP. http://dx.doi.org/10.1007/10688868_17 (accessed Feb 7, 2011).

(40) Scientific Group Thermodata Europe (SGTE). SpringerMaterials: The Landolt–Börnstein Database. Thermodynamic Properties of Compounds, Mn₂O₃ to TaN. http://dx.doi.org/10.1007/10688868_11 (accessed Feb 7, 2011).

(41) Scientific Group Thermodata Europe (SGTE). SpringerMaterials: The Landolt–Börnstein Database. Thermodynamic Properties of Compounds, CdO to PCIF. http://dx.doi.org/10.1007/10705961_16 (accessed Feb 7, 2011).

(42) Jeanne-Rose, V.; Poumellec, B. *J. Phys.: Condens. Matter* **1999**, *11*, 1123–1137.

(43) Honig, J. M.; Zandt, L. L. V. *Annu. Rev. Mater. Sci.* **1975**, *5*, 225–278.

(44) Enjalbert, R.; Galy, J. *Acta Crystallogr. C: Cryst. Struct. Commun.* **1986**, *42*, 1467–1469.

(45) Andersson, G.; Parck, C.; Ulfvarson, U.; Stenhagen, E.; Thorell, B. *Acta Chem. Scand.* **1956**, *10*, 623–628.

(46) Finger, L. W.; Hazen, R. M. *J. Appl. Phys.* **1980**, *51*, 5362–5367.

(47) Magnéli, A.; Andersson, G.; Blomberg, B.; Kihlberg, L. *Anal. Chem.* **1952**, *24*, 1998–2000.

(48) Brandt, B. G.; Skapski, A. C.; Thom, E.; Stoll, E.; Eriksson, G.; Blinc, R.; Pausak, S.; Ehrenberg, L.; Dumanovic, J. *Acta Chem. Scand.* **1967**, *21*, 661–672.

(49) Gerward, L.; Staun Olsen, J.; Petit, L.; Vaitheeswaran, G.; Kanchana, V.; Svane, A. *J. Alloys Compd.* **2005**, *400*, 56–61.

(50) Bärnighausen, H.; Schiller, G. *J. Less-Common Met.* **1985**, *110*, 385–390.

Effects of Biomolecular Flexibility on Alchemical Calculations of Absolute Binding Free Energies

Morgan Lawrenz,^{*,†} Riccardo Baron,^{*,†} Yi Wang,[†] and J. Andrew McCammon^{†,‡}

[†]Department of Chemistry and Biochemistry, Center for Theoretical Biological Physics, and [‡]Department of Pharmacology, Howard Hughes Medical Institute, University of California, San Diego, La Jolla, California 92093, United States

S Supporting Information

ABSTRACT: The independent trajectory thermodynamic integration (IT-TI) approach (Lawrenz, M., et al. *J. Chem. Theory. Comput.* **2009**, *5*, 1106–1116) for free energy calculations with distributed computing is employed to study two distinct cases of protein–ligand binding: first, the influenza surface protein N1 neuraminidase bound to the inhibitor oseltamivir, and second, the *Mycobacterium tuberculosis* enzyme RmlC complexed with the molecule CID 77074. For both systems, finite molecular dynamics (MD) sampling and varied molecular flexibility give rise to IT-TI free energy distributions that are remarkably centered on the target experimental values, with a spread directly related to protein, ligand, and solvent dynamics. Using over 2 μ s of total MD simulation, alternative protocols for the practical, general implementation of IT-TI are investigated, including the optimal use of distributed computing, the total number of alchemical intermediates, and the procedure to perturb electrostatics and van der Waals interactions. A protocol that maximizes predictive power and computational efficiency is proposed. IT-TI outperforms traditional TI predictions and allows a straightforward evaluation of the reliability of free energy estimates. Our study has broad implications for the use of distributed computing in free energy calculations of macromolecular systems.

1. INTRODUCTION

Alchemical free energy methods often employ molecular dynamics (MD) simulations of unphysical intermediate microstates in order to calculate the free energy difference between two physically relevant canonical ensembles. Examples are the relative binding free energy difference between different ligands to a receptor or the free energy change upon transferral of a ligand and protein from the unbound to the bound state. The latter is often referred to as the *absolute binding free energy* described by the thermodynamic cycle in Scheme 1.^{1–7} Although MD-based free energy calculations rely on rigorous statistical mechanics principles,^{5,6,8,9} their practical application is still challenging for systems with numerous degrees of freedom. MD sampling may be trapped in confined regions of conformational space due to the frustrated nature of protein and ligand energy landscapes, thus leading to insufficient statistics.

The use of independent MD simulations recently proved to be an appealing strategy to alleviate this issue, particularly with the rapid and steady increase of computational power in the form of multiple CPU and GPU clusters. This approach was applied to a number of systems in different flavors. Fujitani et al. employed multiple free energy perturbation (FEP) calculations to estimate absolute free energies of FKBP ligand binding.¹⁰ Zagrovic and van Gunsteren used multiple one-step perturbation runs to calculate relative free energies of PDE5 ligand binding.¹¹ In Mobley et al. and Boyce et al., multiple FEP calculations for different docked ligand binding poses were used to predict relative and absolute binding free energies for ligands to engineered binding sites of T4 lysozyme.^{12,13} Lawrenz et al. employed independent trajectory thermodynamic integration (IT-TI) to obtain accurate absolute free energies for peramivir binding to N1 neuraminidase, as well as relative binding free

energies of alchemically modified compounds.¹⁴ The latter study also emphasized the importance of solvent effects in this context. Accurate free energies are needed for all states of the thermodynamic cycle of interest (see Scheme 1) to achieve high predictive power, as realized since the pioneering applications of alchemical approaches.^{2,3,15,16} Here, we use IT-TI to compute absolute binding free energies for two ligands to two protein drug targets with different active site structural and chemical properties.

First, we consider the influenza surface protein N1 neuraminidase binding to oseltamivir.¹⁷ N1 neuraminidase facilitates viral shedding from infected cells and is a key target for treatment of pandemic flu. The N1 active site is composed of flexible loops^{14,18} and is highly solvent exposed (see Figure 1a,c). The ligand oseltamivir has zero net charge but contains one ammonium group and one carboxyl group (Figure 1e); the latter forms salt bridges with the arginine triad binding motif (R118, R292, and R371 in Figure 1c).¹⁷ Electrostatic interactions have been characterized as the dominant contribution to ligand binding.^{14,19} Oseltamivir is flexible due to 10 nonsterically hindered rotatable bonds, including a branched aliphatic tail that occupies a hydrophobic subpocket.^{14,20}

Second, we study the *Mycobacterium tuberculosis* (*M. tuberculosis*) enzyme dTDP-6-deoxy-*d*-xylo-4-hexopyranosid-4-ulose 3,5-epimerase (RmlC), which is crucial for assembly of the mycobacterial waxy, impermeable cell wall and is a viable drug target.²¹ In this case, the bound ligand, compound identifier (CID) 77074, was a top hit from virtual screening, followed by experimental validation.²¹ The RmlC binding site is organized into β -sheets and is smaller and

Received: April 4, 2011

Published: June 02, 2011

narrower than that in N1 (compare Figure 1a,c and Figure 1b,d). Aromatic residues Y138, F26, and H119 stack against the ligand aromatic rings (see Figure 1d). The ligand itself contains seven rotatable bonds, with limited flexibility due to the presence of two bulky ring groups (Figure 1f).

We investigate to which extent protein, ligand, and solvent dynamics influence the reliability of absolute binding free energies computed with TI. Using IT-TI, we see that finite sampling and varied molecular flexibility of the two investigated protein–ligand systems give rise to distributions of free energy estimates. This observation is in line with previous suggestions for N1 neuraminidase based on more reduced statistics.¹⁴ We show that the features of these distributions—while remarkably centered around the target experimental values—are linked to protein, ligand, and solvent dynamic sampling. Additionally, we use statistics from over 2 μ s of overall IT-TI simulation time to compare different approaches for optimal distributed computing and alternative protocols for the practical application of TI. We suggest a protocol that is optimal for two systems with different dynamic properties. Future work will investigate whether this protocol might be optimal for protein–ligand binding in general.

2. MATERIALS AND METHODS

2.1. Molecular Models and Simulations. Initial coordinates were available for N1 bound to the ligand oseltamivir on the basis of X-ray crystallography experiments (PDB: 2HU0).¹⁷ For RmlC, initial coordinates for its complex with CID 77074, or 1-(3-(5-allyl-5H-[1,2,4]triazino[5,6-b]indol-3-ylthio)propyl)-1H-benzo[*d*]imidazol-2(3H)-one, were based on the unbound X-ray structure (PDB: 2IXC) and an experimentally verified computational docking pose.²¹ A monomer of the natively

tetrameric protein N1 was simulated, as in previous studies,¹⁴ while the RmlC protein was simulated as a dimer, for half the N1 simulation time, because its active site spans the interface between two monomers (see Figure 1). Thus, RmlC analyses were performed by concatenating two monomer trajectories for identical overall sampling times for each system. See Table 1 for a summary of MD sampling periods.

Molecular models were based on the AMBER FF99SB force field²² and the compatible TIP3P model for water.²³ The cubic simulation box contained N1, 15 126, and RmlC, 24 305 water molecules, added to the system using AmberTools. Both systems were neutralized with (N1, 1, or RmlC. 24) Na⁺ counterions with AMBER rescaled parameters.²⁴ The importance of a protein-bound Ca²⁺ ion in N1 ligand binding was recently highlighted.²⁵ Ligands were parametrized using the generalized Amber force field (GAFF)²⁶ parameters for angles, bonds, and torsions, and RESP²⁷ fitting of Gaussian03²⁸ calculated electrostatic potentials at the Hartree–Fock/6-31G* level. All simulations were performed using the NAMD software²⁹ (version 2.7b1). A 2 fs time step was employed, with hydrogen-containing protein bonds constrained using RATTLE³⁰ and water geometries constrained using SETTLE.³¹ The particle mesh Ewald (PME) approximation³² (1 \AA^{-3} grid density) was employed for electrostatics. Short-range nonbonded interactions were evaluated every 2 fs and long-range electrostatics every 4 fs (nonbonded interaction cutoff, 12 \AA ; switching distance, 10 \AA).²⁹ After incremental heating to 300 K, the system was equilibrated for 2 ns in the *N, p, T* ensemble with Langevin pressure and temperature controls³³ before each *N, V, T* independent simulation was initialized with a random velocity.

2.2. Free Energy Calculations. Free energy changes along the thermodynamic cycle in Scheme 1 were evaluated using thermodynamic integration (TI) as³⁴

$$\Delta F_{0 \rightarrow 1} = \int_0^1 d\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda} \quad (1)$$

where in this study $\Delta F_{0 \rightarrow 1}$ is either the $\Delta F_{\text{protein}}^{\circ}$ or $\Delta F_{\text{water}}^{\circ}$ standard Helmholtz free energy in Scheme 1 and U is the total potential energy of the system. The ligand is decoupled from

Scheme 1. Thermodynamic Cycle Underlying Alchemical Absolute Binding Free Energy Calculations

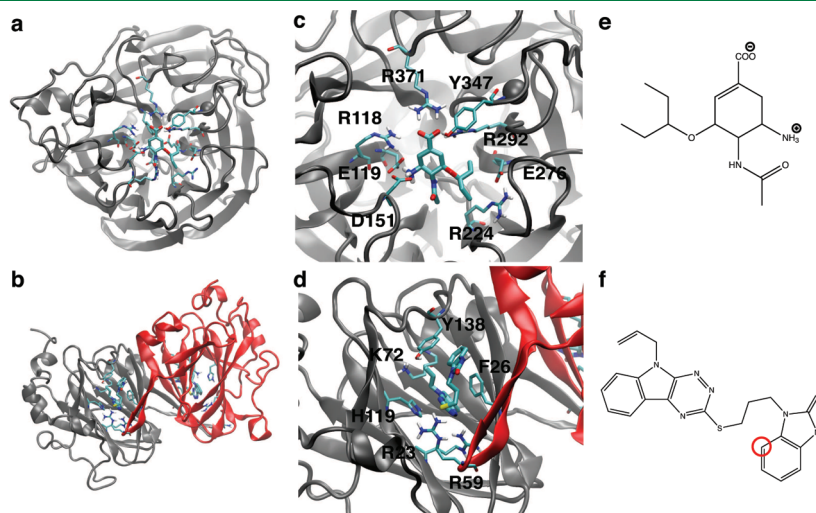
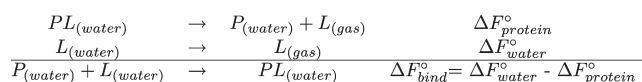


Figure 1. Protein–ligand structures of the investigated systems. Overall views of the N1 monomer (a) and RmlC dimer (b) structures are shown, with the RmlC monomers in b colored to highlight the interface. The active site residues of the two proteins are labeled for N1 in (a) and for RmlC in (d). Ligand chemical structures are depicted for the N1 ligand oseltamivir (e) and the RmlC inhibitor 77074 (f), the latter with the restrained atom (see Materials and Methods) highlighted in red. For oseltamivir, the center of mass was restrained, not a single atom (see Materials and Methods).

Table 1. Protocols for IT-TI Calculations

reference name	elec/vdW no. λ	initialization	nonbonded	decoupling	runs \times time/ λ (ns)	total time (ns)
medium cont/simul/14 λ^a	9/14	continuous	simultaneous	inter + intra	20 \times 1	280
long cont/simul/14 λ	9/14	continuous	simultaneous	inter + intra	10 \times 2	280
medium cont/simul/inter/14 λ	9/14	continuous	simultaneous	inter only	20 \times 1	280
medium cont/sep/14 λ	9/5	continuous	separate	inter + intra	20 \times 1	280
medium cont/sep/19 λ^b	9/10	continuous	separate	inter + intra	20 \times 1	80
medium parall/simul/14 λ	9/14	parallel ^c	simultaneous	inter + intra	20 \times 1	280
medium parall/simul/19 λ	9/19	parallel ^c	simultaneous	inter + intra	20 \times 1	380
medium parall/simul/inter/14 λ	9/14	parallel ^c	simultaneous	inter only	20 \times 1	280
medium parall/sep/14 λ	9/5	parallel ^c	separate	inter + intra	20 \times 1	280
medium parall/sep/19 λ	9/10	parallel ^c	separate	inter + intra	20 \times 1	380
long parall/sep/19 λ	9/10	parallel ^c	separate	inter + intra	10 \times 2	380

^a For N1, 14 λ = [0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]; for RmlC, [0, 0.1, 0.2, 0.25, 0.3, 0.37, 0.45, 0.5, 0.65, 0.75, 0.8, 0.9, 0.97, 1].

^b For N1, 19 λ adds [0.55, 0.65, 0.75, 0.85, 0.95]; for RmlC, [0.05, 0.15, 0.6, 0.7, 0.85]. ^c Protocols are well-suited for distributed computing.

the surrounding environment with the coupling parameter λ that changes from 0 to 1 to linearly scale all ligand nonbonded potential energy terms as

$$U(X; \lambda) = U_{\text{unperturbed}}(X) + \lambda U_{\text{decoupled}}(X) + (1 - \lambda) U_{\text{coupled}}(X) \quad (2)$$

where X denotes the overall system configurational space assuming equilibrium conditions. In all cases the soft-core potential by Zacharias et al. was employed to enhance sampling and eliminate instabilities (shift parameter $\delta = 5$).³⁵ The $\partial U/\partial \lambda$ values of eq 1 were printed for each λ every 0.1 ps, and their forward cumulative average was monitored to evaluate convergence (generally reached within equilibration periods of 500 ps). Numerical integration of eq 1 was performed using an interpolated cubic spline.

A harmonic restraining potential $U(r_L) = (1/2)k_h(r_L - r_0)^2$ was applied to restrict ligand sampling r_L to a finite volume V_{pocket} within the active site throughout the TI calculations of $\Delta F_{\text{protein}}^\circ$. Reasonable k_h values were obtained from average fluctuations of the ligand position ($\langle \delta r^2 \rangle$) during a free 2 ns N, p, T MD run as $k_h = (3RT)/\langle \delta r^2 \rangle$,^{14,36} with R the molar gas constant and T the absolute temperature of 300 K. A $k_h = 2.9 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ was used for restraint of the oseltamivir center of mass and $k_h = 0.74 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ for restraint of a central atom (highlighted in Figure 1f) in the 77074 ligand.

Then, the standard-state free energy was taken into account^{8,9,37} for $\Delta F_{\text{protein}}^\circ$ through an analytical correction for transference of the ligand from the restricted volume V_{pocket} to the bulk V° as

$$\Delta F_{\text{protein}}^\circ = \int_0^1 d\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda + RT \ln \left(\frac{V_{\text{pocket}}}{V^\circ} \right) \quad (3)$$

To reflect protein–ligand binding at a standard ligand concentration of 1 M, $V^\circ = 1661 \text{ \AA}^3$, with $T = 300 \text{ K}$. V_{pocket} was explicitly determined from multiple MD trajectories using the VMD VolMap plugin.³⁸ This procedure gave average ($RT \ln(V_{\text{pocket}}/V^\circ)$) corrections of $-1.25 \text{ kcal} \cdot \text{mol}^{-1}$ for the N1 system and $-1.07 \text{ kcal} \cdot \text{mol}^{-1}$ for the RmlC system. We note that the magnitude of such corrections is significant (up to 10% of the $\Delta F_{\text{bind}}^\circ$ values for both systems) and should not be neglected.^{9,37} For each RmlC calculation, the $\Delta F_{\text{protein}}^\circ$ was halved to obtain an average value for one active site.

One can obtain IT-TI $\Delta F_{\text{bind}}^\circ$ estimates from all combinations of K -independent $\Delta F_{\text{water}}^\circ$ estimates and J -independent $\Delta F_{\text{protein}}^\circ$ estimates as

$$\Delta F_{\text{bind},(k,j)}^\circ = [\Delta F_{\text{water},k}^\circ - \Delta F_{\text{protein},j}^\circ]_{j=1 \dots J}^{k=1 \dots K} \quad (4)$$

Here, a total of $N = KJ$ estimates of $\Delta F_{\text{bind}}^\circ$ are generated and binned in windows of width $RT = 0.6 \text{ kcal} \cdot \text{mol}^{-1}$. The linear average of the N -independent binding free energy estimates, $\Delta \bar{F}_{\text{bind}}^\circ$ is reported throughout the paper.

2.3. Alternative IT-TI Protocols. Alternative approaches for IT-TI distributed computing were investigated by using more, *medium* independent simulations or fewer, *long* independent simulations. Effects of varied user-defined inputs for TI were also explored, as summarized in Table 1. For independent TI calculations, the λ intermediate simulations were either initialized continuously (*cont* protocols) or in parallel (*parall* protocols). In the first case, simulations at $\lambda = 0$ started from the configuration (coordinates and velocities) from a 2 ns N, p, T equilibrated system; at each increasing λ value, the end configuration from the previous λ simulation was used. These IT-TI protocols are less-suited for distributed computing because the MD initialization requires information from sequential runs, but this approach does allow more equilibrated starting structures at successive λ values. Instead, for the *parall* protocols, all λ simulations were independently initialized from the same N, p, T equilibrated structure with a random velocity. This approach is well-suited for distributed computing, because the MD initialization is independent among each λ simulation. Ligand electrostatics and van der Waals interactions were perturbed, as in eq 2, in three alternative ways (see Table 1). First, electrostatic interactions were decoupled for $0 \leq \lambda \leq 0.5$ and van der Waals more slowly for $0 \leq \lambda \leq 1$ (*simul* protocol). Second, the same components were scaled separately, with electrostatic interactions for $0 \leq \lambda \leq 0.5$ and then van der Waals for $0.5 \leq \lambda \leq 1$ (*sep* protocol). Third, only the intermolecular terms were decoupled (*inter* protocol).

2.4. Error Analysis of IT-TI Predictions. We evaluated the *accuracy* and *precision* of our IT-TI estimates. Accuracy is described by the match of $\Delta \bar{F}_{\text{bind}}^\circ$ with respect to a reference experimental value, here assumed to be characterized by zero uncertainty. Precision is reflected in the spread of the IT-TI $\Delta F_{\text{bind}}^\circ$ estimates and is described by the standard deviation σ_{bind} . Here σ_{bind} has two components, σ_{water} from the

$\Delta\bar{F}_{\text{water}}^{\circ}$ calculations and σ_{protein} from the $\Delta F_{\text{protein}}^{\circ}$ calculations. Accuracy is limited by systematic errors, which are due to, for example, empirical force field and water models, as well as numerical approximations in the MD algorithms. Both accuracy and precision is affected by random errors from finite sampling. We can capture the statistical uncertainty on the IT-TI $\Delta\bar{F}_{\text{bind}}^{\circ}$ due to random errors from N independent calculations with the standard error $\delta = \sigma/\sqrt{N}$, as previously suggested.¹⁴ Note that this metric approaches zero for large N . We computed this uncertainty for the J estimates of $\Delta F_{\text{protein}}^{\circ}$ and for K estimates of $\Delta F_{\text{water}}^{\circ}$ (eq 4) and propagate for the uncertainty on $\Delta\bar{F}_{\text{bind}}^{\circ}$ as

$$\delta_{\text{bind}} = \sqrt{\frac{\sigma_{\text{water}}^2}{\sqrt{K}} + \frac{\sigma_{\text{protein}}^2}{\sqrt{J}}}$$

2.5. Analysis of Conformational Sampling. MD snapshots were saved every 2 ps for analysis, with all protein backbone atoms first aligned to a reference structure. Active site residues for both systems were identified as those within 5 Å of the ligand.

For each system, principal component analysis (PCA)^{39,40} of protein fluctuations was performed by calculating the covariance matrix for active site heavy atoms with GROMACS (version 4.0.4 compiled in double precision),⁴¹ using all λ simulations in all $J = 10$ long cont/simul/14 λ calculations, for 280 ns of total simulation time (Table 1). Then, projections for independent λ simulations were generated along 20 out of the total 528 principal components (PCs) of this matrix, accounting for 75% of the protein fluctuations. Projections for trajectories using other IT-TI protocols for a given system are along these same PCs for comparison, with projections along the four most dominant PC described in detail. We also project previously performed $\lambda = 0$ apo and holo MD simulations onto these PCs for reference, with 400 and 10 ns each for apo and holo simulation with N1 and RmlC, respectively. Details of these N1 simulations have been previously reported.²⁵ For a fair comparison of the two systems, all projections were reweighted as w^{-1} , w being the number of active site atoms (N1, 176; RmlC, 161). For hydration analysis, water–water hydrogen bonds within 5 Å of the ligand in the active site were monitored. Hydrogen bonds were defined to have a maximum hydrogen–acceptor distance of 3.5 Å and a minimum donor–hydrogen–acceptor angle of 120°. The software VMD,³⁸ xmgrace, as well as python scripts based on matplotlib and NumPy libraries were used for analysis and graphical representations.

3. RESULTS AND DISCUSSION

3.1. IT-TI Free Energy Distributions and Their Dependence on Biomolecular Flexibility. Because independent TI estimates of $\Delta F_{\text{bind}}^{\circ}$ vary with the specific set of simulations performed, IT-TI generates distributions of free energy estimates and provides an average value $\Delta\bar{F}_{\text{bind}}^{\circ}$ with a reliable measure of uncertainty (δ_{bind}).¹⁴ We evaluate the accuracy of our predicted $\Delta\bar{F}_{\text{bind}}^{\circ}$ values with reference free energies derived from the experimental K_i as $\Delta F_{\text{expt}} = RT \ln(K_i)$. For N1-oseltamivir and RmlC-77074 binding, ΔF_{expt} values of -13.7 and -9.9 kcal·mol⁻¹ were reported, respectively.^{21,42} We compute N IT-TI estimates of $\Delta F_{\text{bind}}^{\circ}$ from K independent calculations of $\Delta F_{\text{water}}^{\circ}$ and J calculations of $\Delta F_{\text{protein}}^{\circ}$ (eq 4). The $K = 20$ $\Delta F_{\text{water}}^{\circ}$ results have a much smaller spread relative to the

Table 2. Summary of IT-TI Results with Varied Protocols

reference name	$\Delta\bar{F}_{\text{bind}}^{\circ} \pm \delta_{\text{bind}}$ (kcal·mol ⁻¹)	
	N1	RmlC
medium cont/simul/14 λ	-12.2 ± 1.0	-11.0 ± 0.4
long cont/simul/14 λ	-10.4 ± 1.6	-9.4 ± 0.4
medium cont/simul/inter/14 λ	-14.9 ± 1.2	
medium cont/sep/14 λ	-10.4 ± 1.2	
medium cont/sep/19 λ	-13.7 ± 1.1	
medium parall/simul/14 λ	-10.6 ± 0.6	
medium parall/simul/19 λ	-11.2 ± 0.6	
medium parall/simul/inter/14 λ	-12.2 ± 0.7	
medium parall/sep/14 λ	-11.1 ± 0.6	
medium parall/sep/19λ	-14.3 ± 0.5	-11.8 ± 0.3
long parall/sep/19λ	-12.8 ± 0.6	-10.8 ± 0.2

$J = 20$ $\Delta F_{\text{protein}}^{\circ}$ results, with $\sigma_{\text{water}} = 0.4$ and 0.2 kcal·mol⁻¹ compared to $\sigma_{\text{protein}} = 4.4$ and 3.2 kcal·mol⁻¹ for the *medium cont/simul/14 λ* N1-oseltamivir and RmlC-77074 calculations, respectively. Thus, the shape of the $\Delta F_{\text{bind}}^{\circ}$ distributions is dominated by the variation of the J $\Delta F_{\text{protein}}^{\circ}$ results, as expected due to the numerous degrees of freedom and complex energy landscape in this state.

Both $J = 20$ medium and $J = 10$ long independent protein simulations were used to compute $\Delta\bar{F}_{\text{bind}}^{\circ}$, to test the computational efficiency of more, shorter runs compared with fewer, longer independent runs. See Table 2 for a summary of all IT-TI predictions. Figure 2 shows distributions of $\Delta F_{\text{bind}}^{\circ}$ estimates obtained using the cont/simul/14 λ protocol in Table 1. The distributions are remarkably different for the two systems investigated. The N1-oseltamivir results in Figure 2a,c have a broad range, from very favorable (-20 kcal·mol⁻¹) to unfavorable (>0 kcal·mol⁻¹). As reported in Table 2, estimates from medium runs give $\Delta\bar{F}_{\text{bind}}^{\circ} = -12.2 \pm 1.0$ kcal·mol⁻¹. Corresponding results for the long simulations display a marked shift of $\Delta\bar{F}_{\text{bind}}^{\circ}$ to 3.3 kcal·mol⁻¹ away from ΔF_{expt} and an increase of δ_{bind} , with $\Delta\bar{F}_{\text{bind}}^{\circ} = -10.4 \pm 1.6$ kcal·mol⁻¹. The use of more, independent simulations improved the free energy results in this case. In contrast, the RmlC-77074 distributions are centered near ΔF_{expt} (see Figure 2b,d), and the use of fewer, independent runs with longer sampling times gave the most accurate and precise results. A close match with experiment is found for the long simulation results in Figure 2d, with $\Delta\bar{F}_{\text{bind}}^{\circ} = -9.4 \pm 0.4$ kcal·mol⁻¹ (Table 2). Overall, the δ_{bind} of the RmlC-77074 results is significantly smaller than the δ_{bind} of the N1-oseltamivir results, due to a much smaller spread σ_{bind} .

To probe underlying causes of the different free energy results for the two systems, we analyzed protein sampling during the independent simulations with PCA of protein fluctuations. Figure 3 shows projections along the two most dominant principal components at 5 λ values, [0, 0.2, 0.5, 0.8, 1], depicting changes in protein sampling along the perturbation in eq 2. For comparison, the same data for longer apo and holo N1 and RmlC simulations with $\lambda = 0$ (see Materials and Methods) onto the same PC are reported (Figure 3). For N1, the J simulations slowly equilibrated into varied portions of phase space, resulting in non-overlapping projections at $\lambda = 1$. Many of the simulations also exclusively sampled motions that are not visited in the reference holo or apo simulations. Our analysis indicates highly frustrated N1 sampling as the λ simulations are continuously

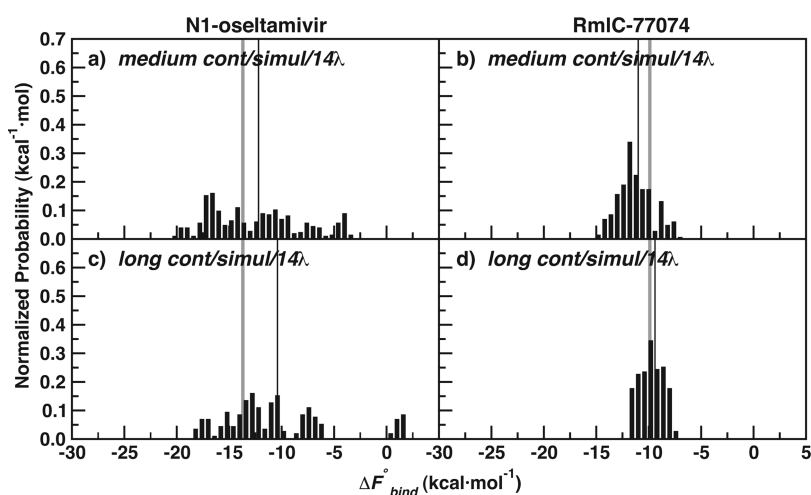


Figure 2. Normalized distributions of N1-oseltamivir (left) and RmlC-77074 (right) IT-TI results for medium and long cont/simul/14 λ TI protocols (Table 1). $\Delta F_{\text{expt}}^{\circ}$ for both systems is also depicted (gray line), along with $\Delta F_{\text{bind}}^{\circ}$ (thin black line).

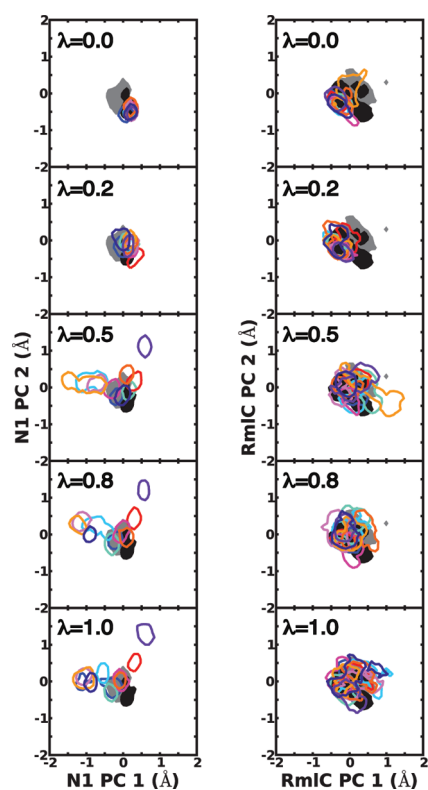


Figure 3. Receptor flexibility for N1 (left) and RmlC (right) as captured by two dominant principal components (PCs) of active site residue fluctuations from long cont/simul14 λ simulations. Contours depicting projections for 90% of the apo (filled gray) and holo (filled black) MD simulations, as well as each of $J = 10$ independent trajectories (unfilled color) are shown at λ values [0, 0.2, 0.5, 0.8, 1]. Projections are reweighted to allow direct comparison between the two systems. See Materials and Methods for details.

initialized, contributing to varied free energy estimates and a large σ_{protein} component of σ_{bind} . A different picture emerges for RmlC-77074, which had a significantly smaller σ_{protein} compared to N1-oseltamivir. In this case, the J independent simulations accessed similar regions of conformational space, as inferred

from overlapping projections (see Figure 3). The RmlC projections also significantly overlap with projections from apo and holo reference simulations. These observations hold similarly when analyzing projections along other, less dominant PC from PCA (not shown) and link the varied spreads of IT-TI free energy estimates, and corresponding uncertainties δ_{bind} to protein conformational sampling.

Differences in N1 and RmlC dynamics are also revealed in the sampling of specific binding site residue torsions. Comparison of torsion sampling at $\lambda = 0$ and at $\lambda = 1$ reveals that 9 out of 15 monitored N1 active site residues, but only 2 out of 11 RmlC residues, increased flexibility and sampled multiple conformations upon ligand decoupling. The torsional angles were also sampled in populations that vary among the $J = 10$ independent runs, particularly for charged N1 residues R224, R371, R118, E277, and E119 (Supporting Information Figure S1). As seen in the PCA, the N1 system is challenged to access its conformational space within a single simulation, while for RmlC, sampling is more complete within and similar among independent IT-TI simulations. We note that, in addition to protein sampling and σ_{protein} , water sampling and σ_{water} contributes to the varied σ_{bind} for the two systems; the more flexible ligand oseltamivir has more diverse sampling than the sterically hindered 77074, reflected in the larger σ_{water} for this ligand (see above). Altogether, these sampling behaviors yield the different uncertainties δ_{bind} on $\Delta F_{\text{bind}}^{\circ}$ estimates for the two systems (see Table 2).

3.2. IT-TI Free Energy Distributions and Their Dependence on Solvent Effects.

Hydration dynamics and solvent fluctuations also contribute to the spread of the IT-TI free energy distributions, in addition to protein and ligand flexibility described in the previous section. Here we report an example from the N1-oseltamivir IT-TI results in closer detail. In Figure 2c, an outlier, unfavorable $\Delta F_{\text{protein}}^{\circ}$ estimate was computed (see histogram bars around $\Delta F_{\text{bind}}^{\circ} = 0$). This result can be linked to pronounced solvent fluctuations during the $\Delta F_{\text{protein}}^{\circ}$ calculation at λ values 0.2 and 0.25. At these intermediate states, water molecules diffuse into the active site, very close to the partially decoupled ligand carboxyl and ammonium groups, and an increased number of active site water–water hydrogen bonds is observed (Figure 4c). This coincides with a shift in the electrostatics component of $\partial U/\partial \lambda$ (Figure 4b), giving a less

positive integrated $\Delta F_{\text{protein}}^{\circ}$ value and unfavorable $\Delta F_{\text{bind}}^{\circ}$ estimates (eq 4).

These observations are fully consistent with the dynamic nature of protein hydration and dewetting fluctuations in binding cavities recently reported in the literature^{43–45} and their thermodynamic relevance.^{46–48} Because time scales of these solvent fluctuations may reach several hundred picoseconds, it is expected that our individual nanosecond TI runs may have diverse solvent behavior among the 10 performed. Here, the advantage of using IT-TI is illustrated, since a single TI calculation could yield a falsely unfavorable $\Delta F_{\text{bind}}^{\circ}$ estimate for N1-oseltamivir. Multiple estimates of $\Delta F_{\text{bind}}^{\circ}$ allow recovery of the probability distribution from multiple, independent simulations that sample both rare and dominant events. With enough independent

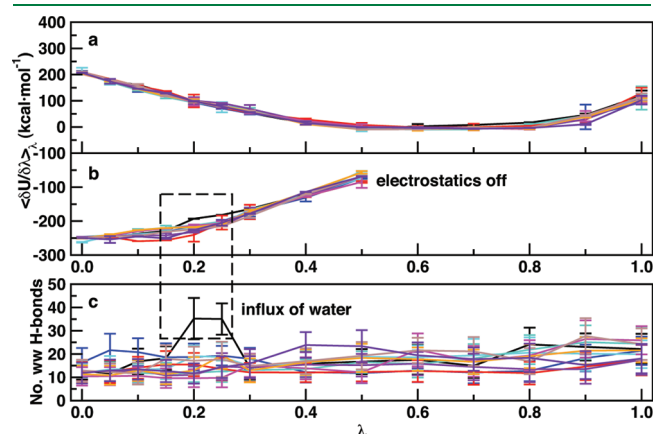


Figure 4. N1 active site hydration behavior and corresponding $\langle \partial U / \partial \lambda \rangle_{\lambda}$ values from the long cont/simul/14 λ protocol. $J = 10$ independent estimates of $\langle \partial U / \partial \lambda \rangle_{\lambda}$ are color-coded and interpolated for (a) van der Waals and (b) electrostatics components, with (c) the average and standard deviation of water–water hydrogen bonds within 5 Å of oseltamivir at each λ . The black curve in all panels indicates the TI calculation which gave an unfavorable $\Delta F_{\text{protein}}^{\circ}$ result. See Materials and Methods for details.

estimates, this distribution should reflect that of the true physical system. We also note that the solvent-exposed N1 has a consistent number of water molecules in the active site throughout the TI calculations (Figure 4c), highlighting the importance of water in both the bound and unbound states. Instead, the RmlC binding site has a more abrupt influx of water near $\lambda = 1$ (Supporting Information Figure S2).

3.3. N1-oseltamivir Protocol Investigation. In an effort to improve consistency of the N1-oseltamivir $\langle \partial U / \partial \lambda \rangle_{\lambda}$ values in Figure 4 and the free energy estimates in Figure 2, we conducted a series of IT-TI protocol changes for the N1-oseltamivir system. The varied N1-oseltamivir medium protocols implemented for the IT-TI calculations are described in Table 1, with the corresponding $\Delta \bar{F}_{\text{bind}}^{\circ} \pm \delta_{\text{bind}}$ listed in Table 2. Here, we aim for improved precision and accuracy over the medium cont/simul/14 λ results (Figure 2a and Figure 5a). We compare the spread, σ_{bind} , of the IT-TI distributions in Figure 5 to the $\sigma_{\text{bind}} = 4.4$ kcal·mol^{−1} of the medium cont/simul/14 λ results.

All protocols with continuous initialization of λ intermediates gave free energy distributions with a large spread, with $4.4 \leq \sigma_{\text{bind}} \leq 5.4$ kcal·mol^{−1} (see Figure 5a,c,e,g). In these cases, the $\Delta \bar{F}_{\text{bind}}^{\circ}$ values were also consistently less favorable than ΔF_{expt} , with the exception of protocol cont/sep/19 λ in Figure 5g. In the latter case, estimates are shifted to more favorable values and $\Delta \bar{F}_{\text{bind}}^{\circ}$ matches the ΔF_{expt} of -13.7 kcal·mol^{−1} with $\sigma_{\text{bind}} = 4.8$ kcal·mol^{−1} (Table 2). PCA of these cont simulations (Supporting Information Figure S3) indicated frustrated sampling, with projections similar to those seen in Figure 3. Decoupling of only ligand intermolecular nonbonded components in both protocols cont/simul/inter/14 λ and parall/simul/inter/14 λ reduced precision and made little difference in accuracy (Figure 5c,d and Table 2).

Overall, the σ_{bind} of the IT-TI distributions is significantly reduced with parallel initialization of each λ simulation. These estimates, in Figure 5b,d,f,h, had σ_{bind} values ≤ 3.0 kcal·mol^{−1}. However, only the protocol parall/sep/19 λ gave an accurate $\Delta \bar{F}_{\text{bind}}^{\circ}$ close to ΔF_{expt} at -14.3 kcal·mol^{−1} with $\sigma_{\text{bind}} = 2.1$ kcal·mol^{−1} (Table 2). This improvement in accuracy is observed

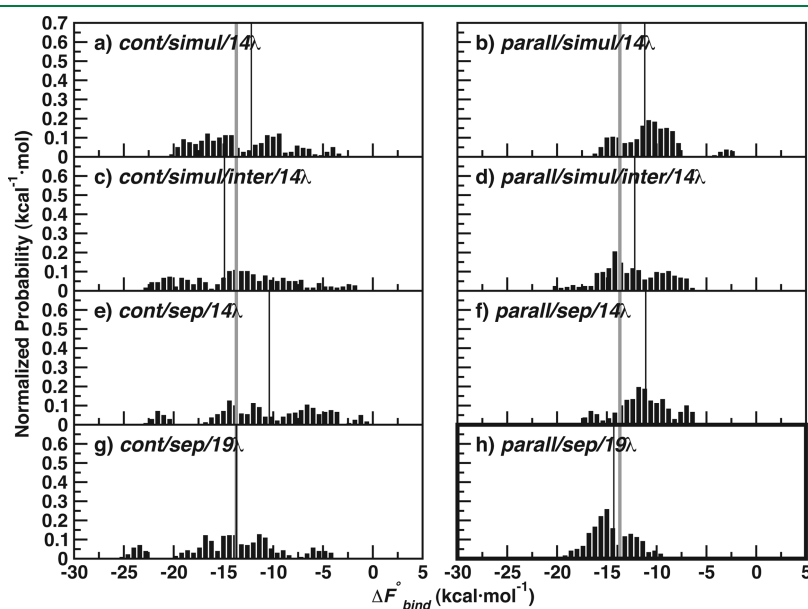


Figure 5. Normalized distributions for N1-oseltamivir IT-TI results with various medium decoupling protocols. Panels are labeled with the procedures from Table 1, and ΔF_{expt} is depicted (gray line), along with $\Delta \bar{F}_{\text{bind}}^{\circ}$ (thin black line). $\Delta \bar{F}_{\text{bind}}^{\circ} \pm \delta_{\text{bind}}$ values are reported in Table 2.

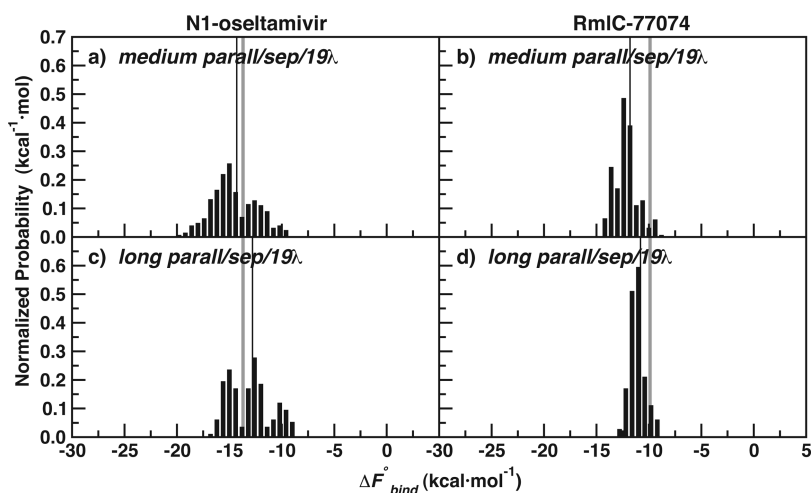


Figure 6. Normalized distributions of N1-oseltamivir (left) and RmlC-77074 (right) IT-TI results for medium and long parallel/sep/19 λ protocols (Table 2). $\Delta F_{\text{expt}}^{\circ}$ for both systems is also depicted (gray line), along with $\Delta \bar{F}_{\text{bind}}^{\circ}$ (thin black line).

for both the parall and cont protocols with separate decoupling of nonbonded components and 19 λ values (compare Figure 5e,f with Figure 5g,h). For protocols cont/sep/14 λ and parall/sep/14 λ , the van der Waals interactions are decoupled with only 5 λ intermediates, while the cont/simul/14 λ and parall/simul/14 λ protocols employed 14 λ (see Table 1). Five additional λ values for the cont/sep and parall/sep protocols, for 10 van der Waals λ intermediates and 19 total λ values, improved interpolation of the van der Waals $\langle \partial U / \partial \lambda \rangle_{\lambda}$ values for more accurate integration. These results are highlighted in bold in Figure 5h and Table 2.

The diverse outcomes obtained using alternative IT-TI protocols show that free energy calculations depend on a broad variety of user-defined choices. An optimal protocol was designed for N1-oseltamivir binding and suggests that (i) TI intermediates can be more conveniently placed at target λ values once a preliminary knowledge of the $\langle \partial U / \partial \lambda \rangle_{\lambda}$ vs λ curve is known; (ii) these λ values may be run in parallel, initialized from a $\lambda = 0$ holo configuration—an approach particularly suited for distributed computing; and (iii) separate decoupling of both inter- and intramolecular nonbonded components gives more accurate free energy estimates. We note that the approach described in (ii) may not be optimal for cases when the apo and holo states are separated by large conformational changes.

3.4. Application of Optimized Protocol to Both N1-Oseltamivir and RmlC-77074 Test Systems. We applied our optimal IT-TI protocol for N1-oseltamivir binding to RmlC-77074 and, again, compared estimates from two approaches for distributed computing. For N1-oseltamivir, both approaches gave more accurate IT-TI results with reduced uncertainty compared to Figure 2. In Figure 6a, we see that more, medium simulations gave a more favorable $\Delta \bar{F}_{\text{bind}}^{\circ}$ estimate of $-14.3 \pm 0.5 \text{ kcal} \cdot \text{mol}^{-1}$, compared to $\Delta \bar{F}_{\text{bind}}^{\circ} = -12.8 \pm 0.6 \text{ kcal} \cdot \text{mol}^{-1}$ computed with fewer, long simulations in Figure 6c. The additional simulations enhanced N1 sampling and improved the reliability of the IT-TI estimates (Table 2). For RmlC-77074, the results in Figure 6b,d reflect similar accuracy and precision compared to Figure 2, particularly for the long simulations. Here the $\Delta \bar{F}_{\text{bind}}^{\circ} = -10.8 \pm 0.2 \text{ kcal} \cdot \text{mol}^{-1}$ (Table 2), with a very small $\delta_{\text{bind}}^{\circ}$ due to consistent sampling among the independent simulations.

We can directly compare the 2-D projections of the long parallel/sep/19 λ simulations for both systems in Figure 7 with

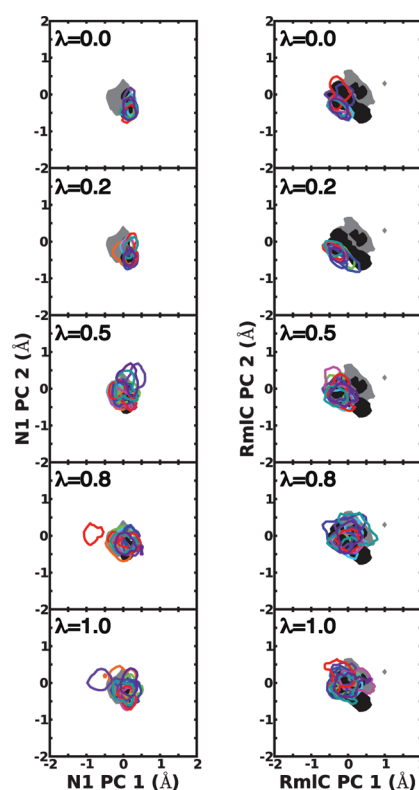


Figure 7. Receptor flexibility for N1 (left) and RmlC (right) as captured by two dominant PCs of active site residue fluctuations from long parallel/sep/19 λ simulations. See Figure 3 for color coding.

those in Figure 3. With the optimized protocol, the highly frustrated N1 sampling of Figure 3 is largely alleviated; the $J = 10$ projections overlap significantly with each other as well as with the reference apo and holo projections. One N1 simulation sampled outlier motions, but these exchanged with motions near the holo state during the same simulation (red contours in Figure 7 at $\lambda = 0.8$). The RmlC projections are similar to those in Figure 3, with consistent overlap among the independent runs and access of both apo and holo motions. In addition to more consistent protein sampling, both systems have reduced

fluctuations of water–water hydrogen bonds in the active site when using the parallel/sep protocol (Supporting Information Figure S2). This is primarily due to parallel initialization, rather than the separate decoupling protocol.

4. CONCLUSION

We investigated the (thermo)dynamics underlying two protein–ligand binding processes that involve very different protein, ligand, and water sampling. The distributions of binding free energy estimates produced from independent-trajectory thermodynamic integration (IT-TI) were remarkably centered on the target experimental values but had very different spreads for the two protein–ligand systems. The solvent exposed active site of N1, with many flexible, charged binding residues, has a larger population of microstates, with nontrivial barriers, that are easily over- or undersampled during a single TI calculation. The frustrated sampling observed for this system resulted in a broad range of free energy estimates from IT-TI, and additional independent runs, rather than longer sampling times, gave more reliable results. The RmlC microstates are more accessible, even within short (ns) simulations. Here, the IT-TI distributions had a smaller spread, and extension of simulation time improved the reliability of the results.

With tests of varied protocols for TI, we find that, for both protein systems, each alchemical intermediate may be run in parallel, each initialized with a $\lambda = 0$ configuration, for more consistent free energy results with maintained accuracy. This parallel approach also allows faster completion of the calculations compared to calculations with continuously initialized intermediates and is ideal for distributed computing. Additionally, separate decoupling of the inter- and intramolecular nonbonded terms gave optimal accuracy and precision overall, but only when employed with adequate intermediates for smooth interpolation of both electrostatics and van der Waals $\langle \partial U / \partial \lambda \rangle_\lambda$ values during integration (eq 1). We suggest an approach for performance of IT-TI calculations that maximizes reliability and computational efficiency with available sampling times.

■ ASSOCIATED CONTENT

S Supporting Information. Analysis of torsional angle sampling for N1 and RmlC active site residues (Figure S1), N1 and RmlC active site hydration behavior for varied IT-TI protocols (Figure S2), and PCA of N1 active site residue fluctuations from varied IT-TI protocols (Figure S3). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mlawrenz@ucsd.edu (M.L.), rbaron@mccammon.ucsd.edu (R.B.). Tel.: +1-858-822-2771 (M.L.), +1-858-534-2913 (R.B.).

■ ACKNOWLEDGMENT

We thank the members of the McCammon group for useful discussions. This work was supported, in part, by the National Institutes of Health, the National Science Foundation, the National Biomedical Computational Resource, and the Howard Hughes Medical Institute. We thank the Center for Theoretical Biological Physics (NSF Grant PHY-0822283) and the Texas Advanced Computer Center (Grant TG-MCA93S013) for

distributed computing resources. We also thank Dr. Ross C. Walker at the San Diego Supercomputing Center for additional computational resources.

■ REFERENCES

- (1) Tembe, B.; McCammon, J. *Comput. Chem.* **1984**, *8*, 281–283.
- (2) Beveridge, D.; DiCapua, F. *Annu. Rev. Biophys. Chem.* **1989**, *18*, 431–492.
- (3) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (4) Jorgensen, W. *Science* **2004**, *303*, 1813–1818.
- (5) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem.* **2010**, *114*, 10235–10253.
- (6) Gilson, M. K.; Zhou, H.-X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
- (7) van Gunsteren, W. F.; Beutler, T. C.; Fraternali, F.; King, P. M.; Mark, A. E.; Smith, P. E. *Computation of Free Energy in Practice: Choice of Approximations and Accuracy Limiting Factors*; ESCOM Science: Leiden, The Netherlands, 1993; Vol. 2.
- (8) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.
- (9) Borech, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.
- (10) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. *J. Chem. Phys.* **2005**, *123*, 084108.
- (11) Zagrovic, B.; van Gunsteren, W. J. *Chem. Theory Comput.* **2007**, *3*, 301–311.
- (12) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118–1134.
- (13) Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. *J. Mol. Biol.* **2009**, *394*, 747–763.
- (14) Lawrenz, M.; Baron, R.; McCammon, J. *J. Chem. Theory Comput.* **2009**, *5*, 1106–1116.
- (15) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050–3054.
- (16) Wong, C.; McCammon, J. *J. Am. Chem. Soc.* **1986**, *108*, 3830–3832.
- (17) Russell, R. J.; Haire, L. F.; Stevens, D. J.; Collins, P. J.; Lin, Y. P.; Blackburn, G. M.; Hay, A. J.; Gamblin, S. J.; Skehel, J. J. *Nature* **2006**, *443*, 45–49.
- (18) Amaro, R. E.; Minh, D. D. L.; Cheng, L. S.; Lindstrom, W. M.; Olson, A. J.; Lin, J.-H.; Li, W. W.; McCammon, J. A. *J. Am. Chem. Soc.* **2007**, *129*, 7764–7765.
- (19) Le, L.; Lee, E. H.; Hardy, D. J.; Truong, T. N.; Schulten, K. *PLoS Comput. Biol.* **2010**, *6*, No. e1000939.
- (20) Stoll, V.; Kent, S. D.; Maring, C. J.; Muchmore, S.; Giranda, V.; Gu, Y. Y.; Wang, G.; Chen, Y.; Sun, M.; Zhao, C.; Kennedy, A. L.; Madigan, D. L.; Xu, Y.; Saldivar, A.; Kati, W.; Laver, G.; Sowin, T.; Sham, H. L.; Greer, J.; Kempf, D. *Biochemistry* **2003**, *42*, 718–727.
- (21) Sivendran, S.; Jones, V.; Sun, D.; Wang, Y.; Grzegorzewicz, A. E.; Scherman, M. S.; Napper, A. D.; McCammon, J. A.; Lee, R. E.; Diamond, S. L.; McNeil, M. *Bioorg. Med. Chem.* **2010**, *18*, 896–908.
- (22) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
- (23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (24) Åqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021–8024.
- (25) Lawrenz, M.; Wereszczynski, J.; Amaro, R.; Walker, R.; Roitberg, A.; McCammon, J. A. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2523–2532.
- (26) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Case, P. A. K. D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (27) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (28) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K.; Burant, J.; Millam, J.; Iyengar, S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.;

Scalmani, G.; Rega, N.; Petersson, G.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J.; Hratchian, H.; Cross, J.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.; Yazyev, O.; Austin, A.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P.; Morokuma, K.; Voth, G.; Salvador, P.; Dannenberg, J.; Zakrzewski, V.; Dapprich, S.; Daniels, A.; Strain, M.; Farkas, O.; Malick, D.; Rabuck, A.; Raghavachari, K.; Foresman, J.; Ortiz, J.; Cui, Q.; Baboul, A.; Clifford, S.; Cioslowski, J.; Stefanov, B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R.; Fox, D.; Keith, T.; Laham, A.; Peng, C.; Nanayakkara, A.; Challacombe, M.; Gill, P.; Johnson, B.; Chen, W.; Wong, M.; Gonzalez, C.; Pople, J. *Gaussian 03*, Revision C.02, 2003; Gaussian, Wallingford, CT, 2004.

(29) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(30) Andersen, H. J. *Comput. Phys.* **1983**, *52*, 24–34.

(31) Shuichi, M.; Peter, A. J. *Comput. Chem.* **1992**, *13* (952–962), No. 148324.

(32) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(33) Feller, S.; Zhang, Y.; Pastor, R.; Brooks, B. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(34) Kirkwood, J. J. *Chem. Phys.* **1935**, 300–313.

(35) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025–9031.

(36) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.

(37) General, I. J. *J. Chem. Theory Comput.* **2010**, *6*, 2520–2524.

(38) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14* (33–38), 27–28.

(39) Garca, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696–2699.

(40) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Bioinf.* **1993**, *17*, 412–425.

(41) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(42) Kati, W. M.; Montgomery, D.; Carrick, R.; Gubareva, L.; Maring, C.; McDaniel, K.; Steffy, K.; Molla, A.; Hayden, F.; Kempf, D.; Kohlbrenner, W. *Antimicrob. Agents Chemother.* **2002**, *46*, 1014–1021.

(43) Setny, P.; Geller, M. *J. Chem. Phys.* **2006**, *125*, 144717.

(44) Baron, R.; McCammon, J. *Biochemistry* **2007**, *46*, 10629–10642.

(45) Young, T.; Hua, L.; Huang, X.; Abel, R.; Friesner, R.; Berne, B. J. *Proteins* **2010**, *78*, 1856–1869.

(46) Baron, R.; Setny, P.; McCammon, J. A. *J. Am. Chem. Soc.* **2010**, *132*, 12091–12097.

(47) Setny, P.; Baron, R.; McCammon, J. A. *J. Chem. Theory Comput.* **2010**, *6*, 2866–2871.

(48) Hummer, G. *Nature Chem.* **2010**, *2*, 906–907.

Theoretical Investigations on the Photoinduced Phase Transition Mechanism of Tetrathiafulvalene-*p*-chloranil

Yutaka Nakatsuka,^{†,‡} Takao Tsuneda,^{*,†,‡} Takeshi Sato,[§] and Kimihiko Hirao^{†,‡}

[†]Advanced Science Institute, RIKEN 2-1, Hirosawa, Wako, Saitama 351-0198, Japan

[‡]CREST, Japan Science and Technology Agency, Saitama 332-0012, Japan

[§]Photon Science Center of the University of Tokyo, Tokyo 113-8656, Japan

ABSTRACT: The photoinduced phase transition (PIPT) mechanism of tetrathiafulvalene-*p*-chloranil (TTF-CA) molecular crystal was theoretically investigated using the long-range corrected time-dependent density functional theory (LC-TDDFT) combined with a local response dispersion (LRD) method, which enables us to quantitatively reproduce charge transfer (CT) excitations of van der Waals clusters. By calculating the excitation spectrum and potential energy surface, we found that the PIPT of TTF-CA crystal may proceed through the angle change of the molecular planes. We also found that the CT excitation of one TTF-CA pair helps other neighboring TTF-CA pairs to become excited. Consequently, we theoretically proposed the initial structural change in the neutral-to-ionic PIPT of TTF-CA crystal, which is consistent with experiments.

I. INTRODUCTION

For the last 3 decades, the neutral–ionic phase transitions of alternatively stacked molecular crystals, which consist of electron donor (D) and acceptor (A) molecules, have been extensively studied due to their potentials for new materials and their unique physical phenomena. These phase transitions have been known to be induced by pressure and temperature. Recently, some phase transitions have also been found to be induced by photoirradiation. These photoinduced phase transitions (PIPT) can be distinguished in that a local change in electric structure leads to a macroscopic change in the crystal structure. The PIPT has been attracting interest as a new physical phenomenon and as a route to promising photoswitching devices.

Tetrathiafulvalene-*p*-chloranil (TTF-CA) molecular crystal, which consists of tetrathiafulvalene (TTF) donor and *p*-chloranil (CA) acceptor molecules, is a typical PIPT material. After Mayerle et al.¹ had observed a mixed-stack structure of TTF-CA crystal, Torrance et al. found that TTF-CA crystal is on the boundary of ionic (I) and neutral (N) states and undergoes pressure-induced² and temperature-induced³ phase transitions: in high temperature or low pressure, TTF-CA crystal is in the N state, where TTF and CA molecules are alternatively stacked along the *a* direction (Figure 1) with equivalent distances. On the other hand, in low temperature or high pressure, TTF-CA crystal is in the I state, where TTF and CA molecules are dimerized. The crystal structures and charge-transfer rates have been determined for each state by neutron diffraction⁴ and molecular vibration analyses.⁵ Koshihara et al.⁶ found that at around the critical temperature (84 K), laser irradiation induces the phase transitions from I to N and from N to I states. To figure out this PIPT mechanism, many studies have been reported. The initial mechanism of the I-to-N PIPT has been especially eagerly studied with picosecond^{7,8} and femtosecond^{9,10} time-resolved spectroscopies. The N-to-I PIPT was also investigated to confirm that no phase transition is induced by temperature increase via the laser irradiation.¹¹ Okamoto et al.¹² investigated the dynamics of both

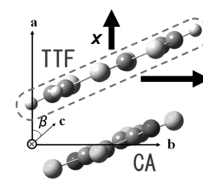


Figure 1. The structure of TTF-CA monomer and the orientations of the coordinates varied.

I-to-N and N-to-I PIPT with a femtosecond time-resolved spectroscopy and found that in the N-to-I PIPT, the one-dimensional domain in the I state is formed within 200 fs by the laser irradiation. These studies showed that the mechanisms of I-to-N and N-to-I PIPT significantly differ. The most significant difference is the threshold behavior only observed in the I-to-N PIPT.⁹ The study of dynamics of N-to-I PIPT has shown the existence of ultrafast generation of an intermediate state, where the electronic and lattice changes are strongly coupled.^{12–16}

In theoretical studies, this phase transition has been investigated on simple models and model Hamiltonians with empirical parameters. For the pressure-induced and temperature-induced phase transitions, Metzger and Torrance¹⁷ showed in semiempirical calculations that the electron transfer energy from donor to acceptor and the Madelung energy play essential roles in the N-to-I phase transition. The importance of the Madelung energy was also stressed on the model calculation with the Hartree–Fock charge distribution by Kawamoto et al.¹⁸ Soos et al.¹⁹ applied a diagrammatic valence bond theory to investigate the CT absorption spectrum. For the PIPT, Nagaosa and Ogawa²⁰ outlined a mechanism in the framework of the kinetic Ising model. Huai et al.²¹ calculated the adiabatic relaxation path on the extended Hubbard model and suggested the critical domain size in the I-to-N PIPT.

Received: December 12, 2010

Published: June 02, 2011

Yonemitsu also explained the I-to-N and N-to-I phase transitions on the extended Hubbard model.^{22–24} Further theoretical works with model Hamiltonians can be found in the review by Yonemitsu and Nasu²⁵ and references therein. These model Hamiltonian calculations have qualitatively discussed the nature of PIPT. As for the quantitative discussion, a few works have been performed on density functional theory (DFT). However, past DFT calculations^{26–29} have lacked both van der Waals (vdW) correlation and long-range exchange interactions. Molecular crystals are known to consist of weak vdW interactions. Also, long-range exchange interactions are reportedly required to give accurate CT excitations, while CT excitations have been suggested to be the main intermediates in the PIPT. Despite these inevitabilities, vdW correlation and long-range exchange have been neglected in molecular crystal calculations mainly due to their high computational cost.

In this study, we use long-range corrected time-dependent DFT (LC-TDDFT)³⁰ combined with the local response dispersion (LRD) correction to reveal the N-to-I PIPT mechanism of TTF-CA crystal. It has been suggested that the long-range correction (LC) scheme³¹ enables TDDFT to quantitatively calculate CT excitation energies and oscillator strengths, which have been significantly underestimated in conventional TDDFT calculations,³² without increasing the computational cost. It has also been reported that the LRD correction³³ gives accurate vdW bonds with no empirical parametrization by combining with the LC scheme.³⁴ We assume that electron itinerancy affects the N-to-I PIPT only through the background charges, because we suppose that in contrast to metals and semiconductors, TTF-CA crystal has only a very narrow conduction band attributing to the lowest CT excitation and therefore itinerant electrons may not directly affect the electronic states. We suppose that the initial N-to-I process of this PIPT can be treated by the single reference method because the starting point (N state) can be well-described with a single reference method. In this study, we focus on the N-to-I PIPT, because as mentioned later, we found that the I phase seems to have a multiconfigurational wave function, which is not appropriately calculated by DFT. As far as we know, this is the first calculation taking into account both long-range exchange and vdW correlation in TDDFT calculations of molecular crystals.

II. COMPUTATIONAL DETAILS

All DFT and TDDFT calculations have been carried out on the generalized Kohn–Sham method³⁵ using the LC scheme applied to Becke 1988 exchange³⁶ + one-parameter progressive correlation functional³⁷ (LC-BOP). In the LC scheme, the only parameter, μ , is set as $\mu = 0.33$, which is the optimum value in excited state calculations.³⁰ As the basis set, we use the 6-31G(d) basis functions.^{38–40} The vdW correlation energy is calculated by the LRD method.³³ In LRD calculations, we assumed that the vdW correlation energies of the CT state, $\Delta E_{\text{LRD}}^{\text{CT}}$, are well-approximated by those of the triplet state, $\Delta E_{\text{LRD}}^{\text{triplet}}$, at the same geometries, i.e.,

$$E_{\text{total}}^{\text{GS}} = E_{\text{LC-BOP}}^{\text{GS}} + \Delta E_{\text{LRD}}^{\text{GS}} \quad (1)$$

$$E_{\text{total}}^{\text{CT}} = E_{\text{LC-BOP}}^{\text{CT}} + \Delta E_{\text{LRD}}^{\text{triplet}} \quad (2)$$

This is because the triplet state has similar electron distribution to the CT state and the vdW energy of the LRD method does not depend on the spin. All the calculations are carried out using the development version of Gaussian 03 program.⁴¹ The molecular structure has been taken from the experimental

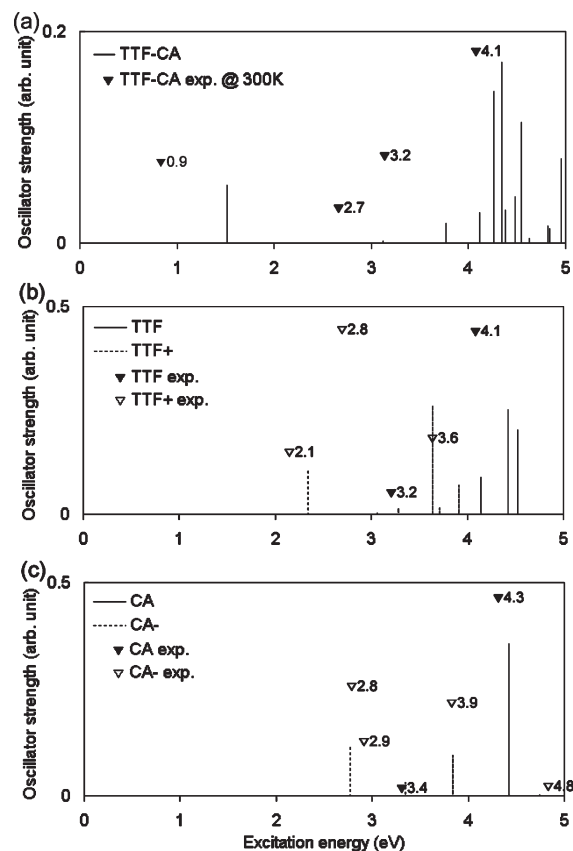


Figure 2. The calculated excitation spectrum of (a) TTF-CA monomer for the experimental structure at 90 K, (b) TTF molecule and TTF⁺ cation, and (c) CA molecule and CA⁻ anion. Experimental peaks of TTF-CA crystal at 300 K,³ TTF, TTF⁺,⁴² CA, and CA⁻⁴³ are also shown. The 6-311++G** basis set is used.

neutron diffraction data at 90 K.⁴ For simplicity, the term “TTF-CA pair” is used to represent the pair of TTF and CA molecules stacked in the *a* direction (Figure 1).

The position of the TTF molecule is varied in the *a* and *b* directions from the original position at the experimental structure, while the intramolecular structures of TTF and CA molecules are fixed at the experimental structures at 90 K.

The excitation spectrum of TTF-CA crystal is calculated using the embedded model of one TTF-CA pair in background point charges which represent surrounding TTF-CA pairs. The Peierls instability is taken into account only for the electrostatic effect by imposing sequential background point charges representing the Peierls-distorted charge distribution. The point charges are positioned on the atoms of the surrounding TTF-CA pairs on both sides of the *a* axis. The structures of neighboring pairs are set to be always parallel to the central TTF-CA pair. To fix the lattice constant, the distances between two TTF molecules are unchanged. The magnitudes of point charges are determined by the Mulliken charges of the CT state of the central TTF-CA pair. In this study, we use only the CT state charges, because we found that the charges of the ground state negligibly affect the excitation spectrum of a TTF-CA pair.

III. RESULTS AND DISCUSSION

TTF-CA Monomer. Before discussing the photoinduced phase transition mechanism of TTF-CA molecular crystal, let us

consider the electronic structure of TTF-CA monomer to make clear the molecular nature of this system. This may also be of help in examining our method for reproducing the excitation energies of such vdW complexes. Experimentally, the photoabsorption spectra of TTF-CA crystal have been observed in film and powder forms at 11 and 300 K.³

Figure 2a shows the calculated excitation spectrum of TTF-CA monomer for the experimental optimal structure at 90 K. In the figure, the lowest excitation at 1.52 eV corresponds to the CT excitation from HOMO to LUMO, which are both localized in TTF and CA, respectively, as shown in Figure 3.

Note that this CT excitation energy is higher than the lowest peak energy of TTF-CA crystal, 0.9 eV, in the experimental absorption spectrum at 300 K.³ In addition, the ionicities of the ground and excited states—calculated as the sum of Mulliken charges—are almost zero and unity, respectively. Later, we will show that this discrepancy can be explained with the electrostatic effect from the surrounding TTF-CA units. The figure also shows strong peaks for 4–5 eV, which are also observed in the experimental absorption spectrum. These peaks correspond to the excitations of constituent TTF and CA molecules, in agreement with the experimental assignments. Figure 2b,c illustrates the calculated excited spectra of TTF, TTF⁺, CA, and CA⁻. In these TDDFT calculations, the larger 6-311++G** basis set^{44–46} has been used because the higher energy excitations (over 4.0 eV) of TTF molecule was not sufficiently reproduced with the 6-31G(d) basis set. These figures indicate that as the experiment assigned, the peaks at 2.3 and 2.9 eV in the absorption spectrum of TTF-CA crystal at 11 K³ correspond to the lowest excitations of TTF⁺ and CA⁻, respectively. Experimental peak of TTF, TTF⁺, CA, and CA⁻ in CH₃CN can be found in the literature.^{42,43} The accuracies in the excitation spectra support the reliability of the present method in excited state calculations of vdW complexes. In the following calculations, smaller 6-31G(d) basis set is used because the lowest CT excitation energy, which plays the main role in PIPT, was not

significantly affected by improving basis set: the calculated CT excitation energies were 1.517 and 1.613 eV with the 6-311++G** and 6-31G(d) basis sets, respectively.

Next, we calculated the potential energy surfaces of TTF-CA monomer to clarify the structural dependence. The structures are varied for *x* and *y* coordinates given in Figure 1. Parts a and b of Figure 4 show the calculated potential energy surfaces of TTF-CA monomer for the ground and lowest CT excited states, respectively.

As shown in Figure 4a, the ground state has double minima: one is near the experimental structure (G), in which the distance between TTF and CA centers is 3.57 Å, and the other is symmetrical to this G structure (G'). These are slipped structures. Between G and G' structures, the ground state has a low barrier with 0.11 eV. In contrast, Figure 4b shows that the lowest CT state has a single minimum at around the top of the barrier, which corresponds to a symmetrical structure and has 3.39 Å for the distance between TTF and CA centers. The G and E structures are illustrated in Figure 5. The low barrier between the G and G' structures in the ground state shows that one TTF-CA pair can easily oscillate between two stable structures by the thermal fluctuations in the high-temperature. However, there are no experimental results concerning the dynamic slipping motion. This could be explained by considering a steric hindrance. Since only one TTF-CA pair is explicitly calculated in this study, we have considered no steric hindrance from neighboring TTF-CA pairs. When we explicitly calculated two TTF-CA pairs, we found that the steric hindrance between two pairs is affected by the interrelation of their structures. The steric hindrance is small if two pairs move in a coherent motion (the structures of the neighboring pairs are always parallel to the central pair). On the other hand, we found that when two pairs move randomly between G and G' structures, the steric effect is much larger than the calculated energy barrier in one pair model and both pairs tend to stay in the energetically optimal G structure. Thus,

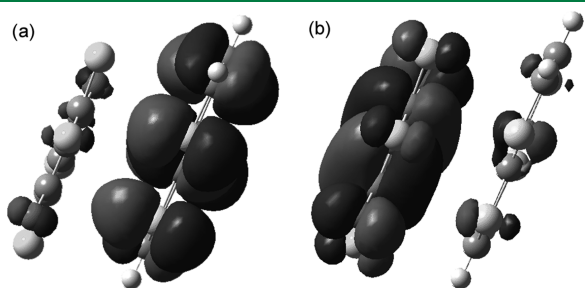


Figure 3. (a) HOMO and (b) LUMO of TTF-CA monomer.

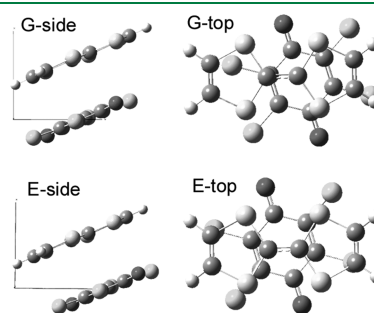


Figure 5. The side and overhead views of the G and E structures of TTF-CA monomer.

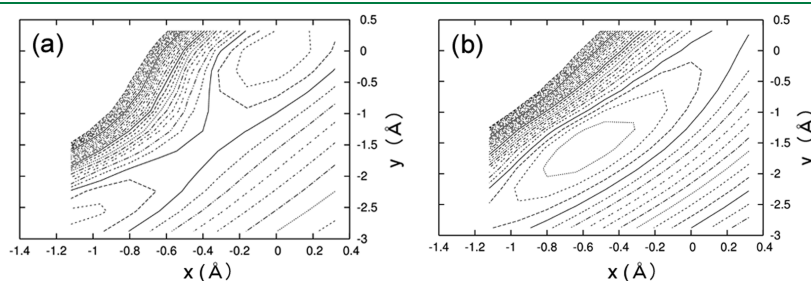


Figure 4. Calculated potential energy surfaces of TTF-CA monomer for (a) the ground state and (b) the lowest CT state. (Isovalues = 0.05 eV).

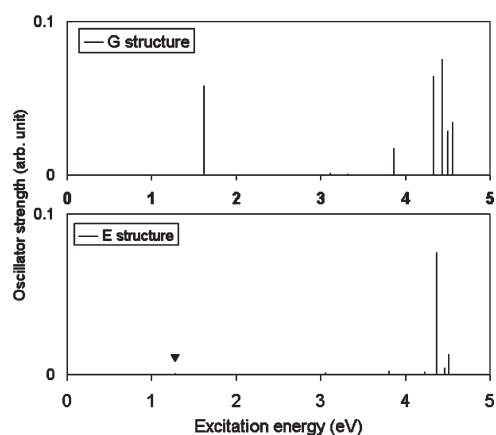


Figure 6. The calculated excitation spectra of TTF-CA monomer in the G and E structures. The black triangle indicates the position of CT excitation energy in the E structure.

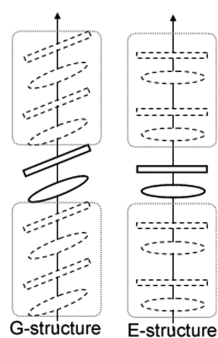


Figure 7. A possible structural change in TTF-CA crystal. Neighboring TTF-CA pairs illustrated with dotted lines are modeled with point charges in the present calculations.

we suppose that the double-well character of the ground state could not be observed unless the coherent motion of many TTF-CA pairs is realized. According to this supposition, we guess that the coherent motion of many TTF-CA pairs would be unlikely to occur in the high-temperature N phase. Note that in the case of TTF-CA crystal, the interstack hydrogen bonds—which are ignored in this study—would make the double-well potential energy surface of the ground state somewhat asymmetric. However, an experimental reflection spectrum study¹² showed that metastatic phases grow in the *a* direction in the early stage of the PIPT, indicating that the interstack bonds may not play a main role in the initial stage of the PIPT. The overhead views of these structures show that the TTF and CA centers overlap with each other in the E structure.

Figure 6 displays the calculated excitation spectra of TTF-CA monomer for the G and E structures. This figure indicates that the oscillator strength of the CT excitation drastically decreases between the G and E structures. The small oscillator strength is expected to make the de-excitation unlikely and to produce a long-lived CT excitation.

TTF-CA Molecular Crystal. On the basis of the above-mentioned features of TTF-CA monomer, let us consider the initial dynamics of PIPT in TTF-CA crystal. Considering that the optimal structure is transformed from the slipped G structure of the ground state to the symmetrical structure of the lowest CT state, we suppose that photoabsorption may change the crystal

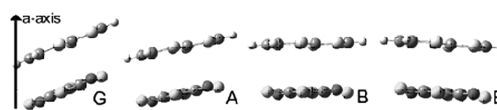


Figure 8. Calculated structures of the TTF-CA unit: G, $x = -0.04$, $y = -0.04$; A, $x = -0.16$, $y = -0.64$; B, $x = -0.32$, $y = -1.28$; E, $x = -0.58$, $y = -1.52$. Coordinates x and y are given in Figure 1.

structure for the molecular plane angle to the stacking axis, as shown in Figure 7.

This angle change is also expected to trigger a wide range of molecular plane angle shifts in neighboring TTF-CA pairs. To examine the effect of the angle change, excitation spectra are calculated for the G and E structures and two other structures between them (A and B) given in Figure 8.

The surrounding excited TTF-CA pairs—20 pairs on each side of the *a* direction—are taken into account as the point charge. Figure 9 shows the excitation spectra of (a) TTF-CA monomer and (b) TTF-CA crystal for these structures.

As shown in the figure, the structural change from G to E monotonically decreases the lowest CT excitation energy and the corresponding oscillator strength. However, for TTF-CA monomer, the structural effect is too small to make the CT excitation energy close to the experimental value (0.9 eV). On the other hand, in the crystal model with surrounding charges, the CT excitation energy is remarkably decreased even with a few pairs of the point charges of TTF-CA pairs. The CT excitation energies for the G and E structures are plotted in Figure 10 as a function of the number of surrounding point charges of TTF-CA pairs.

The structural change in TTF-CA crystal from G to E drastically decreases the CT excitation energy and makes it very close to the experimental value for the A structure. Simultaneously, the oscillator strength is decreased by the structural change from G to E to give a very small value for the E structure. This may cause the long-lived CT excitation mentioned above. The Mulliken charge ionicity of the ground state in G structure is 0.23, which is close to 0.3, the experimental estimation of ionicity in the N phase. On the other hand, the ground state ionicity at the E structure (0.02) is much lower than that of the I phase (0.6–0.7).^{47,48} This ionicity implies that the E structure does not correspond to the I phase. Notice that the experimental structure of the I phase at low temperature is not so different from that of the N phase.⁴ We therefore suppose that the I phase also has a structure close to the G structure in the direction of the slipping mode with the dimerization, although the ionicity cannot be reproduced, probably due to the neglect of explicit TTF-CA pairs.

On the calculated results, let us consider the mechanism of the N-to-I PIPT. First, we would like to emphasize that our model using one TTF-CA model with surrounding point charges quantitatively reproduces the experimental excitation spectra of TTF-CA crystal as mentioned above. We also found that this model gives the ground state optimum structure and ionicity close to those of the N phase. These results may indicate that our model is appropriate to discuss, at least the initial process of the N-to-I PIPT process of TTF-CA crystal. On the basis of the results, we propose that the initial N-to-I PIPT proceeds through the slipping motion from the G to E structures. A recent time-resolved Laue crystallography study has shown the fast formation of an intermediate structure after photoabsorption.¹⁴ This experimental result is consistent with the present result showing

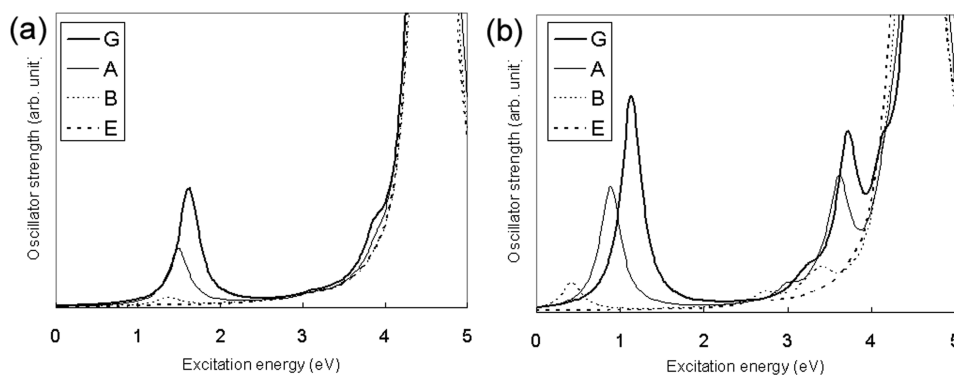


Figure 9. Calculated excitation spectra of (a) TTF-CA monomer and (b) TTF-CA crystal.

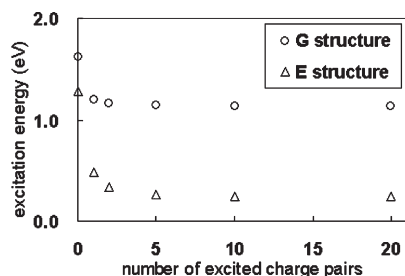


Figure 10. Calculated excitation energies of TTF-CA pair at the G and E structures with respect to the number of background excited TTF-CA pairs.

that TTF-CA transforms to the E structure, which is the optimum structure of the CT state. This would be experimentally confirmed because this transformation immediately affects the inversion symmetry of TTF-CA crystal due to the short distance between TTF and CA at the E structure, while it does not directly affect other diffraction peaks due to the local structural change. On the basis of the calculated result, we suppose that the E structure corresponds to the intermediate structure of the N-to-I PIPT observed in the time-resolved Laue crystallography.

For further analysis of the PIPT, let us consider the nature of the I phase and its generation process. Following previous experimental and theoretical studies, we assume that the I phase would be the long sequence of coherently polarized TTF-CA pairs, which is stabilized by Madelung and Peierls effects. Conceptual potential energy surfaces are illustrated in Figure 11. On the present model using one TTF-CA pair (Figure.11a), the I phase is always less stable than the N phase for any structure. Assuming that a bunch of neighboring TTF-CA pairs move in coherent motion (Figure.11b), we suppose that the I phase would get more stable than the N phase at around the G structure because of the Madelung and Peierls effects. On the other hand, we presume that at around the E structure, the N phase would be more stable than the I phase. This is because the TTF-CA pair is supposed to be less stabilized at the E structure by the background charges than that at the G structure due to the small ionicity. Therefore, we expect that the potential energy surfaces of the I and N phases intersect with each other at between the G and E structures. Although the I phase could not be obtained on the present model, the generation of the I phase could be explained on the structural and electronic changes in the present calculations.

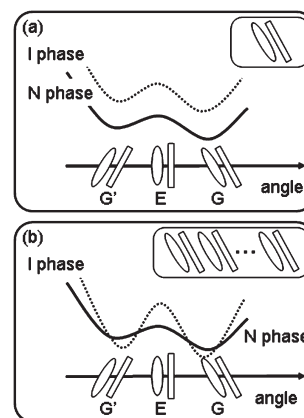


Figure 11. Conceptual potential energy curves of I and N phases along the slipping mode of TTF-CA pair (a) for one TTF-CA pair and (b) for a considerable number of coherent TTF-CA pairs.

With the above discussions, we propose that the N-to-I PIPT proceeds at the initial stage as follows (Figure 12):

- (i) The ground state of TTF-CA cluster has a shallow potential energy surface with double minima (G and G' structures). Due to the steric effect, TTF-CA pairs remain in the vicinity of one side (G) structure. The thermal lattice vibration is also centered at the G structure even at high temperature.
- (ii) At high temperature, a TTF-CA pair in the crystal is photoexcited to the lowest CT state by laser irradiation at a certain point on the PES. The CT excitation induces the drastic change in the charge distribution and consequently triggers the excitations of neighboring TTF-CA pairs.
- (iii) The sequence of excited TTF-CA pairs is then transformed into the E structure because the CT state has a single minimum at the E structure on the potential energy surface. The CT state may be long-lived due to the small oscillator strength at the E structure. Since the intermolecular distance is shortened from G to E structures, TTF-CA pairs are dimerized to break the inversion symmetry after the CT excitation. Then, PIPT may proceed through the lattice oscillation in the direction of the slip. The excited TTF-CA pairs may contribute to subsequent CT excitations of neighboring TTF-CA pairs as background charges. With these effects, the sequence of excited TTF-CA pairs is generated near the E structures.

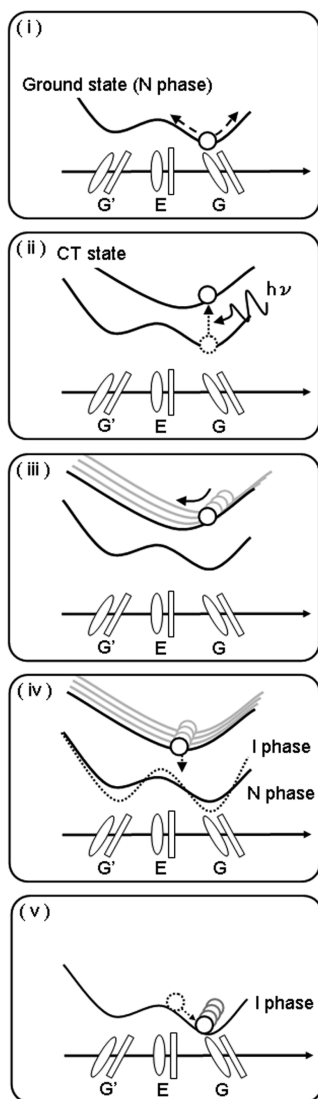


Figure 12. Conceptual potential energy curves of I and N phases along the slipping mode of TTF-CA pair in the initial process of N-to-I PIPT. Each ellipse stands for one TTF-CA pair.

- (iv) The sequence of excited (dimerized) TTF-CA pairs falls into the potential energy surface of the I phase, which is not reproduced in a small number of dimerized TTF-CA pairs. The excited TTF-CA pair at the E structure is supposed to be suitable for the generation of the sequence of TTF-CA pairs in coherent motion because of the large charge bias between TTF and CA molecules and the long lifetime of CT excitation.
- (v) According to the experimental data, the optimal structure of the I phase is supposed to be a dimerized G-like structure. For the sufficiently long sequence of the dimerized pairs, the I phase is expected to be more stable than the N phase. Otherwise, the N phase is expected to be so stable that the sequence of the dimerized pairs is broken. At high temperature, a thermal vibration hinders the generation of the sequence to make the N phase more stable.

The double-laser experiment¹³ has given further information about the characteristic of N-to-I PIPT: The maximum ratio of

photogenerated I phase is determined by the sum of the excitation density of two laser pulses and the reflectance change decays on the ps time scale, while there exists the oscillation of reflectance (the period $T = 0.62$ ps), which is nonlinear to the excitation density. The oscillation can be controlled by the delay time of two lasers: with delay time $\Delta t = T$, the oscillation is enhanced, while the oscillation disappears with $\Delta t = 1.5 T$. This oscillation can be explained with the environment change of neighboring TTF-CA pairs in process ii. This assumption is also consistent with the experimental attribution of reflectance oscillation in ref 12: the fast oscillation with the period of 0.62 ps is due to the lattice phonon corresponding to the dimeric molecular displacement, while the slow oscillation with about 50 ps is due to the shock wave.

The new mechanism explains the initial generation process of the I domain without collective excitation among several TTF-CA pairs. However, there remain several properties to be revealed in our mechanism: the electronic structure of the I phase, the crossing of the I and N PESs, and the multipair PESs with steric effects. The clear description of the I phase requires the excited state calculation of a dispersion-controlled solid state, which is costly and still challenging. For the rigorous discussion, better models containing stacked pairs would also be required. To perform such calculations, we need further computer power and theoretical development.

IV. CONCLUSION

In this study, we have carried out the excited state calculations of TTF-CA molecular crystal to clarify the initial neutral (N) to ionic (I) process of the photoinduced phase transition (PIPT) mechanism. For this purpose, we have used the state-of-the-art long-range corrected time-dependent density functional theory combined with the local response dispersion method.

First, we calculated the excitation spectrum of an isolated TTF-CA pair and succeeded in reproducing an accurate spectrum, which clearly explains the nature of each excitation. We next calculated the potential energy surfaces (PES) of the ground and lowest charge transfer (CT) excitations to make sure of the structural effect on the PIPT. As a result, we found that the ground state has a double minima on the PES and the CT state has a single minimum on the center of the double minimum as for the relative position change in a TTF-CA pair. On the basis of these results, we propose that the PIPT proceeds through the angle change of molecular planes.

To investigate the PIPT of TTF-CA crystal, we adopted an embedded TTF-CA pair in the background point charges for modeling TTF-CA crystal. Consequently, we found that the point charges of surrounding excited TTF-CA pairs obviously decrease the CT excitation energy of the TTF-CA pair and make the CT energy close to the experimental absorption energy (0.9 eV) through the angle change of the molecular planes. These point charges also increase the oscillator strength of this excitation. We therefore suggest that the CT excitation of the TTF-CA pair itself may contribute to the CT excitations of other neighboring TTF-CA pairs.

In summary, we propose that the PIPT of TTF-CA crystal proceeds as follows:

- (i) In the ground state of the TTF-CA crystal, the molecular plane angles vary around the G structure by the thermal vibration.

- (ii) By laser irradiation, the TTF-CA crystal is excited to the CT state. The electrostatic effect of the excited pair enhances the excitation of neighboring TTF-CA pairs. Thus, the sequence of excited TTF-CA pairs is generated in the *a* direction of TTF-CA (Figure 1).
- (iii) The sequence of excited TTF-CA pairs transforms into the parallelistic (E) structure on the single minimum PES in the coherent motion.
- (iv) The sequence of excited TTF-CA pairs is relaxed to the I phase surface, which is not realized in one TTF-CA pair.
- (v) The generated I phase domain varies to the optimal structure. In the N-to-I PIPT, where the temperature is high, the stability of the I phase is insufficient and the thermal vibration breaks the unstable I phase into the N phase.

AUTHOR INFORMATION

Corresponding Author

*E-mail: tsuneda@riken.jp

ACKNOWLEDGMENT

This research was supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) (Grants: 20350002 and 20038012). The numerical calculations were conducted on the RIKEN Cluster of Clusters (RICC). One of the authors (T.T.) would like to thank Canon Inc. and Hitachi Chemical Co. Ltd. for their contributions.

REFERENCES

- (1) Mayerle, J. J.; Torrance, J. B.; Crowley, J. I. *Acta Crystallogr. B* **1979**, *35*, 2988–2995.
- (2) Torrance, J. B.; Vazquez, J. E.; Mayerle, J. J.; Lee, V. Y. *Phys. Rev. Lett.* **1981**, *46*, 253–257.
- (3) Torrance, J. B.; Girlando, A.; Mayerle, J. J.; Crowley, J. I.; Lee, V. Y.; Batail, P.; LaPlaca, S. J. *Phys. Rev. Lett.* **1981**, *47*, 1747–1750.
- (4) Le Cointe, M.; Lemée-Cailleau, M. H.; Cailleau, H.; Toudic, B.; Toupet, L.; Heger, G.; Moussa, F.; Schweiss, P.; Kraft, K. H.; Karl, N. *Phys. Rev. B* **1995**, *51*, 3374–3386.
- (5) Girlando, A.; Marzola, F.; Pecile, C.; Torrance, J. B. *J. Chem. Phys.* **1983**, *79*, 1075–1085.
- (6) Koshihara, S.; Tokura, Y.; Mitani, T.; Saito, G.; Koda, T. *Phys. Rev. B* **1990**, *42*, 6853–6856.
- (7) Luty, T.; Cailleau, H.; Koshihara, S.; Collet, E.; Takesada, M.; Lemée-Cailleau, M. H.; Buron-Le Cointe, M.; Nagaosa, N.; Tokura, Y.; Zienkiewicz, E.; Ouladiaz, B. *Europhys. Lett.* **2002**, *59*, 619–625.
- (8) Guérin, L.; Collet, E.; Lemée-Cailleau, M. H.; Buron-Le Cointe, M.; Cailleau, H.; Plech, A.; Wulff, M.; Koshihara, S.; Luty, T. *Chem. Phys.* **2004**, *299*, 163–170.
- (9) Iwai, S.; Tanaka, S.; Fujimura, K.; Kishida, H.; Okamoto, H.; Tokura, Y. *Phys. Rev. Lett.* **2002**, *88*, 057402.
- (10) Tanimura, K. *Phys. Rev. B* **2004**, *70*, 144112.
- (11) Collet, E.; Lemée-Cailleau, M. H.; Buron-Le Cointe, M.; Cailleau, H.; Wulff, M.; Luty, T.; Koshihara, S.; Meyer, M.; Toupet, L.; Rabiller, P.; Techert, S. *Science* **2003**, *300*, 612–615.
- (12) Okamoto, H.; Ishige, Y.; Tanaka, S.; Kishida, H.; Iwai, S.; Tokura, Y. *Phys. Rev. B* **2004**, *70*, 165202.
- (13) Iwai, S.; Ishige, Y.; Tanaka, S.; Okimoto, Y.; Tokura, Y.; Okamoto, H. *Phys. Rev. Lett.* **2006**, *96*, 057403.
- (14) Messerschmidt, M.; Tschentscher, T.; Cammarata, M.; Meents, A.; Sager, C.; Davaasambuu, J.; Busse, G.; Techert, S. *J. Phys. Chem. A* **2010**, *114*, 7677–7681.
- (15) Uemura, H.; Okamoto, H. *Phys. Rev. Lett.* **2010**, *105*, 258302.
- (16) Guérin, L.; Hébert, J.; Buron-Le Cointe, M.; Adachi, S.; Koshihara, S.; Cailleau, H.; Collet, E. *Phys. Rev. Lett.* **2010**, *105*, 246101.
- (17) Metzger, R. M.; Torrance, J. B. *J. Am. Chem. Soc.* **1985**, *107*, 117–121.
- (18) Kawamoto, T.; Iizuka-Sakano, T.; Shimoi, Y.; Abe, S. *Phys. Rev. B* **2001**, *64*, 205107.
- (19) Soos, Z. G.; Kuwajima, S.; Harding, R. H. *J. Chem. Phys.* **1986**, *85*, 601–610.
- (20) Nagaosa, N.; Ogawa, T. *Phys. Rev. B* **1989**, *39*, 4472–4483.
- (21) Huai, P.; Zheng, H.; Nasu, K. *J. Phys. Soc. Jpn.* **2000**, *69*, 1788–1800.
- (22) Yonemitsu, K. *J. Phys. Soc. Jpn.* **2004**, *73*, 2868–2878.
- (23) Yonemitsu, K. *J. Phys. Soc. Jpn.* **2004**, *73*, 2879–2886.
- (24) Yonemitsu, K. *J. Phys. Soc. Jpn.* **2004**, *73*, 2887–2893.
- (25) Yonemitsu, K.; Nasu, K. *Phys. Rep.* **2008**, *465*, 1–60.
- (26) Oison, V.; Katan, C.; Rabiller, P.; Souhassou, M.; Koenig, C. *Phys. Rev. B* **2003**, *67*, 035120.
- (27) Giovannetti, G.; Kumar, S.; Stroppa, A.; van den Brink, J.; Picozzi, S. *Phys. Rev. Lett.* **2009**, *103*, 266401.
- (28) Ishibashi, S.; Terakura, K. *Phys. B* **2010**, *405*, S338–S340.
- (29) Filatov, M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 144–148.
- (30) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2004**, *120*, 8425–8433.
- (31) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 3540–3544.
- (32) Tawada, Y.; Chiba, M.; Tsuneda, T.; Hirao, K. *J. Chem. Phys.* **2006**, *124*, 144106.
- (33) Sato, T.; Nakai, H. *J. Chem. Phys.* **2009**, *131*, 224104.
- (34) Song, J.-W.; Tsuneda, T.; Sato, T.; Hirao, K. *Org. Lett.* **2010**, *12*, 1440–1443.
- (35) Seidl, A.; Görling, A.; Vogl, P.; Majewski, J. A.; Levy, M. *Phys. Rev. B* **1996**, *53*, 3764–3774.
- (36) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (37) Tsuneda, T.; Suzumura, T.; Hirao, K. *J. Chem. Phys.* **1999**, *110*, 10664–10678.
- (38) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (39) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (40) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
- (41) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision D.02*; Gaussian, Inc.: Wallingford, CT, 2004.
- (42) Torrance, J. B.; Scott, B. A.; Welber, B.; Kaufman, F. B.; Seiden, P. E. *Phys. Rev. B* **1979**, *19*, 730–741.
- (43) Andre, J. J.; Weill, G. *Mol. Phys.* **1968**, *15*, 97–99.
- (44) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (45) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639–5648.
- (46) Blauddau, J.-P.; McGrath, M. P.; Curtiss, L. A.; Radom, L. *J. Chem. Phys.* **1997**, *107*, 5016–5021.
- (47) Jacobsen, C. S.; Torrance, J. B. *J. Chem. Phys.* **1982**, *78*, 112–115.
- (48) Koshihara, S.; Takahashi, Y.; Sakai, H.; Tokura, Y.; Luty, T. *J. Phys. Chem. B* **1999**, *103*, 2592–2600.

A Comparative Study for Molecular Dynamics Simulations of Liquid Benzene

Cen-Feng Fu and Shan Xi Tian*

Hefei National Laboratory for Physical Sciences at the Microscale and Department of Chemical Physics, University of Science and Technology of China, Hefei, Anhui 230026, China

ABSTRACT: The classical equilibrium and nonequilibrium molecular dynamics simulations for liquid benzene, the prototypical aromatic π – π interaction system, are performed using a variety of molecular force fields, OPT-FF, AMBER 03, general AMBER force field (GAFF), OPLS-AA, OPLS-CS, CHARMM27, GROMOS 53A5, and GROMOS 53A6. The simulated results of the molecular structure and thermodynamic properties of liquid benzene are compared with the experimental data available in the literature, accounting for the superiority of each force field in the descriptions of the π – π interaction system. The OPLS-AA force field is recommended to be the best one, which reproduces quite well the properties examined in this work, while the others fail in predicting either the local structure or the thermodynamic properties. Such distinct discrepancies for the above force fields are discussed within the scheme of the pairwise interaction construction of the standard force field, which will stimulate searching for a force field with generally good quality not only in terms of microstructure descriptions but also in the predictions of the thermodynamic properties of the liquids.

1. INTRODUCTION

Interaction between π -aromatic molecules attracts great interest, owing to its crucial significance in the multidisciplinary fields such as DNA base stacking,¹ protein folding and structure,^{2–5} drug design,^{6,7} stereochemistry of organic reactions,⁸ molecular recognitions, host–guest chemistry, and crystal packing.^{9–16} In theoretical studies of complex large systems, the classical molecular dynamics (MD) and Monte Carlo (MC) simulations are frequently applied, owing to their high computational efficiencies. It is of the utmost importance to establish a procedure for qualifying the force field for weak intermolecular interactions, in particular, π – π interactions of DNA and proteins.^{1–5} Molecular clusters consisting of benzene (the typical π -aromatic molecule) have been studied extensively to reveal effects of the intermolecular aromatic π – π interactions on the cluster structures and physicochemical properties. As for the benzene dimer in the gas phase, a lot of experimental^{17–25} and theoretical^{26–34} studies on the stability of different configurations have been reported. What is more, the larger benzene clusters, e.g., trimer, tetramer, and so on, were also investigated to give a detailed description of the aromatic π – π interactions in the cluster.^{19–21,35–37} Nevertheless, the essential demand toward an understanding of π -aromatic interactions in condensed matter calls on us to explore the structure of aromatic molecule liquids. Consequently, liquid benzene, as the prototypical case, has been studied by X-ray diffraction,³⁸ neutron diffraction,^{39–41} and molecular simulations.^{42–50}

High-level quantum chemistry methods, such as the coupled-cluster singles, doubles, and perturbative triples [CCSD(T)] method and density-functional-theory- (DFT-)based symmetry-adapted perturbation method, were employed in the studies of the benzene dimer in the gas phase.^{27–34} The parallel displaced (PD) and T-shaped (T; see Figure 1) configurations were found to be nearly isoenergetic, in which the T configuration has been experimentally confirmed as the global energy minimum.^{22–25} The computationally inexpensive DFT with improvements to the deficiency in the previous exchange-correlation functionals^{51–56} also can provide an

accurate description of the weak π – π interactions. In the classical MD and MC studies of the aromatic molecule systems, the results are heavily dependent on the force fields used in the simulations. At least, the selected force fields must be able to reproduce the intermolecular π – π interaction potentials. Although many MD and MC studies on liquid benzene using the different force fields have been reported,^{42–50} there is no conclusive answer to which field force could be the best to reproduce both the liquid structure and the thermodynamic properties. Recently, Headen et al.⁴¹ obtained a full six-dimensional spatial and orientational picture of liquid benzene by means of high-resolution neutron diffraction and empirical potential structure refinement simulations. They concluded that the first neighbor shell, containing approximately 12 molecules, appears nearly isotropic; at the small molecular separation (<0.50 nm), the favored nearest neighbor geometry is the PD configuration, while at a larger molecular separation (>0.50 nm), the perpendicular Y-shaped configuration (Y; see Figure 1) is predominant.⁴¹ The T configuration, proposed as the global minimum for the gas-phase benzene dimer, occurs only as a saddle point in the liquid state.⁴¹ The force fields used in the MD simulations for liquid benzene become extremely sensitive, namely, the questions of whether they could reproduce the slightly different interaction energies (deviations less than 1 kcal/mol)^{33,34} of four dimer configurations, S (sandwich), PD, T, and Y depicted in Figure 1, and whether the subtle energetic differences of these configurations influence the structure and thermodynamic properties of liquid benzene. Therefore, it is deserving to carry out a comprehensive MD simulation study of liquid benzene with the different force fields, more importantly, evaluating the qualities of these force fields.

Up to date, there are many popular force fields used for protein, DNA, and small organic molecule systems, in particular, widely applicable and recent modified force fields such as

Received: March 28, 2011

Published: June 06, 2011

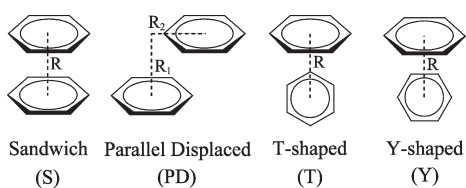


Figure 1. Sandwich (S), parallel displaced (PD), T-shaped, and Y-shaped configurations of benzene dimer.

AMBER 03,⁵⁷ general AMBER force field (GAFF),⁵⁸ OPLS-AA,^{59,60} CHARMM27,^{61,62} and GROMOS 53A5 and GROMOS 53A6.⁶³ More recently, Baker and Grant⁴⁹ developed a charge-separated all-atom force field, OPLS-CS (originated from OPLS-AA force field⁴³), for liquid benzene, on the basis of Hunter and Sanders⁶⁴ conclusion that molecular mechanics calculations without the spatial charge distribution of the π -electron system are unlikely to be successful in modeling aromatic system. Additionally, the symmetry-adapted perturbation theory (SAPT) analysis of the benzene dimer shows that the induction terms should be a relatively minor contribution to the noncovalent interactions involving π -aromatic systems, and the predominant factors are electrostatics, exchange-repulsion, and dispersion.^{29,31,32,34} This suggests that the standard force fields with point-charge electrostatics may be competent in predicting reliable results for such interactions. Hence, Sherrill et al.⁶⁵ recently derived a standard force field OPT-FF by fitting high-level *ab initio* potential curves of the benzene dimer. In the present work, eight sets of the standard force fields, OPT-FF,⁶⁵ AMBER 03,⁵⁷ GAFF,⁵⁸ OPLS-AA,⁴³ OPLS-CS,⁴⁹ CHARMM27,^{61,62} GROMOS 53A5,⁶³ and GROMOS 53A6,⁶³ are used in the MD simulation study of liquid benzene properties, including its long-range structure, the local structure, and thermodynamic properties. The simulation results will be compared with the experimental data available in the literature, aiming at evaluating the superiority of these force fields for the π -aromatic system.

2. COMPUTATIONAL METHODS

2.1. Benzene Force Fields. For the all-atom force fields used here, OPT-FF,⁶⁵ AMBER 03,⁵⁷ GAFF,⁵⁸ OPLS-AA,⁴³ OPLS-CS,⁴⁹ CHARMM27,^{61,62,66} GROMOS 53A5,⁶³ and GROMOS 53A6,⁶³ their intermolecular nonbonding interactions were described by the Lennard-Jones plus Coulomb potential:

$$U(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (1)$$

In the OPT-FF, AMBER 03, GAFF, and CHARMM27 force fields, the parameters ϵ_{ij} and σ_{ij} were calculated following the rules as $\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2}$ and $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, while for the OPLS-AA, OPLS-CS, GROMOS 53A5, and GROMOS 53A6 force fields, the parameters ϵ_{ij} and σ_{ij} were calculated following the rules as $\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2}$ and $\sigma_{ij} = (\sigma_i \sigma_j)^{1/2}$. The parameters of these force fields are listed in Table 1. The parameters of the GAFF force field were generated by ANTECHAMBER.⁶⁷ The effective charges of the GAFF force field were obtained by fitting the quantum-chemistry derived gas-phase electrostatic potentials with the restrained electrostatic potential (RESP) method.⁶⁸ As described in ref 58, both geometry optimization

and single-point calculation were reperformed at the HF/6-31G* level in this work. For the AMBER 03 force field, the effective charges were also obtained by fitting the electrostatic potential with the RESP method.⁶⁸ Here, it is noted that, as described in ref 57, the geometry optimizations were done at the RHF/6-31G** level. Subsequently, the IEFPCM continuum solvent model was used to reproduce an organic solvent environment ($\epsilon = 4$), and a single-point calculation at the B3LYP/cc-pVTZ level was performed to obtain the electrostatic potential. These quantum chemistry calculations were carried out with the Gaussian 03 suite of programs.⁶⁹

2.2. Equilibrium MD Simulations. All of the classical MD simulations were performed with the GROMACS program package.^{70,71} Initially, the system containing 600 benzene molecules was generated in a cubic box (89.0511 nm³) with the density equal to the experimental value (873.8 kg/m³).⁷² The periodic boundary condition was implemented. The system was energy-minimized for the first 5000 steps using the conjugate gradient method with one additional steepest descent step following every conjugate gradient step. Subsequently, a 10.0 ns MD simulation in the *NPT* ensemble was performed at $P = 1$ atm and $T = 298$ K with a 1 fs step to equalize the system. Finally, a 10.0 ns MD simulation in the *NPT* ensemble was carried out at $P = 1$ atm and $T = 298$ K with a 1 fs step for the product collecting simulation. The Nosé–Hoover thermostat^{73,74} with a coupling time of 0.5 ps was employed to regulate the temperature. The Parrinello–Rahman isotropic barostat^{75,76} was used to impose constant pressure with a 2.0 ps coupling time. The cutoff distance for van der Waals interactions was 1.5 nm, with a long-range dispersion correction applied for energy. The particle-mesh Ewald (PME) method⁷⁷ was employed to treat Coulomb interactions, using a 1.3 nm cutoff for real-space, a grid spacing of 0.12 nm, and a PME order of 8. All bonds were constrained using the P-LINCS algorithm.^{78,79}

The calculations of heat capacity at constant pressure ($C_p(l)$) and vaporization enthalpy (ΔH_{vap}) are performed in the usual way.⁴⁴ The intermolecular component $C_p^{\text{inter}}(l)$ of $C_p(l)$ is estimated from the fluctuation of intermolecular energy E_{inter} :

$$C_p^{\text{inter}}(l) = \frac{\langle (E_{\text{inter}} + PV)^2 \rangle - \langle E_{\text{inter}} + PV \rangle^2}{N_A k_B T^2} \quad (2)$$

where N_A is Avogadro's constant and k_B is the Boltzmann constant. $C_p(l)$ is further calculated with

$$C_p(l) = C_p^{\text{inter}}(l) + C_p^{\text{o}}(g) - R \quad (3)$$

where the heat capacity of the ideal gas $C_p^{\text{o}}(g)$ is 19.5 cal mol⁻¹ K⁻¹ at 298 K,⁴⁴ R is the gas constant, and the g in parentheses represents the gas phase. ΔH_{vap} is defined as

$$\Delta H_{\text{vap}} = E_{\text{intra}}(g) - (E_{\text{inter}}(l) + E_{\text{intra}}(l)) + RT \quad (4)$$

where E_{intra} is the intramolecular energy. Assuming $E_{\text{intra}}(l) \approx E_{\text{intra}}(g)$, the above formula can be approximated to $\Delta H_{\text{vap}} \approx -E_{\text{inter}}(l) + RT$.

The self-diffusion coefficients are determined by the Einstein relation⁸⁰

$$D = \lim_{t \rightarrow \infty} \frac{1}{6t} \langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle \quad (5)$$

where $\mathbf{r}(t)$ is the center of mass vector position of each molecule at time t , $\mathbf{r}(0)$ is the position at $t = 0$, and D is the self-diffusion coefficient.

Table 1. Force Field Parameters for Benzene Used in This Work

parameter	OPT-FF ⁶⁵	AMBER 03 ⁵⁷	GAFF ⁵⁸	OPLS-AA ⁴³	OPLS-CS ⁴⁹	CHARMM27 ^{61,62}	GROMOS 53A5 ⁶³	GROMOS 53A6 ⁶³
R_{CC} (Å)	1.40	1.40	1.387	1.40	1.40	1.375	1.39	1.39
R_{CH} (Å)	1.08	1.08	1.087	1.08	1.08	1.080	1.09	1.09
$R_{C\pi}$ (Å)					0.90			
θ_{CCC} (deg)	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0
θ_{CCH} (deg)	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0
$\theta_{CC\pi}$ (deg)					90.0			
Ψ_{CCCC} (deg)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ψ_{CCCH} (deg)	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0
$\Psi_{CC\pi}$ (deg)					90.0/−90.0			
q_C (e)	−0.134	−0.121866 ^a	−0.130889 ^b	−0.115	0.1435	−0.115	−0.146	−0.140
q_H (e)	0.134	0.121866 ^a	0.130889 ^b	0.115	0.1435	0.115	0.146	0.140
q_π (e)					−0.1435			
σ_C (Å)	3.42462	3.39967	3.39967	3.55	3.69	3.55005	3.58118	3.58118
σ_H (Å)	2.19161	2.59964	2.59964	2.42	2.52	2.42004	2.37341	2.37341
ϵ_C (kcal/mol)	0.115	0.086	0.086	0.07	0.07	0.07	0.06630	0.06630
ϵ_H (kcal/mol)	0.011	0.015	0.015	0.03	0.03	0.03	0.02829	0.02829

^a RESP charges obtained at the B3LYP/cc-pVTZ//RHF/6-31G** level. ^b RESP charges obtained at the HF/6-31G*/HF/6-31G* level.

The reorientational dynamics of the liquid were also analyzed. The rotational correlation times (τ) were computed as described by Bonnaud et al.⁵⁰ The rotational correlation times are divided into two cases, $\tau_{2\perp}$ and $\tau_{2\parallel}$, that correspond to the relaxation times of the unitary vectors perpendicular \hat{e}_\perp and parallel \hat{e}_\parallel to the plane of the benzene molecule, respectively (see Figure 2 in ref 50). Two rotational autocorrelation functions can be defined as

$$C_l^i(t) = \langle P_l(\hat{e}_i(t) \cdot \hat{e}_i(0)) \rangle \quad (6)$$

where \hat{e}_i is \hat{e}_\perp or \hat{e}_\parallel as defined above and P_l is a Legendre polynomial with $l = 1$ or 2 for the first and second orders, respectively. Then, these functions are fitted in the format of the Kohlrausch–Williams–Watts exponential function:^{81,82}

$$C_{i,KWW}^l(t) = \exp[-(t/\alpha_{li})^{\beta_{li}}] \quad (7)$$

where $0 < \beta_{li} \leq 1$. After fitting, the rotational correlation time can be obtained as

$$\tau_{il} = \int_0^\infty C_{i,KWW}^l(t) dt = \frac{\alpha_{li}}{\beta_{li}} \Gamma\left(\frac{1}{\beta_{li}}\right) \quad (8)$$

2.3. Nonequilibrium MD Simulations. Nonequilibrium MD simulations were performed to determine the shear viscosity (η) of liquid benzene. A spatially periodic forcing function is imposed on the system, and the shear viscosity can be determined from the response of the system to the applied function. The behavior of the velocity $\mathbf{u}(\mathbf{r}, t)$ of a particle in the liquid is described by the Navier–Stokes equation:⁸³

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho(\mathbf{u} \cdot \nabla) \mathbf{u} = \rho \mathbf{a} - \nabla p + \eta \nabla^2 \mathbf{u} \quad (9)$$

where \mathbf{a} is the external force per unit of mass and volume and ρ is the fluid density. A periodic external force \mathbf{a} is used in cosine form

$$a_x(z) = A \cos(kz) \quad (10)$$

where A is the applied acceleration, l_z is the length of the simulation box along the z direction, and $k = 2\pi/l_z$. The

steady-state velocity profile is in the following form by integrating the Navier–Stokes equation:

$$u_x(z) = V(1 - e^{-t/\tau_r}) \cos(kz) \quad (11)$$

$$V = A\tau_r = A \frac{\rho}{\eta k^2} \quad (12)$$

where τ_r is the macroscopic relaxation time of the liquid. In an MD simulation, the instantaneous $V(t)$ is defined as

$$V(t) = 2 \sum_{i=1}^N m_i v_{i,x}(t) \cos(kr_{i,z}(t)) / \sum_{i=1}^N m_i \quad (13)$$

where $v_{i,x}$ is the x component of the velocity, $r_{i,z}$ is the z coordinate, and m_i is the mass of atom i . Then, the viscosity η can be obtained from eq 14 by calculating V directly from the simulations:

$$\eta = \frac{A\rho}{V k^2} \quad (14)$$

In the nonequilibrium MD simulations, the initial configurations were generated at the equilibrated density for each force field. As mentioned by Hess,⁸³ the wavelength of the imposed acceleration should be at least an order of magnitude larger than the simulation box; three boxes were stacked in the z direction, making the system 2 times larger than that used in the equilibrium MD simulations. Wensink et al.⁸⁴ concluded that the shear viscosity calculated from the nonequilibrium MD simulation in the NVT ensemble was consistent with that calculated in the NPT ensemble. Thus, a 5.0 ns simulation for each force field was performed in the NVT ensemble. An acceleration of 0.01 nm/ps² was imposed on each atom of the system in the x direction. The temperature was coupled to 298 K, using a Berendsen⁸⁵ thermostat with a coupling time of 0.1 ps. The treatments of van der Waals and Coulomb interactions were identical to that for the equilibrium MD simulations. In the simulation analyses, the first 500 ps for each simulation were dropped.

Table 2. Equilibrium Energy and Distance of the Different Configurations of the Benzene Dimer

	sandwich (S)		parallel displaced (PD)			T-shaped		Y-shaped	
	E (kcal/mol)	R (Å)	E (kcal/mol)	R ₁ (Å)	R ₂ (Å)	E (kcal/mol)	R (Å)	E (kcal/mol)	R (Å)
OPT-FF	-1.95	3.70	-2.48	3.50	2.50	-2.92	4.90	-2.90	4.84
AMBER 03	-1.74	3.64	-2.23	3.50	2.40	-2.09	5.10	-2.18	5.00
GAFF	-1.48	3.68	-2.18	3.50	2.66	-2.19	5.10	-2.26	5.00
OPLS-AA	-1.69	3.78	-2.10	3.50	2.68	-2.15	5.10	-2.24	5.02
OPLS-CS	^a	^a	-2.89	3.50	4.00	-3.77	5.14	-3.99	5.02
CHARMM27	-1.83	3.76	-2.22	3.50	2.62	-2.11	5.14	-2.23	5.04
GROMOS 53A5	-0.72	3.94	-1.95	3.50	3.34	-2.44	5.06	-2.43	4.98
GROMOS 53A6	-0.88	3.90	-1.94	3.50	3.28	-2.37	5.08	-2.37	4.98
estd. CCSD(T)/CBS(Δ ha(DT)Z) ^b	-1.70	3.9	-2.71	3.5	1.7	-2.70	5.0		
DFT-SAPT/aug-cc-pVTZ+mb ^c	-1.782	3.816	-2.683	3.480	1.841	-2.698	4.970	-2.441	5.009

^a The sandwich conformation predicted with the OPLS-CS force field is a repulsive state. ^b Cited from ref 33. ^c Cited from ref 34.

3. RESULTS AND DISCUSSION

3.1. Equilibrium Configurations of Benzene Dimer. The interaction energy of the benzene dimer in the gas phase may give a basis to evaluate the different force fields used in the present MD simulations. As listed in Table 2, the interaction energies of benzene dimers in the different configurations are calculated with the force fields and compared to the high-level quantum chemistry results.^{33,34} The equilibrium distances, R , R_1 , and R_2 , are defined in Figure 1. For the S configuration, the interaction energies predicted with the OPT-FF, AMBER 03, OPLS-AA, and CHARMM27 force fields are close to the high-level quantum chemistry results,^{33,34} with the percent deviations less than 15%. For the GROMOS 53A5 and GROMOS 53A6 force fields, the equilibrium interaction energies are much lower than the high-level quantum chemistry results.^{33,34} It is worth noting that the OPLS-CS force field predicts a repulsive state for the S configuration, due to the separated charges out of the molecular plane that were used to represent the π electrons.⁴⁹ As for the PD configuration, the interaction energies predicted with the OPT-FF and OPLS-CS force fields are consistent with the high-level quantum chemistry results,^{33,34} with about 9% and 7% deviations, respectively. However, the equilibrium distance R_2 for the OPLS-CS force field is too large, again due to the separated charges used in its potential function. The AMBER 03, GAFF, OPLS-AA, and CHARMM27 force fields are quite good at the prediction of the interaction energies, while a little worse for the GROMOS 53A5 and GROMOS 53A6 force fields. The interaction energy and R value for the T configuration predicted with the OPT-FF force field are in excellent agreement with the quantum chemical results,^{33,34} while the others, except for OPLS-CS, underestimate the interaction energy but overestimate the equilibrium distance R , with respect to the high-level quantum chemistry results.^{33,34} It is more remarkable that an overestimated energy with about 40% percent deviation is given with the OPLS-CS force field. Regarding to the Y configuration, within 1% and 3% deviations, respectively, the GROMOS 53A5 and GROMOS 53A6 force fields give perfect estimations in comparison with the DFT-SAPT results.³⁴ The interaction energies predicted with the AMBER 03, GAFF, OPLS-AA, and CHARMM27 force fields are lower with values of 0.2–0.3 kcal/mol, but those for the OPT-FF and OPLS-CS force fields are quite larger (their percent deviations are 19% and 63%, respectively).

Generally, the results of the four configurations of the benzene dimer predicted with the AMBER 03, GAFF, OPLS-AA, and CHARMM27 force fields are comparable to each other and acceptable in comparison with the high-level quantum chemical results.^{33,34} Before going on to the details about the effects of such differences on the structural and thermodynamic properties of liquid benzene, we should recall the original characteristics of some force fields. The standard molecular force fields are usually obtained by fitting the *ab initio* pairwise interaction potential energies, while some of them also were adjusted according to the experimental thermodynamic properties of liquids, e.g., OPLS.⁴³ To qualify the OPLS performances, Jorgensen et al. carried out an MC simulation study of liquid benzene and benzene derivatives, indicating the reliability in thermodynamics predictions.⁴⁴ Considering the quadrupole–quadrupole interactions arising from the aromatic π electrons of two benzene molecules, the OPLS parameters were further modified to be OPLS-CS.⁴⁹ On the other hand, the CHARMM, AMBER, OPLS-AA, and MM3 force fields were examined in a comparison of the intermolecular interaction potentials of the benzene dimer.⁶⁵ In that work, the OPT-FF force field used here was obtained with a much better fitness to the *ab initio* results.⁶⁵ However, as pointed out by those authors, the fitted Lennard-Jones parameters hardly reproduce the *ab initio* potentials for four typical configurations of the benzene dimer, and the more flexible electrostatic force fields and the polarization terms should be considered.⁶⁵ The computational demand will be enhanced for the sophisticated force fields including complex parameters. It is appealing that the deficiencies in describing the intermolecular interactions with the standard force fields do not always lead to the fateful failure.⁸⁶ Therefore, in this work, eight sets of the standard force fields, OPT-FF,⁶⁵ AMBER 03,⁵⁷ GAFF,⁵⁸ OPLS-AA,⁴³ OPLS-CS,⁴⁹ CHARMM27,^{61,62} GROMOS 53A5,⁶³ and GROMOS 53A6,⁶³ that were obtained under different specific backgrounds, are worth being further examined in the investigation for the nature of liquid benzene.

3.2. Spatial Structures and Aromatic π – π Interactions. To obtain the overall features of the liquid structure with the different force fields, the radial distribution functions $g(r)$ of both the center of mass (COM) and the carbon atoms of two benzene molecules are calculated and compared with the experimental results cited from refs 38 and 41. As shown in Figure 2, one can find that the deviations from the experimental data are distinctly large for the OPLS-CS and OPT-FF force fields,

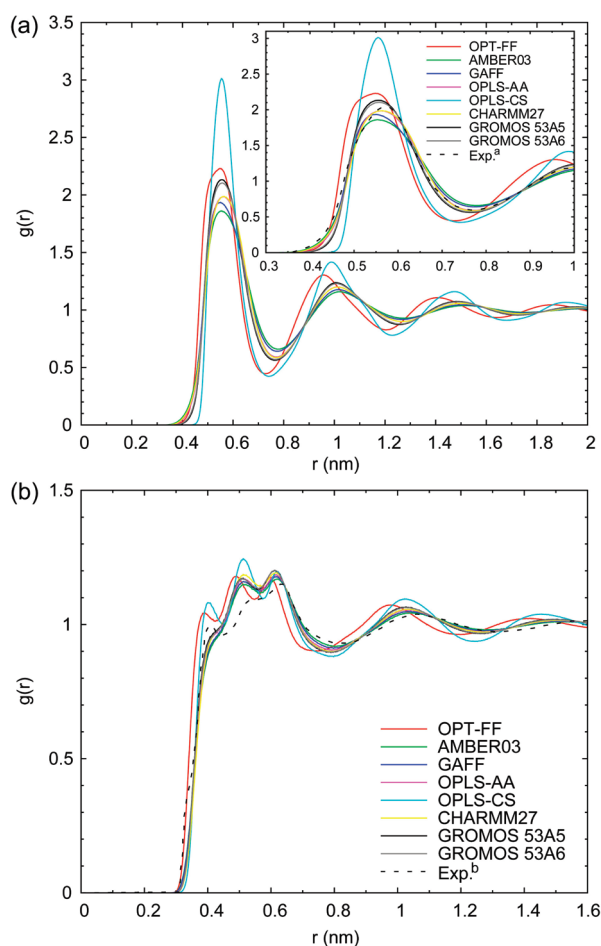


Figure 2. The radial distribution functions for the center of mass of benzene molecules (a) and the C–C atoms of benzene molecules (b). (a) Experimental data cited from ref 41. (b) Experimental data cited from ref 38.

although the OPLS-CS simulations were previously proved to be much better than the OPLS results.⁴⁹ In the inside panel of Figure 2a, one can find more details in comparison with the experimental data⁴¹ for the first shell of liquid benzene. Around the first peak of $g(r)_{\text{COM-COM}}$ (ca. 5.5 Å), two GROMOS force fields exhibit almost identical profiles of $g(r)$, while the AMBER 03 and GAFF force fields predict lower $g(r)$ values. The results of OPLS-AA and CHARMM27 are much closer to the experimental data, and the similar good agreements for $g(r)_{\text{C-C}}$ (see Figure 2b) can be extended to the long-range structures of liquid benzene. In Figure 2b, the OPLS-CS and OPT-FF force fields predict well the first small peak of $g(r)_{\text{C-C}}$, but their simulated $g(r)_{\text{C-C}}$ values are obviously larger than the experimental data.³⁸ All of the simulated results overestimate the intensity of $g(r)_{\text{C-C}}$ at the second and third peaks, but except for OPT-FF force field, the predictions for the positions of the second and third peaks are acceptable. In the long-range region, two diffuse peaks predicted with the OPT-FF and OPLS-CS force fields shift to smaller $r(\text{C-C})$ values, while the other force fields provide quite satisfying results. In general, both OPLS-CS and OPT-FF failed in predicting the liquid structures, although the OPLS-CS force field was always recommended⁴⁹ and applied in simulations of benzene and benzene-like liquids,⁴¹ and the OPT-FF force field can reproduce the benzene–benzene interaction

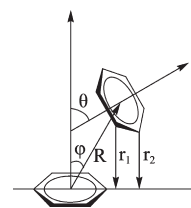


Figure 3. Diagram showing the definitions of θ , φ , R , r_1 , and r_2 .

potentials well.⁶⁵ Here, the other six sets of force fields, AMBER 03, GAFF, OPLS-AA, CHARMM27, GROMOS 53A5, and GROMOS 53A6, are proved to be reliable in the simulations of not only the local structures but also the long-range structures of liquid benzene.

To gain more features of the local structure, the populations of the different configurations of the benzene dimer in the first liquid shell are calculated. At first, five variables, θ , φ , R , r_1 , and r_2 as defined in Figure 3, are used as the criteria to identify the different configurations: R is the distance between the COMs of two benzene molecules; θ is the angle between the two benzene planes; φ is the angle formed by the normal vector pointing from the COM of the bottom benzene to the COM of another benzene; r_1 and r_2 are the two smallest projection distances of two vicinal carbon atoms in the upper benzene onto the molecular plane of the bottom benzene. The criteria for the parallel, perpendicular, T, and Y configurations are given as follows: Parallel (including S and PD configurations), $R \leq R_{\text{min}}$, $0^\circ \leq \theta \leq 40^\circ$; Perpendicular (including T and Y shapes), $R \leq R_{\text{min}}$, $50^\circ \leq \theta \leq 90^\circ$, $0^\circ \leq \varphi \leq 40^\circ$ or $50^\circ \leq \varphi \leq 90^\circ$, where R_{min} is the r value for the first minimum of $g(r)_{\text{COM-COM}}$ outside of the first shell (see Figure 2a), and it is varied with the different force fields. T-shape and Y-shape are further classified with $|r_1 - r_2| \geq R_{\text{CC}}/4$ and $|r_1 - r_2| < R_{\text{CC}}/4$, respectively (R_{CC} is the carbon–carbon bond length of benzene). In Table 3, the differences of R_{max} (the peak position) are less than 0.27 Å for the different force fields, while the differences of R_{min} are less than 0.24 Å, with respect to the corresponding experimental values.⁴¹ The R_{max} value predicted with the CHARMM27 force field and the R_{min} value predicted with the OPLS-CS force field are in the best agreement with the respective experimental values.⁴¹ The coordination numbers predicted with all force fields are generally in accord with the experimental value of 12.⁴¹ The OPLS-CS force field predicts the largest population of the perpendicular configurations (66.89%) and the smallest population for the parallel configurations (15.42%). Here, the parallel configuration is significantly disfavored, due to the separated charges of this force field. Since the larger interaction energies for the perpendicular (T and Y) configurations were calculated with high-level quantum chemistry methods,^{33,34} the OPT-FF force field on the basis of the fittings to the *ab initio* results⁶⁵ certainly predicts the second largest population for the perpendicular configurations (59.59%) and the second smallest population for the parallel configurations (21.84%). The population sequence of the perpendicular configurations for the other force fields, GROMOS 53A5 \sim GROMOS 53A6 $>$ OPLS-AA \sim GAFF $>$ CHARMM27 $>$ AMBER 03, is consistent with the interaction energy order calculated with these force fields (see Table 2). The population distributions of the parallel configurations are in a similar scenario. For the S and PD configurations, since the interaction energies predicted with two GROMOS force fields are much smaller than those with the AMBER 03, GAFF,

Table 3. Location of the maxima (R_{\max}) and the End of the First Peak (R_{\min}) of the Radial Distribution Function $g(r)_{\text{COM}-\text{COM}}$, the Coordination Number for the First Shell, and the Populations of the Different Configurations of the Benzene Dimer in the First Liquid Shell

	R_{\max} (Å)	R_{\min} (Å)	coordination number				
			($r \sim 0.0 - R_{\min}$ Å)	parallel (%)	perpendicular (%)	T-shaped (%)	Y-shaped (%)
OPT-FF	5.48	7.26	12.8	21.84	59.59	25.98	33.60
AMBER 03	5.54	7.74	12.2	24.44	56.33	25.44	30.89
GAFF	5.50	7.74	12.4	24.04	56.83	25.80	31.03
OPLS-AA	5.62	7.68	12.4	23.72	57.07	25.77	31.30
OPLS-CS	5.54	7.38	12.5	15.42	66.89	27.90	38.99
CHARMM27	5.64	7.68	12.5	24.22	56.50	25.61	30.89
GROMOS S3A5	5.54	7.64	12.5	22.25	58.96	26.45	32.50
GROMOS S3A6	5.58	7.66	12.5	22.55	58.58	26.37	32.22
exptl. ⁴¹	5.75	7.50	~12				

OPLS-AA, and CHARMM27 force fields, the populations of the parallel configurations are relatively minor for the former force fields. It is interesting that all force fields predict isoenergetic interactions for the T- and Y-shaped dimers, but the populations of T and Y configurations in the liquid are distinctly different. For all of the force fields, the average ratio of the population, $Y/T \approx 1.25:1$, is consistent with the experimental conclusion that the Y configuration is more favored than the T configuration in liquid benzene.⁴¹ Generally, the population of the perpendicular configuration is 1 to 2 times more than that of the parallel configuration in the first shell. However, as pointed out by Headen et al.,⁴¹ at a small molecular separation (<0.50 nm), the PD configurations are preferred, while at a larger separation (>0.50 nm), the neighboring aromatic rings are predominantly perpendicular to each other. The spatial scale of the first liquid shell discussed above (≤ 0.75 nm) is just beyond the sensitive criterion of 0.50 nm given by Headen et al.⁴¹ Therefore, to further analyze the different configuration populations dependent on the molecular separations in the first shell, we explore the molecular orientations and anisotropy in the first shell.

In Figure 4, the number of benzene molecules in the first shell is plotted as a function of the angle θ between the two molecular planes. The molecular orientations of the first shell of benzene have been proved to be nearly isotropic in the experiment,⁴¹ which is successfully reproduced in the present simulations with the AMBER 03, GAFF, OPLS-AA, CHARMM27, GROMOS S3A5, and GROMOS S3A6 force fields. In Figure 4, only the results of the OPT-FF (the upper panel in Figure 4a) and OPLS-CS (the upper panel in Figure 4e) force fields are distinctly different, indicating the large deviations from the isotropism at $\theta \sim 90^\circ$. When we turn back to the argument of whether the preferred parallel $\pi-\pi$ contacts (PD) at the small molecular separation (<0.50 nm)⁴¹ could be reproduced in the simulations, in the middle panels of Figure 4a–h, one can find that the AMBER 03, GAFF, OPLS-AA, and CHARMM27 force fields used in the simulations are approved to be reliable, supporting the experimental result;⁴¹ the OPT-FF and OPLS-CS results are significantly different, violating Headen's conclusion;⁴¹ the nearly isotropic distribution is found in the range of θ from $0-20^\circ$ for two GROMOS force fields, implying homogeneous configurations on this spatial scale. When the molecular separation is more than 0.50 nm (see the below panels), all of the simulations show that the perpendicular configuration is favored; in particular, the OPLS-CS model exhibits the largest deviation from the isotropic distribution. The molecular orientations in the

liquid must be closely related to the intermolecular interactions. In Figure 5, the contour maps of the benzene–benzene interaction energy can let readers more easily catch the major characteristics of the different force fields. In the calculations of these benzene–benzene interaction energy potentials, the benzene molecule at the bottom (see Figure 3) is fixed while the upper molecule rotates by changing θ from 0 to 90° . Within the interaction energy range of -4.5 to 6.0 kcal/mol, two basins (two local existing minima that correspond to the parallel and perpendicular configurations) are clearly shown on the potential energy surfaces obtained with the OPT-FF, AMBER 03, OPLS-AA, and CHARMM27 force fields. At a small molecular separation ($R < 0.50$ nm), only the perpendicular configurations (T and Y) are permitted for OPLS-CS. As for the OPT-FF force field, its potential well for the perpendicular configuration having a considerable large interaction energy (-2.5 kcal/mol) extends to the small molecular separation (<0.50 nm). This leads to the predominant perpendicular orientations in the small molecular separation region (<0.50 nm) at about 60° , corresponding to a maximum of the $N(\theta)$ values in the middle panel of Figure 4a.

What is the physical factor playing a role in the preference of the parallel configuration at a small separation? According to the compositions of these force fields, see eq 1, the total potential energy at a separation distance of 0.45 nm (plotted in Figure 6a) is further decomposed to the Coulomb contribution (see Figure 6b), the intermolecular Lennard-Jones interactions between carbon and carbon atoms in two benzenes (see Figure 6c), the intermolecular Lennard-Jones interactions between carbon atoms in one benzene and hydrogen atoms in the other (see Figure 6d), and the intermolecular Lennard-Jones interactions between hydrogen and hydrogen atoms in two benzenes (see Figure 6e). As shown in Figure 6b, one can find that the strong Coulomb attractions lead to a preference of the perpendicular configuration for the OPLS-CS force field. For the OPT-FF force field, the remarkable large Lennard-Jones attraction in whole range of θ , shown in Figure 6c, mainly arises from the largest ϵ_C of this force field (see Table 1); in Figure 6d, the much weaker Lennard-Jones repulsive interaction around $\theta \approx 90^\circ$ is due to the smallest values of σ_H and ϵ_H (see Table 1). The intermolecular Lennard-Jones attractions (-1.2 to -1.7 kcal/mol, see Figure 6c) between carbon atoms should be responsible for the formation of the parallel configuration; moreover, as shown in Figure 6d, the perpendicular configuration is disfavored due to the relatively strong repulsive Lennard-Jones interactions between carbon atoms in one benzene and hydrogen atoms in another benzene. As shown in

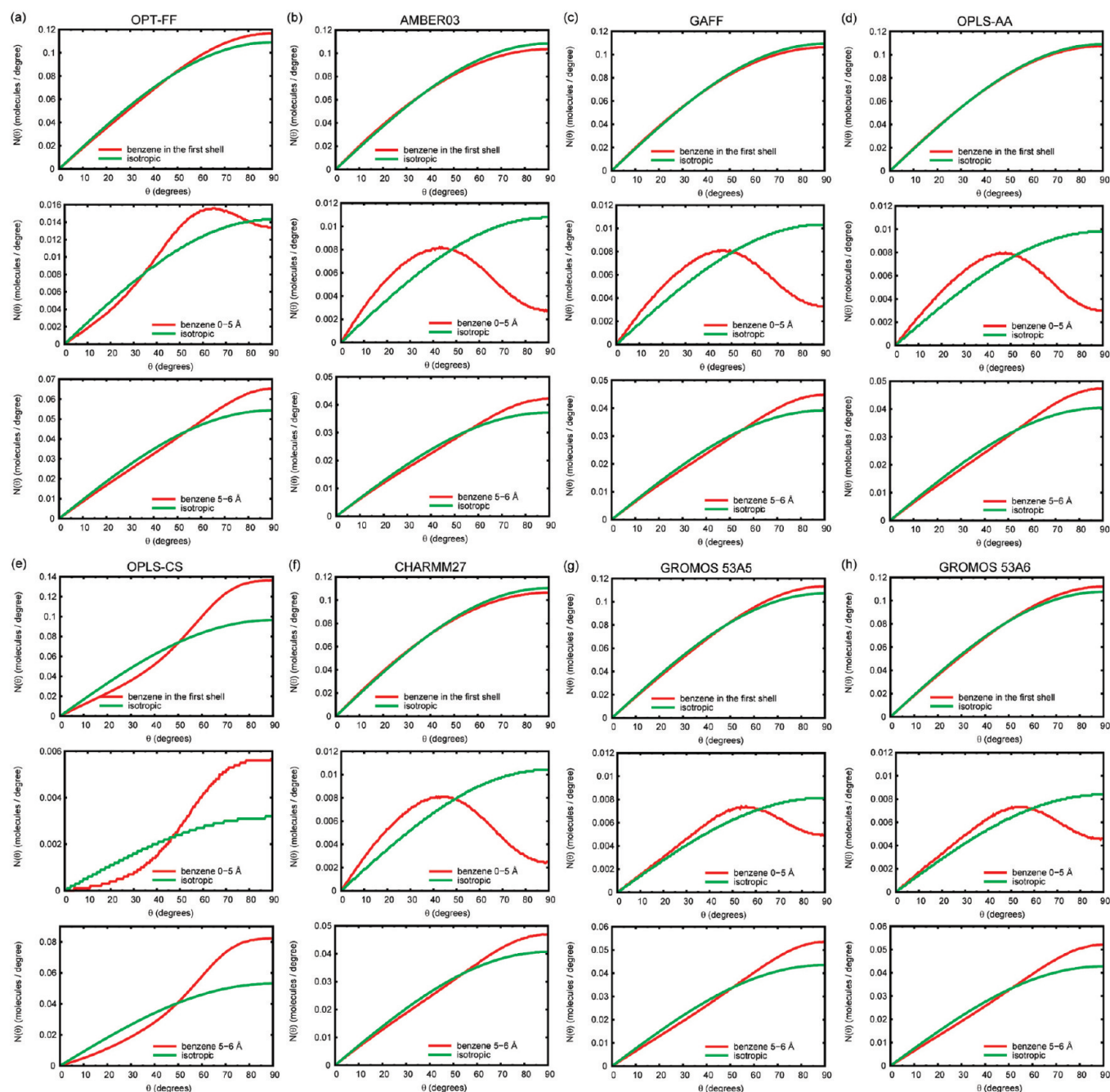


Figure 4. Number of benzene molecules in the first coordination shell as a function of the angle θ between the aromatic planes (red lines). The green lines represents a random isotropic distribution of molecules.

Figure 6e, the intermolecular Lennard-Jones attractions between hydrogen atoms are too weak (<0.2 kcal/mol) to influence the orientations of the benzene molecule.

3.3. Thermodynamic Properties. The accuracy of the structure calculated from MD simulations can discern the excellence of the different force fields. On the other hand, the thermodynamic properties predicted by MD simulations are also important in evaluating the superiority of the force fields. In Table 4, the density (ρ), molecular volume (V), heat capacity at constant pressure ($C_p(l)$), vaporization enthalpy (ΔH_{vap}), shear viscosity (η), self-diffusion coefficient (D), and rotational correlation times of vectors perpendicular and parallel to the aromatic

plane ($\tau_{2\perp}$ and $\tau_{2\parallel}$) for the different force fields used in the simulations are listed and compared with the experimental values cited from refs 72 and 87–90. The statistic errors of the theoretical values are also estimated with the methods proposed before.⁸⁰ The density, $\rho = 870.4 \pm 0.6$ kg/m³, predicted with the CHARMM27 force field is in excellent agreement with the experimental result, while the AMBER 03 result is the smallest. The results of the OPLS-AA, GAFF, and two GROMOS force fields can also provide good results, within percent deviations of less than 3%. The percent deviations for the OPLS-CS and OPT-FF force fields reach 8.4% and 19.6%, respectively. The overestimated density of the OPT-FF force field may be attributed to

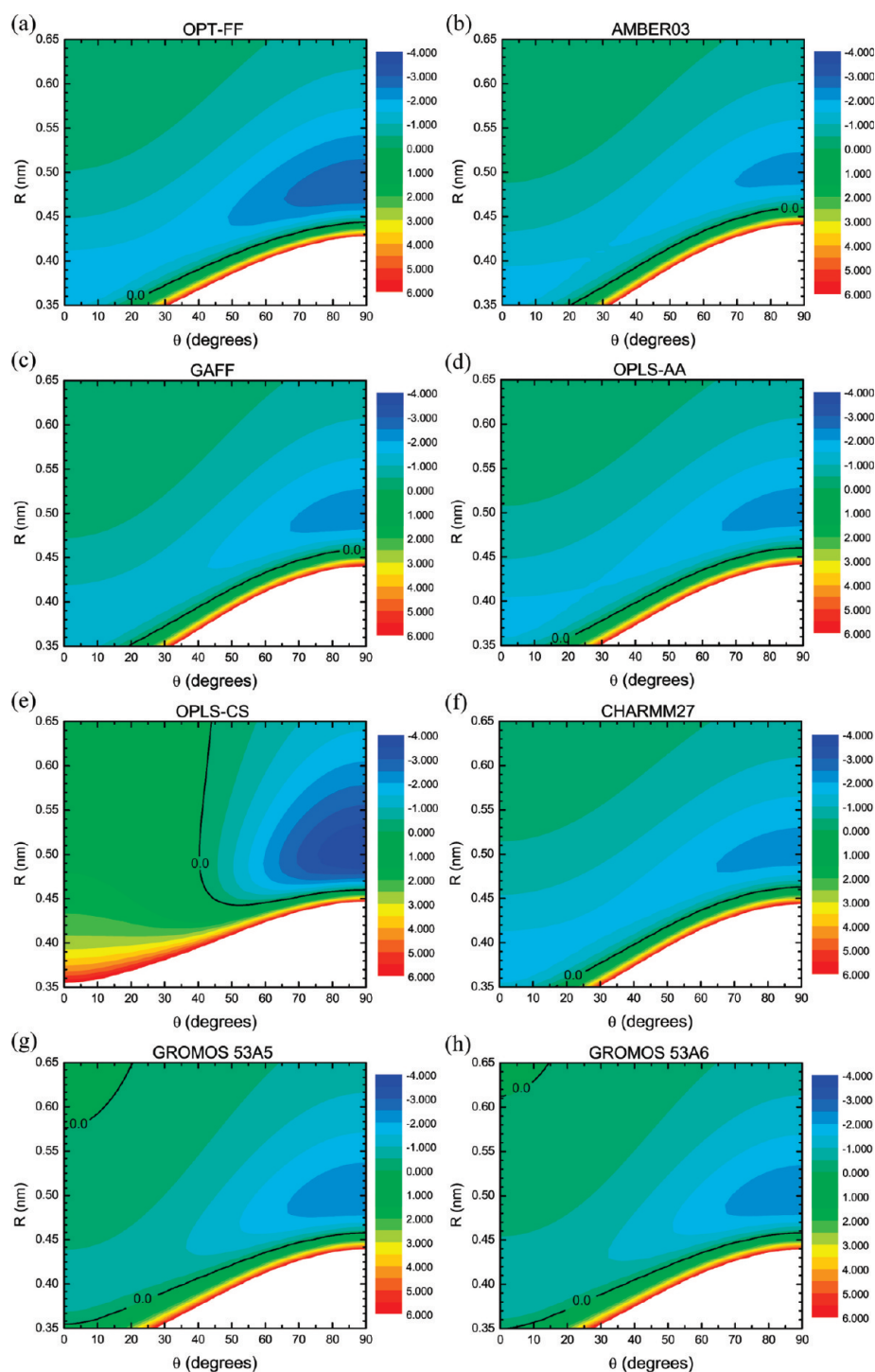


Figure 5. Contour maps ($\theta \sim R_{\text{COM-COM}}$, $\varphi = 0^\circ$) of the interaction potential energy surfaces of the benzene dimer. The black lines represent the contour line of 0 kcal/mol.

its quite small σ_H (see Table 1), which leads to a significant decrease of the molecular volume ($V = 124.2 \pm 0.1 \text{ \AA}^3$).

The $C_p(l)$ and ΔH_{vap} values predicted with the OPLS-AA force field match the experimental data very well.^{72,87} Within the computational uncertainty of $C_p(l)$, the result for the OPT-FF force field is also acceptable. The $C_p(l)$ values of the CHARMM27 and two GROMOS force fields are a little larger than the experimental datum,⁸⁷ but the largest one, $39.1 \pm 1.2 \text{ cal mol}^{-1}$

K^{-1} , is predicted with the GAFF force field, and the smallest one, $28.6 \pm 0.5 \text{ cal mol}^{-1} \text{ K}^{-1}$, is estimated with the OPLS-CS force field. The two largest ΔH_{vap} values are predicted with the OPT-FF and OPLS-CS force fields, while the results with the AMBER 03 and GAFF force fields are much smaller than the experimental datum.⁷² However, the predictions of ΔH_{vap} for most force fields except for the OPT-FF and OPLS-CS force fields are acceptable because of the large experimental uncertainty of

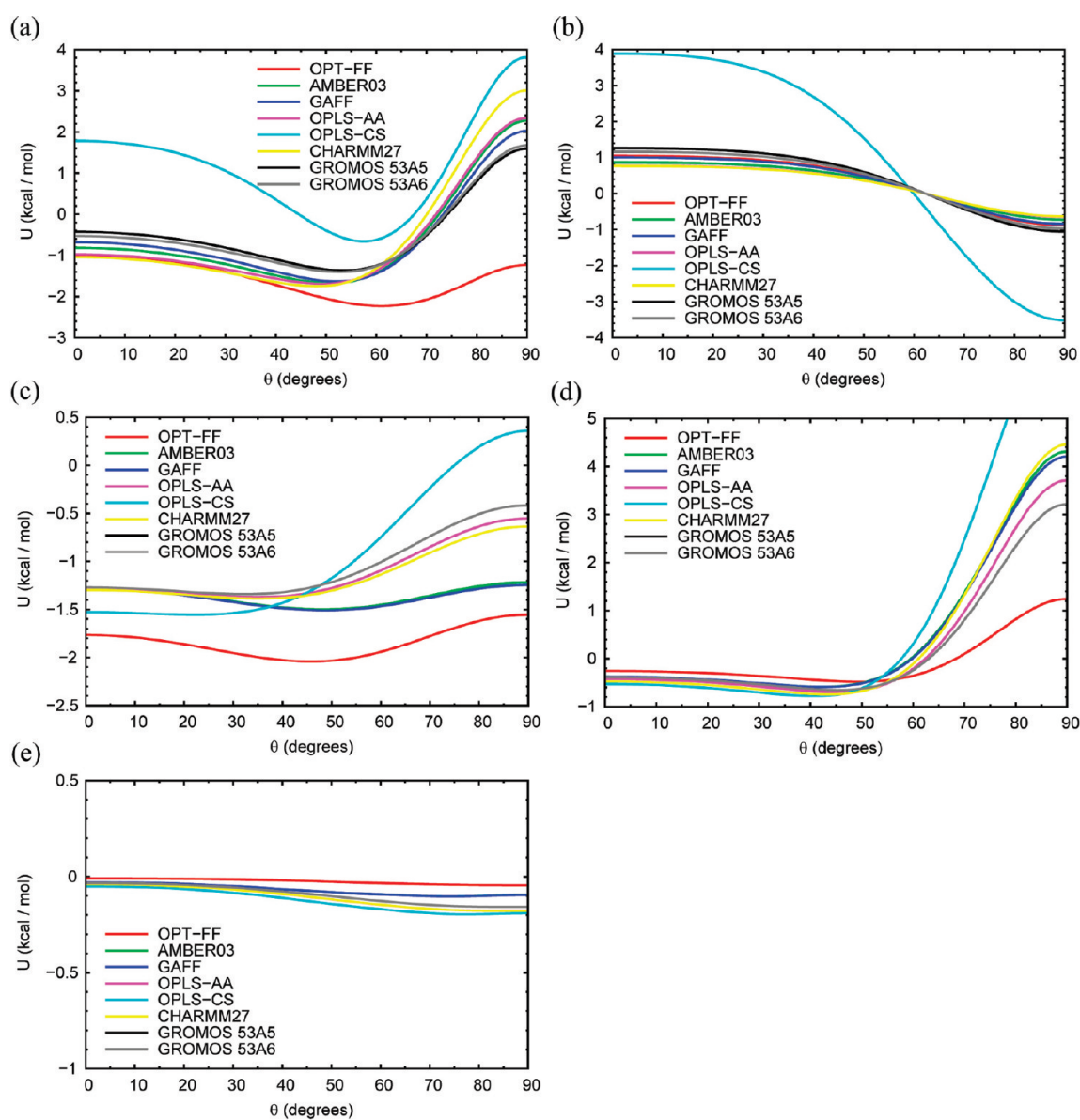


Figure 6. The potential energy curves of the benzene dimer as a function of the angle θ between the aromatic planes ($\varphi = 0^\circ$ and $R_{\text{COM-COM}} = 0.45$ nm). (a) The total intermolecular potential energy of the benzene dimer. (b) The Coulomb potential energy of the benzene dimer. (c) The intermolecular Lennard-Jones potential energy between carbon and carbon atoms in two benzene molecules. (d) The intermolecular Lennard-Jones potential energy between carbon atoms in one benzene and hydrogen atoms in the other benzene. (e) The intermolecular Lennard-Jones potential energy between hydrogen and hydrogen atoms in two benzene molecules.

$\Delta H_{\text{vap}} (\pm 0.48 \text{ kcal/mol})$.⁷² In the scheme of the definitions of $C_p(l)$ and ΔH_{vap} (see eqs 2–4), we think that the distinct differences of these thermodynamic values are basically arising from the fluctuation discrepancies of intermolecular energies E_{inter} for these force fields. The significant differences of the intermolecular π - π interaction energies for these force fields have been discussed above. Moreover, the differences of V shown in Table 4 should be considered. The estimated $C_p(l)$ and ΔH_{vap} can also be influenced by the cutoffs of the long-range potential employed in the simulations.^{91,92} The common values of the cutoff of 1.3 nm for the Coulomb interactions and the cutoff of 1.5 nm for van der Waals interactions were used in the present simulations for these different force fields. For example, in eq 1, the smaller atomic charges (e.g., those of OPLS-AA and CHARMM27, see Table 1) can lead to the quick quenching of the Coulomb interaction, while

the larger charges used in the Coulomb potential calculations may need the larger cutoff value.

The viscosity, as a kinetic property, is not only important for pure liquids, as it also influences the rates of diffusion and conformational change of molecules solvated in the liquid. Thus, it is important that the force field can predict the accurate value of viscosity of liquids. The shear viscosity of liquid benzene for the different force fields calculated from the nonequilibrium MD simulations is compared with the experimental result⁸⁸ in Table 4. The CHARMM27 force field predicts the best value, with respect to the experimental datum.⁸⁸ The results of the OPLS-AA and GROMOS 53A6 force fields are also outstanding, and they have percent deviations of 9.0% and 9.8%, respectively. Unfortunately, quite large values with percent deviations of 39.4%, 28.3%, and 21.3% are predicted with the AMBER 03, GAFF, and GROMOS

Table 4. Some Thermodynamic Properties Calculated from the MD Simulations Using the Different Force Fields

	ρ (kg/m ³)	V (Å ³)	$C_p(l)$ (cal mol ⁻¹ K ⁻¹)	ΔH_{vap} (kcal/mol)	η (10 ⁻³ kg m ⁻¹ s ⁻¹)	D (10 ⁻⁹ m ² /s)	$\tau_{2\perp}$ (ps)	$\tau_{2\parallel}$ (ps)
OPT-FF	1044.6 ± 0.6	124.2 ± 0.1	35.2 ± 1.3	11.30 ± 0.02	3.280 ± 0.031	0.35 ± 0.01	6.16	2.53
AMBER 03	835.9 ± 0.5	155.2 ± 0.1	36.6 ± 1.3	7.24 ± 0.01	0.364 ± 0.003	2.79 ± 0.22	1.40	1.09
GAFF	852.1 ± 0.8	152.2 ± 0.2	39.1 ± 1.2	7.55 ± 0.01	0.431 ± 0.002	2.33 ± 0.01	1.71	1.24
OPLS-AA	867.3 ± 0.4	149.6 ± 0.1	33.3 ± 0.8	8.02 ± 0.01	0.547 ± 0.004	1.97 ± 0.10	1.59	1.19
OPLS-CS	947.2 ± 0.3	136.9 ± 0.1	28.6 ± 0.5	14.86 ± 0.01	^a	^a	^a	^a
CHARMM27	870.4 ± 0.6	149.0 ± 0.2	37.6 ± 1.2	8.17 ± 0.01	0.567 ± 0.003	1.97 ± 0.17	1.41	1.10
GROMOS 53A5	887.3 ± 0.2	146.2 ± 0.1	37.3 ± 0.7	8.44 ± 0.01	0.729 ± 0.004	1.53 ± 0.02	2.07	1.30
GROMOS 53A6	881.5 ± 0.3	147.2 ± 0.1	37.2 ± 0.6	8.25 ± 0.01	0.660 ± 0.005	1.78 ± 0.04	1.79	1.19
exptl.	873.8 ^b	148.4 ^b	32.4 ^c	7.89 ± 0.48 ^b	0.601 ^d	2.20 ^e	1.68 ^f	1.29 ^f

^a The unreliable values of η , D , $\tau_{2\perp}$, and $\tau_{2\parallel}$ are predicted with OPLS-CS force field. ^b Cited from ref 72. ^c Cited from ref 87. ^d Cited from ref 88. ^e Cited from ref 89. ^f Cited from ref 90.

53A5 force fields, respectively. The result of the OPT-FF force field is nearly 5 times higher than the experimental datum.⁸⁸ What is more, the viscosity cannot be derived with the OPLS-CS force field. In the nonequilibrium simulations, see eq 14, the different densities ρ used for the various force fields may numerically influence the shear viscosity. However, for certain density values with the same percent deviation derived with different force fields, their corresponding shear viscosities can be significantly different, thus such a numerical effect of ρ should be neglected. The differences in the fluidity of liquid benzene determined by these force fields should mainly come from the deviations of the descriptions of the intermolecular π - π interactions. The rather large density and interaction energies of the different configurations of the benzene dimer for the OPT-FF force field may significantly impair the fluidity of liquid benzene, namely, result in a larger shear viscosity. The separated charges used in the OPLS-CS force field overemphasize the perpendicular orientations of benzene molecules; thus the fluidity of liquid benzene may be hindered by the strong attractive interaction energies of the perpendicular configurations.

The self-diffusion coefficient D predicted from the MD simulations is another important property to account for the force fields. The GAFF force field can predict quite well the experimental value⁸⁹ with a percent deviation of 5.9%. The OPLS-AA and CHARMM27 force fields also gain reasonable values with percent deviations of 10.5%. Furthermore, within the statistical uncertainties for the OPLS-AA and CHARMM27 force fields, their theoretical values may be more acceptable. Unfortunately, the results of the AMBER 03, GROMOS 53A5, and GROMOS 53A6 force fields have large deviations from the experimental value⁸⁹ with percent deviations of 26.8%, 30.5%, and 19.1%, respectively. The result of the OPT-FF force field is almost 1 order of magnitude smaller than the experimental value.⁸⁹ What is more, the OPLS-CS force field fails again in the prediction of the self-diffusion coefficient of liquid benzene.

The rotational correlation times of the vectors perpendicular and parallel to the aromatic plane ($\tau_{2\perp}$ and $\tau_{2\parallel}$) are evaluated with the different force fields. For all of these force fields, $\tau_{2\perp} > \tau_{2\parallel}$, which is consistent with the experimental datum.⁹⁰ The GAFF force field obtains perfect values of both $\tau_{2\perp}$ and $\tau_{2\parallel}$ with percent deviations of 1.8% and 3.9%, respectively. Although GROMOS 53A5 predicts the best values of $\tau_{2\parallel}$ (only with percent deviation of 0.8%), the value of $\tau_{2\perp}$ evaluated by this force field is much larger than the experimental datum⁹⁰ (with a percent deviation of 23.2%). The results of both $\tau_{2\perp}$ and $\tau_{2\parallel}$ predicted with the OPLS-AA and GROMOS 53A6 force fields are acceptable, with percent deviations of 5.4% and 6.5% for $\tau_{2\perp}$, respectively, and both being 7.8% for $\tau_{2\parallel}$. However, the results

predicted with the OPT-FF, AMBER 03, and CHARMM27 force fields show quite large percent deviations, namely, 266.7%, 16.7%, and 16.1% for $\tau_{2\perp}$ and 96.1%, 15.5%, and 14.7% for $\tau_{2\parallel}$, respectively. Unfortunately, the OPLS-CS force field fails again in predicting either $\tau_{2\perp}$ or $\tau_{2\parallel}$.

As mentioned above, the dynamic properties D and τ_2 show significant differences for these sets of force fields, which should be closely related to the distinct differences in description of the intermolecular interactions. For instance, the relatively strong attractive interactions in the perpendicular configuration of the benzene dimer that are predicted with the OPT-FF and OPLS-CS force fields assuredly lead to much smaller self-diffusion coefficients (nearly zero for OPLS-CS); moreover, the relaxation time $\tau_{2\perp}$ is seriously overestimated or irregularly larger than $\tau_{2\parallel}$, e.g., for OPLS-CS, $\tau_{2\perp} \sim 3.02 \times 10^{33}$ ps while $\tau_{2\parallel} \sim 2.14 \times 10^3$ ps. The overestimation of the π - π attractive interactions and the overemphasis on a certain configuration (e.g., the perpendicular configuration for OPLS-CS) will remarkably influence the simulation results of the structural and dynamic properties of liquid benzene.

4. CONCLUSION

In the present study, the classical MD simulations of liquid benzene are performed with the recently developed force fields OPT-FF,⁶⁵ AMBER 03,⁵⁷ GAFF,⁵⁸ OPLS-AA,⁴³ OPLS-CS,⁴⁹ CHARMM27,^{61,62} GROMOS 53A5,⁶³ and GROMOS 53A6.⁶³ To evaluate the qualities of these force fields, our strategy was as follows: First, the potential energy curves of the benzene dimer at four different configurations, S, PD, T, and Y-shaped, were predicted with these force fields and compared with the high-level quantum chemistry results.^{33,34} Second, with the references of the experimental studies,^{38,41} the local and long-range structures of liquid benzene were carefully examined on the basis of the present simulations. Third, the thermodynamic properties were calculated and compared with the experimental datum.^{72,87-90} In the last two parts, the significant differences of the simulations using the various force fields were discussed in detail, in which we particularly pursued the physical diversities in the description of aromatic π - π interactions for these force fields. Conclusive remarks are summarized here:

- (1) On the local structures of liquid benzene, the parallel PD configuration at a small separation is favored.⁴¹ As shown in Figure 6, this can be reasonably interpreted with the intermolecular Lennard-Jones attractions between carbon and carbon atoms in two benzene molecules in the parallel configuration, while the intermolecular Lennard-Jones

interactions between carbon and hydrogen atoms in the perpendicular configuration are strongly repulsive.

- (2) The AMBER 03, GAFF, OPLS-AA, and CHARMM27 force fields are reliable in the description of the orientational distribution of benzene molecules in the first coordination shell. However, the preference for the parallel configurations at a small molecular separation (<0.50 nm) cannot be reproduced in the simulations with the OPT-FF, GROMOS 53A5, and GROMOS 53A6 force fields, which is due to the largest ϵ_C and the smallest σ_H and ϵ_H used in OPT-FF and the relatively large point charges on the C and H atoms used in the two GROMOS force fields. The separated charges used in the OPLS-CS force field result in the perpendicular configuration being strongly preferred. Through careful examinations of the simulated results using these different force fields, we recommend that the OPLS-AA force field is the best one, not only in the descriptions of the microstructures of liquid benzene but also in the thermodynamic properties investigated in this work. The OPLS-AA force field is promising for the applications to the other aromatic $\pi-\pi$ interaction systems.
- (3) For the aromatic $\pi-\pi$ interaction systems, neither the exaggeration of the quadrupole-quadrupole interaction in the force field function (e.g., OPLS-CS⁴⁹) nor merely fitting the *ab initio* pairwise interaction potentials (e.g., OPT-FF⁶⁵) can simply warrant the reliability of these force fields. Thus, a highly qualified force field should be good at the descriptions on not only the microstructure but also at the predictions of the thermodynamic properties of the liquids.

AUTHOR INFORMATION

Corresponding Author

*E-mail: sxtian@ustc.edu.cn.

ACKNOWLEDGMENT

This work is partially supported by MOST (Grant Nos. 2007CB815204 and 2011CB921401).

REFERENCES

- (1) Hunter, C. A. Aromatic Interactions in Proteins, DNA and Synthetic Receptors. *Philos. Trans. R. Soc. London, Ser. A* **1993**, 345, 77.
- (2) Burley, S. K.; Petsko, G. A. Aromatic-Aromatic Interaction: A Mechanism of Protein Structure. *Science* **1985**, 229, 23.
- (3) Hunter, C. A.; Singh, J.; Thornton, J. M. $\pi-\pi$ Interactions: The Geometry and Energetics of Phenylalanine-Phenylalanine Interactions in Proteins. *J. Mol. Biol.* **1991**, 218, 837.
- (4) Ranganathan, D.; Haridas, V.; Gilardi, R.; Karle, I. L. Self-Assembling Aromatic-Bridged Serine-Based Cyclodepsipeptides (Serinophanes): A Demonstration of Tubular Structures Formed through Aromatic $\pi-\pi$ Interactions. *J. Am. Chem. Soc.* **1998**, 120, 10793.
- (5) Vondrášek, J.; Bendová, L.; Klusák, V.; Hobza, P. Unexpectedly Strong Energy Stabilization Inside the Hydrophobic Core of Small Protein Rubredoxin Mediated by Aromatic Residues: Correlated *Ab Initio* Quantum Chemical Calculations. *J. Am. Chem. Soc.* **2005**, 127, 2615.
- (6) Frederick, C. A.; Williams, L. D.; Ughetto, G.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Wang, A. H.-J. Structural Comparison of Anticancer Drug-DNA Complexes: Adriamycin and Daunomycin. *Biochemistry* **1990**, 29, 2538.
- (7) Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angew. Chem., Int. Ed.* **2003**, 42, 1210.
- (8) Evans, D. A.; Chapman, K. T.; Hung, D. T.; Kawaguchi, A. T. Transition State π -Solvation by Aromatic Rings: An Electronic Contribution to Diels-Alder Reaction Diastereoselectivity. *Angew. Chem., Int. Ed. Engl.* **1987**, 26, 1184.
- (9) Quijcho, F. A.; Vyas, N. K. Novel Stereospecificity of the L-arabinose-Binding Protein. *Nature* **1984**, 310, 381.
- (10) Vyas, N. K.; Vyas, M. N.; Quijcho, F. A. A Novel Calcium Binding Site in the Galactose-Binding Protein of Bacterial Transport and Chemotaxis. *Nature* **1987**, 327, 635.
- (11) Benzing, T.; Tjivikua, T.; Wolfe, J.; Rebek, J., Jr. Recognition and Transport of Adenine Derivatives with Synthetic receptors. *Science* **1988**, 242, 266.
- (12) Vyas, N. K.; Vyas, M. N.; Quijcho, F. A. Sugar and Signal-Transducer Binding Sites of the Escherichia Coli Galactose Chemoreceptor Protein. *Science* **1988**, 242, 1290.
- (13) Muehldorf, A. V.; van Engen, D.; Warner, J. C.; Hmlton, A. D. Aromatic-Aromatic Interactions in Molecular Recognition: A Family of Artificial Receptors for Thymine That Shows Both Face-to-Face and Edge-to-Face Orientations. *J. Am. Chem. Soc.* **1988**, 110, 6561.
- (14) Desiraju, G. R.; Gavezzotti, A. From Molecular to Crystal Structure; Polynuclear Aromatic Hydrocarbons. *J. Chem. Soc., Chem. Commun.* **1989**, 621.
- (15) Ferguson, S. B.; Sanford, E. M.; Seward, E. M.; Diederich, F. Cyclophane-Arene Inclusion Complexation in Protic Solvents: Solvent Effects versus Electron Donor-Acceptor Interactions. *J. Am. Chem. Soc.* **1991**, 113, 5410.
- (16) Vallée, R.; Damman, P.; Dosière, M.; Toussaere, E.; Zyss, J. Nonlinear Optical Properties and Crystalline Orientation of 2-Methyl-4-nitroaniline Layers Grown on Nanostructured Poly(tetrafluoroethylene) Substrates. *J. Am. Chem. Soc.* **2000**, 122, 6701.
- (17) Janda, K. C.; Hemminger, J. C.; Winn, J. S.; Novick, S. E.; Harris, S. J.; Klemperer, W. Benzene Dimer: A Polar Molecule. *J. Chem. Phys.* **1975**, 63, 1419.
- (18) Steed, J. M.; Dixon, T. A.; Klemperer, W. Molecular Beam Studies of Benzene Dimer, Hexafluorobenzene Dimer, and Benzene-Hexafluorobenzene. *J. Chem. Phys.* **1979**, 70, 4940.
- (19) Hopkins, J. B.; Powers, D. E.; Smalley, R. E. Mass-Selective Two-Color Photoionization of Benzene Clusters. *J. Phys. Chem.* **1981**, 85, 3739.
- (20) Langridge-Smith, P. R. R.; Brumbaugh, D. V.; Haynam, C. A.; Levy, D. H. Ultraviolet Spectra of Benzene Clusters. *J. Phys. Chem.* **1981**, 85, 3742.
- (21) Krause, H.; Ernstberger, B.; Neusser, H. J. Binding Energies of Small Benzene Clusters. *Chem. Phys. Lett.* **1991**, 184, 411.
- (22) Henson, B. F.; Hartland, G. V.; Venturo, V. A.; Felker, P. M. Raman-Vibronic Double-Resonance Spectroscopy of Benzene Dimer Isotopomers. *J. Chem. Phys.* **1992**, 97, 2189.
- (23) Venturo, V. A.; Felker, P. M. Intermolecular Raman Bands in the Ground State of Benzene Dimer. *J. Chem. Phys.* **1993**, 99, 748.
- (24) Erlekam, U.; Frankowski, M.; Meijer, G.; von Helden, G. An Experimental Value for the B_{1u} C-H Stretch Mode in Benzene. *J. Chem. Phys.* **2006**, 124, 171101.
- (25) Erlekam, U.; Frankowski, M.; von Helden, G.; Meijer, G. Cold Collisions Catalyze Conformational Conversion. *Phys. Chem. Chem. Phys.* **2007**, 9, 3786.
- (26) Hobza, P.; Selzle, H. L.; Schlag, E. W. Floppy Structure of the Benzene Dimer: *Ab Initio* Calculation on the Structure and Dipole Moment. *J. Chem. Phys.* **1990**, 93, 5893.
- (27) Hobza, P.; Selzle, H. L.; Schlag, E. W. Potential Energy Surface for the Benzene Dimer. Results of *Ab Initio* CCSD(T) Calculations Show Two Nearly Isoenergetic Structures: T-Shaped and Parallel-Displaced. *J. Phys. Chem.* **1996**, 100, 18790.
- (28) Sinnokrot, M. O.; Valeev, E. F.; Sherill, C. D. Estimates of the *Ab Initio* Limit for $\pi-\pi$ Interactions: The Benzene Dimer. *J. Am. Chem. Soc.* **2002**, 124, 10887.

- (29) Sinnokrot, M. O.; Sherill, C. D. Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations. *J. Phys. Chem. A* **2004**, *108*, 10200.
- (30) Hesselmann, A.; Jansen, G.; Schütz, M. Density-Functional Theory-Symmetry-Adapted Intermolecular Perturbation Theory with Density Fitting: A New Efficient Method to Study Intermolecular Interaction Energies. *J. Chem. Phys.* **2005**, *122*, 014103.
- (31) Podeszwa, R.; Bukowski, R.; Szalewicz, K. Potential Energy Surface for the Benzene Dimer and Perturbational Analysis of π - π Interactions. *J. Phys. Chem. A* **2006**, *110*, 10345.
- (32) Sinnokrot, M. O.; Sherill, C. D. High-Accuracy Quantum Mechanical Studies of π - π Interactions in Benzene Dimers. *J. Phys. Chem. A* **2006**, *110*, 10656.
- (33) Sherrill, C. D.; Takatani, T.; Hohenstein, G. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H₂S. *J. Phys. Chem. A* **2009**, *113*, 10146.
- (34) van der Avoird, A.; Podeszwa, R.; Szalewicz, K.; Leforestier, C.; van Harrevelt, R.; Bunker, P. R.; Schnell, M.; von Helden, G.; Meijer, G. Vibration-Rotation-Tunneling States of the Benzene Dimer: An Ab Initio Study. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8219.
- (35) Engkvist, O.; Hobza, P.; Selzle, H. L.; Schlag, E. W. Benzene Trimer and Benzene Tetramer: Structures and Properties Determined by the Nonempirical Model (NEMO) Potential Calibrated from the CCSD(T) Benzene Dimer Energies. *J. Chem. Phys.* **1999**, *110*, 5758.
- (36) Tauer, T. P.; Sherrill, C. D. Beyond the Benzene Dimer: An Investigation of the Additivity of π - π Interactions. *J. Phys. Chem. A* **2005**, *109*, 10475.
- (37) Mahadevi, A. S.; Rahalkar, A. P.; Gadre, S. R.; Sastry, G. N. Ab Initio Investigation of Benzene Clusters: Molecular Tailoring Approach. *J. Chem. Phys.* **2010**, *133*, 164308.
- (38) Narten, A. H. X-Ray Diffraction Pattern and Models of Liquid Benzene. *J. Chem. Phys.* **1977**, *67*, 2102.
- (39) Misawa, M.; Fukunaga, T. Structure of Liquid Benzene and Naphthalene Studied by Pulsed Neutron Total Scattering. *J. Chem. Phys.* **1990**, *93*, 3495.
- (40) Cabaço, M. L.; Danten, Y.; Besnard, M.; Guissani, Y.; Guillot, B. Neutron Diffraction and Molecular Dynamics Study of Liquid Benzene and Its Fluorinated Derivatives as a Function of Temperature. *J. Phys. Chem. B* **1997**, *101*, 6977.
- (41) Headen, T. F.; Howard, C. A.; Skipper, N. T.; Wilkinson, M. A.; Bowron, D. T.; Soper, A. K. Structure of π - π Interactions in Aromatic Liquids. *J. Am. Chem. Soc.* **2010**, *132*, 5735.
- (42) Evans, D. J.; Watts, R. O. On the Structure of Liquid Benzene. *Mol. Phys.* **1976**, *32*, 93.
- (43) OPLS: Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657. OPLS-AA: Jorgensen, W. L.; Severance, D. L. Aromatic-Aromatic Interactions: Free Energy Profiles for the Benzene Dimer in Water, Chloroform, and Liquid Benzene. *J. Am. Chem. Soc.* **1990**, *112*, 4768.
- (44) Jorgensen, W. L.; Laird, E. R.; Nguyen, T. B.; Tirado-Rives, J. Monte Carlo Simulations of Pure Liquid Substituted Benzenes with OPLS Potential Functions. *J. Comput. Chem.* **1993**, *14*, 206.
- (45) Errington, J. R.; Panagiotopoulos, A. Z. New Intermolecular Potential Models for Benzene and Cyclohexane. *J. Chem. Phys.* **1999**, *111*, 9731.
- (46) Cacelli, I.; Cinacchi, G.; Prampolini, G.; Tani, A. Modeling >Benzene with Single-Site Potentials from Ab Initio Calculations: A Step toward Hybrid Models of Complex Molecules. *J. Chem. Phys.* **2004**, *120*, 3648.
- (47) Cacelli, I.; Cinacchi, G.; Prampolini, G.; Tani, A. Computer Simulation of Solid and Liquid Benzene with an Atomistic Interaction Potential Derived from Ab Initio Calculations. *J. Am. Chem. Soc.* **2004**, *126*, 14278.
- (48) Contreras-Camacho, R. O.; Ungerer, P.; Boutin, A.; Mackie, A. D. Optimized Intermolecular Potential for Aromatic Hydrocarbons Based on Anisotropic United Atoms. I. Benzene. *J. Phys. Chem. B* **2004**, *108*, 14109.
- (49) Baker, C. M.; Grant, G. H. The Structure of Liquid Benzene. *J. Chem. Theory Comput.* **2006**, *2*, 947.
- (50) Bonnaud, P.; Nieto-Draghi, C.; Ungerer, P. Anisotropic United Atom Model Including the Electrostatic Interactions of Benzene. *J. Phys. Chem. B* **2007**, *111*, 3730.
- (51) Milet, A.; Korona, T.; Moszynski, R.; Kochanski, E. Anisotropic Intermolecular Interactions in van der Waals and Hydrogen-Bonded Complexes: What can We Get from Density Functional Calculations? *J. Chem. Phys.* **1999**, *111*, 7727.
- (52) Cybulski, S. M.; Seversen, C. E. Critical Examination of the Supermolecule Density Functional Theory Calculations of Intermolecular Interactions. *J. Chem. Phys.* **2005**, *122*, 014117.
- (53) Zhao, Y.; Truhlar, D. G. How Well Can New-Generation Density Functional Methods Describe Stacking Interactions in Biological Systems? *Phys. Chem. Chem. Phys.* **2005**, *7*, 2701.
- (54) Sato, T.; Tsuneda, T.; Hirao, K. A Density-Functional Study on π -Aromatic Interaction: Benzene Dimer and Naphthalene Dimer. *J. Chem. Phys.* **2005**, *123*, 104307.
- (55) Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. Hybrid Density Functional Theory for π -Stacking Interactions: Application to Benzenes, Pyridines, and DNA Bases. *J. Comput. Chem.* **2006**, *27*, 491.
- (56) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2006**, *2*, 364.
- (57) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999.
- (58) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157.
- (59) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (60) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474.
- (61) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586.
- (62) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400.
- (63) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, F. A. Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656.
- (64) Hunter, A. C.; Sanders, J. K. The Nature of π - π Interactions. *J. Am. Chem. Soc.* **1990**, *112*, 5525.
- (65) Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. Assessment of Standard Force Field Models Against High-Quality Ab Initio Potential Curves for Prototypes of π - π , CH/ π , and SH/ π Interactions. *J. Comput. Chem.* **2009**, *30*, 2187.

- (66) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Bess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6*, 459.
- (67) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247.
- (68) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269.
- (69) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. Dapprich, G.; S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (70) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701.
- (71) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435.
- (72) NIST Chemistry WebBook. <http://webbook.nist.gov/chemistry> (accessed June 2011).
- (73) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255.
- (74) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695.
- (75) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182.
- (76) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055.
- (77) Essman, U.; Perela, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577.
- (78) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463.
- (79) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116.
- (80) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquid*; Clarendon Press: Oxford, U. K., 1987.
- (81) Kohlrausch, F. Ueber die elastische Nachwirkung bei der Torsion. *Ann. Phys.* **1863**, *119*, 337.
- (82) Williams, G.; Watts, D. C. Non-Symmetrical Dielectric Relaxation Behaviour Arising from a Simple Empirical Decay Function. *Trans. Faraday Soc.* **1970**, *66*, 80.
- (83) Hess, B. Determining the Shear Viscosity of Model Liquids from Molecular Dynamics Simulations. *J. Chem. Phys.* **2002**, *116*, 209.
- (84) Wensink, E. J. W.; Hoffmann, A. C.; van Maaren, P. J.; van der Spoel, D. Dynamic Properties of Water/Alcohol Mixtures Studied by Computer Simulation. *J. Chem. Phys.* **2003**, *119*, 7308.
- (85) Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684.
- (86) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049.
- (87) Grolier, J.-P. E.; Roux-Desgranges, G.; Berkane, M.; Jiménez, E.; Wilhelm, E. Heat Capacities and Densities of Mixtures of very Polar Substances 2. Mixtures containing *N,N*-dimethylformamide. *J. Chem. Thermodynamics* **1993**, *25*, 41.
- (88) Dymond, J. H.; Robertson, J.; Isdale, J. D. Transport Properties of Nonelectrolyte Liquid Mixtures – IV. Viscosity Coefficients for Benzene, Perdeuterobenzene, Hexafluorobenzene, and an Equimolar Mixture of Benzene + Hexafluorobenzene from 25 to 100 °C at Pressures up to the Freezing Pressure. *Int. J. Thermophys.* **1981**, *2*, 223.
- (89) McCool, M. A.; Collings, A. F.; Woolf, L. A. Pressure and Temperature Dependence of the Self-Diffusion of Benzene. *J. Chem. Soc., Faraday Trans. 1* **1972**, *68*, 1489.
- (90) Witt, R.; Sturz, L.; Dölle, A.; Müller-Plathe, F. Molecular Dynamics of Benzene in Neat Liquid and a Solution Containing Polystyrene. ¹³C Nuclear Magnetic Relaxation and Molecular Dynamics Simulation Results. *J. Phys. Chem. A* **2000**, *104*, 5716.
- (91) Alper, H. E.; Bassolino, D.; Stouch, T. R. Computer Simulation of a Phospholipid Monolayer-Water System: The Influence of Long Range Forces on Water Structure and Dynamics. *J. Chem. Phys.* **1993**, *98*, 9798.
- (92) Alper, H. E.; Bassolino-Klimas, D.; Stouch, T. R. The Limiting Behavior of Water Hydrating a Phospholipid Monolayer: A Computer Simulation Study. *J. Chem. Phys.* **1993**, *99*, 5547.

Quantum Mechanical Origins of the Iczkowski–Margrave Model of Chemical Potential

Steven M. Valone

Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States

ABSTRACT: Charge flow in materials at the atomistic level is controlled through chemical potential equalization among its constituents. Consequently employing this concept in a simulation requires some model of chemical potential. Current atomistic models of chemical potential, such as the Iczkowski–Margrave (IM) model, are built largely on heuristic arguments and depend linearly on the net charge of each constituent. To gain new insight into the IM model, a many-electron model Hamiltonian is constructed at the atomistic level that is commensurate with the IM model, as opposed to one designed at the one-electron level. For a three-state, two-fragment system, the essential electronegativity and the chemical hardness energies are recovered. However, the model Hamiltonian imparts new charge dependencies not found in the IM model. Decidedly nonlinear, transitional or hopping contributions in those new dependencies are shown to be critical to regulating charge flow. Other modifications to the IM model are illustrated with simple two- and three-fragment systems, involving as many as five states, that act as paradigms for general materials models. Including more than three states in the three-fragment example introduces local bonding refinements to the Mulliken electronegativity and chemical hardness.

1. INTRODUCTION

Demands for more refined materials models are driving efforts to ground atomistic levels of those models more fundamentally in the underlying electronic structure. Among the refinements of most intense interest are those pertaining to the chemical potential¹ that determines charge flow among the constituents of a material as they encounter different chemical and mechanical environments.

Numerous variable charge models have been constructed for the purpose of modeling atomic-level chemical potential. The principle model used for this purpose has been that of Iczkowski and Margrave (IM).² Specifically, the IM model concerns itself with how the energy of each atom in the material changes with charge fluctuations. Thus, the IM may be viewed as a model of a site energy. The model comes from the simple idea that a site can occupy certain integer charge states (Figure 1),^{3–5} most typically the neutral, cation, and anion states. One is naturally inclined to interpolate among those points with a polynomial.² With these available states, for an atom A with charge q , a quadratic polynomial suffices to yield

$$E_A^{\text{IM}}(q) = E_A^{(0)} + \chi_A q + 1/2\eta_A q^2 \quad (1)$$

With $E_A^{(0)}$ as the energy of the isolated atom, a pleasing interpretation is available by identifying the linear expansion coefficient as the Mulliken electronegativity χ_A ⁴ and the quadratic coefficient as the chemical hardness η_A .⁶ Before defining χ_A and η_A , first define $E_A(N_A)$ as the energy of isolated A with N_A electrons. If A has nuclear charge Z_A , then $E(Z_A)$, $E(Z_A - 1)$, and $E(Z_A + 1)$ correspond to the energies of the neutral atom, cation, and anion, respectively. The electronegativity is the average of the (first) ionization potential I_A and electron affinity \mathcal{E}_A , $\chi_A \equiv (I_A + \mathcal{E}_A)/2$, and while the hardness is the difference $\eta_A \equiv I_A - \mathcal{E}_A$. The ionization and electron affinity energies are defined by differences in the energies for isolated atom A with different

numbers of electrons, namely $I_A \equiv E(Z_A - 1) - E(Z_A)$ and $\mathcal{E}_A \equiv E(Z_A) + E(Z_A + 1)$. For future reference, note the finite difference relations $\chi_A = (E(Z_A - 1) - E(Z_A + 1))/2$ and $\eta_A = E(Z_A - 1) + E(Z_A + 1) - 2E(Z_A)$.

The implied chemical potential in the IM model is linear in q :

$$-\mu_A^{\text{IM}}(q) \equiv dE_A/dq = \chi_A + \eta_A q \quad (2)$$

For this reason, some have characterized this model of chemical potential as metallic in nature.⁷ Consequently, charge flow is unimpeded when atoms dissociate. This facet of the IM model has led to many variations on the basic theme.^{8–22} For the most part, definitions of the coefficients have been made more sophisticated to allow more realistic charge flow over wider ranges of conditions. Nevertheless most of these approaches adhere to a quadratic expansion in q and retain the essentially metallic nature of the model.

Perdew et al. (hereafter referred to as PPLB) showed that as A dissociates from the rest of the material, the interpolation among integer charge states becomes piecewise linear (Figure 1), instead of a smooth function.²³ The associated model of chemical potential $\mu_A^{\text{PPLB}}(q)$ is piecewise constant, in marked contrast to the linear IM model. Reconciling these different models has proven difficult. A more fundamental understanding of the origins of the charge dependence of $E_A(q)$ should prove fruitful. Initial efforts at reconciliation have yielded some progress,^{24–26} although they have not yet resulted in a general model of chemical potential.

To this end, a relationship between chemical potential and the stationary principle of density functional theory (DFT) has been proven.²⁷ That is, suppose that $E_v[\rho]$ is the energy for an N - and v -representable electron density ρ according to the Hohenberg–Kohn theorem²⁸ or for an N -representable density by constrained

Received: April 24, 2011

Published: May 23, 2011

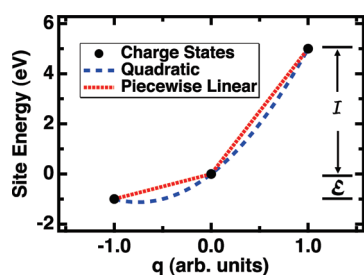


Figure 1. Three-state model of site energies underlying the IM model. Both quadratic and piecewise linear interpolations are possible. The resulting models of chemical potential differ dramatically.

search theory,²⁹ for some fixed external potential v . Then, for that density to correspond to the ground state of the system, it is necessary that $\delta E_v / \delta \rho$, the functional derivative of $E_v[\rho]$ with respect to ρ :

$$\mu^{\text{DFT}} \delta E_v / \delta \rho(\mathbf{r}) \quad (3)$$

be a constant, over all points in space \mathbf{r} (neglecting spin), and where μ^{DFT} is the global chemical potential of the system. The DFT point of view is oriented toward the electronic structure of a molecule or material and provides a value of chemical potential at each point in space. However, the DFT approach to chemical potential is not pursued here.

Rather, we pursue generalizations of the IM definition of chemical potential, a “per atom” quantity, $\mu_A = \delta E / \delta N_A$, where N_A is the number of electrons associated with A. It is generally accepted that $\mu_A = \mu^{\text{DFT}}$, although that relationship is not explored here either. We will, however, return to the distinction between finite-difference and differential definitions of electronegativity and hardness, as they are related to μ_A , and to the relationship of μ_A^{IM} to μ_A .

Our desire in analyzing the chemical potential is to connect with a certain genre of atomistic models that are founded on eq 2, and where the states of the atoms are the central focus, rather than the states of the electrons defined by eq 3. To attain such a goal, several features of the DFT lines of thinking must be recognized at the outset: (1) Implicit in the vast majority of atomistic materials models is the notion that one can actually define atoms or, more generally, fragments in a material. (2) To make these identifications, the total electron density is often decomposed into fragment quantities, as warranted by DFT^{8,30–32} and tight-binding (TB)-based³³ approaches to these models. (3) In both cases, single electrons and their states constitute the basic entity, or finest level of granularity, in the model. Coarser entities, such as atoms and states of atoms, are then reconstituted from these more basic entities. As a result, feature 3 is inconsistent with features 1 and 2.

Consequently, we adopt a different tactic here with an atomistic focus from the outset. A model Hamiltonian is constructed with atoms as the fundamental level of granularity. This model Hamiltonian does share certain commonalities with TB Hamiltonians:^{34,35} The total energy decomposes similarly, and descriptions of the basic components will sound familiar. However, there are important differences in actuality. The present development relies on many-electron states of the material. The states of the atoms are derived from the many-electron states, rather than single-electron states. The model Hamiltonian will be defined well-enough that all of its components (described below)

could, in principle, be calculated from wave function-based electron structure methods. Finally, both quadratic-like and piecewise-linear limiting behaviors (compare, for instance, refs 8 and 36–40) may be derived from the present approach.

2. FRAGMENT MODEL HAMILTONIAN FORMALISM

The broad framework of the fragment model Hamiltonian requires a combination of three essential elements, these being certain characteristics of many-electron wave functions, Hamiltonian decomposition, and charge-state variables. A fragment is a collection of nuclei and electrons that deemed helpful for the material and problem of interest. Thus, a fragment may be anything from a single atom or a functional group to a unit cell or even a molecule in a molecular fluid or crystal. The charge-state variables pertain to the charge states of the fragments and form the connection back to both the many-electron wave functions and the Hamiltonian decomposition, as described here.

First, for an arbitrary, unnormalized state ψ , take $\psi = \sum_i c_i |i\rangle$, where the $|i\rangle$ are many-electron, nonadiabatic states, and the c_i are expansion coefficients. The only required characteristics are that the $|i\rangle$ be normalized antisymmetric, and, for each fragment, the number of electrons be assigned an integer value. There are a number of definitions of charge that might be used. Here it is only necessary to choose a definition and to apply it consistently in constructing the nonadiabatic states. For convenience at this stage of conceptual development, we also require strong orthogonality, meaning that certain one-particle transition densities between any two nonadiabatic states are neglected. Nothing about orbital basis sets, reference states, density decompositions, or even representations is obligatory. No further specificity in the nonadiabatic states is required at this time.

Now, for any fragment A, define a set of its states by the relations:

$$|A\rangle = \sum_{\zeta} C_{A\zeta} |A^{\zeta}\rangle \quad (4)$$

$$|A^{\zeta}\rangle = \sum_i \delta_{\zeta i^A} c_i |i\rangle / C_{A\zeta} \quad (5)$$

and

$$C_{A\zeta}^2 = \sum_i \delta_{\zeta i^A} c_i^2 \quad (6)$$

The Kronecker δ functions select which nonadiabatic contributes to the charge state ζ of A. For any set of fragments defining our total system, one has available the identities $\psi = |A\rangle = |AB\rangle$, for any fragment A or fragment pair AB. Note that $|AB\rangle$ is not necessarily a product of fragment wave functions. We note here that the c_i^2 and $C_{A\zeta}^2$ are equivalent to occupation numbers for the various states. Thus, the value of $C_{A\zeta}^2$ defines what fraction of the time that A spends in the ζ charge state. Finally the fractional number of electrons is $N_A \equiv \sum_{\zeta} N_{A\zeta} C_{A\zeta}^2$, where $N_{A\zeta}$ is the integer number of electrons for the ζ charge state.

One familiar prototype for these states is the valence bond states for the neutral hydrogen molecule. The Heitler–London or covalent state places one electron on each atom. For example, $|\text{cov}\rangle = \phi_L(1)\phi_R(2) + \phi_L(2)\phi_R(1)$, where ϕ is an arbitrarily chosen orbital and L and R label the left and right atoms of the molecule, respectively, when the bond axis is oriented horizontally and the spin is neglected. The labels “1” and “2” signify the spatial coordinates of electron 1 and 2, respectively. Each of two

ionic states places both electrons on one atom and none on the other. For example, $|\text{ion}_L\rangle = \phi_L(1)\phi_L(2)$. Each of these states may be augmented with factors that include explicit correlation. This example can be extended to arbitrary systems and numbers of nonadiabatic states.

Second, recall the well-known fact that any electronic Hamiltonian can be decomposed into a sum of fragment Hamiltonians \hat{H}_A and purely Coulombic interactions \hat{V}_{AB} between fragment pairs.⁴¹ Some readers may be familiar with this idea from the field of “atom-in-molecule” (AIM) approaches to atomistic modeling. The principle differences are that fragments do not have to be atoms, and only the concept of Hamiltonian decomposition is adopted from that work. Other concepts associated with the AIM approach, such as orbital representations, are *not* adopted here.

Each fragment Hamiltonian is the same operator as applies to that fragment in isolation, except for one caveat to be explained shortly. In typical fashion, \hat{H}_A is the sum of kinetic energy, electron–nuclear, and electron–electron operators for its electrons. That is, $\hat{H}_A = \hat{T}_A + \hat{v}_A + \hat{v}_A^{\text{ee}}$, where

$$\hat{T}_A = \sum_{j \in \text{electrons of A}} \hat{T}_j$$

$$\hat{v}_A = \sum_{j \in \text{electrons of A}} \sum_{\kappa \in \text{nuclei in A}} \hat{v}_{j\kappa}$$

and

$$\hat{v}_A^{\text{ee}} = 1/2 \sum_{j \in \text{electrons of A}} \sum_{k \in \text{electrons of A} \neq j} \hat{v}_{jk}^{\text{ee}} \quad (7)$$

The Coulombic interactions in \hat{V}_{AB} are only those that come from electrons on A interacting with the nuclei of B and the electrons of B, plus electrons of B interacting Coulombically with electrons and nuclei of A. These Coulombic terms account for all of the interaction operators not appearing in the fragment Hamiltonians themselves. In the end, this is merely a book-keeping exercise. Thus, from Moffitt,⁴¹ in the general case, the fragment Hamiltonian is simply the total electronic Hamiltonian rewritten as

$$\hat{H}^{(f)} \sum_A \hat{H}_A + \frac{1}{2} \sum_{AB} \hat{V}_{AB} \quad (8)$$

\sum' means a double sum excluding terms with B = A. To complete the definition of $\hat{H}^{(f)}$, and this is the caveat mentioned before, each fragment A must be assigned an integer number of electrons N_A . Traditionally the assignment $N_A = Z_A$ has always been adopted. In point of fact, there is no physical requirement for this assignment. In the context of charge-transfer events, a variable distribution of electrons among the fragments seems more appropriate.

As an example of the fragment form of an electronic Hamiltonian, suppose the total system consists of two fragments, A and B, that are atoms, meaning that each fragment has only one nucleus. The total Hamiltonian is $\hat{H}^{(f)} = \hat{H}_A + \hat{H}_B + \hat{V}_{AB}$. Assume that the system is neutral, so that $N = Z_A + Z_B$, where N is the total number of electrons. Since there are only two fragments in this example, \hat{V}_{AB} always contains all of the electrons. It need not be discussed here any further. $\hat{H}_A(N_A)$ is the Hamiltonian for isolated A with N_A electrons, and $\hat{H}_B(N_B)$ is the Hamiltonian for isolated B with N_B electrons. To maintain neutrality, it is only necessary that $N_A + N_B = N$. It is *not* necessary that $N_A = Z_A$. Electron indistinguishability falls under the definitions of the

nonadiabatic states. It is perfectly acceptable to arbitrarily assign electrons 1 through N_A to A and electrons $N_A + 1$ through N to B. For instance, for the kinetic energy operator of A, $\hat{T}_A(N_A) = \sum_{j=1}^{N_A} \hat{T}_j$ and, for B, $\hat{T}_B(N_B) = \sum_{j=N_A+1}^N \hat{T}_j$. Similar expressions can be constructed for the other operators in the fragment Hamiltonian. This example begs the question of how many electrons to assign to each fragment.

The crucial and unique notion of the present approach is that each state $|i\rangle$ defines an electron distribution on $\hat{H}^{(f)}$. Symbolically, $\hat{H}^{(f)}|i\rangle = \hat{H}_i^{(f)}|i\rangle$ with the meaning that $(\hat{H}_A)_i$ (more explicitly $\hat{H}_{\zeta'}^A$) is assigned $N_A^{\zeta'}$ electrons, the number of electrons appropriate to A for state i . The contracted representation of the wave function for any fragment A leads to its energy being expressed as

$$\bar{E}_A \sum_{\zeta\zeta'} C_{A\zeta'} C_{A\zeta} H_{\zeta\zeta'}^A \quad (9)$$

where $\hat{H}_{\zeta\zeta'}^A \equiv \langle A^{\zeta'} | \hat{H}_{\zeta\zeta'}^A | A^{\zeta} \rangle$ with $\hat{H}_{\zeta\zeta'}^A = (\hat{H}_{\zeta'}^A + \hat{H}_{\zeta}^A)/2$. In density matrix form:

$$\bar{E}_A = \text{tr}(\mathbf{H}_A \Gamma_A) \quad (10)$$

where tr is the matrix trace operator, the density matrix Γ_A has elements $C_{A\zeta'} C_{A\zeta}$ ⁴² and $\text{tr} \Gamma_A = 1$ is assumed. Averaging the fragment Hamiltonians in the energy matrix elements ensures that they are hermitian. The fragment–fragment interactions can be brought to an analogous form, namely $V_{AB} \equiv \text{tr}(\mathbf{V}_{AB} \Gamma_{AB})$.

Let us return once more to the diatomic example. In the definition of $\hat{H}_{\zeta\zeta'}^A$, the kinetic energy for the fragment Hamiltonian of A becomes

$$\hat{T}_{\zeta\zeta'}^A = 1/2 \left(\sum_{j=1}^{Z_A - \zeta_A} \hat{T}_j + \sum_{j=1}^{Z_A - \zeta'_A} \hat{T}_j \right) \quad (11)$$

It depends explicitly on the values of ζ_A and ζ'_A . The kinetic energy operators for B are balanced with those of A so that charge neutrality is maintained at all times. As before, analogous expressions may be constructed for the other operators contributing to $\hat{H}_{\zeta\zeta'}^A$. The Coulombic operators for the A–B atom interaction may be expressed as

$$\begin{aligned} \hat{V}_{\zeta\zeta'}^{AB} = & 1/2 \left[\left(\sum_{j=1}^{Z_A - \zeta} \hat{v}_{jB} + \sum_{j=1}^{Z_A - \zeta'} \hat{v}_{jB} \right) + \left(\sum_{j=Z_A - \zeta + 1}^N \hat{v}_{jA} \right. \right. \\ & + \sum_{j=Z_A - \zeta' + 1}^N \hat{v}_{jA} \left. \right) + \left(\sum_{j=1}^{Z_A - \zeta} \sum_{k=Z_A - \zeta + 1}^N \hat{v}_{jk}^{\text{ee}} \right. \\ & \left. \left. + \sum_{j=1}^{Z_A - \zeta'} \sum_{k=Z_A - \zeta + 1}^N \hat{v}_{jk}^{\text{ee}} \right) \right] \quad (12) \end{aligned}$$

Thus, V_{AB} itself is determined by the matrix of distributions of electrons between A and B, reflected in the $\hat{V}_{\zeta\zeta'}^{AB}$ operators, even though AB is neutral by assumption. Only the charge state of A is needed, because of the neutrality assumption. In the more general, multifragment case, more complex notation is required. Summing the fragment contributions for a given distribution of electrons recovers the total-system Hamiltonian.

The fragment density matrices defined in eq 10 are of a different ilk⁴² than those familiar from contemporary energy density, one-matrix, and two-matrix functionals. One-matrices pertaining to energy functionals, for example, come from integrating the N -particle density matrix of the system over $N - 1$ of its electrons. The fragment density matrices are comprised of

coefficients for contracted, N -electron, nonadiabatic states, organized by the charge states of each fragment. For this reason, there are no problems with representability issues, as arise in *reduced* density matrices.²⁹

Combining the various fragment contributions, one obtains a total variational energy:

$$\bar{E} = \sum_A \bar{E}_A + \frac{1}{2} \sum'_{AB} \bar{V}_{AB} \quad (13)$$

Systematic definitions of the fragment contributions, particularly the transitional or hopping elements of the fragment energy matrices, were absent from earlier work.⁴³ This expression conforms to the ansatz for most empirical and semiempirical potentials in common use. However, it originates from a model Hamiltonian with atomic granularity, rather than from a one-electron Hamiltonian,^{16,34,44,45} time-averaged values of fragment charges,⁴⁶ or perturbation theory estimates.^{47,48} Furthermore, as is common among methods that decompose a many-body Hamiltonian, there are no explicit bond and dihedral angle dependencies. Such dependencies come about implicitly from the properties of the nonadiabatic states and from optimization of \bar{E} .

Third, the charge q on A may be defined from the fragment charge occupation numbers,^{24,26,49} so that $q \equiv Z_A - N_A$ in the present notation. If three charge states, 0, +, and −, are accessible to fragment A, then the charge reduces to the expression:

$$q = C_{A^+}^2 - C_{A^-}^2 \quad (14)$$

for $\text{tr } \Gamma_A = 1$. That is, the net charge of A is determined by the balance in occupancy of the cation and anion states. The charge range depends only on which nonadiabatic states are admitted to the model.

Note that ordinarily there are more occupation numbers than charges. One method for eliminating the undefined coefficients and defining the energy for a chosen set of fragment charges \mathbf{q} consists of minimizing the variational energy eq 13 over all possible sets of nonadiabatic coefficients \mathbf{c} that yield those fragment charges.⁴³ That is

$$E(\mathbf{q}) \equiv \min_{\mathbf{c} \rightarrow \mathbf{q}} \bar{E}(\mathbf{c}) \quad (15)$$

where $\mathbf{c} \rightarrow \mathbf{q}$ means any choice of \mathbf{c} that yields \mathbf{q} . This definition may be motivated by constrained search density functional theory²⁹ and viewed as a special form of more general constrained functionals.⁵⁰

Spin variables could be added in a more complete treatment. However, as the IM model does not address spin dependencies in an explicit manner, they are neglected here.

3. CHARGE–FLOW REGULATION IN A TWO-STATE, TWO-FRAGMENT MODEL

The development above completes the definition of the model Hamiltonian. The simplest illustration of its use in understanding chemical potential consists of a neutral, two-fragment system, where the fragments are atoms, or more simply a diatomic molecule AB. It is described by two states $|0\rangle = |A^0B^0\rangle$ and $|+\rangle = |A^+B^-\rangle$. Clearly, since there are only two states, the resulting model and the IM model cannot be compatible. Neither electronegativity nor hardness as defined by the finite difference relations associated with eq 2 can

play a role in this simplest application of the fragment model Hamiltonian. On the other hand, new second- and higher-order contributions in the charge dependence play a vital role in regulating charge flow.

To see the divergence between the models, begin with eq 14.^{24,25} Since, by choice $C_{A^-} = 0$, one arrives at the relations $C_{A^+} = (q)^{1/2}$ and $C_{A^0} = (1 - C_{A^+}^2)^{1/2} = (1 - q)^{1/2}$. The coefficients for B have the same values as for A in this case. The subscript A is dropped from q since only one charge is needed to define the system. Thus its density matrix, as a function of q , becomes

$$\Gamma_A = \begin{pmatrix} 1 - q & \sqrt{q(1 - q)} \\ \sqrt{q(1 - q)} & q \end{pmatrix}_A \quad (16)$$

The essential charge dependence of $E_A(q) = \text{tr}(\mathbf{H}_A \Gamma)$ differs considerably from the IM model. These differences can be brought to light by expanding the fragment energy to obtain

$$E_A(q) = H_{00}^A + (H_{++}^A - H_{00}^A)q + 2H_{0+}^A \sqrt{q(1 - q)} \quad (17)$$

where the energy matrix elements are defined in eq 9. The asterisk indicates that these energies come from the system states, rather than the properties of the isolated atoms. A term linear in q appears with an energy scale determined by a generalized ionization energy $I_A^* \equiv H_{++}^A - H_{00}^A$. The associated chemical potential μ_A^* may be identified as

$$\mu_A^*(q) = I_A^* + H_{0+}^A \frac{1 - 2q}{\sqrt{q(1 - q)}} \quad (18)$$

It is evident that a Taylor series representation of $\mu_A^*(q)$ cannot be centered on either $q = 0$ or 1. Furthermore, neither the Mulliken electronegativity nor the chemical hardness appear in this model. Instead, first- and higher-order terms appear that have nothing to do with the finite-difference definition of chemical hardness. Whether or not a constant term appears depends on the expansion point used to determine this contribution. At $q = 1/2$, no constant appears to combine with the ionization potential. At any other permissible expansion point, a nonzero constant does appear.

Thus, the square-root dependence from the transition matrix elements constitutes a unique facet to the model.^{24,25} These dependencies correspond to rigorous density functional²³ and perturbative⁵¹ analyses of nonlinear behavior in the fragment chemical potentials. This model of transitional or hopping contributions to the fragment energies provides a physical basis for the nonlinearities in recently proposed models of chemical potential.^{19,20} Hopping contributions constitute departures from thermodynamic models of fractional charge.^{40,52}

Transitional contributions do not appear in the IM model and yet constitute the heart of charge flow regulation. To see this, note the following illustration: There is only one charge state for AB so that $C_{(AB)^0} = 1$. However, the interaction energy $V_{(AB)^0}$ depends on q and may be expressed as a trace over the same density matrix $\Gamma_{(AB)^0} = \Gamma_A$. Consequently the total energy simplifies to $E(q) = \text{tr}(\mathbf{H}_A + \mathbf{H}_B + \mathbf{V}_{AB})\Gamma = \text{tr}\mathbf{H}\Gamma$. The spectral form of the variational energy²⁶ is more insightful and can be expressed as

$$E(q) = E_0 + \omega(q)(E_1 - E_0) \quad (19)$$

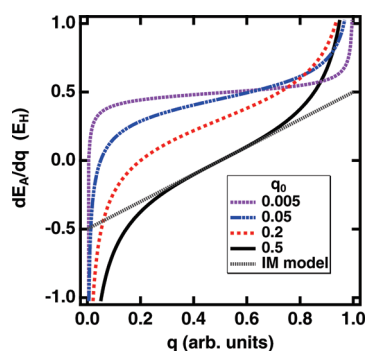


Figure 2. Demonstration of charge flow regulation from differentiating $E_A(q)$ and compared to the IM model. E_A is determined from representative values of the \mathbf{H}_A matrix elements. The ground state charge q_0 is related to the off-diagonal or hopping element of \mathbf{H}_A .

where E_0 and E_1 are the ground- and excited-state energies derived from these two basis states, and

$$\omega(q) = \left(\sqrt{(1-q_0)q} - \sqrt{q_0(1-q)} \right)^2 \quad (20)$$

is the occupation number for the excitation. The ω depends explicitly on q_0 , the ground-state charge. Because the ground- and excited-state energies are just a special case of the variational energy, each fragment energy and the pair interaction may be recast in a spectral form. That is, $E_A(q) = E_A^0 + \omega(q)(E_1^A - E_0^A)$ and $E_B(q) = E_B^0 + \omega(q)(E_1^B - E_0^B)$. A similar expression applies to the pair interaction.

To illustrate how charge transfer is regulated, consider the situation where the bond length d is large. For this illustration, the spectral forms for each fragment energy are retained. However, the pair interaction is approximated as $V_{AB}(d; q) \approx V_{AB}^{\text{sr}}(d) - q^2/d$, where V_{AB}^{sr} is a charge-independent short-range contribution and $-q^2/d$ is a long-range, electrostatic contribution. Similarly, for large d , all of the atomic contributions, E_{A^0} , E_{B^0} , $E_1^A - E_0^A \approx \mu_{A^+}^*$, and $E_1^B - E_0^B \approx -\mu_{B^-}^*$, are constant. Define the chemical potential difference $\mu_{A^+}^* - \mu_{B^-}^* \equiv \Delta\mu$. These energy gaps come from adiabatic states as opposed to nonadiabatic states, where ionization and electron affinity energies apply. The ground-state charge depends on d , as denoted by $q_0(d)$, and therefore $\omega = \omega(d; q)$. Combining these approximations, the illustration becomes

$$E(d; q) \approx E_{A^0} + E_{B^0} + \omega(d; q)\Delta\mu + V_{AB}^{\text{sr}}(d) - q^2/d \quad (21)$$

Setting the partial derivative of $E(d; q)$ with respect to q to 0 yields the condition $2q/d = \omega'(d; q)\Delta\mu$, where

$$\omega' = 1 - 2q_0 - (1 - 2q)\sqrt{\frac{q_0(1-q_0)}{q(1-q)}} \quad (22)$$

In Figure 2, the point at which the solid, black line crosses one of the nonsolid, colored lines determines the new optimum value of q . In that Figure, $\Delta\mu$ is $0.5 E_H$ and $2/d = 0.35 E_H$. As q_0 increases from 0 to 1, more charge is allowed to transfer due to the electrostatic interaction.

The corresponding IM model is

$$E^{\text{IM}}(d; q) \approx E_{A^0} + E_{B^0} + (\chi_A - \chi_B)q + 1/2(\eta_A + \eta_B)q^2 + V_{AB}^{\text{sr}}(d) - q^2/d \quad (23)$$

Its charge equilibration condition may be expressed as $2q/d = \chi_A - \chi_B + (\eta_A + \eta_B)q$. E^{IM} is shown in Figure 2 for $\eta_A + \eta_B = 1 E_H$ and $\chi_A - \chi_B = 0.5 E_H$. As expected, it transfers the same amount of charge for given atomic properties for all d . The difference stems from the charge–flow regulation afforded by nonlinear components of ω . In turn, those nonlinearities are regulated by the reference charge function q_0 . That function is related to the off-diagonal elements in the fragment energy matrices, that is the propensity for A to transition between charge states.

The chemical potential from eq 18 as derived from $E_A(q)$ may be compared to the DFT representations.^{6,8,37,53} From the present perspective, higher-order contributions originate from the off-diagonal elements of the density matrix. The physics of these transitions pertains to the process of passing from one charge state to another, rather than the energy of the individual charge states per se. These contributions are attributed to environmental effects, an outlook shared by the DFT approach. Significantly, a different, identifiable energy scale is controlling the charge flow, compared to anything available in the IM model. Even more significantly, from the fragment Hamiltonian point of view, the fundamental limitation of the IM model is the basic, nonpolynomial nature of the charge dependence, rather than the spatial dependencies of the energies associated the charges. These contributions enter differential definitions of electronegativity and hardness^{6,52–54} but not finite-difference definitions commonly employed by some atomistic models.

4. FRAGMENT HAMILTONIAN INTERPRETATION OF THE IM MODEL

To make more direct contact with the IM model, we examine the charge dependence of the fragment energy for the three-state case of eq 10 for diatomic AB. The state $|-\rangle = |A^-B^+\rangle$ is readmitted to the set of allowed states.^{4,25,43} The fragment density matrix is

$$\Gamma_A(C_+, C_-) = \begin{pmatrix} C_0^2 & C_0C_+ & C_0C_- \\ + & C_+^2 & C_+C_- \\ + & + & C_-^2 \end{pmatrix}_A \quad (24)$$

assuming normalization, while the fragment Hamiltonian, rearranged compared to eq 10, is

$$\mathbf{H}_A = H_{00} + \begin{pmatrix} 0 & H_{0+} & H_{0-} \\ + & H_{++} - H_{00} & H_{+-} \\ + & + & H_{--} - H_{00} \end{pmatrix}_A \quad (25)$$

This form for \mathbf{H}_A is used because it more readily highlights the differences with polynomial expansions.^{19,20,25} Along the diagonal, both I^* and a generalized electron affinity $\mathcal{G}^* \equiv H_{00} - H_{--}$ appear. The generalized electronegativity and hardness have the same relationships to the atomic properties I and \mathcal{G} as in eq 1, namely in terms of the energy matrix elements $\chi^* = (H_{++} - H_{--})/2$ and $\eta^* = H_{++} + H_{--} - 2H_{00}$.

To connect the present model to the IM model, let $\varepsilon_A^0 \equiv H_{00}^A$. Considering just the diagonal contributions ε_A of E_A , one obtains

$$\varepsilon_A = \varepsilon_A^0 + I_A^* C_{A^+}^2 - \mathcal{G}_A^* C_{A^-}^2 \quad (26)$$

These occupation numbers are not conducive to our analysis, since we want to describe the fragment energies in terms of the net charge as much as possible. As shown previously,⁴³ these

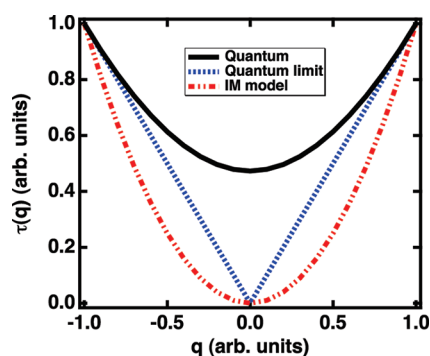


Figure 3. Charge dependence of the hardness term in eq 28 at an OH bond length of 3 Å (solid black) and at the dissociation limit (dashed blue). The IM hardness model (dot-dashed red) is shown for comparison.

occupation numbers can be transformed into functions of q and an auxiliary variable Λ , such that $C_{A^+}^2 = (q + \Lambda^2)/(1 + \Lambda^2)$ and $C_{A^-}^2 = \Lambda^2(1 - q)/(1 + \Lambda^2)$.⁴³ $C_{A^0}^2$ may be determined from normalization. By applying the minimization procedure in eq 15, one can determine $\Lambda = \Lambda(q)$.

Instead of Λ , we consider a new variable that complements the net charge. This variable can be identified by rearranging eq 26 according to the definitions of Mulliken electronegativity and chemical hardness. This rearrangement yields

$$\varepsilon_A(C_+, C_-) = \varepsilon_A^0 + \chi_A^*(C_{A^+}^2 - C_{A^-}^2) + \frac{1}{2}\eta_A^*(C_{A^+}^2 + C_{A^-}^2) \quad (27)$$

The new variable is now obvious: It is the total ionic character $\tau = C_{A^+}^2 + C_{A^-}^2$. In terms of these variables:

$$\varepsilon_A(q, \tau) = \varepsilon_A^0 + \chi_A^*q + \frac{1}{2}\eta_A^*\tau \quad (28)$$

The net charge dependence associated with the electronegativity is the same as in the IM model. The IM approximation for $\tau(q)$ is always just $\tau(q) \approx q^2$, clearly too simplistic. Alternatively, τ can be transformed into a function of q through eq 15, in the same manner that Λ was made a function of q , and that is by taking the value of τ that minimizes the variational energy for a chosen q . When this is done, Figure 3 exhibits a charge dependence that contains contributions other than quadratic. Specifically, Figure 3 contrasts the parabolic behavior of the IM model with the present quantum-based arguments, residing in $\tau(q)$. The data for the black curve in the Figure were generated from a VB2000 calculation^{55,56} for an OH molecule at 3 Å.⁴³ The dissociation-limit properties are known analytically.

Figure 3 shows how the different fragment energy contributions change as the charge dependence of the site hardness term changes with its environment. When bonds are composed of comparable amounts of covalent and ionic character, the hardness charge dependence is essentially quadratic. The balance of covalent and ionic character changes with bond length though, and $\tau(q)$ changes shape accordingly. From Figure 3 one can observe that as the bond stretches, $\tau(q)$ becomes “V-” shaped. Higher-order terms in q must contribute to the chemical potential simply from the term associated with the chemical hardness. The singularities caused by transforming to charge-dependent variables make it difficult to find uniform polynomial approximations to $\tau(q)$. Specifically, the expansion point cannot

in general be at $q = 0$. There is no natural alternative expansion, leading to considerable ambiguity in any polynomial representation of the chemical potential.

Presently, no general expression for $\tau(q)$ is known for finite bond lengths. Based on the limited experience represented in Figure 3, it appears that a reasonable representation of this function might be piecewise polynomials. Then, all but the linear coefficients can be forced to go to 0 as a bond is stretched, so as to match at $q = 0$. Assuming that a piecewise representation is indeed appropriate, it is then tempting to hope that substituting a model of $\tau(q)$ into existing variable-charge models would be sufficient to overcome the deficiencies in the IM chemical potential. This is not so. The transition contributions remain essential in regulating charge flow, just as in the two-state case. A more complete model of chemical potential inherits its properties from these terms as well as those in eq 28.

Finally, we rewrite the three-state fragment density matrix as a function of q and τ . The fragment density matrix of eq 24 becomes

$$\Gamma_A(q, \tau) = \begin{pmatrix} 1 - \tau & \sqrt{(1 - \tau)(\tau + q)/2} & \sqrt{(1 - \tau)(\tau - q)/2} \\ + & (\tau + q)/2 & \sqrt{\tau^2 - q^2}/2 \\ + & + & (\tau - q)/2 \end{pmatrix}_A \quad (29)$$

Viewed in this way, the net charge and ionicity work together in a very specific way to represent the state of A. This is the representation used in the final Section.

5. MANY-BODY INTERACTIONS

Because of the fragmentation properties of the many-electron Hamiltonian, any system can be decomposed into a pair of fragments, casting the total system in the form of a pseudodiatom. Obviously the interactions for the pseudopair must differ in a fundamental way from a simple diatomic molecule because of many-body (polyatomic) effects. The question as to how the IM model should be modified for many-body effects is rarely raised. The fragment model Hamiltonian provides one means of addressing this question. The hallmark of a materials environment compared to an isolated diatomic is the increase in coordination. As long as the allowed charge states of the fragments remain the same, the fragment density matrices have the same structure as eq 29. The principle modifications due to many-body effects appear in the fragment energy matrices.

To investigate such many-body effects, we consider two three-fragment systems. In one case, the system is decomposed into an atom by two fragments that consist of a one or more atoms of a different type. The two fragments are assumed to be equivalent to each other. This decomposition corresponds to an impurity view of an atom embedded in a chain. In the other case, the system is decomposed as in the first case, one atom and two fragments. However, the two fragments are bound to each other, and the atom is bound to only one fragment. This decomposition corresponds to an atom of one kind terminating a chain of atoms of another kind. This time the two fragments are not equivalent to each other. One fragment corresponds to the unperturbed bulk chain, while the other fragment corresponds to one or more atoms that are under the influence of the terminating atom.

For the impurity prototype, consider the linear chain A–B–A, where B is the impurity atom in the chain. When it is necessary to distinguish the left- and right-hand fragments, the notation LBR will be used for the chain. Suppose that the unnormalized wave function of ABA involves just five nonadiabatic states, $|000\rangle$, $|+-0\rangle$, $|+0\rangle$, $|0+-\rangle$, and $|0-\rangle$. Charge transfer between the end fragments is neglected. In keeping with the three-state theme, any fragment can only be singly charged, but they can be either cationic or anionic. The ionic states of B and the neutral states of the end fragments are the most interesting

For B in its many-body environment, it now has two neighbors instead of one. It has contracted states, $|B^0\rangle = |000\rangle$, $|B^+\rangle = (|+-0\rangle, |0+-\rangle)$, and $|B^-\rangle = (|+0\rangle, |0-\rangle)$. Assuming that the two end fragments are equivalent, the transition energies H_{0+}^B and H_{0-}^B have values that are $\sqrt{2}$ times the value for one resonance. The factor of $\sqrt{2}$ is the result of normalization considerations.

More generally though, these two expectation values have a dependence on the ratio of coefficients for the charge states of the end fragments. Taking H_{0-}^B as the primary example, the ratio of interest is c_{+-0}/c_{0+-} . This ratio may be thought of as being related to the ratio of cation occupancies for the end fragments so that

$$c_{+-0}/c_{0+-} = \sqrt{\frac{n_{L^+}}{n_{R^+}}} \equiv \gamma_{B^+} \quad (30)$$

Recall however that these occupancies are related to the preferred variables q and τ . In terms of these values, $n_{L^+} = (\tau_L + q_L)/2$ and similarly for the right-hand fragment occupancy. The ratio of interest becomes

$$\gamma_{B^+} = \sqrt{\frac{\tau_L + q_L}{\tau_R + q_R}} \quad (31)$$

Analogous expression, with appropriate notational changes in occupation number ratios, apply to the cation state of B. Specifically, $\gamma_{B^+} = \sqrt{\frac{\tau_L + q_L}{\tau_R + q_R}}$.

Moreover, the diagonal energies H_{++}^B and H_{--}^B depend on these ratios as well. That is, the state of B and transition among its accessible charge states depends on the state of its environment (defined by the end fragments). Note that the two-state nature of these energies can be exploited. Specifically,

$$H_{++}^B \sim \varepsilon_{B^+} + \omega(\gamma_{B^+}, 1)\Delta\varepsilon_{B^+} \quad (32)$$

and

$$H_{--}^B \sim \varepsilon_{B^-} + \omega(\gamma_{B^-}, 1)\Delta\varepsilon_{B^-} \quad (33)$$

$\Delta\varepsilon_{B^+}$ and $\Delta\varepsilon_{B^-}$ are characteristic excitations to the charge states of B that would come from the two fragments being forced to be inequivalent. If these excitations are set to zero, then one reverts to a simpler model of the fragment energy. The excitation occupancy is the same function as eq 20 except that it is used in its more seminal form:²⁶

$$\omega(\gamma, \gamma_0) = \frac{(\gamma - \gamma_0)^2}{(1 + \gamma^2)(1 + \gamma_0^2)} \quad (34)$$

In the above, $\gamma_0 = 1$ because the balanced charge distributions correspond to the ground state in the present case, where equivalence of the two A fragments was assumed.

Collectively, from eq 28, the diagonal energy contributions of B fragment ε_B may be expressed as

$$\varepsilon_B = \varepsilon_B^0 + \chi_B^* q_B + \frac{1}{2} \eta_B^* \tau_B \quad (35)$$

Now, however, because of the many-body nature of the environment of B, both the effective Mulliken electronegativity and hardness depend on the balances between the individual charge-state occupancies of the end fragments. In detail, $\chi_B^* = (H_{++}^B(\gamma_{B^+}) - H_{--}^B(\gamma_{B^-}))/2$, while $\eta_B^* = H_{++}^B(\gamma_{B^+}) + H_{--}^B(\gamma_{B^-}) - 2H_{00}^B$. The effective representation of electronegativities derived here suggests a possible path connecting the Mulliken⁴ and Phillips⁵⁷ scales of electronegativity.

The neutral states of the end fragments are also interesting. For specificity, consider $|L^0\rangle = (|000\rangle, |0-\rangle, |0+-\rangle)$, while each of its ionic states involves only one resonance. Recall that, in contrast, B has only one resonance contributing to its neutral state, while the other four contribute to its total ionic character. Thus, in terms of the number of contributions, the neutral state of L⁰ is being stabilized by these other two resonances, in comparison to the neutral state of B. Thus, we see that the fragment Hamiltonian view is cognizant of rebalancing ionic character, depending on different bonding environments.

For the chain termination prototype, suppose that A–A–B is described by the same five resonances as in previous example. Now, however, the A fragments are not equivalent, reducing the symmetry in the system. We consider how that asymmetry modifies the effective electronegativities and chemical hardnesses in fragment R (in the LRB notation). Note that the state vector $|R\rangle$ has the same elements as $|B\rangle$ in the previous example. Thus the environmental modifications to $H_{++}^R(\gamma_{R^+})$ and $H_{--}^R(\gamma_{R^-})$ may be modeled by expressions analogous to eqs 32 and 33, respectively. The difference is that the reference value γ_0 for R is no longer unity and does not need to be the same for the cation and anion states of R. These differences stem from the fact that, for example, for the cation state of R, $\gamma_{R^+} = \sqrt{\frac{\tau_L + q_L}{\tau_B + q_B}}$. Thus, its reference value will need to take on one that is characteristic of the occupancy ratio of A and B.

Given the environmental dependencies of these fragment energy matrix elements, chemical potential equilibration takes on a new dimension. Equilibration is achieved by optimizing eq 13, subject to the constraint of charge neutrality, $0 = \sum_A q_A$. We use μ as the Lagrange multiplier associated with enforcing that neutrality. Then, optimization of each charge q_κ leads to $\mu = -\partial\bar{E}/\partial q_\kappa$. Expanding $\partial\bar{E}/\partial q_\kappa$, one obtains

$$\begin{aligned} \partial\bar{E}/\partial q_\kappa = & \text{tr} \sum_A (\mathbf{H}_A \partial\Gamma_A / \partial q_\kappa + \partial\mathbf{H}_A / \partial q_\kappa \Gamma_A) \\ & + \sum_{B \neq \kappa} \partial(\mathbf{V}_{AB} \Gamma_{AB}) / \partial q_\kappa \end{aligned} \quad (36)$$

In defining $\partial\bar{E}/\partial q_\kappa$, one has two choices. In one choice, not only are the other net charges held constant but also the total ionic character variables of each fragment are held constant. Then, a separate optimization process is required to determine those variables. In the other choice, one could assume that the τ are functions of their respective q . The first term is conventional as far as its physical content goes, being directly analogous to the IM model. For either definition of $\partial\bar{E}/\partial q_\kappa$, each fragment density matrix Γ_A will depend only on its q_A . Consequently, the κ term is the sole survivor in the sum over the fragment energies.

The second term, $\sum_A \text{tr} \Gamma_k \partial \mathbf{H}_A / \partial q_{\kappa}$, is unprecedented. Ordinarily, it is simply set to 0, for lack of a more complete model. Physically $\partial \mathbf{H}_A / \partial q_{\kappa}$ describes how excitations of A influence the chemical potential of κ , an effect originating from the many-body effects introduced here. These and other effects described in this section are unknown and unavailable in the IM model. The present approach both describes a physical origin to new effects and allows for systematically incorporating them as needed for a particular application. One is then able to select the appropriate level of approximation for a particular problem. Extensions to more complex systems proceed along parallel lines of reasoning.

6. CONCLUSION

A new class of model Hamiltonians is constructed where atoms or collections of atoms referred to as fragments, rather than single electrons, are the central entities. Transformation of the wave function expansion variables for the model Hamiltonian provides new models of the charge dependence of a fragment energy and chemical potential. The model smoothly transitions between approximately locally quadratic (in charge) and piecewise linear behaviors. The results stand in contrast to the strictly quadratic behavior of the IM model and help justify other empirical models of charge dependence. Of equal significance, transitions between charge states of the atoms possess distinct energy scales that are wholly absent from the IM model. These transitions bear the highly nonlinear aspects of the chemical potential that regulate charge flow and conform to the analysis of Perdew et al.²³ For simple pairs of fragments, the chemical hardness is shown to be the coefficient for a new variable, the total ionic character. A separate process is required if one wants to make the total ionic character a function of net charge. For systems of multiple fragments, the electronegativity and hardness depend on the balances in charge occupancies of surrounding fragments, as typified by a three-fragment linear chain. The results have broad implications for development of materials models and simulations that have need of dynamical charge variation.

AUTHOR INFORMATION

Corresponding Author

E-mail: smv@lanl.gov.

ACKNOWLEDGMENT

Work was performed at Los Alamos National Laboratory under the auspices of the U.S. Department of Energy, under contract no. DE-AC52-06NA25396 with funding provided by the U.S. Department of Energy, Laboratory Directed Research and Development Program (first half) and the Center for Materials at Irradiation and Mechanical Extremes, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under award number 2008LANL1026 (second half). Thanks is given to the Institute for Mathematics and Its Applications, University of Minnesota, for its support during the earliest stages of this conceptual development. The author thanks Donald G. Truhlar, Susan R. Atlas, Heinz Siedentop, Blas P. Uberuaga, Eric Cancès, and Helen G. Telila.

REFERENCES

- (1) Sanderson, R. T. An Interpretation of Bond Lengths and a Classification of Bonds. *Science* **1951**, *114*, 670–672.
- (2) Iczkowski, R. P.; Margrave, J. L. Electronegativity. *J. Am. Chem. Soc.* **1961**, *83*, 3547–3551.
- (3) Pauling, L. The Nature of the Chemical Bond. IV. The Energy of Single Bonds and the Relative Electronegativity of Atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570–3582.
- (4) Mulliken, R. S. A New Electroaffinity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. *J. Chem. Phys.* **1934**, *2*, 782–793.
- (5) Pritchard, H. O.; Skinner, H. A. The Concept Of Electronegativity. *Chem. Rev.* **1955**, *55*, 745–786.
- (6) Parr, R. G.; Pearson, R. G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516.
- (7) Smalø, H. S.; Åstrand, P.-O.; Jensen, L. Nonmetallic Electronegativity Equalization and Point-Dipole Interaction Model Including Exchange Interactions for Molecular Dipole Moments and Polarizabilities. *J. Chem. Phys.* **2009**, *131*, 044101.
- (8) Mortier, W. J.; Ghosh, S. K.; Shankar, S. Electronegativity-Equalization Method for the Calculation of Atomic Charges in Molecules. *J. Am. Chem. Soc.* **1986**, *108*, 4315–4320.
- (9) Rappé, A. K.; Goddard, W. A. Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358–3363.
- (10) Alavi, A.; Alvarez, L. J.; Elliott, S. R.; McDonald, I. R. Charge-Transfer Molecular Dynamics. *Philos. Mag. B* **1992**, *65*, 489–500.
- (11) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical Fluctuating Charge Force Fields: Application to Liquid Water. *J. Chem. Phys.* **1994**, *101*, 6141–6156.
- (12) Streitz, F. H.; Mintmire, J. W. Electrostatic Potentials for Metal-Oxide Surfaces and Interfaces. *Phys. Rev. B* **1994**, *50*, 11996–12003.
- (13) van Beest, B. W. H.; Kramer, G. J.; van Santen, R. A. Force Fields for Silicas and Aluminophosphates Based on Ab Initio Calculations. *Phys. Rev. Lett.* **1990**, *64*, 1955–1958.
- (14) Huang, L.; Kieffer, J. Molecular Dynamics Study of Cristobalite Silica Using a Charge Transfer Three-Body Potential: Phase Transformation and Structural Disorder. *J. Chem. Phys.* **2003**, *118*, 1487–1498.
- (15) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.
- (16) Nistor, R. A.; Polihronov, J. G.; Müser, M. H.; Mosey, N. J. A generalization of the charge equilibration method for nonmetallic materials. *J. Chem. Phys.* **2006**, *125*, 094108.
- (17) Mathieu, D. Split Charge Equilibration Method with Correct Dissociation Limits. *J. Chem. Phys.* **2007**, *127*, 224103.
- (18) Chen, J.; Martínez, T. J. QTPIE: Charge Transfer with Polarization Current Equalization. A Fluctuating Charge Model with Correct Asymptotics. *Chem. Phys. Lett.* **2007**, *438*, 315–320.
- (19) Yu, J.; Sinnott, S. B.; Phillpot, S. R. Charge Optimized Many-Body Potential for the Si/SiO₂ System. *Phys. Rev. B* **2007**, *75*, 085311.
- (20) Mikulski, P. T.; Knippenberg, M. T.; Harrison, J. A. Merging Bond-Order Potentials with Charge Equilibration. *J. Chem. Phys.* **2009**, *131*, 241105.
- (21) Elsener, A.; Politano, O.; Derlet, P. M.; Swygenhoven, H. V. A Local Chemical Potential Approach within the Variable Charge Method Formalism. *Model. Simul. Mat. Sci. Eng.* **2008**, *16*, 025006.
- (22) Verstraelen, T.; Speybroeck, V. V.; Waroquier, M. The Electronegativity Equalization Method and the Split Charge Equilibration Applied to Organic Systems: Parametrization, Validation, and Comparison. *J. Chem. Phys.* **2009**, *131*, 044127.
- (23) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **1982**, *49*, 1691–1694.
- (24) Valone, S. M.; Atlas, S. R. An Empirical Charge Transfer Potential with Correct Dissociation Limits. *J. Chem. Phys.* **2004**, *120*, 7262–7273.

- (25) Morales, J.; Martínez, T. J. A New Approach to Reactive Potentials with Fluctuating Charges: Quadratic Valence-Bond Model. *J. Phys. Chem. A* **2004**, *108*, 3076–3084.
- (26) Valone, S. M.; Atlas, S. R. Energy Dependence on Fractional Charge for Strongly Interacting Subsystems. *Phys. Rev. Lett.* **2006**, *97*, 256402.
- (27) Levy, M.; Perdew, J. P.; Sahni, V. Exact Differential Equation for the Density and Ionization Energy of a Many-Particle System. *Phys. Rev. A* **1984**, *30*, 2745–2748.
- (28) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864.B871.
- (29) Levy, M. Universal Variational Functionals of Electron Densities, First-order Density Matrices, and Natural Spin-Orbitals and Solution of the V-representability Problem. *Nat. Acad. Sci. Proc. U.S.A.* **1979**, *76*, 6062–6065.
- (30) Parr, R. G.; Ayers, P. W.; Nalewajski, R. F. What Is an Atom in a Molecule? *J. Phys. Chem. A* **2005**, *109*, 3957–3959.
- (31) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-Consistent-Charge Density-Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- (32) Ayers, P. W. Atoms in Molecules, An Axiomatic Approach. I. Maximum Transferability. *J. Chem. Phys.* **2000**, *113*, 10886–10898.
- (33) Foulkes, W. M. C.; Haydock, R. Tight-Binding Models and Density-Functional Theory. *Phys. Rev. B* **1989**, *39*, 12520–12536.
- (34) Bloch, F. Über die Quantenmechanik der Elektronen in Kristallgittern. *Zeits. Phys. A* **1929**, *52*, 555–600.
- (35) Slater, J. C.; Koster, G. F. Simplified LCAO Method for the Periodic Potential Problem. *Phys. Rev.* **1954**, *94*, 1498–1524.
- (36) Nistor, R. A.; Müser, M. H. Dielectric Properties of Solids in the Regular and Split-Charge Equilibration Formalisms. *Phys. Rev. B* **2009**, *79*, 104303.
- (37) Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. Electronegativity: The Density Functional Viewpoint. *J. Chem. Phys.* **1978**, *68*, 3801–3807.
- (38) Nalewajski, R. F.; Korchowicz, J.; Zhou, Z. Molecular Hardness and Softness Parameters and Their Use in Chemistry. *Int. J. Quantum Chem.* **1988**, *34*, 349–366.
- (39) Cohen, M. H.; Wasserman, A. Revisiting N-continuous Density-functional Theory: Chemical Reactivity and Atoms in Molecules. *Isr. J. Chem.* **2003**, *43*, 219–227.
- (40) Yang, W.; Zhang, Y.; Ayers, P. W. Degenerate Ground States and a Fractional Number of Electrons in Density and Reduced Density Matrix Functional Theory. *Phys. Rev. Lett.* **2000**, *84*, 5172–5175.
- (41) Moffitt, W. Atoms in Molecules and Crystals. *Proc. R. Soc. London, Ser. A* **1951**, *210*, 245–268.
- (42) Feynman, R. P. In *Statistical Mechanics: A Set of Lectures*; notes taken by Kikuchi, R., Feiveson, H.A., Shaham, J., Ed.; W.A. Benjamin: Reading, MA, 1972; Chapter 2, pp 70–80.
- (43) Valone, S. M.; Li, J.; Jindal, S. Quantum-Based Models of Charge-Dependent Potential Energy Surfaces: Three-state Models. *Int. J. Quantum Chem.* **2008**, *108*, 1452–1464.
- (44) Hubbard, J. Electron Correlations in Narrow Energy Bands. *Proc. R. Soc. London, Ser. A* **1963**, *276*, 238–257.
- (45) Kotliar, G.; Savrasov, S. Y.; Haule, K.; Oudovenko, V. S.; Parcollet, O.; Marianetti, C. A. Electronic Structure Calculations with Dynamical Mean-Field Theory. *Rev. Mod. Phys.* **2006**, *78*, 865–951.
- (46) Fano, U. Description of States in Quantum Mechanics by Density Matrix and Operator Techniques. *Rev. Mod. Phys.* **1957**, *29*, 74–93.
- (47) Finnis, M. W.; Sinclair, J. E. A Simple Empirical N-Body Potential for Transition Metals. *Philos. Mag. A* **1984**, *50*, 45–55.
- (48) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry. *J. Phys. Chem. A* **2001**, *105*, 293–307.
- (49) Morales, J.; Martínez, T. J. Classical Fluctuating Charge Theories: The Maximum Entropy Valence Bond Formalism and Relationships to Previous Models. *J. Phys. Chem. A* **2001**, *105*, 2842–2850.
- (50) Pan, X.-Y.; Sahni, V.; Massa, L. Determination of a Wave Function Functional. *Phys. Rev. Lett.* **2004**, *93*, 130401.
- (51) Cioslowski, J.; Stefanov, B. B. Electron Flow and Electronegativity Equalization in the Process of Bond Formation. *J. Chem. Phys.* **1993**, *99*, 5151–5162.
- (52) Ayers, P. The Dependence on and Continuity of the Energy and Other Molecular Properties with Respect to the Number of Electrons. *J. Math. Chem.* **2008**, *43*, 285–303.
- (53) Ayers, P. W. An Elementary Derivation of the Hard/Soft-Acid/Base Principle. *J. Chem. Phys.* **2005**, *122*, 141102.
- (54) Liu, S.; De Proft, F.; Parr, R. G. Simplified Models for Hardness Kernel and Calculations of Global Hardness. *J. Phys. Chem. A* **1997**, *101*, 6991–6997.
- (55) Li, J.; McWeeny, R. VB2000: Pushing Valence Bond Theory to New Limits. *Int. J. Quantum Chem.* **2002**, *89*, 208–216.
- (56) Li, J.; Duke, B.; McWeeny, R. VB2000, version 1.8 R2; SciNet Technologies: San Diego, CA, 2006.
- (57) Phillips, J. C. Dielectric Definition of Electronegativity. *Phys. Rev. Lett.* **1968**, *20*, 550–553.

Self-Consistent-Charge Density-Functional Tight-Binding Parameters for Cd–X (X = S, Se, Te) Compounds and Their Interaction with H, O, C, and N

Sunandan Sarkar,[†] Sougata Pal,^{†,‡} Pranab Sarkar,^{*,†} A. L. Rosa,[§] and Th. Frauenheim[§]

[†]Department of Chemistry, Visva-Bharati University, Santiniketan- 731235, India

[‡]Institute of Theoretical Physics and Astrophysics, University of Kiel, D-24098, Kiel, Germany

[§]BCCMS, University of Bremen, Am Fallturm 1, 28359, Bremen, Germany

ABSTRACT: Parameters for CdX, SeX, and TeX (X = H, C, N, O, S, Se, Te, and Cd) have been generated within the self-consistent-charge density-functional tight-binding (SCC-DFTB) framework. The approach has been tested against ab initio density-functional theory calculations for the relevant bulk phases, surfaces, nanowires, and small molecular systems. The SCC-DFTB approach reproduces structural, electronic, and energetic properties very well, demonstrating that the developed parameters are fully transferable among different chemical environments.

1. INTRODUCTION

During the past couple of decades, nanostructured materials have been a very exciting field of research.^{1–5} The impact of this research on both fundamental science and potential applications has been tremendous and is still rapidly growing. The interest in this particular class of materials stems from the fact that their properties can be tuned by varying either their size or shape. The emphasis has been mainly on II–VI and III–V semiconductors, which show strong confinement effects. II–VI semiconductors are particularly interesting and have been the focus of many experimental^{6–23} and theoretical studies^{24–47} in the recent past. The strong size- and shape-dependent optical properties of these materials have led to their use in optoelectronic devices^{8,11–18,18–20} and as biological tags.^{9,21–23} CdS is one of the direct wide-gap semiconductor with good stability and high luminescence properties. Because of its high photosensitivity and attractive applications in photoconducting cells, CdS nanoparticles have been studied as model photoconductors. CdS nanostructures such as nanowires, nanorods, and nanodots have been grown experimentally, and they have been applied in photovoltaic cells, photonic switches, and optoelectronic devices.^{6,7} Recently, CdS nanostructures have also been used in biolabeling, imaging, drug delivery, and other biotechnological areas.^{8–10} CdSe is one of the most extensively studied quantum nanostructured materials due to its strong size-tunable properties. Hence, it is a good candidate for optic-related applications, including solar cells and photoelectrochemical (PEC) cells.^{11–18} CdTe has proven to provide good optical performance across a wide range of temperatures and has provided adequate mechanical robustness to be used as substrate materials. Self-assembled CdTe nanodots have high efficiency in terms of photoluminescence and hence are appropriate candidates for the purpose of making LEDs.²⁰ Very recently, semiconductor quantum dots of CdSe and CdTe with their tunable band edges offer new opportunities for harvesting light energy in the visible region of the solar spectrum.^{48–50} The use of CdSe and CdTe quantum dots facilitates charge separation and the generation of photo currents

under visible light irradiation. A further area of concentration focuses on shape-controlled II–VI nanomaterials with recently reported structures comprising nanowires, nanocables, nanoribbons, and nanorods.^{51–54} There are relatively few reports involving the preparation of ternary II–II'–VI nanomaterials.^{55–60} In these alloyed compounds, manipulation of the band gap energy can be achieved by changing both particle size and composition (i.e., the ratio of M to M'). Recently, this has been demonstrated with the synthesis of a series of highly luminescent Zn_xCd_{1–x}Se nanocrystals whose emission energy can be tuned across the visible spectrum by increasing the Zn/Cd ratio.⁶¹ Thus, the development of new approaches to access ternary nanoparticles is an attractive pursuit. So exploration of size-, composition-, and shape-tunable electronic structures of this class of materials is still an active field of research from both experimental and theoretical points of view.

Due to the interest in the variation of the electronic properties as a function of the system size and shape, there have been a number of theoretical studies devoted to the properties of various kinds of cadmium chalcogenide nanostructures. Obviously, the highest accuracy is achieved by the *first-principle* methods, but because of their high computational demands, they become prohibitive to simulating a large number of atoms. In addition to numerous traditional quantum mechanical methods, self-consistent-charge density-functional tight-binding (SCC-DFTB) schemes have also been very successful^{62–66} in treating large systems. In many cases, the results of these two-center-oriented schemes deviate only slightly from those of more sophisticated methods. Therefore, the SCC-DFTB method has been applied to simulate a large number of atoms quantum mechanically in the fields of nanotechnology and solid state physics. Very recently, Bhattacharya et al. developed TB parameters for CdS and ZnS by fitting the TB band structure to the full potential linear augmented plane wave (FP-LAPW) band

Received: April 16, 2011

Published: June 16, 2011

structure.⁶⁷ In this work, we report the development of SCC-DFTB parameters for various cadmium chalcogenides [CdX, X = S, Se, and Te] and their interactions with C, O, N, and H. Our parameters have been validated against density-functional theoretical (DFT) calculations for bulk, surfaces, nanowires, and small molecules. We also compare the SCC-DFTB results with the available experimental results. We show that our parameters are able to reproduce atomic geometries, binding energies, and energy dispersion quite well in comparison with first-principle calculations, demonstrating good transferability among different chemical environments.

2. COMPUTATIONAL METHODOLOGY

2.1. SCC-DFTB Approach. The self-consistent charge density-functional method used in this work has been described in detail elsewhere.^{62–66} Therefore, we just give a brief overview of the method. In SCC-DFTB method, the total energy can be expressed as a second order expansion of the DFT Kohn–Sham^{68,69} total energy with respect to charge density fluctuations.

$$E_{\text{tot}}^{\text{DFTB}} = \sum_i^{\text{occ}} n_i \langle \Psi_i | \hat{H}_0 | \Psi_i \rangle + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + E_{\text{rep}} \\ = E_0[n_0] + E_2[n_0, \delta n] + E_{\text{rep}} \quad (1)$$

The first term of eq 1 is the sum over the occupied eigenstates Ψ_i of the effective Kohn–Sham Hamiltonian \hat{H}_0 , derived under the approximation that the initial electronic density of the many-atom system can be represented as a superposition of corresponding neutral atomic charge densities (n_0). This Hamiltonian is also subjected to the frozen-core approximation, i.e., only the valence electrons are treated explicitly whereas the inner electrons are represented by an effective (pseudo) potential. The second term, E_2 , corresponds to the second-order expansion of the exchange-correlation functional with respect to charge density fluctuations δn (the first-order terms in this expansion vanish for any arbitrary n_0), approximated as atomic point like charges (Δq) together with an analytical interpolating function $\gamma_{\alpha\beta}$. This term is important for our cadmium chalcogenide systems, as there are electronegativity differences between the cadmium and chalcogenides. Finally, the third term, E_{rep} , accounts for the “double-counting” terms and the ion–ion core interaction in a set of distance-dependent pairwise repulsive potentials, modeled as the difference between the SCC-DFTB electronic energy ($E_{\text{Elec}}^{\text{DFTB}} = E_0 + E_2$) and the total DFT energy for some reasonably chosen reference system.

The wave functions (Ψ_i) are expanded as a linear combination of atomic orbitals:

$$\Psi_i = \sum_n c_n \phi_n \quad (2)$$

The Kohn–Sham atomic orbitals ϕ_m in eq 2 are constructed as a linear combination of Slater-type orbitals, obtained by solving the following Kohn–Sham equation for the spin-unpolarized free atom self-consistently:

$$\left[\hat{T} + v_0 + \left(\frac{r}{r_0} \right)^2 \right] \phi_n(r) = \varepsilon_n \phi_n(r) \quad (3)$$

The modified Hamiltonian in eq 3 consists of a kinetic energy operator \hat{T} , the potential energy for the neutral atom v_0 , and an

additional harmonic potential $(r/r_0)^2$ used to enforce the localization of the atomic orbitals and to improve the quality of the energy band structures. Having defined the atomic orbitals (ϕ) and an initial set of expansion coefficients (c_n), we estimated the atomic charge fluctuations (Δq) via Mulliken population analysis, and the DFTB total energy problem is solved self-consistently by using the following secular equation:

$$\sum_i c_{in} (H_{nm} - \varepsilon_i S_{nm}) = 0 \quad \forall n, i \quad (4)$$

with

$$H_{nm} = H_{nm}^0 + \frac{1}{2} S_{nm} \sum_{\eta}^N (\gamma_{\alpha\eta} + \gamma_{\beta\eta}) \Delta q_{\eta} \\ H_{nm}^0 = \langle \phi_n | \hat{H} | \phi_m \rangle \quad \forall n \in \alpha, m \in \beta \\ S_{nm} = \langle \phi_n | \phi_m \rangle$$

In eq 4, the diagonal zero-order matrix elements H_{nm}^0 are taken as the eigenvalues obtained from the free atom calculations, while the distance-dependent nondiagonal elements H_{nm}^0 and S_{nm} are calculated within the two-center approximation. These values are tabulated over a large number of interatomic distances that allows one to obtain interpolated values at any distance and thus avoids integral evaluations during DFTB calculations leading to a computational efficiency comparable to those of traditional semiempirical methods while retaining the accuracy. The charge transfer among different atoms takes into account their respective chemical hardness (Hubbard parameters), calculated as the first derivative of the total atomic energy with respect to the electronic occupation around the neutral-atom electronic density.

The necessary parameters to represent a system within the DFTB method include the Hamiltonian and overlap matrix elements (H_{nm}^0 and S_{nm}), Hubbard parameters for every chemical element, and the repulsive pairwise potentials for all interacting neighbors. The successful SCC-DFTB parametrization implies that the main energetic and structural properties are well reproduced with respect to either DFT or the experiment. This can be achieved by adjusting the wave function and density confinement radii for a given element as well as by accurate construction of repulsion profiles. The required properties of any atom in the SCC-DFTB framework include the LCAO basis functions χ_{μ} , the reference input density (n_0), the wave function confinement radius (r_0), and the chemical hardness parameters, U_A . The LCAO basis functions χ_{μ} are obtained from atomic Kohn–Sham calculations with the PBE exchange-correlation functional, which was derived by Perdew, Burke, and Ernzerhof, in the presence of the harmonic potential, $(r/r_0)^2$.⁷⁰ It forces the wave functions to avoid areas far away from the nucleus, thus resulting in an electron density that is compressed in comparison to the free atom. The good choice for r_0 values is $r_0 \approx 2r_{\text{cov}}$, where r_{cov} is the covalent radius of the element. The r_0 values are chosen to be 3.0, 2.8, and 2.8 bohr corresponding to s, p, and d functions of cadmium. For selenium, r_0 values are 3.4, 5.3, and 5.4 bohr corresponding to s, p, and d functions, and the r_0 values of tellurium are 3.0, 6.8, and 6.4 bohr corresponding to s, p, and d functions. Another atomic property, the reference input density (n_0), is chosen to be 8.2, 10.0, and 11.0 bohr for the corresponding elements Cd, Se, and Te. These parameters have been selected out of a large number of trials and ensure that SCC-DFTB reproduces accurate DFT electronic band structures for solid-state elements to the highest possible degree. In the case of

Table 1. Parametrization Details of the Pairwise Repulsive Potentials for Cd–X, Se–X, and Te–X Interactions (X = H, C, N, O, S, Se, Te, and Cd)

interaction	reference	repulsive	bond distance (Å)	
	system	cutoff (Å)	SCC-DFTB	PP-PBE
Cd–H	CdH ₂	1.90	1.76	1.76
Cd–C	CdC	2.38	2.16	2.19
Cd–N	CdN	2.38	2.22	2.39
Cd–O	CdO	2.12	1.89	1.94
Cd–S	zb-CdS	2.80	2.56	2.60
Cd–Se	zb-CdSe	2.96	2.61	2.71
Cd–Te	zb-CdTe	3.17	2.80	2.89
Cd–Cd	fcc-Cd	4.17	3.25	3.21
Se–H	SeH	1.59	1.51	1.51
Se–C	SeC	2.12	1.71	1.71
Se–N	SeN	2.12	1.71	1.70
Se–O	SeO	3.02	1.72	1.71
Se–S	SeS	2.86	2.05	2.10
Se–Se	Se ₂	3.23	2.34	2.53
Te–H	TeH	1.80	1.70	1.72
Te–C	TeC	2.65	1.94	1.95
Te–N	TeN	2.59	1.93	1.95
Te–O	TeO	3.00	1.95	1.94
Te–S	TeS	3.00	2.33	2.32
Te–Se	TeSe	3.17	2.50	2.49
Te–Te	Te ₂	3.81	3.02	3.10

other elements (H, C, N, O, S), we have employed the same initial atomic parameters as have been reported earlier.⁶³

Repulsion profiles for Cd–X (X = H, C, N, O, S, Se, Te, and Cd), Se–X (X = H, C, N, O, S, Se), and Te–X (X = H, C, N, O, S, Se, and Te) pairs at a series of interatomic distances (*R*) were generated by fitting them with cubic splines of the difference of DFT total energy and the electronic DFTB energy against distance in a simple reference system. Repulsion profiles for different pairs are characterized by different cutoff values, indicating the distance at which the repulsion energy approaches zero. The brief overview of parametrization details used for the generation of Cd–X, Se–X, and Te–X repulsion profiles are summarized in Table 1.

The SCC-DFTB calculations were carried out with the DFTB + code.⁶⁶ The bulk properties of rock-salt CdO and zinc-blende (zb) CdX (X = S, Se, and Te) are calculated using the periodic boundary condition (PBC) and used an 8 × 8 × 8 Monkhorst–Pack (MP)⁷¹ grid for k-point sampling. For hcp-Cd and wurtzite (wz) CdX (X = S, Se, and Te), we used an 8 × 8 × 4 MP grid for k-point sampling. We used 5 × 5 × 5 MP grids for k-point sampling to study trigonal bulk Se and Te.

2.2. Benchmark DFT Calculations. As a reference for the performance of the SCC-DFTB methods, we use the GGA-PBE functional, as implemented in SIESTA package⁷² using a double- ζ plus polarization function (DZP) basis set for the Cd atom; a double- ζ function (DZ) basis set for H, C, N, O, S, Se, and Te atoms; and norm-conservative Troullier–Martins pseudopotentials (PP)⁷³ for representing the valence and inner electrons, and also a B3LYP functional^{74,75} with an SBK basis set for Cd, Se, and Te and 6-31G(d,p) for H, C, N, O, and S, as implemented in the Gaussian 03 package.⁷⁶ In the case of bulk systems like

Table 2. Structural and Energetic Properties of Bulk Phases of Cd, Se, and Te Calculated Using SCC-DFTB and PP-PBE (Experimental Values Are Also Given)

	<i>c</i> (Å)	<i>a</i> (Å)	<i>c/a</i>	<i>B</i> (Gpa)	<i>E_c</i> (eV/atom)	<i>E_g</i> (eV)
hcp Cd						
SCC-DFTB	5.58	2.96	1.88	76.40	1.41	
PP-PBE	5.72	3.08	1.86	42.93	1.24	
exptl ^a	5.62	2.98	1.89	46.70	1.16	
trigonal Se						
SCC-DFTB	5.18	4.39	1.18	49.15	2.63	1.49
PP-PBE	5.18	4.54	1.14	39.48	2.26	0.84
exptl ^b	4.96	4.37	1.14	14.90	2.25	1.98
trigonal Te						
SCC-DFTB	6.26	4.85	1.29	50.95	2.21	0.68
PP-PBE	6.05	4.55	1.33	28.33	2.19	0.19
exptl ^c	5.93	4.45	1.33	23.00	2.19	0.33

^a Refs 77–79. ^b Refs 79 and 80. ^c Refs 81 and 82.

face-centered (fcc), hexagonal close packing (hcp), trigonal, rock-salt, zinc-blende (zb), and wurtzite (wz), PP-PBE calculations were performed with periodic boundary conditions (PBC). For the fcc-Cd, rock-salt CdO, and zb-CdX (X = S, Se, and Te) solid-state reference systems, the k-points were sampled with an 8 × 8 × 8 grid. For hcp-Cd and wz-CdX (X = S, Se, and Te), the k-points were sampled with an 8 × 8 × 4 grid and, for trigonal Se and Te, a 5 × 5 × 5 Monkhorst–Pack (MP)⁷¹ grid and using the 225 Ry Mesh cutoff. For the calculation of (10 $\bar{1}$ 0) and (11 $\bar{2}$ 0) surfaces, the k-points were sampled with 6 × 1 × 4 and 1 × 4 × 4 MP grids, respectively, with a 225 Ry Mesh cutoff. The (10 $\bar{1}$ 0) and (11 $\bar{2}$ 0) faceted nanowires were calculated using a 1 × 1 × 4 MP grid for k-point sampling. For molecular reference systems, the calculations were performed by employing large supercells, including a 15 Å vacuum region in all directions to isolate the molecules from their periodic replicas.

3. RESULTS AND DISCUSSION

3.1. Bulk hcp-Cd and Trigonal Se and Te. We have determined the lattice parameters and elastic properties of bulk hcp-Cd and bulk trigonal selenium and tellurium by calculating energy-volume profiles in a ±15% range around the experimental equilibrium volumes and fitting the results to the Murnaghan equation of state. Like other studies we have not fixed the value of the *c/a* ratio; rather we have also optimized it. In our calculation, the internal parameter *u* (in Å) for selenium and tellurium was taken as 0.2285 and 0.2640, respectively. The equilibrium lattice parameters and energetic properties of the cadmium hcp bulk phase and trigonal phases of selenium and tellurium obtained using SCC-DFTB and PP-PBE together with experimental values are shown in Table 2. As is evident from Table 2, the DFTB parameters except the bulk modulus for all considered bulk systems are in fairly good agreement with the PP-PBE and experimental results.^{77–82} As expected, the values of cohesive energies are little higher than the corresponding DFT and experimental results because the DFTB calculation overestimates this property. The large error in bulk modulus is because of the fact that we have used strong wave function compression for these elements, which lowers their electronic energy wells.

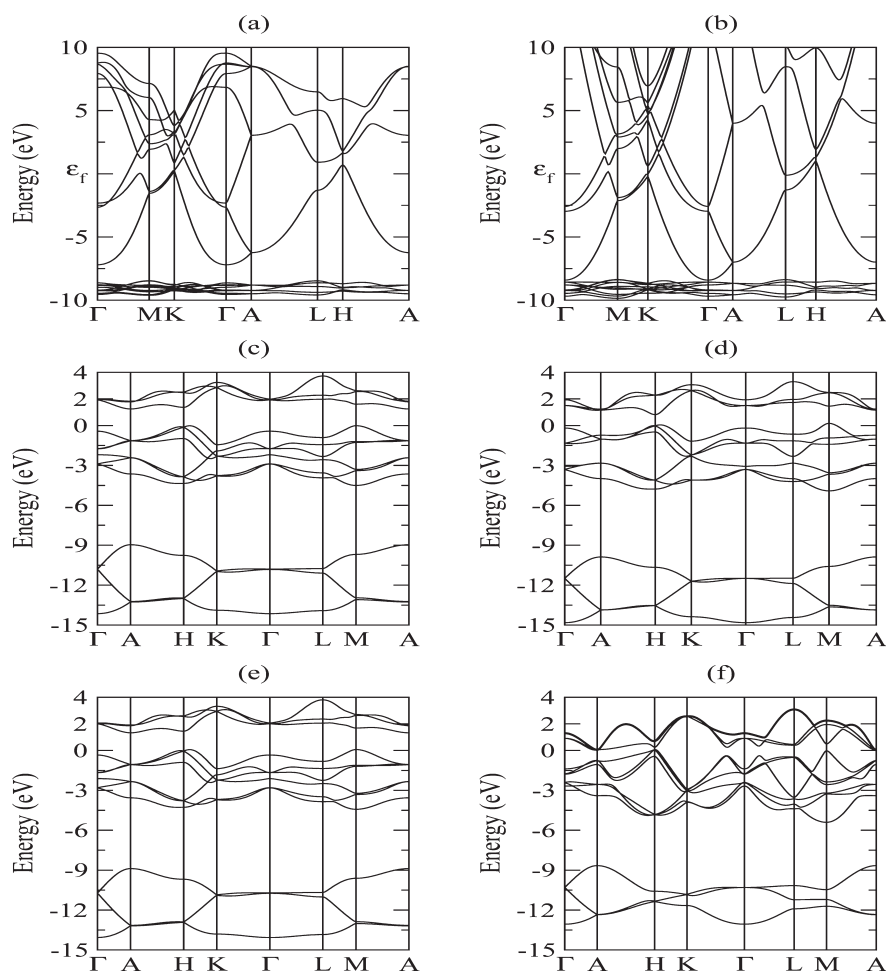


Figure 1. Electronic band structure of hcp-Cd (a, b), trigonal Se (c, d), and trigonal Te (e, f) calculated by SCC-DFTB (left panel) and PP-PBE (right panel) methods as implemented in the SIESTA package. For hcp-Cd, ϵ_f denotes the Fermi level for each case. For Se and Te, the zero of energy is set at the top of the valence band.

This strong wave function compression is essential in ensuring a reasonable band structure for these materials.

The electronic band structures of hcp-Cd and trigonal Se and Te obtained using the SCC-DFTB and PP-PBE methods (as implemented in SIESTA) are shown in Figure 1. From the figure, it is seen that SCC-DFTB bands are compressed in comparison to the bands obtained from DFT while retaining their shape. Such behavior is typical for the SCC-DFTB method due to the use of a minimal basis set. As a general feature, relevant states lying close to the Fermi level are reasonably described using SCC-DFTB. The trigonal Se and Te band gap (E_g) values obtained with the SCC-DFTB method (1.49 and 0.68 eV, respectively) are closer to the experimental values (1.98 and 0.33 eV, respectively) than the PP-PBE result (0.84 and 0.19 eV, respectively). It is important to mention that, usually, DFT underestimates the band gap. The good agreement for the band gap values obtained using SCC-DFTB is therefore fortuitous and due to the use of a minimum basis set. In any case, this opening of the gap using DFTB might be useful when treating systems with states lying inside the band gap.

3.2. Bulk CdO (Rock-Salt), Zinc-Blende, and Wurtzite Phase of CdS, CdSe, and CdTe. Although, the most stable phase of CdO is rock-salt, cadmium chalcogenides adopt either a zinc-blende (zb) or wurtzite (wz) crystal structure. We have

calculated the bulk lattice parameters, energetics, and band structure of CdO and also of cadmium chalcogenides of both zb and wz modifications. The Brillouin-zone integration of the superstructures is calculated using the MP special k points scheme. For cadmium chalcogenides, a large number of k points and high Mesh-cutoff energies for the basis functions are used to ensure that the total energy difference between the wz and zb phases is converged to within 0.001 eV/Å. All of the structural parameters are fully relaxed to minimize the total energy.

The bulk properties of rock-salt CdO calculated by our derived SCC-DFTB parameters and PP-PBE (SIESTA) methods are given in Table 3 (top). The available experimental results^{83,84} are also shown. The SCC-DFTB results for most of the bulk properties agree well with the PP-PBE as well as experimental results. However, because of the strong wave function compression used for Cd, the value of the bulk modulus shows a large deviation. Again, the difference in cohesive energy is due to overestimation of the energy calculation in DFTB method. The band structure of bulk CdO is depicted in Figure 2 together with the reference PP-PBE one. From the figure, it is seen that DFTB reproduces the shape of the bands very well.

In Table 3, we have also presented the results of bulk properties for zinc-blende phase of CdS, CdSe, and CdTe. From the table, it is clear that bulk properties calculated using our derived

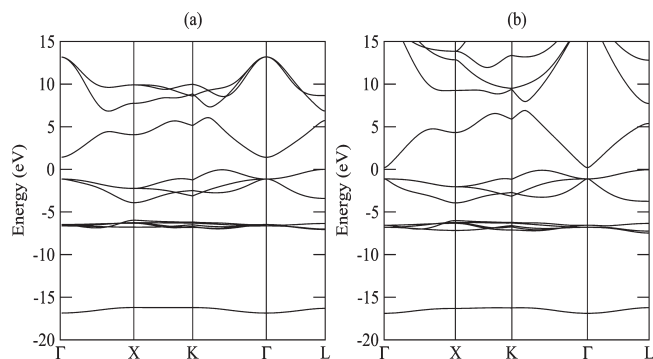


Figure 2. Electronic band structure of bulk rock-salt CdO calculated by (a) SCC-DFTB and (b) PP-PBE methods as implemented in the SIESTA package. Zero energy is set at the top of the valence band.

parameters agree well with those of PP-PBE and also experimental values.^{85–87} It is important to note that, although there were large differences in the calculated and experimental bulk modulus of elemental Cd, Se, and Te, the calculated bulk modulus for the cadmium chalcogenides shows very good agreement with experimental values. The band structures of the zinc-blende phase of cadmium chalcogenides calculated with the current SCC-DFTB parameter set and with PP-PBE are shown in Figure 3. From the figure, it is seen that, although SCC-DFTB bands are little compressed, it generally reproduces the band structure features of PP-PBE. Our SCC-DFTB band gap values are higher than the corresponding PP-PBE values and are close to the experimental ones.

In our SCC-DFTB calculation, for the wurtzite structures of cadmium chalcogenides (CdX; X = S, Se, and Te), we have also varied the c/a ratios to find the minimum energy lattice parameters. Another important parameter in periodic calculation is the internal parameter u . The assumption of equal nearest-neighbor bond lengths for the wurtzite crystal leads to the ideal value

$$u = \frac{1}{3} \left(\frac{a}{c} \right)^2 + \frac{1}{4} \quad (5)$$

In our calculation, the internal parameters u (in Å) are 0.3757, 0.3756, and 0.3754 for CdS, CdSe, and CdTe, respectively, very close to the ideal value of 0.375. This is due to the splittings of Cd-anion nearest neighbor bond lengths in the wz structure being very small. The slight decrease in the u parameter is consistent with the slight increases of the c/a ratio as the anion atomic number increases. This is because in wz semiconductors, due to the competition between bond-bending and bond-stretching forces, the c/a ratio and the u parameter always move in opposite directions. In Table 3 (bottom), we have also presented the lattice parameters and energetic properties of wurtzite cadmium chalcogenides. The SCC-DFTB lattice parameters a and c are more or less within 5% of the experimental values and very close to the PP-PBE calculations. The calculated bulk moduli using the DFTB method are in good agreement with experimental data as well as the PP-PBE method. It is observed that the bulk modulus decreases as the anion atomic number increases. At the equilibrium lattice constants, the calculated bulk moduli in the SCC-DFTB approach are 60.59, 51.02, and 38.08 GPa for the zb phase of CdS, CdSe, and CdTe, respectively. It is predicted that the bulk moduli for the wz structures are slightly smaller than their zb counterparts. The cohesive energies calculated by DFTB method

Table 3. Structural and Energetic Properties of Rock-Salt CdO and zb and wz Phases of CdS, CdSe, and CdTe Calculated Using SCC-DFTB and PP-PBE (Experimental Values Are Also Given)

	c (Å)	a (Å)	c/a	B (GPa)	E_c (eV/atom)	E_g (eV)
rock-salt CdO						
SCC-DFTB		4.68		168.11	8.65	2.56
PP-PBE		4.82		124.11	6.04	1.31
exptl ^d		4.70		130.10	6.08	2.38
zb-CdS						
SCC-DFTB		5.90		61.59	7.62	3.03
PP-PBE		5.98		52.54	5.49	1.22
exptl ^b		5.83		64.30	5.56	2.55
zb-CdSe						
SCC-DFTB		6.04		51.02	6.81	2.49
PP-PBE		6.26		44.34	5.07	0.69
exptl ^b		6.05		55.00	4.91	1.90
zb-CdTe						
SCC-DFTB		6.46		38.08	5.83	1.91
PP-PBE		6.68		35.92	4.52	1.02
exptl ^b		6.48		44.50	4.45	1.61
wz-CdS						
SCC-DFTB	6.80	4.17	1.63	58.71	7.62	3.04
PP-PBE	6.92	4.24	1.63	52.96	5.49	1.29
exptl ^c	6.71	4.14	1.621	62.00		2.58
wz-CdSe						
SCC-DFTB	6.96	4.27	1.63	50.72	6.82	2.51
PP-PBE	7.22	4.43	1.63	44.37	5.07	0.75
exptl ^c	7.01	4.30	1.63	53.00		1.83
wz-CdTe						
SCC-DFTB	7.44	4.56	1.63	37.94	5.84	1.89
PP-PBE	7.71	4.71	1.63	35.44	4.51	1.09
exptl ^c	7.45	4.55	1.637			1.60

^a Refs 83 and 84. ^b Refs 85–87. ^c Refs 88 and 91.

are in good agreement with the PP-PBE approach and experimental ones. The electronic band structures calculated using SCC-DFTB method and PP-PBE method for the wz bulk phase of cadmium chalcogenides are shown in Figure 4. The SCC-DFTB method reproduces the PP-PBE bands very well. Our calculated band structures for both zb and wz phases of all of the chalcogenides are in good agreement with other studies available in the literature.^{85,88–92} The band gap values we obtained in DFTB calculations are also close to the experimental values. Due to an incorrect description of the self-interaction in the exchange-correlation potential in the PP-PBE (or LDA), the band gaps are underestimated compared to the experimental values. With the slight differences between the zb and wz structures, the total energies, cohesive energies, and direct band gaps at the Γ point are expected to be similar for these two structures. Our SCC-DFTB calculations reproduce the energetic near-degeneracy of the two crystal structures within 5–8 MeV/Cd–X, which reflects the accuracy of the SCC-DFTB parameters.

3.3. CdX (X = S, Se, and Te) (10 $\bar{1}0$) and (11 $\bar{2}0$) Surfaces. The (10 $\bar{1}0$) and (11 $\bar{2}0$) surfaces were constructed from the

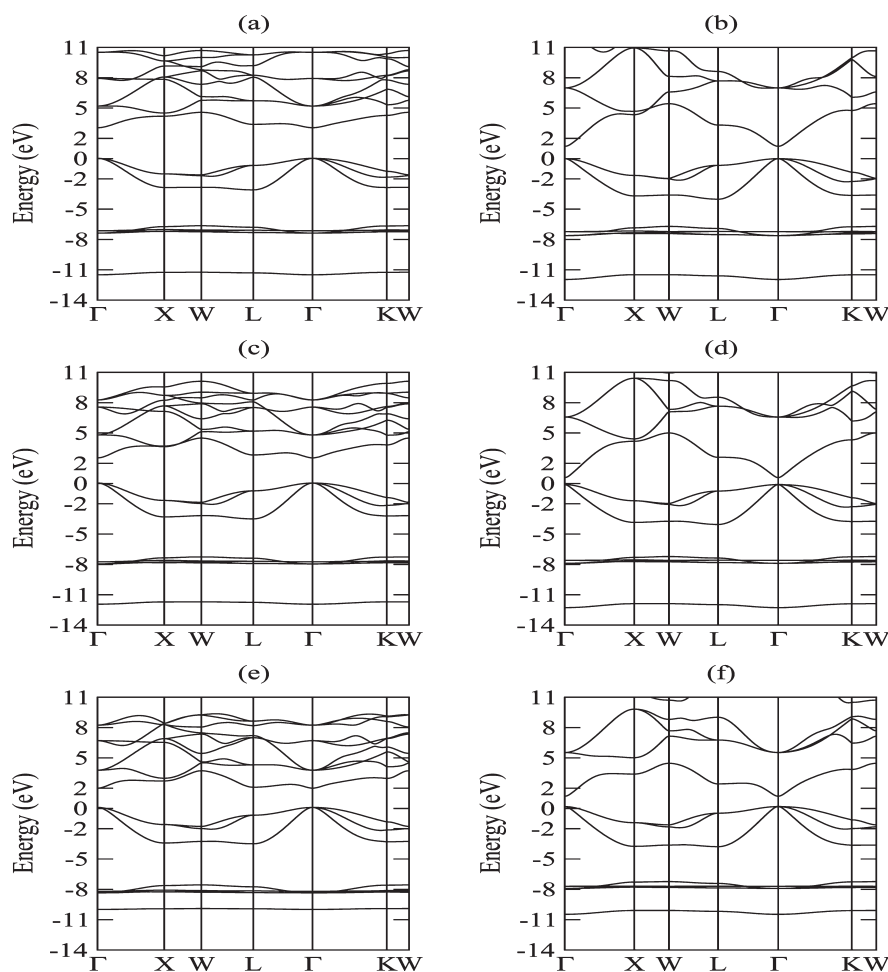


Figure 3. Electronic band structure of zb-CdS (a, b), zb-CdSe (c, d), and zb-CdTe (e, f) calculated by the SCC-DFTB (left panel) and PP-PBE (right panel) methods, as implemented in the SIESTA package. Zero energy is set at the top of the valence band for each case.

optimized bulk wz-CdX ($X = S, Se, \text{ and } Te$) structures. The optimized side views of the hexagonal $(10\bar{1}0)$ and $(11\bar{2}0)$ surfaces of one of the representative CdX system are shown in Figure 5. The results of surface relaxation obtained from SCC-DFTB and PP-PBE for all of the chalcogenides are shown in Table 4. From the table, it is seen that the results of surface relaxation obtained through SCC-DFTB and PP-PBE results are very close to each other. The surface relaxation for all systems normally causes the chalcogen atoms to move outward while Cd atoms move inward, and this behavior is in accord with other studies.^{93–98} Both the surfaces show strong surface relaxations. Thus, for $(10\bar{1}0)$ surfaces of all chalcogenides, the bond lengths d_{Cd-X} are largely shortened compared with the bulk values, while the bond lengths between the second and the third layers are little larger compared to the bulk value. For example, the Cd–S bond lengths (d_{Cd-S}) for the $(10\bar{1}0)$ surface of CdS is 2.22 Å compared to the bulk value of 2.56 Å. The Cd–S bond lengths (d'_{Cd-S}) between the second and third layers is 2.60 Å, which is slightly larger than the bulk value. The X–Cd–X angle (α) changes from the 109° bulk value to around 120° at the top surface layer for all systems, while the same angles in the inner layers (β) show very little change. However, for $(11\bar{2}0)$ surfaces of cadmium chalcogenides, Cd–X ($X = S, Se, \text{ and } Te$) bond lengths (d_{Cd-X}) of the top surface layer shortened, and also the Cd–X (d'_{Cd-X} and d''_{Cd-X}) bond lengths between the surface

and the second layer change from the bulk value. For example, Cd–S the bond length (d_{Cd-S}) in the top surface layer shortened from its bulk value of 2.56 Å to 2.45 Å, while the Cd–S lengths (d_{Cd-S} and d''_{Cd-S}) between the surface and the second layer are 2.60 Å and 2.59 Å, respectively, compared to their bulk value of 2.56 Å. However, all of the bond lengths (for both top surface layers and in the second layers) for other chalcogenides (CdSe and CdTe) shortened compared to the bulk value. The change in bond lengths Cd–X in the surface layer from the bulk values is relatively more for CdTe compared to that for CdSe and CdS. The X–Cd–X bond angles (α) at the top $(11\bar{2}0)$ surface layer increases while the Cd–X–Cd bond angles (β) decrease to a large extent from their bulk values. This contrasting behavior in the bond angle values is due to the inward movement of Cd atoms and outward movement of chalcogen atoms. However, in the inner layer, these angles remain almost the same as those of bulk values. The universal character of the surface atomic geometries for each cleavage surface was also obtained by Duke and Wang.^{93–96} The prediction of the surface relaxation of cadmium chalcogenides is similar to the results of a low-energy electron diffraction (LEED) study of related semiconductor ZnO.⁹⁷ The surface energies of both $(10\bar{1}0)$ and $(11\bar{2}0)$ surfaces of all of the chalcogenides calculated by SCC-DFTB and PP-PBE methods are also shown in Table 4. As expected, the SCC-DFTB method overestimates the surface energies compared to the

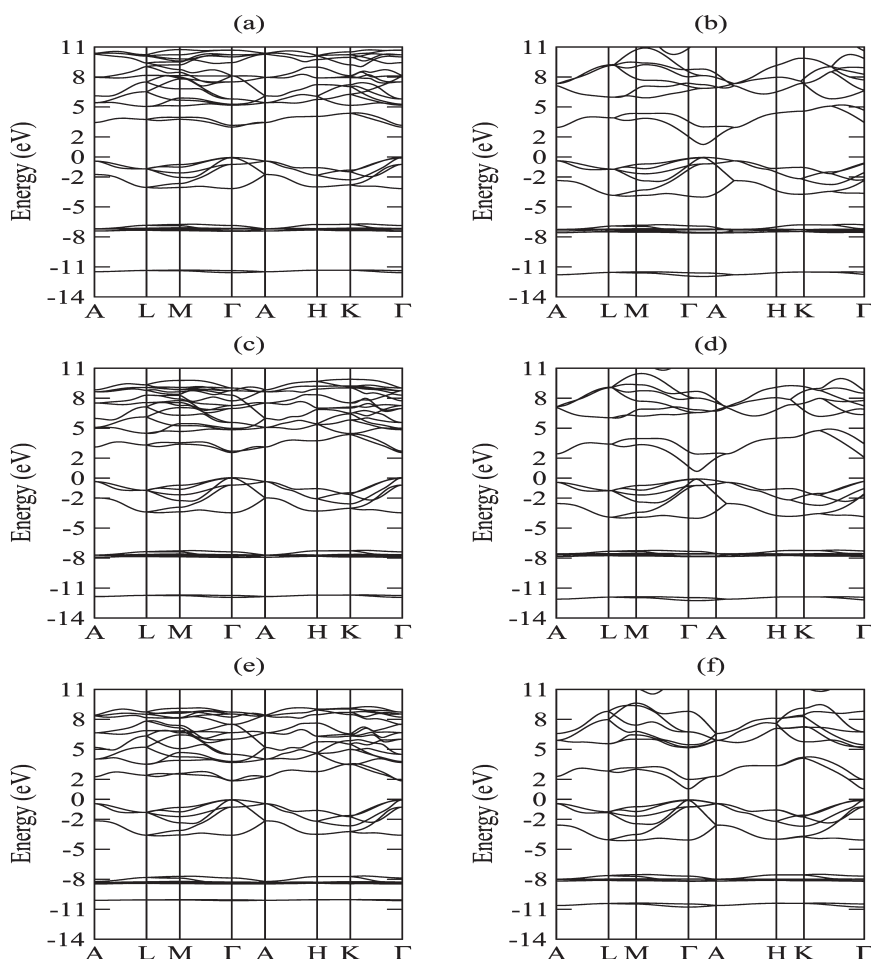


Figure 4. Electronic band structure of wz-CdS (a, b), wz-CdSe (c, d), and wz-CdTe (e, f) calculated by SCC-DFTB (left panel) and PP-PBE (right panel) methods as implemented in SIESTA package. Zero energy is set at the top of the valence band for each case.

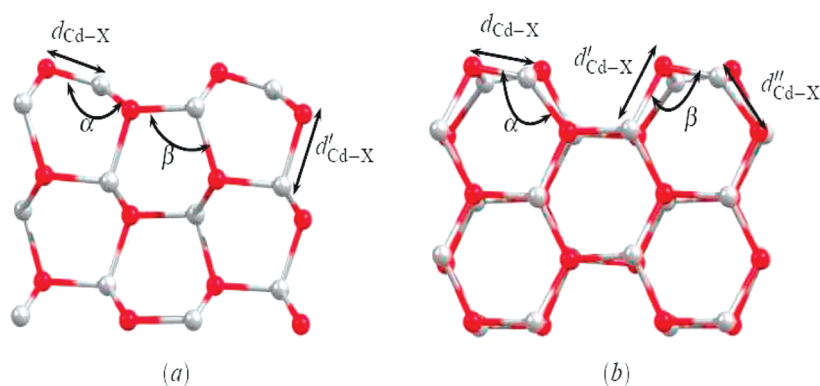


Figure 5. Optimized structure of representative CdX (a) $(10\bar{1}0)$ and (b) $(11\bar{2}0)$ surfaces (side view).

PP-PBE method. The calculated surface energies are in excellent agreement with the very recent study of Li et al.⁹⁸ The surface energy of CdX decreases as X changes from S to Te. The surface energy of all of these systems suggests that $(10\bar{1}0)$ surfaces are relatively more stable compared to corresponding $(11\bar{2}0)$ surfaces. In Figures 6 and 7, we have shown the band structures of $(10\bar{1}0)$ and $(11\bar{2}0)$ surfaces, respectively, calculated by both SCC-DFTB and PP-PBE methods. As is evident from the figures, the SCC-DFTB band dispersion agrees very well with those of

PP-PBE. As expected, the calculated SCC-DFTB band gap values in all cases are higher than the corresponding PP-PBE values. Our calculated band structure of $(10\bar{1}0)$ surfaces of CdS and CdSe are also in good agreement with those studied by Duke and Wang.^{93–96}

3.4. CdX (X = S, Se, and Te) $(10\bar{1}0)$ and $(11\bar{2}0)$ Faceted Hexagonal Nanowires. Semiconductor nanowires have been the subject of intense investigation because of their diverse application in the field of nanotechnology.^{15,16} So, for further

validation of SCC-DFTB parametrization, we investigated (10 $\bar{1}0$) and (11 $\bar{2}0$) faceted hexagonal nanowires for all cadmium chalcogenides. In Figure 8, we have shown the optimized cross sections of one of the representative (10 $\bar{1}0$) and (11 $\bar{2}0$) faceted hexagonal nanowires of one chalcogenide. All of the nanowires

Table 4. Calculated Geometry (Bond Lengths Are in Å and Bond Angles Are in deg) and Surface Energy γ (in J/m²) of Both (10 $\bar{1}0$) and (11 $\bar{2}0$) Surfaces of Cadmium Chalcogenides

	(10 $\bar{1}0$)					(11 $\bar{2}0$)					
	$d_{\text{Cd-X}}$	$d'_{\text{Cd-X}}$	α	β	γ	$d_{\text{Cd-X}}$	$d'_{\text{Cd-X}}$	$d''_{\text{Cd-X}}$	α	β	γ
CdS											
SCC-DFTB	2.22	2.60	122	107	0.59	2.45	2.60	2.59	113	97	0.63
PP-PBE	2.46	2.62	123	104	0.35	2.50	2.56	2.57	120	92	0.34
CdSe											
SCC-DFTB	2.38	2.71	121	104	0.39	2.45	2.58	2.57	117	97	0.40
PP-PBE	2.59	2.74	124	102	0.34	2.64	2.68	2.69	121	90	0.28
CdTe											
SCC-DFTB	2.55	2.94	121	102	0.22	2.61	2.68	2.68	118	96	0.23
PP-PBE	2.77	2.90	125	103	0.21	2.82	2.85	2.86	122	88	0.21

follow a similar relaxation in the surface layers to that of the corresponding (10 $\bar{1}0$) and (11 $\bar{2}0$) surfaces. Thus, the surface cadmium atoms move relatively inward while the chalcogenides move outward. In Table 5, we have presented the details of the surface relaxation calculated using both the SCC-DFTB and PP-PBE methods of all (10 $\bar{1}0$) and (11 $\bar{2}0$) faceted CdX nanowires. As is found for surfaces, the Cd–X bond lengths ($d_{\text{Cd-X}}$) for (10 $\bar{1}0$) faceted nanowires are shortened compared with their bulk values. However, the change in Cd–X bond lengths ($d'_{\text{Cd-X}}$) between the second and third layers is very small for CdS, while those of CdSe and CdTe are shortened compared to their bulk values. This is in contrast to the behavior of the corresponding surfaces where the bond lengths were expanded compared to their bulk values. The X–Cd–X angles of both the top surface layer (α) and the inner layer (β) for all systems show a large deviation from the bulk value (109°). The results of surface relaxation of the (11 $\bar{2}0$) surfaces are very much similar to those of the corresponding surfaces. In the table, we have also presented the value of the formation energy of the nanowires. The formation energy (E_f) of a given nanowire is defined by

$$E_f = \left(\frac{E_{\text{NW}}}{n} - \frac{E_{\text{bulk}}}{2} \right) \quad (6)$$

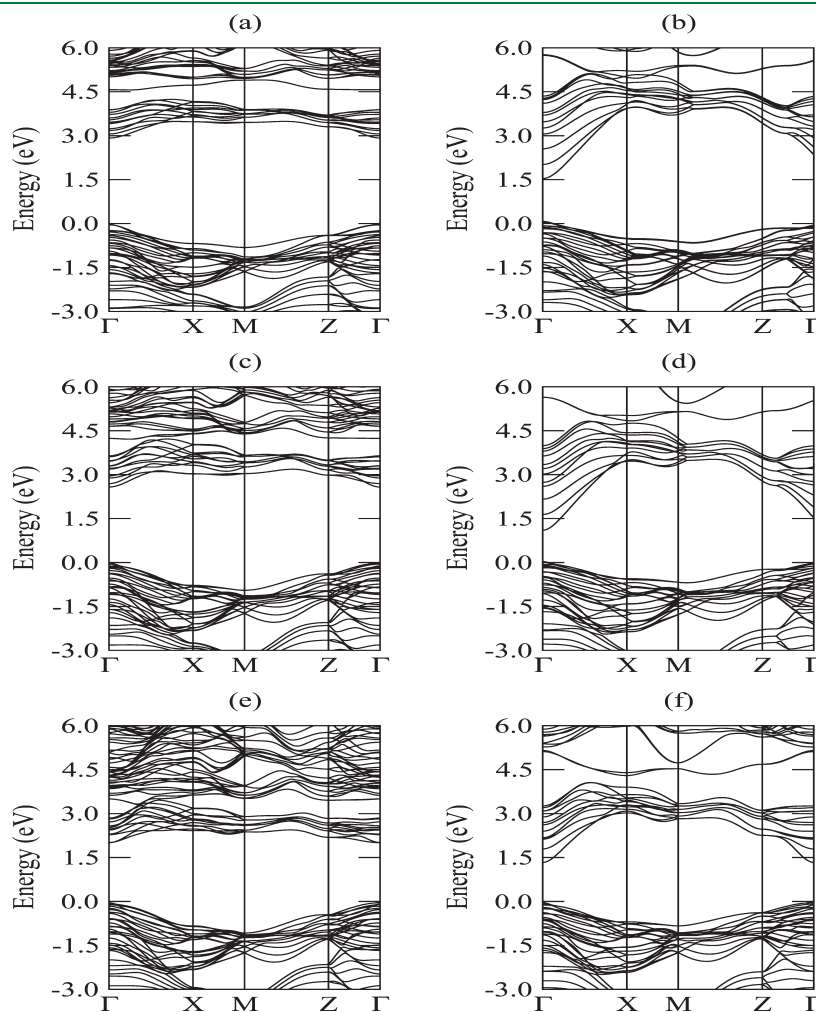


Figure 6. Electronic band structures of CdX [X = S (a, b), Se (c, d), and Te (e, f)] (10 $\bar{1}0$) surfaces, calculated with SCC-DFTB (left panel) and PP-PBE (right panel) methods, as implemented in the SIESTA package. Zero energy is set at the top of the valence band for each case.

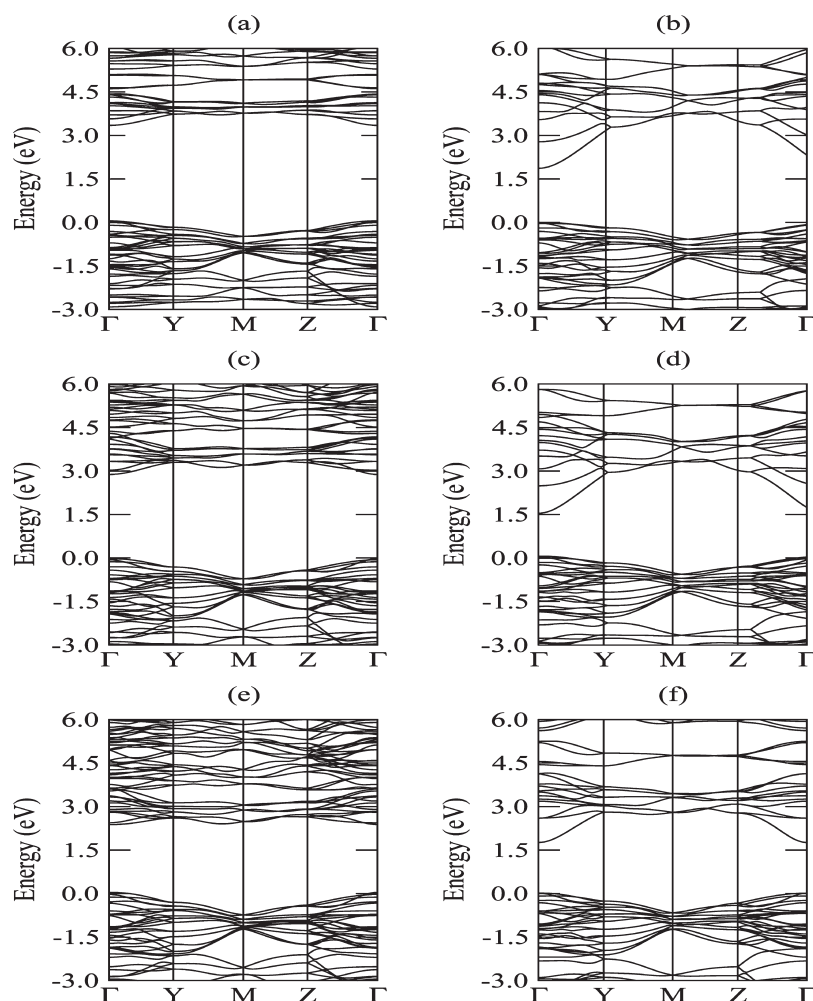


Figure 7. Electronic band structures of CdX [$X = \text{S}$ (a, b), Se (c, d), and Te (e, f)] ($11\bar{2}0$) surfaces, calculated with SCC-DFTB (left panel) and PP-PBE (right panel) methods, as implemented in the SIESTA package. Zero energy is set at the top of the valence band for each case.

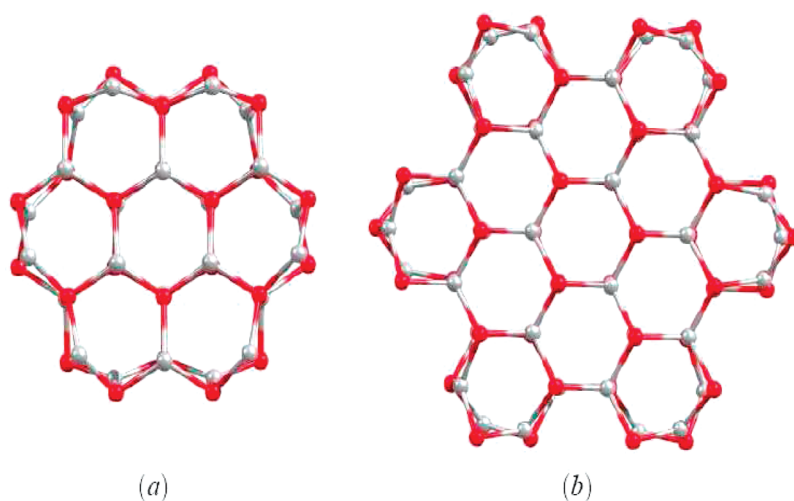


Figure 8. Optimized structure of representative CdX (a) ($10\bar{1}0$) and (b) ($11\bar{2}0$) nanowires (top view).

where E_{NW} is the total energy of the nanowire, n is the total number of Cd–X units in the CdX nanowire, and E_{bulk} is the energy of the bulk wz-CdX. From the values of the formation

energy, we can infer that ($10\bar{1}0$) faceted nanowires are relatively more stable compared to ($11\bar{2}0$) faceted nanowires, although the differences in formation energy of these two forms of nanowire

are very small. As expected, the value of the formation energy of both $(10\bar{1}0)$ and $(11\bar{2}0)$ faceted nanowires decreases from CdS to CdTe. In most of the cases, our SCC-DFTB results are close to the corresponding PP-PBE results. In Figure 9, we have shown the band structure of $(10\bar{1}0)$ faceted nanowires of CdX ($X = S, Se, \text{ and } Te$) respectively calculated with both SCC-DFTB and PP-PBE methods. In Figure 10, we have shown the same for

Table 5. Calculated Geometry (Bond Lengths Are in Å and Bond Angles Are in deg) and Formation Energy E_f (in eV/pair) of Both $(10\bar{1}0)$ and $(11\bar{2}0)$ Nanowires of Cadmium Chalcogenides

	$(10\bar{1}0)$					$(11\bar{2}0)$					
	d_{Cd-X}	d'_{Cd-X}	α	β	E_f	d_{Cd-X}	d'_{Cd-X}	d''_{Cd-X}	α	β	E_f
CdS											
SCC-DFTB	2.29	2.57	124	118	0.48	2.53	2.58	2.54	111	96	0.42
PP-PBE	2.49	2.55	120	118	0.27	2.52	2.56	2.53	116	91	0.24
CdSe											
SCC-DFTB	2.44	2.55	122	119	0.32	2.48	2.57	2.50	116	96	0.29
PP-PBE	2.63	2.68	120	119	0.24	2.65	2.68	2.65	118	89	0.21
CdTe											
SCC-DFTB	2.60	2.66	121	120	0.20	2.64	2.68	2.65	116	95	0.19
PP-PBE	2.81	2.86	120	120	0.21	2.83	2.86	2.83	118	87	0.18

$(11\bar{2}0)$ faceted nanowires. As is evident from the figures, SCC-DFTB reproduces all of the important bands very well. Because of the use of a minimal basis set, the conduction bands show little deviation from PP-PBE results. The other feature is that the SCC-DFTB bands are relatively more compressed compared to PP-PBE bands. The bands of all studied nanowires are more localized compared to the corresponding surfaces. Our calculated SCC-DFTB electronic band structures are also similar to that obtained by Huang et al.⁹⁹ The band gap values obtained by using the SCC-DFTB method are higher than those of the PP-PBE method. As mentioned earlier, this is due to the fact that the SCC-DFTB method generally overestimates this property. Interestingly, the band gap values of studied nanowires are larger compared to their bulk values, consistent with the quantum confinement effects. There are experimental studies on CdS and CdSe nanowires with diameters of around 70–100 nm, and their band gap values are almost close to bulk band gap values.^{100,101} So, nanowires of smaller diameters will show a quantum confinement effect and correspondingly have higher band gap values. Huang et al. also reported similar band gap behavior in their recent study on CdS and CdSe NWs.⁹⁹ The SCC-DFTB prediction of the stability order of these nanowires also is the same as that of PP-PBE results.

3.5. Small Molecular Systems. Finally, we have tested the transferability of our derived SCC-DFTB parameters with respect to DFT by calculating the structural and energetic

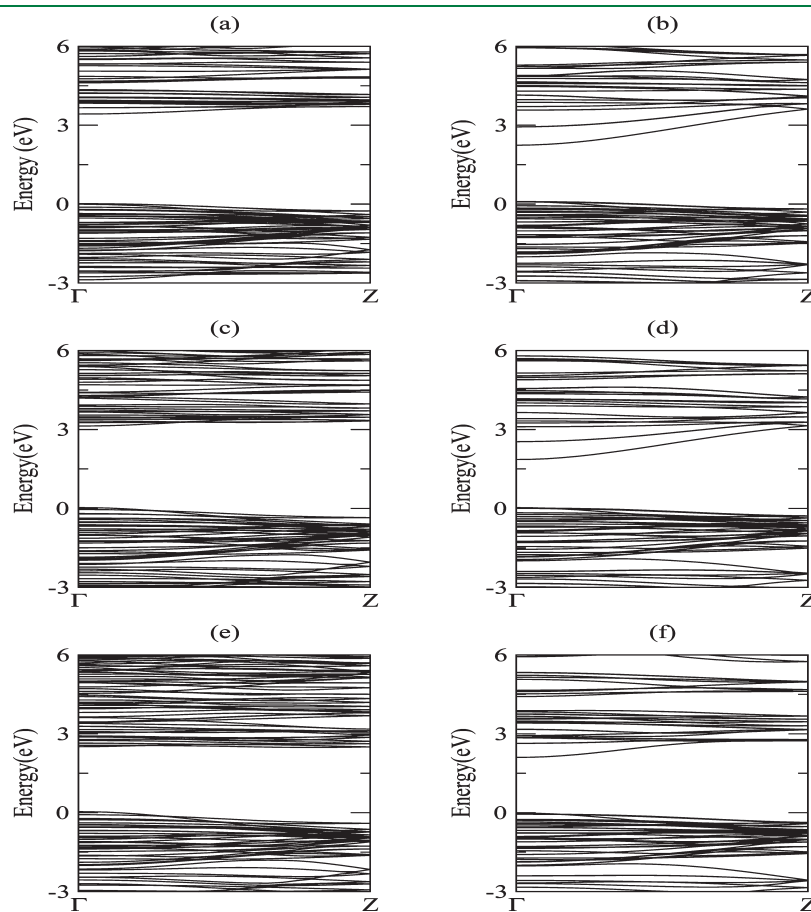


Figure 9. Electronic band structure of $(10\bar{1}0)$ faceted nanowires for CdS (a, b), CdSe (c, d), and CdTe (e, f) calculated with the SCC-DFTB (left panel) and PP-PBE (right panel) methods, as implemented in the SIESTA package. Zero energy is set at the top of the valence band for each case.

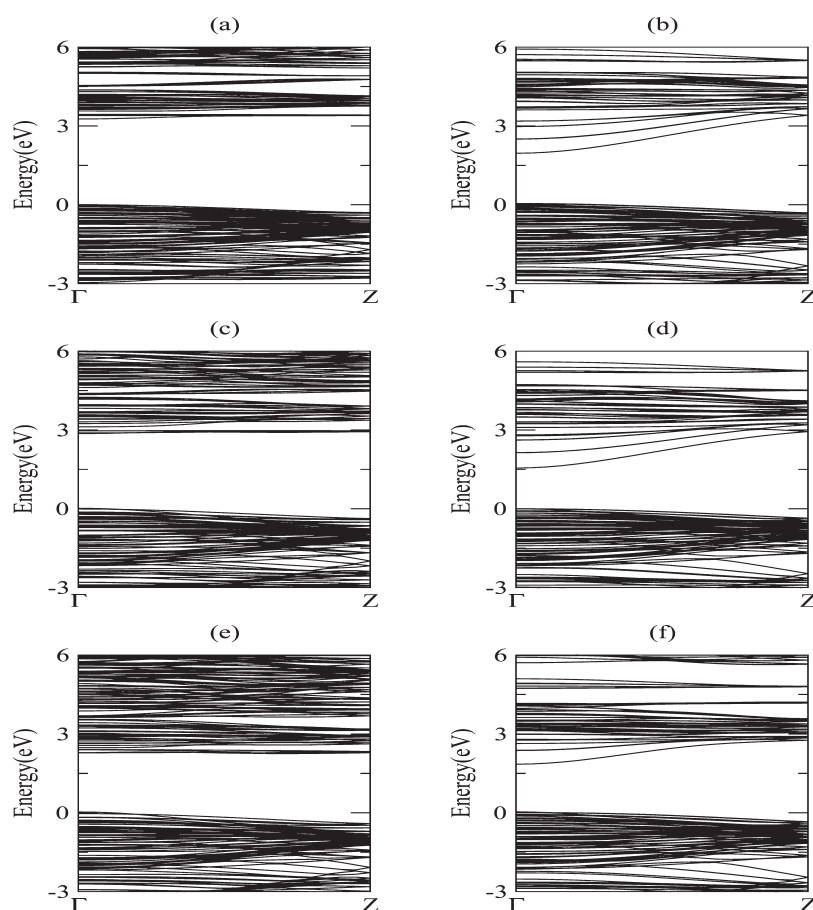


Figure 10. Electronic band structure of (11 $\bar{2}$ 0) faceted nanowires for CdS (a, b), CdSe (c, d), and CdTe (e, f) calculated with the SCC-DFTB (left panel) and PP-PBE (right panel) methods, as implemented in the SIESTA package. Zero energy is set at the top of the valence band for each case.

Table 6. Atomic-Spin-Dependent Constants $W_{Al'l}$ of Se and Te (in au)

$W_{Al'l}$			Se	Te
W_{ss}	W_{sp}	W_{sd}	-0.018	-0.014
W_p	W_{pp}	W_{pd}	-0.014	-0.014
W_{ds}	W_{dp}	W_{dd}	-0.000	-0.096
			-0.014	-0.011
			-0.011	-0.011
			-0.000	-0.083

properties of some simple diatomic molecules and small molecular systems. As all systems studied so far are closed-shell systems, we used the nonspin polarized SCC-DFTB method. However, for diatomic molecules, we have used the spin-polarized SCC-DFTB method of Köhler et al.¹⁰² The atomic-spin-dependent constants $W_{Al'l}$ needed for the spin-polarized SCC-DFTB method are calculated by taking the second derivatives of the total atomic energy with respect to the spin density; at the point where the spin density is zero, this derivative reduces to

$$W_{Al'l} = \frac{1}{2} \left(\frac{\partial \epsilon_{Al}^\uparrow}{\partial n_l^\uparrow} - \frac{\partial \epsilon_{Al}^\downarrow}{\partial n_l^\downarrow} \right)_{\rho=0} = W_{Al'l} \quad (7)$$

where n_l and $n_{l'}$ are the occupation numbers of atomic shells l and l' , respectively, and ϵ_{Al}^\uparrow is the atomic Kohn–Sham orbital energy for alpha (\uparrow) spin. The second derivative values are computed

Table 7. Comparison between SCC-DFTB, PP-PBE, and B3LYP Results on the Binding Energy (in eV) and Equilibrium Bond Length (in Å) of Simple Diatomic Molecules^a

	dimer	SCC-DFTB		PP-PBE		B3LYP/SBK+6-31G(d,p)	
		E_b	r	E_b	r	E_b	r
d	Cd–H	0.86	1.93	0.86	1.86	0.82	1.85
t	Cd–C	2.48	2.16	1.54	2.19	2.66	2.23
q	Cd–N	1.67	2.22	0.44	2.39	2.97	2.48
s	Cd–O	3.15	1.89	1.21	1.94	3.21	1.92
d	Se–H	3.15	1.51	3.08	1.51	3.71	1.49
s	Se–C	6.70	1.71	6.44	1.71	8.89	1.69
d	Se–N	5.48	1.71	3.89	1.70	7.42	1.69
t	Se–O	4.95	1.72	4.52	1.71	4.17	1.70
t	Se–S	3.92	2.05	3.98	2.10	3.48	2.11
d	Te–H	2.84	1.70	2.67	1.72	3.12	1.68
s	Te–C	5.10	1.94	5.05	1.95	7.16	1.92
d	Te–N	3.70	1.93	2.81	1.95	6.47	1.91
t	Te–O	4.35	1.95	3.86	1.94	3.62	1.87
t	Te–S	3.77	2.33	3.54	2.32	3.94	2.31
t	Te–Se	3.16	2.50	3.16	2.49	3.38	2.49

^as, d, t, and q denote singlet, doublet, triplet, and quartet spin states, respectively.

using the finite difference method. We here only calculated the values of atomic spin constants W_{AIF} for Se and Te, and they are given in Table 6. In case of the Cd atom, it is not required due to the closed-shell nature of it. For all other atoms like H, C, N, O, and S, we have used the spin constants calculated previously.¹⁰²

The binding energies and equilibrium bond lengths calculated by the SCC-DFTB method, PP-PBE method, and also B3LYP/SBK+6-31G(d,p) of several diatomic systems are shown in Table 7. The SCC-DFTB results in most of the cases are very

Table 8. Comparison between SCC-DFTB, SIESTA (PP-PBE), G03 (B3LYP), and Experimental Results on the Binding Energy (in eV) and Equilibrium Bond Angle (in deg) and Length (in Å) of AB₂-Type Molecules^a

Molecule	SCC-DFTB			PP-PBE			B3LYP/- SBK+6-31G- (d,p)			Exp.		
	E _b	θ	r	E _b	θ	r	E _b	θ	r	E _b	θ	r
CdH ₂	3.88	180	1.76	4.23	180	1.71	4.10	180	1.75	3.06	180	1.75 ^b
SeH ₂	3.36	97	1.50	3.17	90	1.51	4.05	91	1.48	3.42	91.5	1.48 ^c
SeO ₂	5.00	115	1.71	3.97	111	1.69	3.87	112.7	1.65	4.37	114	1.61 ^d
SeS ₂	3.45	119	2.09	3.05	115	2.13	2.81	115	2.14			
TeH ₂	2.98	95	1.70	2.77	89	1.71	3.55	89.6	1.68	2.82	90	1.69 ^e
TeO ₂	4.49	116	1.93	3.43	109	1.92	3.67	111.5	1.81	3.37	102	1.90 ^f
TeS ₂	3.57	117	2.34	2.88	112	2.33	2.65	113	2.31		103.4	2.39 ^g
TeSe ₂	2.88	118	2.52	2.50	113	2.52	3.07	113	2.51			2.50 ^g

^a All molecules have been calculated in the singlet state. ^b Ref 107. ^c Refs 108 and 109. ^d Ref 110. ^e Ref 106. ^f Refs 104 and 110. ^g Ref 104.

promising. Experimental results for some diatomic molecules are also available, and we have compared our theoretical results with these. For example, CdH is known to have an experimental binding energy of 0.71 eV and an equilibrium bond length of 1.86 Å.¹⁰³ Our calculated binding energy (0.86 eV) and bond length (1.93 Å) are close to experimental values. The ground state of the CdC molecule is a triplet, and the experimental binding energy and bond length are 1.91 eV and 2.17 Å, respectively.¹⁰⁴ Our SCC-DFTB calculation predicts the binding energy to be 2.48 eV, slightly overestimated, but it predicts the bond length (2.16 Å) more accurately (within 0.01 Å). The bond length of the diatomic molecule CdO in its singlet ground state calculated by the SCC-DFTB method (1.90 Å) also agrees well with the experimental value of 1.90 Å.¹⁰⁵ In the cases of Se–B (B = O and S) and Te–B (B = O, S and Se), the bond lengths increase while the binding energies decrease with an increasing size of B. This fact is in good agreement with the other DFT calculations and also with the experimental results.^{104,106}

In Table 8, we presented our results on AB₂ types of molecules. The equilibrium bond lengths of several systems calculated by SCC-DFTB are close to the PP-PBE and B3LYP/SBK+6-31G(d,p) results. However, the bond angles are larger than the corresponding DFT values. As an illustrative example, we consider the CdH₂ molecule. The CdH₂ molecule in its singlet ground state is known experimentally to have a bond length of 1.75 Å and a bond angle of 180° from infrared spectroscopy of the matrix isolated molecule.¹⁰⁷ The equilibrium bond length and bond angle calculated by SCC-DFTB parametrization reproduces these geometrical features very well, and they are also close to the values obtained from PP-PBE and B3LYP/SBK+6-31G(d,p) calculations. The good agreement among SCC-DFTB, PP-PBE,

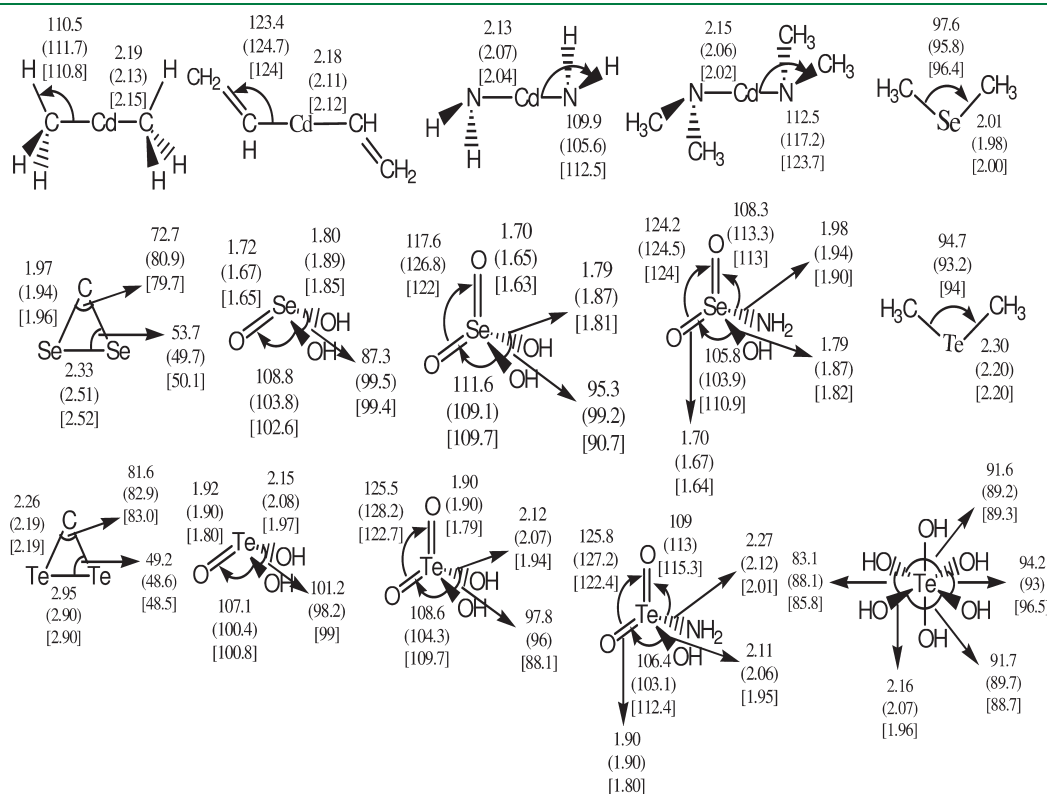


Figure 11. Schematic view of a few small molecules together with their equilibrium geometric parameters obtained with SCC-DFTB (without brackets), PP-PBE in parentheses, and B3LYP/SBK+6-31G(d,p) in brackets. The bond lengths are in Å and angles are in deg.

B3LYP/SBK+6-31G(d,p), and experimental^{108–111} results on bond lengths, bond angles, and binding energies (with usual DFTB overestimation) of several AB₂ type molecules, as presented in Table 8, makes us believe that the derived parameters are highly transferable, and one can study either the adsorption of small molecules on the cadmium chalcogenide surfaces or different types of surface passivation effects.

In addition, we have calculated the geometric features of several model molecules with SCC-DFTB and PP-PBE and also with B3LYP/SBK+6-31G(d,p). The results are shown in Figure 11. The bond lengths and bond angles of several model molecules calculated using the SCC-DFTB method are very close to the other DFT calculations. The binding energies per bond are also in good agreement with those of PP-PBE and B3LYP/SBK+6-31G(d,p) calculations. The experimental results of a few molecules studied here are also available in the literature. Thus, the experimental Cd–C bond length in Cd(CH₃)₂ is 2.21 Å. The bond length obtained with the SCC-DFTB method is very close to the experimental value.¹¹² The experimental bond length and bond angle of Se(CH₃)₂ are found to be 1.98 Å and 98°, respectively.^{113,114} These values are also in very good agreement with our SCC-DFTB values.

4. CONCLUSION

A complete SCC-DFTB parametrization for cadmium and its interaction with oxygen, carbon, hydrogen, sulfur, selenium, tellurium, etc. has been developed in order to get the best possible description of periodic and molecular systems involving these elements. Using this parameter set, we have calculated the bulk properties (*i.e.*, lattice constants, bulk moduli, cohesive energies, band structures, etc.) of hcp-Cd, trigonal Se and Te, rock-salt CdO, and zb and wz cadmium chalcogenides (CdS, CdSe, CdTe). Our calculated values are in good agreement with the reference DFT calculations and also with experimental results. We have also checked the transferability of SCC-DFTB parameters by calculating the electronic structure of several nanostructures such as surfaces and nanowires. Finally, we have tested the transferability of our SCC-DFTB parameters by calculating the structural and energetic properties of some simple diatomic molecules and small molecular systems. The real world application of semiconductor nanoparticles requires the integration of smaller nanoparticles and 1-d nanostructures to form large nanostructures. However, to date, theoretical studies in this particular field are restricted to only isolated nanoparticles and 1-d nanotubes or wires. This is primarily because of the unavailability of appropriate methods to handle nanostructures composed of a large number of atoms. We do hope that, with the derived SCC-DFTB parameter of the cadmium chalcogenides set and taking advantage of the computational efficiency of the SCC-DFTB method, one may be able to study large nanostructures of cadmium chalcogenides.

AUTHOR INFORMATION

Corresponding Author

*E-mail: pranab.sarkar@visva-bharati.ac.in.

ACKNOWLEDGMENT

The financial support from DST, New Delhi [SR/NM/NS-49/2007] and UGC, New Delhi, through research grants are

gratefully acknowledged. S.S. is grateful to CSIR, New Delhi for the award of a Junior Research Fellowship (JRF). A.L.R. acknowledges the German Science Foundation under the grants SPP-1165 and DIP-40100474.

REFERENCES

- (1) Heath, J. R.; Shiang, J. J. *Chem. Soc. Rev.* **1998**, *27*, 65–71.
- (2) Bruchez, M.; Morronne, M.; Gin, P.; Weiss, S.; Alivisatos, A. P. *Science* **1998**, *281*, 2013–2016.
- (3) Alivisatos, A. P. *Science* **1996**, *271*, 933–937.
- (4) Hoffman, M. R.; Martin, S. T.; Choi, W.; Bahnemann, D. W. *Chem. Rev.* **1995**, *95*, 69–96.
- (5) Burda, C.; Chen, X.; Narayanan, R.; El-Sayed, M. A. *Chem. Rev.* **2005**, *105*, 1025–102.
- (6) Fu, X.; Peng, Z.-J.; Li, D.; Zhang, L.; Xiao, J.-Y.; Fang, Z.-Y. *Nanotechnology* **2011**, *22*, 175601.
- (7) Pan, H.; Poh, C. K.; Zhu, Y.; Xing, G.; Chin, K. C.; Feng, Y. P.; Lin, J.; Sow, C. H.; Ji, W.; Wee, A. T. S. *J. Phys. Chem. C* **2008**, *112*, 11227.
- (8) Willner, I.; Willner, B. *Pure Appl. Chem.* **2002**, *74*, 1773–1783.
- (9) Nisman, R.; Dellaire, G.; Ren, Y.; Li, R.; Bazett-Jones, D. P. *J. Histochem. Cytochem.* **2004**, *52*, 13.
- (10) Mansson, A.; Sundberg, M.; Balaz, M.; Bunk, R.; Nicholls, I. A.; Omling, P.; Tagerud, S.; Montelius, L. *Biochem. Biophys. Res. Commun.* **2004**, *314*, 529.
- (11) Nishiuchi, K.; et al. *Jpn. J. Appl. Phys.* **1998**, *37*, 2163.
- (12) Klein, D. L.; Roth, R.; Lim, A. K. L.; Alivisatos, A. P.; McEuen, P. L. *Nature* **1997**, *389*, 699.
- (13) Hu, J.; Li, L.-S.; Yang, W.; Manna, L.; Wang, L.-W.; Alivisatos, A. P. *Science* **2001**, *292*, 2060.
- (14) Murray, C. B.; Kagan, C. R.; Bawendi, M. G. *Science* **1995**, *270*, 1335.
- (15) Cui, Y.; Wei, Q. Q.; Park, H. K.; Lieber, C. M. *Science* **2001**, *293*, 1289.
- (16) Wang, D. M.; Song, J.; Mu, Y.; Lieber, C. M. *Nano Lett.* **2003**, *3*, 1255.
- (17) Wang, G. X.; Park, M. S.; Liu, H. K.; Wexler, D.; Chen, J. *Appl. Phys. Lett.* **2006**, *88*, 193115.
- (18) Colvin, V. L.; Schlamp, M. C.; Alivisatos, A. P. *Nature* **1994**, *370*, 354–357.
- (19) Radloff, C.; Moran, C. E.; Jackson, J. B.; Halas, N. J. *Mol. Nanoelectron.* **2003**, 229–262.
- (20) Yu, A. M.; Meiser, F.; Cassagneau, T.; Caruso, F. *Nano Lett.* **2004**, *4*, 177.
- (21) Pinaud, F.; King, D.; Moore, H. P.; Weiss, S. J. *Am. Chem. Soc.* **2004**, *126*, 6115.
- (22) Sukhanova, A.; Devy, M.; Venteo, L.; Kaplan, H.; Artemyev, M.; Oleinikov, V.; Klinov, D.; Pluot, M.; Cohen, J. H. M.; Nabiev, I. *Anal. Biochem.* **2004**, *324*, 60.
- (23) Derfus, A. M.; Chan, W. C. W.; Bhatia, S. N. *Nano Lett.* **2004**, *4*, 11.
- (24) Wang, L.-W.; Zunger, A. *Phys. Rev. B* **1996**, *53*, 9579.
- (25) Mizel, A.; Cohen, M. L. *Phys. Rev. B* **1997**, *56*, 6737.
- (26) Eichkorn, K.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *288*, 235.
- (27) Ramkrishna, M. V.; Friesner, R. A. *J. Chem. Phys.* **1991**, *95*, 8309.
- (28) Hill, N. A.; Whaley, K. B. *J. Chem. Phys.* **1994**, *100*, 2831.
- (29) Tomasulo, A.; Ramkrishna, M. V. *J. Chem. Phys.* **1996**, *105*, 3612.
- (30) Wang, L.-W.; Zunger, A. *J. Phys. Chem. B* **1998**, *102*, 6449.
- (31) Leung, K.; Pokrant, S.; Whaley, K. B. *Phys. Rev. B* **1998**, *57*, 12291.
- (32) Pokrant, S.; Whaley, K. B. *Eur. Phys. J. D* **1999**, *6*, 255.
- (33) Troparevsky, M. C.; Chelikowsky, J. R. *J. Chem. Phys.* **2001**, *114*, 943.
- (34) Deglmann, P.; Ahlrichs, R.; Tsereteli, K. *J. Chem. Phys.* **2002**, *118*, 1585.

- (35) Joswig, J.-O.; Springborg, M.; Seifert, G. *J. Phys. Chem.* **2000**, *104*, 2617.
- (36) Sarkar, P.; Springborg, M. *Phys. Rev. B* **2003**, *68*, 235409.
- (37) Spano, E.; Hamad, S.; Catlow, C. R. *J. Phys. Chem. B* **2003**, *107*, 10337.
- (38) Hamad, S.; Catlow, C. R.; Spano, E.; Matxain, J. M.; Ugalde, J. M. *J. Phys. Chem. B* **2005**, *109*, 2703.
- (39) Sapra, S.; Sarma, D. D. *Phys. Rev. B* **2004**, *69*, 125304.
- (40) Drager, E. W.; Grossman, J. C.; Williamson, A. J.; Galli, G. *Phys. Rev. Lett.* **2003**, *90*, 167402.
- (41) Drager, E. W.; Grossman, J. C.; Williamson, A. J.; Galli, G. *J. Chem. Phys.* **2003**, *120*, 10807.
- (42) Huang, X.; Lindgren, E.; Chelikowsky, J. R. *Phys. Rev. B* **2005**, *71*, 165328.
- (43) Roy, S.; Springborg, M. *J. Phys. Chem. A* **2005**, *109*, 1324.
- (44) Pal, S.; Goswami, B.; Sarkar, P. *J. Chem. Phys.* **2005**, *123*, 044311.
- (45) Goswami, B.; Pal, S.; Sarkar, P.; Seifert, G.; Springborg, M. *Phys. Rev. B* **2006**, *73*, 205312.
- (46) Joswig, J.-O.; Seifert, G.; Niehaus, T. A.; Springborg, M. *J. Phys. Chem. B* **2003**, *107*, 2897.
- (47) Pal, S.; Sharma, R.; Goswami, B.; Sarkar, P.; Bhattacharyya, S. P. *J. Chem. Phys.* **2009**, *130*, 214703.
- (48) Rogach, A. L.; Franzl, T.; Klar, T. A.; Feldmann, J.; Gaponik, N.; Lesnyak, V.; Shavel, A.; Eychmüller, A.; Rakovich, Y. P.; Donegan, J. F. *J. Phys. Chem. C* **2007**, *111*, 14628.
- (49) Kamat, P. V. *J. Phys. Chem. C* **2007**, *111*, 2834.
- (50) Plass, R.; Pelet, S.; Krueger, J.; Gratzel, M.; Bach, U. *J. Phys. Chem. B* **2002**, *106*, 7578–7580.
- (51) Peng, X. G.; Manna, L.; Yang, W. D.; Wickham, J.; Scher, E.; Kadavanich, A.; Alivisatos, A. *Nature* **2000**, *404*, 59–63.
- (52) Tang, K.; Qian, Y.; Zeng, Z.; Yang, X. *Adv. Mater.* **2003**, *15*, 448–450.
- (53) Liu, Y.; Xu, Y.; Li, J.-P.; Zhang, B.; Wu, D.; Sun, Y.-H. *Chem. Mater.* **2004**, *33*, 1162–1163.
- (54) Ma, C.; Ding, Y.; Moore, D.; Wang, X.; Wang, Z. L. *J. Am. Chem. Soc.* **2004**, *126*, 708–709.
- (55) Korgel, B. A.; Monbouquette, H. G. *Langmuir* **2000**, *16*, 3588–3594.
- (56) Petrov, D. V.; Santos, B. S.; Pereira, G. A. L.; Donegat', C. D. M. *J. Phys. Chem. B* **2002**, *106*, 5325–5334.
- (57) Zhong, X.; Liu, S.; Zhang, Z.; Li, L.; Wei, Z.; Knoll, W. *J. Mater. Chem.* **2004**, 2790–2794.
- (58) Zhong, X.; Han, M.; Dong, Z.; White, T. J.; Knoll, W. *J. Am. Chem. Soc.* **2003**, *125*, 8589–8594.
- (59) Zhong, X.; Zhang, Z.; Liu, S.; Han, M.; Knoll, W. *J. Phys. Chem. B* **2004**, *108*, 15552–15559.
- (60) Rakovich, Y. P.; Volkov, Y.; Sapra, S.; Susha, A. S.; Döblinger, M.; Donegan, J. F.; Rogach, A. L. *J. Phys. Chem. C* **2007**, *111*, 18927.
- (61) Corrigan, J. F.; DeGroot, M. W. Large Semiconductor Molecules. In *The Chemistry of Nanomaterials: Synthesis, Properties and Applications*, 1st ed.; Rao, C. N. R., Mueller, A., Cheetham, A. K., Eds.; Wiley-VCH: Weinheim, Germany, 2004; Vol. 2, pp 418–451.
- (62) Porezag, D.; Frauenheim, Th.; Köhler, Th.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947–12957.
- (63) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Fraunheim, Th.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- (64) Niehaus, Th.; Suhai, S.; DellaSala, F.; Lugli, P.; Elstner, M.; Seifert, G.; Frauenheim, Th. *Phys. Rev. B* **2001**, *63*, 085108.
- (65) Seifert, G. *J. Phys. Chem. A* **2007**, *111*, 5609–5613.
- (66) Aradi, B.; Hourahine, B.; Frauenheim, Th. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.
- (67) Bhattacharya, S. K.; Deodhar, P. A.; Viswanatha, R.; Kshirsagar, A. *J. Phys.: Condens. Matter* **2010**, *22*, 295304.
- (68) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.
- (69) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.
- (70) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (71) Monkhorst, H. J.; Pack, J. D. *Phys. Rev. B* **1976**, *13*, 5188–5192.
- (72) Soler, J. M.; Artacho, E.; Gale, J. D.; Garcia, A.; Junquera, J.; Ordejon, P.; Sanchez-Portal, D. *J. Phys.: Condens. Matter* **2002**, *14*, 2745–2779.
- (73) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993–2006.
- (74) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1993**, *37*, 785–789.
- (75) Becke, A. D. *J. Chem. Phys.* **1998**, *88*, 1053–1062.
- (76) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; G. E. Scuseria, Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision E.01; Gaussian, Inc.: Wallingford, CT, 2004.
- (77) Stark, R. W.; Falicov, L. M. *Phys. Rev. Lett.* **1967**, *19*, 795.
- (78) Daniuk, S.; Jarlborg, T.; Kontrym-Sznajd, G.; Majsnerowski, J.; Stachowiak, H. *J. Phys.: Condens. Matter* **1989**, *1*, 8397–8406.
- (79) Kittel, C. *Crystal Binding. In Introduction to Solid State Physics*, 7th ed.; Wiley: New York, 1995; pp 57.
- (80) McCann, D. R.; Cartz, L. *J. Appl. Phys.* **1972**, *43*, 4473.
- (81) Lingelbach, W.; Stuke, J.; Weiser, G.; Treusch, J. *Phys. Rev. B* **1972**, *5*, 243.
- (82) Ghosh, P.; Upadhyay, K.; Waghmare, U. V. *Phys. Rev. B* **2007**, *75*, 245437.
- (83) Biering, S.; Hermann, A.; Furthmüller, J.; Schwerdtfeger, P. *J. Phys. A* **2009**, *113*, 12427–12432.
- (84) Dou, Y.; Egdell, R. G.; Law, D. S. L.; Harrison, N. W.; Searle, B. G. *J. Phys.: Condens. Matter* **1998**, *10*, 8447–8458.
- (85) Reshak, A. H. *J. Chem. Phys.* **2006**, *124*, 104707–6.
- (86) Huang, M.-Z.; Ching, W. Y. *Phys. Rev. B* **1993**, *47*, 9449.
- (87) Huang, M.-Z.; Ching, W. Y. *Phys. Rev. B* **1993**, *47*, 9464.
- (88) Wei, S.-H.; Zhang, S. B. *Phys. Rev. B* **2000**, *62*, 6944.
- (89) Hannachi, L.; Bouarissa, N. *Physica B* **2009**, *404*, 3650–3654.
- (90) Christensen, N. E.; Christensen, O. B. *Phys. Rev. B* **1986**, *33*, 47466–47474.
- (91) Zakharov, O.; Rubio, A.; Blase, X.; Cohen, M. L.; Louie, S. G. *Phys. Rev. B* **1994**, *50*, 10780–10787.
- (92) Hosseini, S. M. *Physica B* **2008**, *403*, 1907–1915.
- (93) Duke, C. B.; Wang, Y. R. *J. Vac. Sci. Technol. B* **1988**, *6*, 1440–1443.
- (94) Wang, Y. R.; Duke, C. B. *Surf. Sci.* **1987**, *192*, 309.
- (95) Wang, Y. R.; Duke, C. B. *Phys. Rev. B* **1987**, *36*, 2763.
- (96) Wang, Y. R.; Duke, C. B. *Phys. Rev. B* **1988**, *37*, 6417.
- (97) Siemens, B.; Domke, C.; Ebert, Ph.; Urban, K. *Phys. Rev. B* **1997**, *56*, 12321–12326.
- (98) Li, S.; Yang, G. W. *J. Phys. Chem. C* **2010**, *114*, 15054–15060.
- (99) Huang, S.-P.; Cheng, W.-D.; Wu, D.-S.; Hu, J.-M.; Shen, J.; Xie, Z.; Zhang, H.; Gong, Y.-J. *Appl. Phys. Lett.* **2007**, *90*, 031904.
- (100) Xiong, Y.; Xie, Y.; Yang, J.; Zhang, Z.; Wu, C.; Du, G. *J. Mater. Chem.* **2002**, *12*, 3712.
- (101) Kwon, S. J.; Choi, Y. J.; Park, J. H.; Hwang, I. S.; Park, J. G. *Phys. Rev. B* **2005**, *72*, 205312.
- (102) Köhler, C.; Seifert, G.; Gerstmann, U.; Elstner, M.; Overhof, H.; Frauenheim, T. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5109.
- (103) Simões, J. A. M.; Beauchamp, J. L. *Chem. Rev.* **1990**, *90*, 629–688.
- (104) Dutton, W. A.; Cooper, W. C. *Chem. Rev.* **1966**, *66*, 657–675.

(105) Madelung, O. M. Tetrahedrally bonded elements and compounds. In *Semiconductors: Data Hand Book*, 3rd ed.; Springer: Berlin, 2004; pp 218.

(106) Wired Chemist. <http://www.wiredchemist.com/chemistry/data> (accessed Dec 8, 2010).

(107) Greene, T. M.; Brown, W.; Andrews, L.; Downs, A. J.; Chertihin, G. V.; Runeberg, N.; Pyykko, P. *J. Phys. Chem.* **1995**, *99*, 7925–7934.

(108) Greenwood, N. N.; Earnshaw, A. Selenium, Tellurium and Polonium. In *Chemistry of the Elements*, 2nd ed.; Pergamon Press: Oxford, U.K., 1997; p 766.

(109) Balasubramanian, K.; Liao, M. Z. *J. Phys. Chem.* **1988**, *92*, 4595–4599.

(110) Gabuda, S. P.; Kozlova, S. G. *J. Phys. Chem. B* **2006**, *110*, 18091–18092.

(111) Gibson, S. T.; Greene, J. P.; Berkowitz, J. *J. Chem. Phys.* **1986**, *85*, 4815–4824.

(112) Carson, A. S.; Hartley, K.; Skinner H. A. The Bond Dissociation Energies of Cd-CH₃ in Cd(CH₃)₂, and of CH₃-I in CH₃ I. *Proc. R. Soc. London, Ser. A* **1949**, *195*, 500–512. <http://www.jstor.org/stable/98237> (accessed 02/07/2010).

(113) Scott, J. D.; Causley, G. C.; Russell, B. R. *J. Chem. Phys.* **1973**, *59*, 6577.

(114) Goldish, E.; Hergberg, K.; Marsh, R. E.; Schomaker, V. *J. Am. Chem. Soc.* **1955**, *779*, 2948–2949.

Combined Metadynamics and Umbrella Sampling Method for the Calculation of Ion Permeation Free Energy Profiles

Yong Zhang[†] and Gregory A. Voth^{*}

Department of Chemistry, James Franck Institute, Institute for Biophysical Dynamics, and Computation Institute, University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, United States

ABSTRACT: Free energy calculations are one of the most useful methods for the study of ion transport mechanisms through confined spaces such as protein ion channels. Their reliability depends on a correctly defined reaction coordinate (RC). A straight line is usually not a proper RC for such complicated processes, so in this work a combined metadynamics/umbrella sampling (MTD/US) method is proposed. In the combined method, the ion transport pathway is first identified by the MTD method, and then the free energy profile or potential of mean force (PMF) along the path is calculated using umbrella sampling. This combined method avoids the discontinuity problem often associated with normal umbrella sampling calculations that assume a straight line RC, and it provides a more physically accurate potential of mean force for such processes. The method is demonstrated for the proton transport process through the protein channel of aquaporin-1.

I. INTRODUCTION

Ion channels are a large class of proteins that regulate ion flow across the cell membrane. Nearly all cells depend on the proper functioning of these ion channels to transfer ions in and out of the cell. This class of proteins has been widely studied using various experimental techniques (see, e.g., refs 1–5) as well as theoretical calculations (see, e.g., refs 6–9).

Free energy calculations are one of the most important techniques in theoretical and computational chemistry and have been proven to be a powerful tool to study biological systems.¹⁰ The calculated free energy profile along the channel described by a certain reaction coordinate (RC), i.e., the potential of mean force (PMF), contains essential information about the ion transport process and provides detailed insight into the ion transport mechanism.^{11–17}

In general, to perform such calculations, the channel is aligned along one axis of the simulation system (usually the *Z*-axis) and the coordinate component along this axis is used as the RC. This is a convenient setup, but, as shown below, it may provide a calculated PMF that differs significantly from the one along a more realistic and more complicated RC and consequently gives misleading information about the ion transport process.

As an example, we calculated the PMF of a classical hydronium with no Grothuss proton shuttling transported through a kinked carbon nanotube model, as shown in Figure 1. The classical hydronium has a partial charge $-0.5 e$ on the oxygen atom and $+0.5 e$ on each of the three hydrogen atoms. The tube has a pore of 8 Å in diameter. All of the carbon atoms are fixed in space, as was done in previous studies, and should not affect the results.¹⁸ The carbon atoms interact with water oxygens through a Lennard-Jones potential with $\epsilon = 0.12895$ kcal/mol and $\sigma = 3.2752$ Å. This setup allows a single water wire to form in the aperture.¹⁹

The PMFs were obtained along the reaction coordinate (*Z*-coordinate of the hydronium oxygen) using umbrella sampling (US) for different orientations of the same carbon tube, as shown in Figure 1. For this simple model, it is reasonable to

believe that the hydronium transports along the kinked tube central axis. The PMF associated with the real hydronium transport pathway was thus also calculated and shown in Figure 1. As expected, the PMF curves for the two tube orientations using the *Z*-coordinate as reaction coordinate are different from the correct one, demonstrating that using the *Z*-coordinate as the reaction coordinate fails to properly describe the kink and the change in channel radius (due to the kink) even for such a simple model.

Ion channel proteins consist of amino acids and are heterogeneous in nature. The local environment in the channels is much more complicated than the carbon tube model described above. The PMF calculated assuming a straight line RC may describe the ion transport processes improperly and thus yield a misleading description of the underlying biological processes. Here, we propose a combined metadynamics (MTD) and US method, in which the ion transport pathway is first identified using the MTD method and then the US calculation is carried out along the pathway to obtain the PMF along a more physically reasonable ion transport RC.

The metadynamics method and the combined MTD/US method are briefly described in the next section, and the combined method is demonstrated in more detail in section III on a protein channel. Discussion and concluding remarks are then given in the last section.

II. METHODS AND SIMULATION DETAILS

Metadynamics^{20–23} is a free energy method that is based on a biasing of the potential surface and is similar to umbrella sampling in this sense. But, in contrast to umbrella sampling, the bias or “hills” are dynamically placed on top of the underlying potential energy landscape and discourages the system from visiting the same points in the configurational space. The hill can

Received: February 10, 2011

Published: June 02, 2011

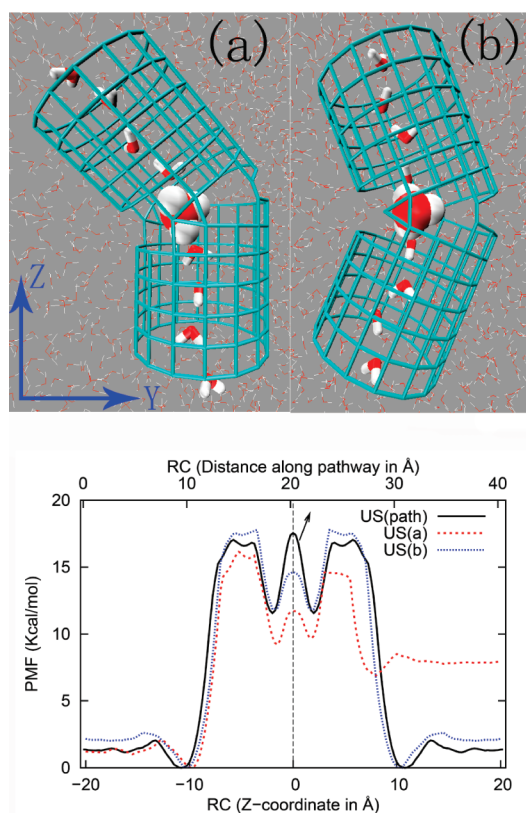


Figure 1. (Upper panels) Carbon tube model with a single water wire formed in the pore. The carbon atoms are fixed in space during the simulation and interact with water oxygen atoms through a van der Waals potential. (Lower panel) Calculated PMFs using the Z-coordinate as the reaction coordinate for different orientations (US(a), dotted line, and US(b), short dashed line, respectively) of the tube in space. The PMF calculated along the real ion transport pathway is also shown for comparison (US(path), solid line, up axis) and aligned at $Z = 0$ Å with the other two indicated by a dashed line. The statistical error was estimated to be ± 0.2 kcal/mol or smaller. The difference in the calculated PMF profiles clearly demonstrates in this simple case that using a straight line along the Z-coordinate fails to properly describe the kink and radius change of the channel.

have any form, but a Gaussian potential is usually used:

$$V_{\text{bias}}(s, t) = \sum_{t_i} H \exp\left(-\frac{|s - s(t_i)|^2}{2w^2}\right) \quad (1)$$

where H is the height of the hill, w is the width, t is time, and s is the collective variable (CV) or RC. The hills are added at a time interval of δt . With the accumulation of hills, the local potential energy well is flattened and the system escapes the initial minimum over the lowest barrier. When applied properly, the MTD method not only accelerates the sampling of configuration space but also maps out the free energy surface or the potential of mean force as the negative of the sum of all the bias hills, $-V_{\text{bias}}(s, t)$.²⁰ The accuracy of the calculated PMF from MTD simulation depends on the parameters H , w , and δt , which in turn depend on the system properties. The selection of these parameters is not trivial²⁴ although successful applications have been reported.^{25–30} The other issue that limits the application of the MTD method is a convergence problem caused by the lack of a straightforward way to determine at which point a MTD simulation should stop, and

this causes error in the final PMF.^{25,31–35} The well-tempered MTD method has solved this problem theoretically.³⁶ However, its successful application requires knowledge about the PMF of interest which is generally not known beforehand. Improperly chosen parameters may thus make the MTD PMF calculation inefficient or result in large errors.

Since MTD is able to find the next lowest barrier, it is also used as a path sampling method.^{37,38} In the present work, instead of generating the PMF using MTD directly, we use it to find the ion transport pathway through an ion channel protein. Once the pathway is identified, the PMF along it can then be rigorously calculated using umbrella sampling. This combined MTD/US approach provides a powerful means to determine a complicated ion transport RC and then calculate a more physically meaningful PMF for the ion transport along that RC. The convergence problem sometimes associated with the MTD method for calculating the PMF is avoided.

For an arbitrarily curved pathway, instead of using the weighted histogram analysis method (WHAM),³⁹ the PMF or the free energy difference from the starting point can be calculated as^{40,41}

$$F(s(\alpha)) - F(s(0)) = \int_0^\alpha d\alpha' \sum_{i=1}^N \frac{ds_i(\alpha')}{d\alpha'} \frac{\partial F(s(\alpha'))}{\partial s_i} \quad (2)$$

where $\alpha \in [0, 1]$ is the parameter used to parametrize the path and the sum in the integral is over N RCs or CVs ($N = 1$ in the present work). The mean force can be calculated for each window by

$$\frac{\partial F(s)}{\partial s_j} \approx \frac{k}{T} \int_0^T dt (s_j - s_j(x(t))) \quad (3)$$

where k is the force constant in umbrella sampling, T is the simulation time, s_j is the equilibration value of the reaction coordinate, and $s_j(x(t))$ is its instantaneous value.

All simulations in this work were carried out at 308 K in the canonical ensemble (constant NVT) using DL_POLY package⁴² or its modified version (see below) with the Amber force field.⁴³ The Nose–Hoover thermostat was used with relaxation time 0.5 ps. A MD integration time step of 1.0 fs was used, and the electrostatics were calculated by the smooth particle mesh Ewald method with a real space cutoff of 9.5 Å. The PLUMED package⁴⁴ was interfaced with DL_POLY for the MTD calculations.

III. RESULTS

The combined MTD/US method is demonstrated here for the protein channel aquaporin-1 (AQP1), which is a relatively simple channel protein but is more complicated than the carbon nanotube model discussed earlier. The exact ion transport pathway through AQP1 is not known. AQP1 is a channel protein that conducts water but effectively blocks protons and other positively charged ions through a high free energy barrier to ion permeation. Its selectivity mechanism has been extensively studied using various methods including free energy calculations.^{1,11,16,45–51}

The PMF associated with proton transport through the AQP1 channel was calculated using the MTD/US method described above. The proton was treated either as a classical hydronium model or by using the multistate empirical valence bond (MS-EVB) model^{52,53} that includes the full physics of Grotthuss proton shuttling and charge defect delocalization. The initial configuration containing the protein monomer and solvation

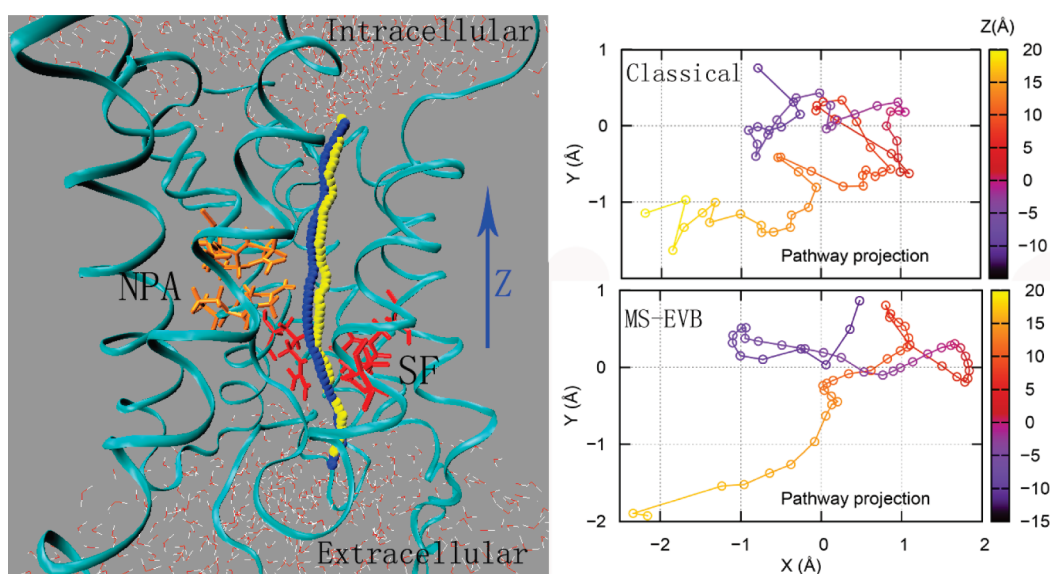


Figure 2. (Left) Structure of AQP1. The proton transport pathways identified by the metadynamics method are shown as yellow points for the classical hydronium model case and blue points for the excess proton described by the MS-EVB model. The NPA motif and the SF region, which form the two barriers for the proton transport, are shown in orange and red, respectively. (Right) Projection of the identified pathways onto the X – Y plane with the Z -coordinate indicated by color. The pathways crossed > 3 Å in the X -direction and > 2 Å in the Y -direction. The pathways are twisted and clearly deviate from a straight line along the Z -coordinate, which would simply be a dot in the X – Y plane.

water on both ends of the channel was taken from a previous study.⁵¹ The protein channel was aligned along the Z -axis. The lipid bilayer was then removed from the simulation box to save computation time.⁵¹ To keep the protein structure near its initial configuration, all protein α carbons, except those which are within 6 Å of any of the embedded water molecules, were tethered with a force constant of 2.4 (kcal/mol)/Å² during the simulation. This way, the protein residues that have close contact with the water wire are allowed to reorganize during the proton transport process, while the tethered atoms fluctuate in a way similar to the full unconstrained simulation.^{17,51}

For the classical hydronium case, a MTD trajectory was generated using the Z -coordinate of the hydronium oxygen atom as the reaction coordinate. Gaussian-shaped hills with $H = 0.1$ kcal/mol and $w = 0.5$ Å were added to the underlying potential surface every 500 steps. It is worth noting that in the combined MTD/US method the final PMF is not sensitive to the choice of these parameters, as long as they are in reasonable range,²⁴ since the PMF is not directly calculated from MTD. Soft walls were added on both ends at $Z = 17$ Å and $Z = -20$ Å, respectively, to limit sampling inside the channel. A 5 ns trajectory was generated, during which a “one-way” trip through the channel was made for the hydronium. Fifty snapshots were obtained from this trajectory at a 100 ps interval. These snapshots sampled diverse positions of the hydronium along the hydronium transport pathway. New MTD simulations were then initiated from each of these configurations to obtain better sampling of the pathway. For 32 of these MTD trajectories, at least one transport through the channel, from either direction, was observed within 3 ns. The transport part from each trajectory, the first one-way trip, was averaged to get the most possible hydronium transport pathway through the channel whereas other parts, which mainly sampled either end of the channel, were discarded. The path in each trajectory was found to be similar, indicating that no competitive pathways exist in this protein channel. For more complicated protein channels that do have more than one ion transport

pathway having a similar free energy barrier, all pathways can be identified accordingly and the PMFs calculated separately.

A similar procedure was followed for the simulations with the excess proton treated using the MS-EVB model. In this case, the reaction coordinate is defined as the Z -coordinate of the center of the excess proton net positive charge defect (CEC), which is defined as⁵²

$$\mathbf{r}_{\text{CEC}} = \sum_i^{N_{\text{EVB}}} c_i^2 \mathbf{r}_{\text{COC}}^i \quad (4)$$

where c_i^2 is the EVB amplitude for state i and $\mathbf{r}_{\text{COC}}^i$ is the center of charge vector

$$\mathbf{r}_{\text{COC}}^i = \frac{1}{Q_{\text{tot}}^i} \sum_k^{\{i\}} q_k \mathbf{r}_k \quad (5)$$

where

$$Q_{\text{tot}}^i = \sum_k^{\{i\}} |q_k| \quad (6)$$

with q_k being the charge on atom k and \mathbf{r}_k its position.

The averaged proton transport pathways through the channel are shown as yellow points for the classical hydronium model and blue points for the excess proton CEC in the MS-EVB model in the left panel of Figure 2. Both pathways are basically along the Z -coordinate but obviously deviate from a straight line. The right two panels of Figure 2 show that the pathways have excursions of more than 3 and 2 Å in the X - and Y -directions, respectively. (The projected pathways onto the X – Y plane in these panels of Figure 2 would simply be dots in that plane if the pathways were straight lines along Z .) The MS-EVB model proton transport pathway is smoother than the classical hydronium model pathway, which is consistent with the fact that the MS-EVB model includes the Grotthuss mechanism of proton hopping explicitly so that multiple protons from different water molecules can

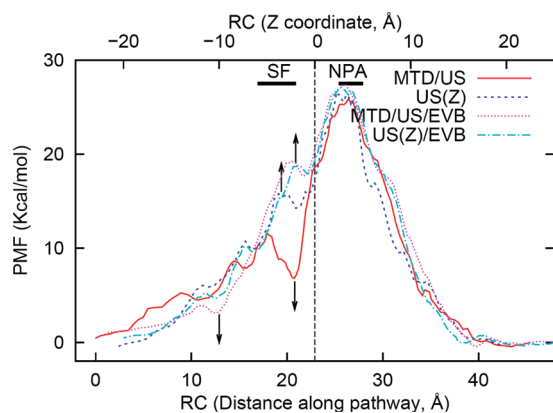


Figure 3. Free energy profile (or potential of mean force, PMF) calculated for the proton transport through the APQ1 channel using the Z-coordinate (up axis) or the displacement along the identified pathways (low axis) as the reaction coordinate, respectively. The proton was treated using either the classical hydronium model or the reactive MS-EVB model for the Grotthuss shuttling excess proton charge defect. The statistical error was estimated to be ± 0.5 kcal/mol or smaller. The PMFs are aligned at $Z = 0$ Å as indicated by a dashed line. The position of NPA motif and the SF region was based on the Z-coordinate (upper axis).

participate in translocating the net positive excess proton charge defect.

It has been found previously that there are two barriers for proton transport through the AQP1 channel, corresponding to the selectivity filter region (SF, formed mainly by the Phe56, His180, and Arg195 residues) and the Asn-Pro-Ala (NPA) motifs.^{16,45–51} The NPA motifs are shown in orange with the licorice helices model in Figure 2, and the SF region is in red. It is clear that the pathways are twisted in the X – Y plane around the NPA motifs (red regions of the points in right panels of Figure 2), which was not directly noted before. While the proton transport pathways identified with different proton transport models seem similar for much of the channel, important differences are observed around the SF and NPA motif regions. Relative to the MS-EVB model case, the classical hydronium model pathway is repulsed further from the NPA motif residues, while the MS-EVB model appears to have the excess proton charge defect distorted further away from the center of the SF region.

The PMFs calculated using eq 2 along the MTD pathways and the ones calculated with the straight line Z-coordinate as the RC are shown in Figure 3. About 150 umbrella sampling windows are used for each case (0.25 Å interval for the channel region and 0.50 Å for the bulk region), and a constraining force constant of 30 (kcal/mol)/Å² was used. A 1 ns trajectory was generated for each window. The first 200 ps was considered equilibration and discarded when calculating the mean force using eq3. The dominant barrier due to the NPA motifs was found to be ~ 27 kcal/mol for each case, similar to previous simulation results⁵¹ of ~ 28 kcal/mol using straight line RC with umbrella sampling and the weighted histogram analysis method. However, the smaller barrier at the selectivity filter region, as indicated in Figure 3, shows different features between the results. For the classical hydronium model simulations, when using the straight Z-coordinate as the RC, the barrier at SF is ~ 16 kcal/mol, consistent with previous results⁵¹ where the same RC was used, proving the accuracy of the mean force method used in this work. On the other hand, the barrier is only 11 kcal/mol when the PMF was

calculated along the classical hydronium transport pathway identified by MTD. In addition, the minimum between the two barriers is significantly less deep when the Z-coordinate is used as the reaction coordinate. On the other hand, when the proton is treated using the MS-EVB model, the barrier at the SF region is ~ 18 – 19 kcal/mol, which agrees well with previous MS-EVB results.⁵¹ The two PMFs obtained with different RCs are different, but not as different as the classical hydronium model simulations.

In the calculations using the Z-coordinate as the RC, the system is free in the X – Y plane. Previous studies found that the radius of the channel varies from more than 5 Å at the ends to less than 1 Å in the SF region.^{45,51} In the simulation, the RC (Z-coordinate) windows were distributed evenly with a 0.25 Å interval; however, the actual distance between the sampled regions of two nearby windows can be much larger. Figure 4 shows the distance between the neighboring windows on the basis of the averaged position sampled in each window for the classical hydronium model case, in the Z-direction and absolute distance in 3D, respectively. In spite of a distance of less than 0.4 Å in the Z-direction, the absolute 3D distance between neighboring windows can be as large as 2.2 Å due to the distribution in the X – Y plane. For pairs with the largest distance, the sampled region in the X – Z and Y – Z planes is shown in Figure 4. It is clear in the figure that the sampled regions are far away from each other in either the X - or Y -direction, despite good overlap in the Z-direction, causing a discontinuity in configurational space. It is obvious by comparing Figures 3 and 4 that the difference in the PMF is directly related to this discontinuity. The way the PMF is calculated in this work using eq 2 does not explicitly depend on the overlap between windows as it does in WHAM. However, such discontinuities introduce errors into the calculated PMF. When the proton hopping is described explicitly using the MS-EVB model, fewer bad overlap pairs were found (results not shown), but it still suffers from this problem when using the Z-coordinate as the RC.

The importance of a constraint in the direction perpendicular to the RC axis has been suggested before.¹³ However, in the current PMF calculation using the Z-coordinate as the RC, no constraint was added for the proton in the X – Y plane unless it is close to the bulk solvent (beyond $Z = \pm 10.0$ Å) because the equilibrium positions in the X – Y plane for each umbrella sampling window in the channel region were not known explicitly. In addition, for a curved pathway, a constraint in the X – Y plane perpendicular to the improperly chosen reaction coordinate along the Z-axis would cause errors as well, as demonstrated in the carbon nanotube model described earlier.

In the PMF calculation along the identified pathway described in the current work, a weak harmonic force of 5 (kcal/mol)/Å² was applied in the direction perpendicular to the normal of a given point on the pathway. As shown in Figure 4, the distance between neighboring windows are all reasonable (< 1 Å). The fact that a weak force of 5 (kcal/mol)/Å² can constrain the system around the identified pathway highlights the reliability of the proposed combined MTD/US method.

IV. DISCUSSION AND CONCLUSIONS

A calculated ion permeation PMF is meaningful only if the reaction coordinate is correctly identified for the problem of interest. In the present work we have demonstrated using a simple artificial carbon tube model that a small kink or change in

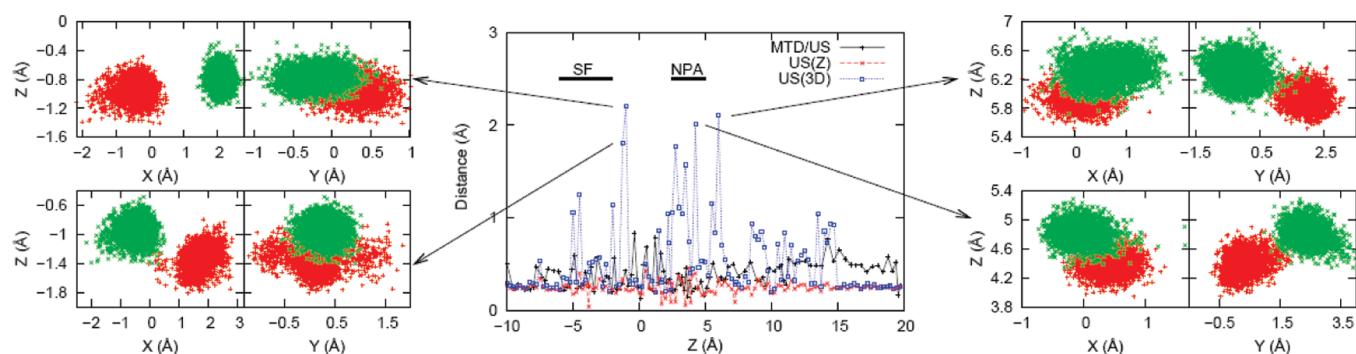


Figure 4. Distance between neighboring windows in the Z-direction (US(Z), dashed line) or in three dimensions (US(3D), dotted line) for the classical hydronium model simulation. The position for each window is calculated by taking the average of the sampled points, and only the pore part is shown. In the calculations using the Z-coordinate as the reaction coordinate, the RC is evenly distributed in the Z-direction with a 0.25 Å interval. However, the absolute distance between neighboring windows can be as large as 2.2 Å, causing discontinuities in the configuration space as shown for the four largest distance pairs. Such discontinuities introduce errors into the calculated PMF and may cause a misleading description of the ion permeation process. The simulation with the excess proton treated using the MS-EVB model suffers from a similar issue (not shown). The combined MTD/US (solid line) method avoids such problems.

channel radius can make an otherwise trivial calculation very complicated. The calculated PMF fully depends on the orientation of the tube, or in other words, it depends on the specific definition of the reaction coordinate.

Ion channel proteins help to establish the environment inside and outside of the cell by controlling the ion flow across the cell membrane. These heterogeneous pores are formed to perform various functions. Many of the ion channels are selective for some species over others by a specific gating mechanism. This feature makes the channel environment more complicated. For example, in AQP1, which conducts water but blocks all positively charged cation species, there are two gating motifs, the NPA motif and the SF region. Furthermore, the radius of the pore and the local environment varies greatly. As a result, the assumption of a straight ion transport pathway along a certain axis is highly questionable and introduces errors into the calculated PMF.

In the present study of AQP1, using the Z-coordinate as the RC overestimates the SF barrier by about 5 kcal/mol when the proton is treated using the classical hydronium model (cf. Figure 3). The NPA domain is still the dominant barrier region, and the barrier heights are similar when either RC definition was used. This is consistent with the fact that the channel radius at the SF region has a large variation from other parts of the channel,^{45,51} highlighting the importance of a better definition of the RC in the PMF calculation. On the other hand, when the excess proton charge defect was treated using the MS-EVB model that allows proton (Grotthuss) shuttling, the PMFs obtained with both RC definitions are similar even in the SF region although small differences can be seen. This is due to the fact that the excess charge defect is delocalized over multiple waters in the MS-EVB model via the known proton hopping physics and results in a much smoother and continuous pathway as shown in Figure 2.

The results shown in Figure 3 for the AQP1 proton PMFs relative to those of the classical hydronium cation are, in fact, likely to be highly significant. This is because they support the notion that the AQP1 SF region is a specially evolved region of the channel designed to help block proton (shuttling H^+) flux (and it does so no matter what the chosen RC). On the other hand, the SF region it is not as important for blocking the flux by a “classical” hydronium cation that cannot shuttle protons (i.e., even though

the “best” RC gives a small barrier for classical hydronium at the SF, a large barrier at the NPA region is still present). These ideas on the importance of the AQP1 SF region for modulating proton flux have recently been proposed from a combined computational/experimental study⁵⁴ of two AQP1 SF mutants, and the present results help to confirm this proposal.

It is further reasonable to anticipate that ion channel permeation studies for other ions (Na^+ , Cl^- , etc) and proteins would also suffer the inaccuracies described earlier when an improper (e.g., straight line Z-coordinate) RC is used to calculate a PMF. The proposed combined MTD/US method however provides a computationally affordable and technically feasible solution to circumvent this problem.

In this work, we have used the MTD method to identify the ion transport pathway, but other pathway searching methods^{55–63} could be used instead. For example, one such approach is the string method,^{40,64} one version of which has been applied to biological systems by Roux and co-workers to study the conformational change of a small protein.⁶⁵ This method can also be applied to ion channel systems to identify the ion transport pathway that connects the two ends of the channel as well. However, the string method may not guarantee the identified pathway is the free energy minimum, which is the path most likely associated with the real processes in the cell. In MTD simulations, starting from one configuration, the system is more likely to cross the next lowest barrier and reach the other end of the channel along the free energy minimum pathway.

In the example described above, we have used the Z-coordinate as the RC in the MTD simulation, which is similar to an experimental setup where the ion concentration or an electrochemical gradient is applied in one direction to drive the ion flow. However, this is fundamentally different from the Z-coordinate RC used in normal umbrella sampling calculations. In the latter case, the simulation in each window is independent from the next window and can cause the discontinuity as observed above due to not enough sampling or other reasons. The MTD simulation, however, is a dynamical process and the successful transport through the channel guarantees continuous sampling. In addition, using the data from the first “one-way trip” assures the sampled path is the minimum corresponding to the most possible ion transport pathway observed in experiments.

The continuity of the sampled region can be further assured by applying a weak harmonic force in the perpendicular direction at a given point along the pathway when calculating the PMF.

In summary, the assumption that the ions transport through the nonuniform protein channel along a straight line can fail due to the complexity of the channel structure so that the PMF calculated along this path can be misleading. To get more reliable results, a combined MTD/US method has been proposed, in which the ion transport pathway is first identified using the MTD method. The PMF for the ion transport processes is then calculated along this pathway using the umbrella sampling method. This combined method also avoids the convergence problem sometimes found with the direct MTD approach for calculating a PMF. We have also applied the proposed method to the proton transport process through the AQP1 channel and demonstrated its feasibility, reliability, and ability to provide important new insight.

AUTHOR INFORMATION

Corresponding Author

*E-mail: gavoith@uchicago.edu.

Present Addresses

[†]Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556.

ACKNOWLEDGMENT

This research was supported by the National Institutes of Health (Grant R01-GM53148). Computational resources were provided by the National Science Foundation through Teragrid computing resources administrated by the Texas Advanced Computing Center and the Pittsburgh Supercomputing Center, and by a grant of computer time from the Department of Defense High Performance Computing Modernization Program at the Maui High Performance Computing Center. We thank Dr. Takefumi Yamashita for helpful discussions.

REFERENCES

- (1) Tajkhorshid, E.; Nollert, P.; Jensen, M. O.; Miercke, L. J. W.; O'Connell, J.; Stroud, R. M.; Schulten, K. *Science (Washington, DC, U. S.)* **2002**, *296*, 525.
- (2) Dutzler, R.; Campbell, E. B.; MacKinnon, R. *Science (Washington, DC, U. S.)* **2003**, *300*, 108.
- (3) Beitz, E.; Wu, B.; Holm, L. M.; Schultz, J. E.; Zeuthen, T. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 269.
- (4) Wang, N.; Miller, C. J. *Mol. Biol.* **2006**, *362*, 682.
- (5) Ho, J. D.; Yeh, R.; Sandstrom, A.; Chorny, I.; Harries, W. E. C.; Robbins, R. A.; Miercke, L. J. W.; Stroud, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 7437.
- (6) Faraldo-Gomez, J. D.; Roux, B. *J. Mol. Biol.* **2004**, *339*, 981.
- (7) Coleman, R. G.; Sharp, K. A. *Biophys. J.* **2009**, *96*, 632.
- (8) Lin, Y.; Cao, Z.; Mo, Y. *J. Phys. Chem. B* **2009**, *113*, 4922.
- (9) Phongphanphane, S.; Yoshida, N.; Hirata, F. *J. Phys. Chem. B* **2010**, *114*, 7967.
- (10) *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C.; Pohorille, A., Eds.; Springer: Berlin, Heidelberg, New York, 2007; Vol. 86.
- (11) Burykin, A.; Warshel, A. *Biophys. J.* **2003**, *85*, 3696.
- (12) Berneche, S.; Roux, B. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 8644.
- (13) Allen, T. W.; Andersen, O. S.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 117.
- (14) Roux, B.; Allen, T.; Berneche, S.; Im, W. *Q. Rev. Biophys.* **2004**, *37*, 15.
- (15) Allen, T. W.; Andersen, O. S.; Roux, B. *Biophys. J.* **2006**, *90*, 3447.
- (16) Chen, H.; Wu, Y.; Voth, G. A. *Biophys. J.* **2007**, *93*, 3470.
- (17) Wang, D.; Voth, G. A. *Biophys. J.* **2009**, *97*, 121.
- (18) Dellago, C.; Hummer, G. *Phys. Rev. Lett.* **2006**, *97*, 245901.
- (19) Hummer, G.; Rasaiah, J. C.; Noworyta, J. P. *Nature* **2001**, *414*, 188.
- (20) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.
- (21) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302/1.
- (22) Ensing, B.; Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. *Acc. Chem. Res.* **2006**, *39*, 73.
- (23) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 090601.
- (24) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714.
- (25) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600.
- (26) Gervasio, F. L.; Parrinello, M.; Ceccarelli, M.; Klein, M. L. *J. Mol. Biol.* **2006**, *361*, 390.
- (27) Domene, C.; Klein, M. L.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Am. Chem. Soc.* **2008**, *130*, 9474.
- (28) Mantz, Y. A.; Branduardi, D.; Bussi, G.; Parrinello, M. *J. Phys. Chem. B* **2009**, *113*, 12521.
- (29) Pfaendtner, J.; Dranduardi, D.; Parrinello, M.; Pollard, T. D.; Voth, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12723.
- (30) Barducci, A.; Bonomi, M.; Parrinello, M. *WIREs Comput. Mol. Sci.* **2011**, DOI: 10.1002/wcms.31.
- (31) Micheletti, C.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2004**, *92*, 170601.
- (32) Ensing, B.; Klein, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6755.
- (33) Wu, Y.; Schmitt, J. D.; Car, R. J. *Chem. Phys.* **2004**, *121*, 1193.
- (34) Babin, V.; Roland, C.; Darden, T. A.; Sagui, C. *J. Chem. Phys.* **2006**, *125*, 204909.
- (35) Min, D.; Liu, Y.; Carbone, I.; Yang, W. J. *Chem. Phys.* **2007**, *126*, 194104.
- (36) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
- (37) Nishihara, Y.; Hayashi, S.; Kato, S. *Chem. Phys. Lett.* **2008**, *464*, 220.
- (38) Ceccarelli, M.; Anedda, R.; Casu, M.; Ruggerone, P. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 1231.
- (39) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
- (40) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 024106/1.
- (41) Gan, W.; Yang, S.; Roux, B. *Biophys. J.* **2009**, *97*, L08.
- (42) Smith, W.; Forester, T. R. The DL_POLY Molecular Simulation Package; http://www.dl.ac.uk/TCSC/Software/DL_POLY/main.html, Vol. 2009, 1999.
- (43) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (44) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961.
- (45) de Groot, B. L.; Grubmuller, H. *Science (Washington, DC, U. S.)* **2001**, *294*, 2353.
- (46) de Groot, B. L.; Frigato, T.; Helms, V.; Grubmuller, H. *J. Mol. Biol.* **2003**, *333*, 279.
- (47) Ilan, B.; Tajkhorshid, E.; Schulten, K.; Voth, G. A. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 223.
- (48) Chakrabarti, N.; Roux, B.; Pomes, R. *J. Mol. Biol.* **2004**, *343*, 493.

- (49) Burykin, A.; Warshel, A. *FEBS Lett.* **2004**, *570*, 41.
- (50) de Groot Bert, L.; Grubmuller, H. *Curr. Opin. Struct. Biol.* **2005**, *15*, 176.
- (51) Chen, H.; Ilan, B.; Wu, Y.; Zhu, F.; Schulten, K.; Voth, G. A. *Biophys. J.* **2007**, *92*, 46.
- (52) Day, T. J. F.; Soudackov, A. V.; Cuma, M.; Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **2002**, *117*, 5839.
- (53) Swanson, J. M. J.; Maupin, C. M.; Chen, H.; Petersen, M. K.; Xu, J.; Wu, Y.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4300.
- (54) Li, H.; Chen, H.; Steinborn, C.; Wu, B.; Beitz, E.; Zeuthen, T.; Voth, G. A. *J. Mol. Biol.* **2011**, *407*, 607.
- (55) Jonsson, H.; Mills, G.; Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B. J., Ciccotti, G., Coker, D. F., Eds.; World Scientific: Singapore, 1998; pp 385.
- (56) Fischer, S.; Karplus, M. *Chem. Phys. Lett.* **1992**, *194*, 252.
- (57) Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **1999**, *111*, 7010.
- (58) Isralewitz, B.; Gao, M.; Schulten, K. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224.
- (59) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291.
- (60) Elber, R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 151.
- (61) van der Vaart, A.; Karplus, M. *J. Chem. Phys.* **2007**, *126*, 164106/1.
- (62) Yang, H.; Wu, H.; Li, D.; Han, L.; Huo, S. *J. Chem. Theory Comput.* **2007**, *3*, 17.
- (63) Zheng, L.; Chen, M.; Yang, W. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20227.
- (64) E, W.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *66*, 052301/1.
- (65) Pan, A. C.; Sezer, D.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 3432.

Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values

Chresten R. Søndergaard,* Mats H. M. Olsson, Michał Rostkowski, and Jan H. Jensen*

Department of Chemistry and Center for Computational Molecular Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark

S Supporting Information

ABSTRACT: The new empirical rules for protein pK_a predictions implemented in the PROPKA3.0 software package (Olsson et al. *J. Chem. Theory Comput.* **2010**, *7*, 525–537) have been extended to the prediction of pK_a shifts of active site residues and ionizable ligand groups in protein–ligand complexes. We present new algorithms that allow pK_a shifts due to inductive (i.e., covalently coupled) intraligand interactions, as well as noncovalently coupled interligand interactions in multiligand complexes, to be included in the prediction. The number of different ligand chemical groups that are automatically recognized has been increased to 18, and the general implementation has been changed so that new functional groups can be added easily by the user, aided by a new and more general protonation scheme. Except for a few cases, the new algorithms in PROPKA3.1 are found to yield results similar to or better than those obtained with PROPKA2.0 (Bas et al. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 765–783). Finally, we present a novel algorithm that identifies noncovalently coupled ionizable groups, where pK_a prediction may be especially difficult. This is a general improvement to PROPKA and is applied to proteins with and without ligands.

1. INTRODUCTION

The interactions between biological molecules are fundamental to all biological phenomena. The often high degree of specificity of the intermolecular interactions offers a high level of control in basic biological processes such as the humoral immune response (antibody–antigen interactions), gene regulation (protein–DNA interactions), and enzyme catalysis (initial enzyme–substrate binding). Similarly, most drugs are small molecules that bind to protein targets. In the early steps of the drug discovery process, thousands of candidate molecules are often screened for activity on the protein target in an expensive and time-consuming process. Accurate modeling of protein–ligand interactions is therefore of great importance and has practical applications in the pharmaceutical industry where virtual screening for drug candidates is of increasing importance.¹

A great number of computer algorithms have been developed for modeling interactions between ligands and protein molecules.² However, much less effort has been devoted to the prediction of titrational events during protein–ligand binding.^{3–5} When a ligand binds to a protein receptor, the chemical environment of the ligand and the active site residues is changed. This can result in titrational events on the ligand molecule⁶ and in the active site, which influences the binding affinity of the ligand molecule to the protein receptor. Hence, the prediction of ligand pK_a values can be of great importance for the accurate modeling of these events.

In this Article, we announce the new version 3.1 of PROPKA,^{5,7,8} which features a number of improvements with particular focus on prediction of ligand pK_a values. The previous version 2.0 of PROPKA models interactions between protein and ligand chemical groups. However, PROPKA2.0 has a number of shortcomings that we will seek remedy in the present PROPKA3.1.

While the previous version 2.0 of PROPKA supported modeling of interactions between protein and ligand molecules, no

modeling of interactions between ligand groups was available. This shortcoming can lead to errors in predicted pK_a values for systems with multiple ligand molecules complexed with a protein, systems containing large ligand molecules including several titratable groups, or systems containing ligand molecules that are coordinated with monatomic ions.

Additionally, PROPKA3.1 takes inductive interactions between covalently coupled ligand groups into account. For two covalently coupled groups in an aromatic ring, it can be the case that a titrational event at one of the covalently coupled groups renders a titrational event at the other group much less likely. As the molecular environment of the two groups is often similar, it is often the chemical environment of the groups that determines which group will titrate. PROPKA3.1 models covalently coupled titrational events by identifying groups that are not titrating due to coupling with another titrational event and restricting these groups to the neutral state.

In PROPKA3.1, we have furthermore implemented algorithms for the modeling of noncovalently coupled titrational events. Two titratable groups, residing in a protein or ligand, that are spatially proximate can influence the titration of each other such that the two groups titrate in a coupled fashion and that the shape of the titration curves can deviate from the standard Henderson–Hasselbalch shape.^{9,10} Such noncovalently coupled titration curves have, for example, been observed for *Bacillus circulans* xylanase¹¹ and thioredoxins.^{12–14} These noncovalently coupled titration events can be considered a combination of two microscopic titrational events where either of the two groups can titrate first. PROPKA3.1 models noncovalently coupled titrational events by predicting the resulting pK_a values in both of the two possible titration orders.

Received: February 23, 2011

Published: May 27, 2011

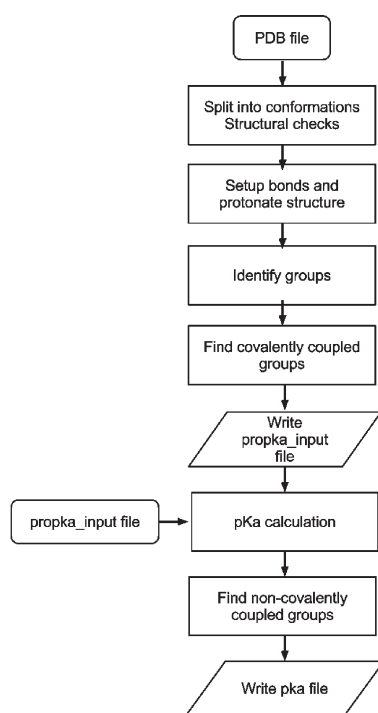


Figure 1. Flowchart of the calculation procedure in PROPKA3.1.

Finally, the implementation of the PROPKA versions 3.0 and 3.1 in Python¹⁵ makes it a relatively easy task to modify and extend the code to include new chemical groups.

2. MATERIALS AND METHODS

2.1. Calculation Workflow. The workflow of PROPKA3.1 is sketched in Figure 1. When initialized with a .pdb file as argument, the workflow of PROPKA3.1 is as follows: (i) The structure PDB file is read into PROPKA3.1, taking into account multiple models and alternative locations, and a series of checks is done to ensure structural consistency, for example, that known protein residues contain the expected number of atoms. The assignment of Sybyl atom types to ligand atoms is done internally in PROPKA3.1 based solely on atom elements and coordinates. For a list of recognized Sybyl types, see Table 1. (ii) Chemical bonds are initialized on the basis of atom coordinates and element types. Atoms separated by less than 2.0 Å are generally assumed to be bonded. However, if one of the two atoms is a hydrogen atom, the threshold is 1.5 Å, and for disulfides the threshold is 2.5 Å. Information on π -electrons and conjugated bonds is set up using table values for protein atom names and ligand Sybyl types. Protons are added as described below. (iii) The chemical groups needed for the pK_a calculation are identified from the atoms that have been read in. For protein atoms, the identification of chemical groups is based on atom names, and for heterogeneous atoms, the identification is based on predicted Sybyl types. Besides the ligand groups recognized by PROPKA2.0,⁵ two additional groups are recognized in PROPKA3.1: titratable oxygen atoms in phosphate groups (default model pK_a value 6.0 was set on the basis of a series of experimental pK_a values for molecules containing phosphate groups¹⁶) and thiol groups (default model pK_a value 10.0¹⁷); confer Table 2. (iv) Covalently coupled groups are identified as described below. (v) A .propka_input file is written out, allowing the user to modify the standard PROPKA3.1 setup. (vi) The pK_a value calculation is done using the algorithm described

Table 1. Recognized Sybyl Types

Sybyl name	atom
C.3	sp^3 carbon
C.2	sp^2 carbon
C.1	sp carbon
C.ar	aromatic carbon
N.3	sp^3 nitrogen
N.1	sp nitrogen
N.ar	aromatic nitrogen
N.am	nitrogen in amide
N.pl3	trigonal planar nitrogen
O.3	sp^3 oxygen
O.2	sp^2 oxygen
O.co2	oxygen in carboxylate
S.o2	sulfur in sulfone
S.3	sp^3 sulfur
P.3	sp^3 phosphorus

in ref 7 expanded to also include ligand atoms as described below. (vii) Noncovalently coupled groups are identified as described below. (viii) A resulting .pka file is written out.

2.2. pK_a Shifts Due to Ligand Interactions. In the following, a brief summary of the underlying algorithms in PROPKA3.1 is given with focus on the extension of the parameter set to include ligand groups. For more details on the theoretical considerations that PROPKA3.1 is based on, please consult ref 7.

In PROPKA3.1, pK_a values are modeled as:

$$pK_a = pK_a^{\text{water}} + \Delta pK_a^{\text{water} \rightarrow \text{protein}} \quad (1)$$

where pK_a^{water} is the pK_a value of the titratable group in water (often referred to as the model pK_a value), and $\Delta pK_a^{\text{water} \rightarrow \text{protein}}$ is the contribution to the pK_a value that stems from the protein environment of the group. While the former value is a table value for protein residues, it will, for ligand groups, depend on the chemical properties of the molecule that the group is a part of. Hence, the default model pK_a values in Table 2 for ligand groups are only proximate and should be replaced if a more precise value is known for the considered ligand molecule. The latter value is predicted by PROPKA3.1 on the basis of the supplied structure and consists of

$$\Delta pK_a^{\text{water} \rightarrow \text{protein}} = \Delta pK_a^{\text{desolv}} + \Delta pK_a^{\text{HB}} + \Delta pK_a^{\text{RE}} + \Delta pK_a^{\text{QQ}} \quad (2)$$

where $\Delta pK_a^{\text{desolv}}$ is the contribution due to desolvation effects, ΔpK_a^{HB} is the contribution due to hydrogen-bond interactions, ΔpK_a^{RE} is the contribution due to unfavorable electrostatic reorganization energies, and ΔpK_a^{QQ} is the contribution due to Coulombic interactions. Of these four terms only the hydrogen bond and Coulombic contributions are calculated using group-specific parameters. The remaining terms are calculated using generic parameters and do therefore not need to be extended with new parameters to allow modeling of ligand groups.

2.2.1. Intrinsic Electrostatic Interactions. Hydrogen-bond interaction energies are modeled as

$$\Delta pK_a^{\text{HB}} = \begin{cases} c^{\text{HB}} w(r) \cos \theta & \text{if } \theta \geq 90^\circ \\ 0 & \text{else} \end{cases} \quad (3)$$

Table 2. Ligand Chemical Groups Recognized by PROPKA3.1 and Their Default Model pK_a Values

Chemical name	Group ID	Model pK_a value	Structure
Guanidinium	CG	11.50	
Amidinium	C2N	11.50	
Ammonium	N30	10.00	
sp ³ primary nitrogen	N31	10.00	
sp ³ secondary nitrogen	N32	10.00	
sp ³ tertiary nitrogen	N33	10.00	
Aromatic nitrogen	NAR	5.00	
Carboxyl	OCO	4.50	
Thiol	SH	10.00	
Phosphate	OP	6.00	
Amide	NAM	-	
Trigonal planar NH ₂	NP1	-	
sp ³ oxygen	O3	-	
Chlorine	CL	-	
Flourine	F	-	
sp nitrogen	N3	-	
sp ² oxygen	O2	-	
Hydroxyl	OH	-	

where c^{HB} is the maximal hydrogen-bond interaction energy (fitted to 0.85 pK_a units), r is the interatomic distance between

the hydrogen-bond donor and acceptor, θ is the angle formed by the donor, the hydrogen atom, and the acceptor, and w is a

Table 3. Distance Parameters for Calculation on Intrinsic Electrostatic Interactions for Ligand Groups That Were Set On the Basis of Parameters of Equivalent Protein Groups^a

chemical name	group ID	similar protein group
guanidinium	CG	arginine guanidinium
amidinium	C2N	arginine guanidinium
ammonium	N30	lysine/N-terminus nitrogen
sp ³ primary nitrogen	N31	lysine/N-terminus nitrogen
sp ³ secondary nitrogen	N32	lysine/N-terminus nitrogen
sp ³ tertiary nitrogen	N33	lysine/N-terminus nitrogen
aromatic nitrogen	NAR	histidine imidazole nitrogen
carboxyl	OCO	aspartic/glutamic acid carboxyl
phosphate	OP	
thiol	SH	cysteine thiol
hydroxyl	OH	threonine/serine hydroxyl
trigonal planar NH ₂	NP1	
sp ³ oxygen	O3	
chlorine	CL	
fluorine	F	
amide nitrogen	NAM	asparagine/glutamine amide nitrogen
sp nitrogen	NI	
sp ² oxygen	O2	

^a Table S.1 in the Supporting Information contains a list of all ligand intrinsic electrostatic interaction distance parameters. Interactions between ligand groups are allowed if the interacting ligand groups are separated by at least four chemical bonds.

distance-dependent weight function given by

$$w(r) = \begin{cases} 1 & \text{if } r \leq r_{\min} \\ \frac{r_{\max} - r}{r_{\max} - r_{\min}} & \text{if } r_{\min} < r \leq r_{\max} \\ 0 & \text{else} \end{cases} \quad (4)$$

where r_{\min} and r_{\max} are group-specific cutoff distances found by analysis of hydrogen-bond distances found in X-ray protein structures.

The generic maximal intrinsic interaction energy, c^{HB} , was set to 0.85 for all interactions involving ligand groups. However, to achieve consistency with PROPKA3.0, strong, buried interactions between histidine residues and ligand carboxylic groups are assigned an initial interaction energy, c^{HB} , of 1.6 pK_a units similarly to interactions between Asp/Glu and His residues.^{5,7}

Because of the relatively limited amount of structures containing hydrogen-bond interactions between protein and specific ligand groups, cutoff distances for ligand groups have been set by extrapolating the existing parameters in PROPKA3.0 for protein groups. Interaction distance cutoffs are set to the values of similar protein residue groups, confer Table 3; for example, the distance parameters of guanidinium and amidinium groups are set to the values of the arginine residue. In those cases where no equivalent protein group exists, the default distance parameters of $c_{\min} = 3.0 \text{ \AA}$ and $c_{\max} = 4.0 \text{ \AA}$ were applied.

2.2.2. Coulombic Interactions. The contribution to the pK_a value of group i due to the Coulombic interaction between groups i and j is calculated as⁷

$$\Delta pK_{a,i}^{\text{QQ}} = \sigma_{ij} \frac{244}{\epsilon r_{ij}} w(r_{ij}) \quad (5)$$

where 244 is the coefficient of Coulomb's law converted into pK_a units, $w(r_{ij})$ is a weight function, and r_{ij} is the distance between groups i and j . The dielectric constant, ϵ , is in this context a function of the degree of solvent exposure as explained in detail in ref 7. A unit step function, σ_{ij} , is determining the sign of the contribution to the pK_a value of group i . The unit step function for the Coulombic interaction between groups i and j is defined as

$$\sigma_{ij} = \begin{cases} -1 & \text{if } i \text{ is acid and } j \text{ is base or} \\ & i \text{ and } j \text{ are bases and } pK_{a,i} < pK_{a,j} \\ +1 & \text{if } i \text{ is base and } j \text{ is acid or} \\ & i \text{ and } j \text{ are acids and } pK_{a,i} > pK_{a,j} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

In some cases, the contribution of Coulombic interactions to the pK_a value depends on the total pK_a values of the interacting groups. An iterative approach is taken to resolve the total pK_a values in these cases. Table S.2 in the Supporting Information lists the interaction type (normal or iterative) for each possible interaction.

2.3. Protonation Scheme. A new generic protonation scheme has been implemented. The protonation scheme involves four steps. (i) Formal charges are set. For protein atoms, ARG-NH1, HIS-ND1, LYS-NZ, and the N-terminus are given a positive charge. For ASP-OD2, GLU-OE2, and the C-terminus, a negative charge is given. For heterogeneous atoms, charges are assigned on the basis of Sybyl atom types. Atoms assigned the Sybyl types N.pl3, N.3, N.4, and N.ar are given a positive charge, and one of the atoms in carboxyl groups (both oxygen atoms are assigned the Sybyl type O.co2) is given a negative charge. These charges are only used internally in the protonation algorithm and not in the pK_a calculation. (ii) The number of protons that should be added to each atom is set on the basis of the number of valence electrons, the number of bonds, the number of π -electrons in double and triple bonds, and the formal charge. (iii) The steric number and number of lone pairs are set for each atom. (iv) Protons are added on the basis of the assigned steric numbers using either trigonal or tetrahedral geometrical considerations. Only ligand molecules and protein groups included in the pK_a calculation are protonated by default.

2.4. Coupling Effects. Two types of coupling effects, covalent and noncovalent, are modeled in PROPKA3.1. The coupling effects are described in the following subsections and demonstrated using specific examples on structures including HIV protease, ATP, and a pyrazine derivative in the Results and Discussion.

2.4.1. Noncovalent Coupling. Titratable groups that are proximate in the protein 3D structure can titrate in a coupled fashion. This phenomenon can manifest itself in macroscopic titration curves that do not conform to the standard Henderson–Hasselbalch shape. Microscopically, the titration of two noncovalently coupled acidic groups can occur with either of the two groups titrating first and the other group titrating at an elevated pH value. PROPKA3.1 identifies noncovalently coupled groups and models their titrational behavior. For a pair of noncovalently coupled titratable groups, PROPKA3.1 models both orders of titration by assigning the shift in pK_a value due to the interaction between the groups to either of the two groups.

The criteria listed in Table 4 must all be fulfilled before two groups are treated as being noncovalently coupled in PROPKA3.1. The difference of the intrinsic pK_a values of the two groups, $\Delta pK_{a,\max}^{\text{int}}$ must not be higher than 2.0 pK_a units to ensure

Table 4. Criteria for Detection of Noncovalently Coupled Groups

name	symbol	value [pH units]
max difference in intrinsic pK _a	$\Delta pK_{a,\max}^{\text{int}}$	2.0
min interaction energy	E_{min}	0.5
max free energy difference	ΔG_{max}	1.0
min pK _a shift on swap	$\Delta pK_{a,\text{min}}$	1.0
min pK _a	$pK_{a,\text{min}}$	0.0
max pK _a	$pK_{a,\text{max}}$	10.0

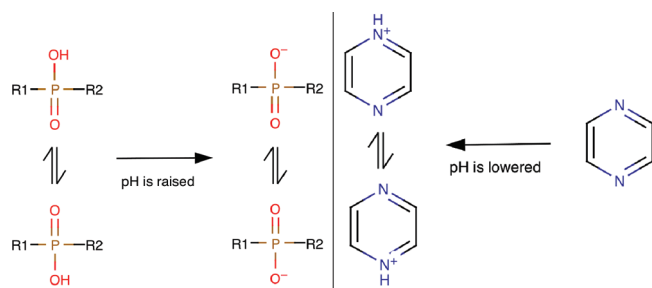


Figure 2. Examples of covalently coupled groups. Left panel: Phosphate group containing two oxygen atoms of which one can titrate and the other is double bonded to the phosphorus atom. Right panel: Pyrazine ring containing two nitrogen atoms for which titration of one nitrogen atom will significantly lower the pK_a value of the other nitrogen atom. Prepared using MarvinSketch.¹⁹

that the groups titrate at similar pH values. The intrinsic pK_a is defined as the theoretical pK_a value a titratable group would have in the protein environment if no electrostatic interactions with other titratable groups were present.¹⁸ The interaction energy, E_{min} , between the two interacting groups must correspond to at least 0.5 pK_a units, and the difference in free energy between the two possible protonation states, ΔG_{max} , must not be greater than 1.0 pK_a units. The shift in pK_a value when swapping to the alternative protonation state, $\Delta pK_{a,\text{min}}$, must be greater than 1.0 pK_a unit. Only titratable groups with pK_a values between 0.0 and 10.0 are considered.

2.4.2. Covalent Coupling. We define the covalent coupling effect as the titration of one group that is significantly influencing the probability of titration of another group proximate in the chemical structure. Two examples of such systems of covalently coupled groups are given in Figure 2. It is the interactions with the environment that determines which of the groups in a covalently coupled system of groups that will titrate. We model covalent coupling effects with a simple scheme that hinders the titration of groups for which titration has been rendered unlikely by a covalently coupled group. The groups for which titration has been penalized are restricted to the neutral form.

Ligand titratable groups are as default assumed to be covalently coupled if they are separated by no more than three chemical bonds and assigned the same Sybyl atom type. However, the user is free to define systems of covalently coupled titratable groups.

An initial pK_a calculation is done including all groups in the molecular structure. Next, for each system of covalently coupled groups, the groups that are to be penalized are found. For bases, the group with the highest predicted pK_a (the most stable group in the charged form) will be the first one to adopt a proton as pH

is lowered, and this group is allowed to titrate. The remaining groups in the system are penalized so that they are restricted to the neutral form and not allowed to titrate. For acids, the group with the highest pK_a value (the least stable group in the charged form) will be the last group to lose the proton as pH is raised and will be penalized so that it is restricted to the neutral form and not allowed to titrate. The remaining groups in the system are allowed to titrate. Finally, a new pK_a calculation is performed disregarding the penalized groups in each covalently coupled system.

In some cases where the charge is distributed between more than one titratable atom (see the below example of aminochloro-pyrazine in complex with cell division protein kinase 2), a common charge center can be set. The common charge center, r_{ccc} , is defined as the average of the positions of the covalently coupled groups:

$$r_{\text{ccc}} = n^{-1} \sum_{i=1}^n r_i \quad (7)$$

where n is the number of covalently coupled groups in the system, and r_i is the position of group i .

2.5. Prediction of Ligand Model pK_a Values. For the calculations of protonation state change on protein–ligand binding, experimental and previously calculated ligand model pK_a values were used.⁵ The Schrödinger program Epik,²⁰ which is part of the Schrödinger Software Suite, was used for the prediction of ligand model pK_a values for the covalently coupling examples presented in the Results and Discussion.

2.6. Calculation of Protonation State Changes. The overall protonation change, n , is calculated as described in ref 5:

$$n = \sum_i^N n_i n = \sum_i^N \left(\frac{10^{pK_{a,i}^c - \text{pH}}}{1 + 10^{pK_{a,i}^c - \text{pH}}} - \frac{10^{pK_{a,i}^f - \text{pH}}}{1 + 10^{pK_{a,i}^f - \text{pH}}} \right) \quad (8)$$

where N is a set containing all titratable groups in the system, and $pK_{a,i}^c$ and $pK_{a,i}^f$ represent pK_a values of titratable group i in the complexed and free forms of the protein and the ligand molecule, respectively.

2.7. Protein Structures. A list of the structures from the Protein Data Bank²¹ used in this study is presented in the Supporting Information in Table S.3.

In the plasmepsin II (PDB id: 1pfz and 1sme) and cathepsin D (1lyw and 1lyb) structures, the N-terminus undergoes a large displacement between the apo and holo structures. This movement can be attributed to the autocatalytic behavior of aspartic acids. To limit the effect of the N-terminus movement on the calculation results, we adopt the procedure of Alexov²² and remove the N-terminus section from these structures. Furthermore, a flap region (residues 76–80) in the plasmepsin II holo structure was not resolved, and the corresponding region in the apo structure (1sme) was removed. All structures were repaired using the corall method of WHAT IF.²³

Reduced cysteine bridges were manually reconstructed in the PROPKA3.1 input files for the complexes trypsin–1bMe (Cys173–Cys197 and Cys25–Cys41) and trypsin–1c (Cys173–Cys197). Ligand model pK_a values were set in accordance with experimentally determined values.²⁴

3. RESULTS AND DISCUSSION

In the following sections, we will demonstrate the implemented covalent and noncovalent coupling algorithms with pK_a calculations on HIV protease, a pyrazine derivative, and ATP.

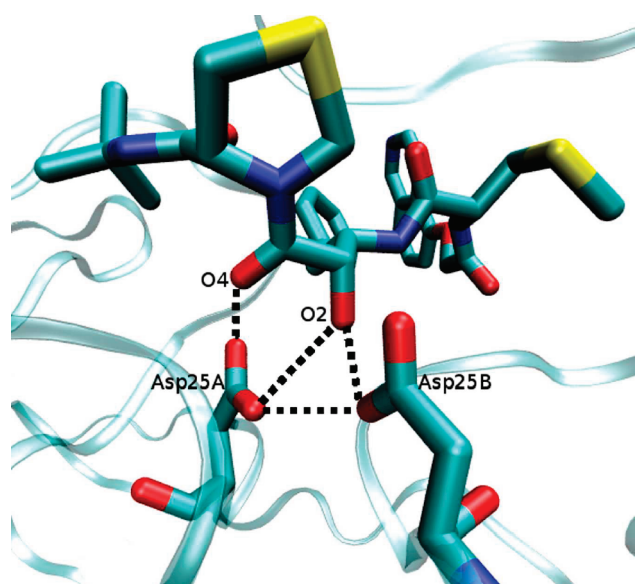


Figure 3. HIV protease complexed with the inhibitor KNI-272 (1hpx). The catalytic active aspartic acids at position 25 in the A and B chains and the ligand molecule are represented as sticks. Side-chain interactions for catalytic active aspartic acids are indicated with dotted lines. Prepared using VMD²⁵ with the PROPKA GUI plugin.²⁶

3.1. Noncovalent Coupling: HIV Protease. PROPKA3.1 detects the catalytic active aspartic acids in HIV protease at position 25 in chains A and B to be noncovalently coupled. Figure 3 is a representation of HIV protease in complex with the KNI-272 inhibitor (1hpx) where the side-chain interactions predicted by PROPKA are represented as dotted lines. PROPKA predicts Asp25A to have side-chain interactions with Asp25B and the two oxygen atoms (O2 and O4) in the allophenylnorstatine fragment of the inhibitor. These side-chain interactions contribute with -0.85 (O2), -0.85 (O4), and -0.84 (Asp25B) pK_a units to the total pK_a value of Asp25A. For the pK_a value of Asp25B, O2 contributes -0.85 pK_a units, whereas Asp25A contributes with $+0.84$ pK_a units. This gives rise to the total pK_a values of 5.07 (Asp25A) and 9.28 (Asp25B). However, the two aspartic residues are identified as being coupled, indicating the existence of an alternative protonation state. If the contributions to the pK_a values due to the interactions between the aspartic acids are swapped, the alternative total pK_a values of 8.64 (Asp25A) and 5.70 (Asp25B) are found.

Experimental results show that the pK_a value of one of the catalytic active aspartic acids is 6.6 when in complex with KNI-272.²⁷ As the dyad is found to be monoprotonated, the experimental value should be compared to the highest of the predicted pK_a values. While these values (9.28 or 8.64) are somewhat higher than the experimental value, the PROPKA2.0 value of 10.28 has a greater discrepancy. The pK_a calculation on HIV protease complexed with KNI-272 is discussed in more detail later.

3.2. Covalent Coupling: Aromatic Nitrogen Atoms in a Pyrazine Derivative. For some chemical compounds, the titration of individual titratable sites is far from independent. Pyrazine rings, for example, contain two potentially titratable nitrogen atoms. However, once one of the nitrogen atoms has adopted a hydrogen atom, the probability of the other nitrogen atom adopting a hydrogen atom is significantly reduced. For the pyrazine

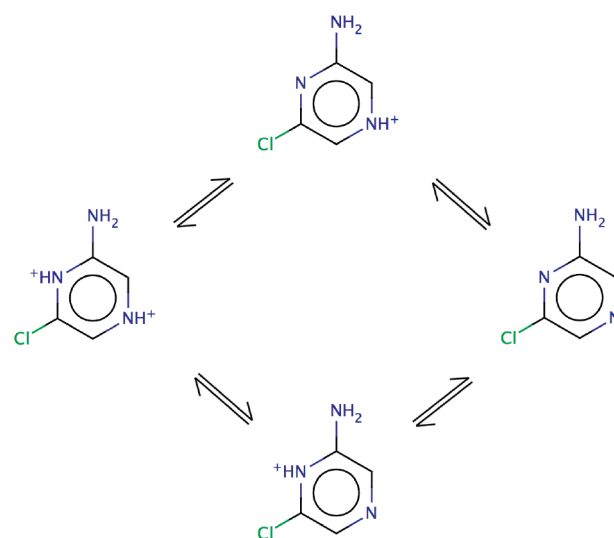


Figure 4. Modeled protonation states of 2-amino-6-chloro-pyrazine.

Table 5. Resulting pK_a Values for the Aromatic Nitrogen Atoms in 2-Amino-6-chloro-pyrazine^a

calculation	predicted pK_a values		model pK_a values		CCC
	N4	N8	N4	N8	
A	1.91	0.44	5.00	5.00	no
B	-0.98	-4.99	2.09	-0.43	no
C	-0.78	-4.76	2.09	-0.43	yes

^a Calculation A, the PROPKA default model pK_a value of 5.00; calculation B, custom ligand model pK_a values; calculation C, custom model pK_a values and common charge center (CCC) approach.

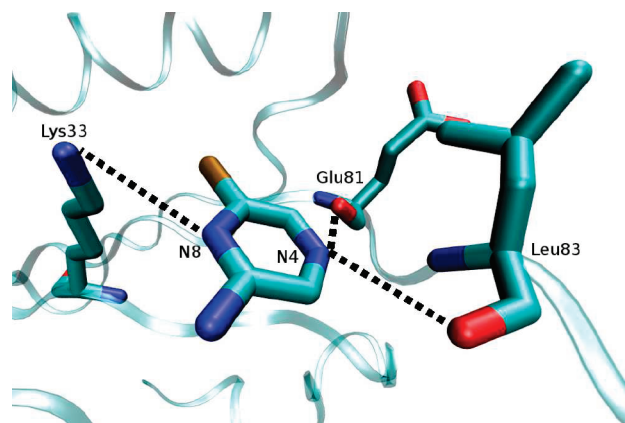


Figure 5. Modeled interactions between 2-amino-6-chloro-pyrazine and CDK2. Prepared using the PROPKA GUI plugin for VMD.

molecule, pK_a values of the first and second protonation have been observed to be $pK_a(1)$ 0.65 and $pK_a(2)$ -6.6 .²⁸ To model such compounds in PROPKA3.1, we must take these covalent couplings into account.

To illustrate this principle, we will investigate cyclin-dependent kinase 2 (CDK2) in complex with amino-chloro-pyrazine (1wcc²⁹). A schematic overview of the protonation states of amino-chloropyrazine is given in Figure 4.

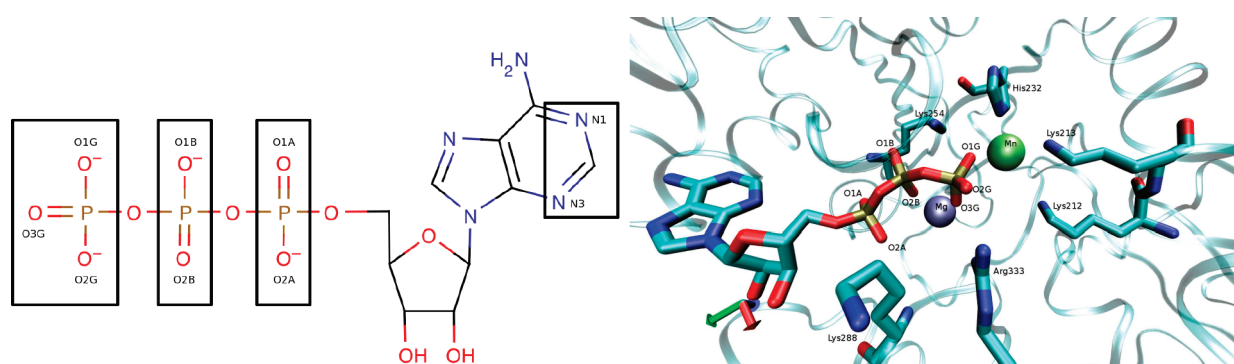


Figure 6. ATP. Left panel: Structure of ATP with covalently coupled groups marked by boxes. Prepared using MarvinSketch. Right panel: ATP in the binding site of phosphoenolpyruvate carboxykinase. Residues with strong interactions with the ATP phosphate groups are represented as sticks. The magnesium and manganese ions are represented as spheres. Prepared using the PROPKA GUI plugin for VMD.

Three calculations on the complex between CDK2 and amino-chloro-pyrazine have been done; see Table 5. In calculation A, the PROPKA default ligand model pK_a value of 5.00 for aromatic nitrogen atoms is applied. In calculation B, custom ligand model pK_a values of 2.09 (N4) and -0.43 (N8) calculated with Schrödinger's Epik program were utilized. Finally, in calculation C, the custom Epik ligand model pK_a values utilized and electrostatic interactions were calculated using the common charge center approach; see Materials and Methods. For both aromatic nitrogen atoms, the predicted pK_a value is significantly lowered due to large desolvation contributions (-2.51 pK_a units for N8 and -3.41 pK_a units for N4 in calculations A and B, and -2.84 pK_a units for both nitrogen atoms in calculation C). A protonation of the N8 nitrogen would result in unfavorable interactions with the proximate Lys33; see Figure 5. The lysine residue therefore contributes with a side-chain interaction contribution of -0.36 pK_a units in all calculations and a Coulombic interaction contribution of -1.56 pK_a units (calculations A and B) or -0.99 pK_a units (calculation C) to the predicted pK_a value of N8. The pK_a value of N4 is, on the other hand, increased due to interactions with the backbone carbonyl of Glu81 (0.24 pK_a units) and Leu83 (0.86 pK_a units). Hence, in all three calculations, N8 has the lowest pK_a value and is excluded from the final calculation.

PROPKA2.0 applied to the same structure wrongly predicts that both aromatic nitrogen atoms titrate at similar pH values. The pK_a values are predicted to be 0.99 (N4) and 0.76 (N8).

3.2.1. ATP. A more complicated example is adenosine-5'-triphosphate (ATP), which with its three phosphate groups urges for careful considerations during the setup of a PROPKA calculation. Each of the phosphate groups contain two or three oxygen atoms that could potentially adopt a proton. However, only four of the total of seven phosphate oxygen atoms can titrate as the remaining three phosphate oxygen atoms have double bonds to a phosphorus atom. The challenge is therefore to determine which of the phosphate oxygen atoms should be allowed to titrate.

The new algorithm for treatment of covalently coupled titratable groups allows the users to let PROPKA3.1 make this decision. The oxygen atoms in each phosphate group are set up to be covalently coupled as shown in the left panel in Figure 6. The ability of titration is hence removed for one oxygen atom in each of the phosphate groups.

Applying this approach to ATP complexed with phosphoenolpyruvate carboxykinase (1aq2,³⁰ Figure 6, right panel)

identifies the double-bonded, nontitratable phosphate oxygen atoms as O3G with the potential pK_a value -11.66 (as compared to O1G, -12.92 and O2G, -12.82), O2B with pK_a value -4.30 (as compared to O1B, -5.31), and O1A with pK_a value 1.69 (as compared to O2A, 0.57); confer the left panel of Figure 6.

The potential pK_a values for the oxygen atoms that are restricted to the neutral state are consistently around one pK_a unit higher than the pK_a values of the oxygen atoms that are found to titrate. This difference in pK_a value is relatively small, and the choice of nontitratable oxygen atoms can seem somewhat ambiguous. We do, however, not consider this ambiguity important. Instead, it is important to note that PROPKA automatically has assigned the double bonds to the oxygen atoms for which it is predicted to be most energetically favorable for the entire system. Additionally, the ambiguity stemming from a user-defined setup of the ATP molecule is removed.

3.3. Protonation State Changes on Complexation. The following sections describe the results using PROPKA3.1 on calculations originally done with PROPKA2.0. For more in-depth information on the experimental results, etc., please consult ref 5.

3.3.1. Trypsin and Thrombin. Table 6 contains the results for calculations of protonation state changes done on serine protease trypsin complexed with a set of related inhibitors derived from N^α -(2-naphthylsulphonyl)-L-3-amidino-phenylalanine and thrombin complexed with four inhibitors. Klebe and co-workers^{24,31} have measured protonation state changes and calculated pK_a changes using the PEOE_PB scheme for these complexes. Also included for comparison are the calculations originally done with PROPKA2.0.

For the binding of the ligand 1b to trypsin, an overall protonation state change of 0.66 was calculated in accordance with the experimental value of 0.90. This is an improvement as compared to the calculated value by PROPKA2.0 of 0.09. The majority of the overall calculated protonation state change has its origin in a pK_a shift of His57 from 7.47 in the apo form to 9.35 in the holo form. The shift in pK_a value stems from interactions between the ligand carboxyl group and His57. While the increased desolvation contribution gives rise to a negative contribution to the pK_a shift of His57 (-2.69 in the holo form as compared to -1.82 in the apo form), the desolvation contribution also activates the COO-HIS exception (confer Materials and Methods) for both Asp102 and the ligand carboxyl giving a contribution of 1.60 for each carboxyl group. For trypsin complexed with the ligands 1bMe, 1c, and 1cMe, negligible overall protonation state changes were predicted in accordance

Table 6. Experimental and Calculated Total Changes in Protonation State on Complexation for the Trypsin and Thrombin Complexes Studied^a

name	n_{exp}	PROPKA		n (pK_a) complexed			pK_a uncomplexed		
		3.1	2.0	His57	ligCOO	ligAMINO	His57	ligCOO	ligAMINO
Trypsin									
1b (1kdi ²⁴)	0.90	0.66	0.09	0.65(9.35)	0.00 (0.99)	n/a	7.47	3.21	n/a
1bMe	0.00	0.12	0.06	0.06(7.57)	n/a	n/a	7.45	n/a	n/a
1c	0.00	-0.05	0.15	-0.01 (7.49)	0.00 (4.19)	n/a	7.50	4.17	n/a
1cMe (1klj ²⁴)	0.00	-0.30	-0.24	-0.21(6.85)	n/a	n/a	7.46	n/a	n/a
1d (1klf ²⁴)	-0.53	-0.54	-0.55	-0.24(6.32)	n/a	-0.26(6.65)	7.37	n/a	7.49
1dAc (1klm ²⁴)	0.00	-0.22	-0.23	-0.18(6.89)	n/a	n/a	7.42	n/a	n/a
2	0.93	0.71	0.69	0.68(9.39)	0.00 (1.30)	n/a	7.43	3.40	n/a
3 (1kln ²⁴)	0.00	0.00	-0.05	-0.09 (7.18)	0.00(3.14)	n/a	7.39	3.84	n/a
4 (1klo ²⁴)	0.00	0.21	0.03	-0.16(6.93)	0.00(0.94)	0.28(7.99)	7.38	2.65	7.48
5 (1klp ²⁴)	0.00	0.22	0.07	-0.12(7.15)	0.00(1.09)	0.27(8.56)	7.43	2.51	7.95
Thrombin									
2	0.88	0.61	0.34	0.80(9.22)	0.00 (-0.90)	n/a	7.08	3.40	n/a
3 (1ypk ³²)	0.00	-0.15	-0.01	-0.13(5.92)	0.00 (1.65)	n/a	7.01	3.84	n/a
4 (1k21 ²⁴)	0.00	0.12	-0.07	-0.15(6.46)	0.00(0.75)	0.27(7.96)	7.18	2.65	7.48
5 (1k22 ²⁴)	0.00	-0.08	0.04	-0.15(6.37)	0.00(0.61)	0.20(8.36)	7.15	2.51	7.95
rmsd		0.18	0.29						

^a For His57 and ligand carboxyl (ligCOO) and amino (ligAMINO) groups (where applicable), pK_a values in the complexed and uncomplexed structures are listed along with contributions to the change in protonation state on complexation.

with experimental results. For the ligand 1d, the calculated value of -0.54 corresponds very well to the experimental value of -0.53. Similarly, the calculated values for trypsin-2 (0.71) and thrombin-2 (0.61) correspond well with the experimental values of 0.93 and 0.88, respectively. For thrombin-2, a clear improvement to the value calculated using PROPKA2.0 of 0.34 is observed.

In summary, an improvement in the predictions of PROPKA3.1 as compared to the predictions of PROPKA2.0 was observed for trypsin-1b and thrombin-2. The remaining results were similar to those obtained with PROPKA2.0. The overall rms deviations between experimental protonation state changes on complexation and calculated values were 0.18 for PROPKA3.1 and 0.29 for PROPKA2.0.

3.3.2. Proteins Complexed with Pepstatin. The protonation state change on ligand binding for three proteins (plasmepsin II, cathepsin, and endiothiapepsin) complexed with pepstatin is calculated and compared to experimental data and PROPKA2.0 results, cf., Table 7.

For plasmepsin II (1pfz) complexed with pepstatin (1sme), PROPKA3.1 predicts the overall protonation state change on ligand binding to be 2.67. This value is higher than the PROPKA2.0 value of 1.50 but still in accord with the experimental value of 1.7. The main predicted contributions to the overall protonation state change stem from Asp303 (1.00), Asp214 (0.99), and His318 (0.54). While the pK_a shifts of Asp303 and Asp12 can be attributed to increased desolvation on ligand binding and to conformational changes, the pK_a shift of His318, which is distant from the active site, stems from an interaction with Asp190 found in the holo structure but not in the apo structure. His164 was found to give a small contribution (-0.20) in the opposite direction of what would be expected from the experimental pK_a values.

Figure 7 shows predicted protonation state changes for the complexation of plasmepsin II with pepstatin at various pH values. While the prediction has a clear discrepancy with the experimental values, it is comparable to previously calculated results.^{5,22}

For cathepsin (1lyw) complexed with pepstatin (1lyb), the overall protonation state change is predicted to be 2.39, which is in good agreement with the experimental value of 2.9 and very close to the PROPKA2.0 value of 2.49. The main contributions stem from Asp323 (0.92) and His56 (0.68). The pK_a shift of Asp323 is caused by an increased desolvation due to the binding of pepstatin, whereas the pK_a shift of His56 is due to conformational changes more distant from the active site.

For endiothiapepsin (4ape) complexed with pepstatin (4er2), only a small protonation state change is predicted (0.08), whereas the PROPKA2.0 value is 0.77 and the experimental value is 1.06. The original PROPKA2.0 calculation found the pK_a value of Asp30 to be shifted from 6.69 to 7.21 (corresponding to a protonation state change of 0.29) and the pK_a value of Asp12 to be shifted from 5.55 to 9.26 (corresponding to a protonation state change of 0.96) during complexation at pH 7.0. These shifts were mainly due to desolvation effects and not reproduced in PROPKA3.1, which has an improved desolvation model.

To summarize, the main result of this subsection is that the overall protonation state values predicted by PROPKA3.1 are similar to the PROPKA2.0 values for cathepsin, but worse for plasmepsin II and endiothiapepsin.

3.3.3. HIV Protease. For HIV protease complexed with pepstatin (Shvp), the pK_a values of the catalytic dyad consisting of Asp25A and Asp25B are predicted to be 10.16 and 5.73, respectively. This is partly in accord with experimental findings where one of the aspartic acids has a pK_a value less than 2.5 and the other has a pK_a value greater than 6.5 as measured with

Table 7. Experimental (Where Available) and Calculated Protonation State Changes for Other Protein–Ligand Complexes Studied^a

name	n_{exp} (PROPKA3.1; 2.0)	experimental pK_a (PROPKA3.1) for selected residues			
plasmepsin II (1pfz ³³)		Asp34A/Asp214A 4.7 (4.90/3.67)	Asp34A/Asp214A 4.7 (4.90/3.67)	His164A 6.0 (6.45)	Asp303A (3.69)
+ pepstatin (1sme ³⁴)	1.7 ^b (2.67; 1.50)	3.0 (4.73/8.77)	6.5 (4.73/8.77)	7.5 (6.06)	(9.14)
cathepsin D (1lyw ³⁵)		Asp33A/Asp231B <6.5 (4.66/7.51)	His77A <6.5 (6.80)	Glu260B <6.5 (4.58)	Asp323B (5.34)
+ pepstatin (1lyb ³⁶)	2.9 ^b (2.39; 2.49)	>6.5 (4.42/8.98)	>6.5 (6.76)	>6.5 (5.87)	(8.42)
endothiapepsin (4ape ³⁷)		Asp32A/Asp215A (9.23/4.42)	Asp12A (6.70)	Asp30A (8.07)	
+ pepstatin (4er ³⁷)	1.06 (0.08; 0.77)	>7.0 (9.33/4.54)	(7.23)	(8.35)	
HIV protease (Shvp ³⁸)		Asp25A/Asp25B (4.78/9.56)	Asp30A/Asp30B (5.10/3.92)	His69A/His69B (6.96/6.25)	
+ pepstatin (Shvp)	NMR	<2.5/>6.5 (10.16/5.73)	(5.40/3.52)	(6.96/6.25)	
HIV protease (1hhp ³⁹)		(5.32/7.19)	(4.38/4.38)	(7.23/7.23)	
+ pepstatin (Shvp)	NMR	<2.5/>6.5 (10.16/5.73)	(5.40/3.52)	(6.96/6.25)	
HIV protease (3hvp ⁴⁰)		(7.60/4.79)	(4.78/4.78)	(6.83/6.83)	
+ pepstatin (Shvp)	NMR	<2.5/>6.5 (10.16/5.73)	(5.40/3.52)	(6.96/6.25)	
HIV protease (1hpx ⁴¹)		6.0 (8.89/5.17)	4.8 (4.42/4.63)	4.8	
+ KNI-272 (1hpx)	−0.23 ^c (0.01; 0.18)	6.6 (5.07/9.28)	3.88/3.78 (4.62/4.91)	2.9 (4.40)	
HIV protease (1hhp)		6.0 (5.32/7.19)	4.8 (4.38/4.38)	4.8	
+ KNI-272 (1hpx)	−0.23 ^c (0.04; 1.27)	6.6 (5.07/9.28)	3.88/3.78 (4.62/4.91)	2.9 (4.40)	
HIV protease (3hvp)		6.0 (7.60/4.79)	4.8 (4.78/4.78)	4.8	
+ KNI-272 (1hpx)	−0.23 ^c (−0.24; 0.97)	6.6 (5.07/9.28)	3.88/3.78 (4.62/4.91)	2.9 (4.40)	
HIV protease (1qbs ⁴²)		6.0 (8.21/6.00)	(4.82/4.82)		
+ DMP-323 (1qbs)	NMR	8.19 (5.88/8.52)	3.99 (5.07/5.09)		
HIV protease (1hhp)		6.0 (5.32/7.19)	(4.38/4.38)		
+ DMP-323 (1qbs)	NMR	8.19 (5.88/8.52)	3.99 (5.07/5.09)		
HIV protease (3hvp)		6.0 (7.60/4.79)	(4.78/4.78)		
+ DMP-323 (1qbs)	NMR	8.19 (5.88/8.52)	3.99 (5.07/5.09)		
chymotrypsin (7gch ⁴³)		His57F/His57I 7.5 (7.33/7.33)			
+ N-acetyl-L-Leu-DL-Phe-CF ₃ (7gch)	NMR	12.0 (8.07/8.07)			
chymotrypsin (6gch ⁴³)		7.5 (7.54/7.48)			
+ N-acetyl-DL-Phe-CF ₃ (6gch)	NMR	10.8 (9.46/9.41)			
xylanase (1bv ⁴⁴)		Glu172A 6.7 (7.52)			
+ 2FXb (1bv)	NMR	4.2 (6.80)			
xylanase N35D (1c5i ⁴⁵)		Asp35A/Glu172A 3.7/8.4(5.58/11.45)			
+ 2FXb (1c5i)	NMR	1.9–3.4 or >9.0 (10.83/5.70)			
hydroxynitrile lyase (2yas ⁴⁶)		His235A 2.5 (6.72)			
+ thiocyanate (2yas)	NMR	8.0 (8.30)			
DHFR (4dfr ⁴⁷)		Asp27B 6.6 (5.82)	methotrexate N1 5.7		
+ methotrexate (4dfr)	NMR	(−0.76)	10.7 (5.97)		
DHFR (4dfr)		6.6 (5.82)	4.4		
+ methotrexate (4dfr)	NMR	(−0.76)	10.7 (4.67)		

^a Unless otherwise stated, pH is 7.0. ^b pH is 6.5. ^c pH is 5.0. Experimental and calculated pK_a values are listed for selected residues.

NMR.⁴⁸ PROPKA3.1 correctly predicts that Asp25A and Asp25B are noncovalently coupled and alternative pK_a values (by swapping the interaction between the two aspartic residues) are 5.81 and 10.08 for Asp25A and Asp25B, respectively. In the apo form, the pK_a

values of Asp25A and Asp25B are predicted to be 4.78 and 9.56, respectively. Utilization of two alternative structures (1hhp and 3hvp) to describe the apo structure lowers the higher pK_a value in the catalytic dyad, making the pK_a shift on pepstatin binding larger.

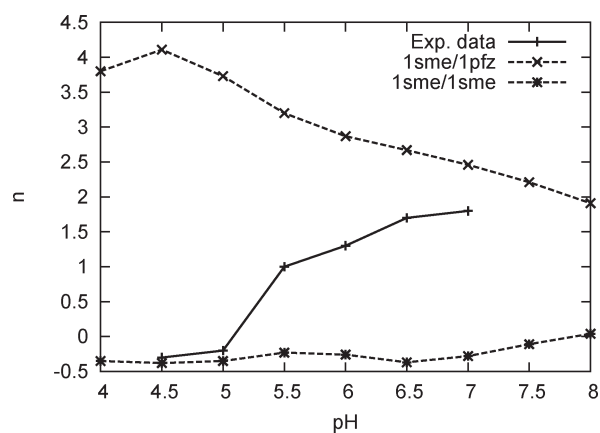


Figure 7. Protonation state change as a function of pH for complexation of plasmepsin II with pepstatin. Predictions were done solely on the basis of the plasmepsin II–pepstatin complex structure (1sme) as well as in combination with the plasmepsin II apo structure (1pfz).

For HIV-1 protease complexed with the inhibitor KNI-272 (1hpx), experimental results show that the pK_a value of one of the catalytic active aspartic acids changes from 6.0 in the apo form to 6.6 when in complex with KNI-272.²⁷ As the dyad is found to be monoprotonated both before and after complexation, we must compare these experimental values to the highest of the predicted pK_a values for the dyad, which are 8.89 (apo) and 9.28 (holo). While these predictions are higher than the experimental value, they are an improvement as compared to the PROPKA2.0 values of 9.26 (apo) and 10.28 (holo). The catalytic dyad is predicted to be noncovalently coupled; see Noncovalently Coupling: HIV Protease. The pK_a value of the isoquinoline nitrogen on KNI-272 has been observed to shift from 4.8 to 2.9 during complexation. Using the experimental model pK_a value of 4.8, PROPKA3.1 predicts the pK_a value after complexation to be 4.40. This is comparable to the PROPKA2.0 value of 4.52.

The overall protonation state change was predicted to be 0.01, 0.04, and -0.24 when the apo structure was represented by 1hpx (KNI-272 removed), 1hhp, and 3hvp, respectively. These values agree well with the experimental value of -0.23 . For comparison, the values predicted with PROPKA2.0 were 0.18, 1.27, and 0.97, respectively.

Experimental protonation state changes as a function of pH for HIV protease complexed with the inhibitor KNI-272 are presented in Figure 8 along with calculated values. It is seen that calculations done with 1hpx representing the apo form underestimate the protonation state changes due to pepstatin binding, whereas better accordance with experimental results is achieved for calculations where the apo form is represented by the structure with the PDB id 3hvp.

For HIV protease complexed with DMP-323 (1qbs), the pK_a values of the catalytic dyad are predicted to be 5.88 and 8.52. This is in accord with the experimental result that the dyad has a pK_a value of 8.19.⁴⁹ While the experimental finding that both aspartic acids have a pK_a value above 7.2 is not reproduced, there is a better agreement than for the PROPKA2.0 values of 3.26 and 8.56.

In summary, a better agreement with the experimental overall protonation state change for the complexation of HIV protease with KNI-272 was achieved with PROPKA3.1 as compared to the previously reported PROPKA2.0 values. Furthermore, an

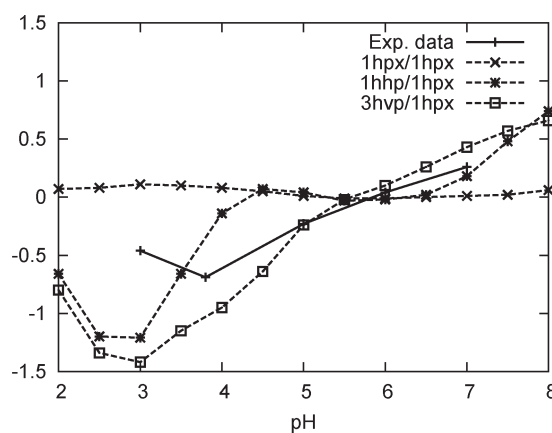


Figure 8. Protonation state change for HIV protease on complexation with the inhibitor KNI-272 as a function of pH. The predictions were done solely on the basis of the structure of the HIV protease–KNI-272 complex (1hpx) as well as in combination with HIV protease apo structures (1hhp and 3hvp).

improvement in the predicted pK_a values of the catalytic dyad was observed for HIV protease complexed with KNI-272 and DMP-323.

3.3.4. Ligands Covalently Bound to the Protein Receptor. We have applied PROPKA3.1 to chymotrypsin covalently bound, via its Ser195 residue, to two peptidyl trifluoromethyl ketone derived inhibitors. The oxygen in the ligand hemiketal adduct, formed during complexation from the ligand ketone and the serine hydroxyl, is negatively charged. For chymotrypsin, His57 undergoes a shift from 7.54 (chain F)/7.48 (chain I) to 9.46 (chain F)/9.41 (chain I) during complexation with the NAc–Phe–CF₃ inhibitor (6gch). This is in accordance with the experimental shift from 7.5 to 10.8. For the NAc–Leu–Phe–CF₃ inhibitor, the shift is predicted to be smaller (7.33/7.33 to 8.07/8.07 for chains F/I) in contrast to the experimental shift (7.5 to 12.0) but in agreement with the PROPKA2.0 shift (6.94 to 8.90).

For the covalent complexation of xylanase (1bv) via its Glu78 with 2-deoxy-2-fluoro- β -xylobioside (2FXb), a pK_a shift of Glu172 from 7.52 to 6.80 was predicted. While smaller than the experimental shift (6.7 to 4.2), it is in agreement with the PROPKA2.0 shift (7.45 to 6.64). Using the xylanase Asn35Asp mutant structure (1c5i), the pK_a shift of the Asp35/Glu172 dyad is predicted to be from 5.58/11.45 to 10.83/5.70. The Asp35/Glu172 dyad is correctly detected as being noncovalently coupled.

3.3.5. Ligands Containing Thiol Groups. The pK_a value of His235 of hydroxynitrile lyase has been found to shift from 2.5 to ~ 8 during binding of thiocyanate.⁵⁰ The PROPKA3.1 predictions of the pK_a value of His235 of 8.30 (holo form) and 6.72 (apo form) capture the direction of the shift and the experimental pK_a value in the holo form; however, the pK_a of the apo form is significantly higher than the experimental value. The nitrile group elevates the pK_a value of His235 by 0.41 pK_a units due to an intrinsic electrostatic interaction, and the thiol group elevates the pK_a value of His235 by 0.26 pK_a unit (intrinsic electrostatic interaction) and 1.05 pK_a units (Coulombic interaction).

3.3.6. Dihydrofolate Reductase. We apply PROPKA3.1 to dihydrofolate reductase complexed with methotrexate. NMR experiments have shown that the aromatic N1 nitrogen of

methotrexate is protonated in the complex up to pH values in excess of 10.^{51–54} In the complex, the N1 nitrogen forms a salt bridge with the deprotonated aspartic acid Asp27. PROPKA3.1 predicts the pK_a value of Asp27 to be 5.82 (experimental value 6.6) in the apo form and –0.76 in holo form. The pK_a value of the N1 nitrogen is predicted to shift from 5.7 to 5.97 during complexation. PROPKA3.1 does therefore not capture a large shift observed experimentally. PROPKA2.0 predicted the pK_a value of the N1 nitrogen in the holo form to be 6.89. Setting the model pK_a value of the N1 nitrogen to a calculated value of 4.4 does not significantly change the predicted shift in pK_a value.

4. CONCLUSION

In the preparation of PROPKA3.1, we have focused on improving the treatment of ligand molecules and coupling effects. It is our hope that the improvements will assist users in PROPKA calculations on complicated systems. The implemented algorithm for noncovalent coupling will automatically detect coupled groups and notify the user of the existence of alternative pK_a values. The implemented algorithm for covalent coupling improves the PROPKA treatment of complicated ligand molecules where titrational events are strongly linked. Furthermore, the algorithm will remove some of the ambiguity that arises when users manually set up complicated molecules for PROPKA calculations.

We have tested new implementations on a set of ligand-containing structures also used for the preparation of PROPKA2.0. Except for a few cases, the new algorithms in PROPKA3.1 were found to yield results similar to or better than those obtained with PROPKA2.0. With the improved treatment of complicated ligand molecules, pK_a calculation on structures containing DNA is now within reach, and this will be the goal of our continued efforts.

The PROPKA3.1 code is available from <http://propka.ki.ku.dk>.

■ ASSOCIATED CONTENT

Supporting Information. Three tables: Table S.1, a list of ligand interaction parameters; Table S.2, types of interactions (normal or iterative) used for all combinations of interacting groups; and Table S.3, Protein Data Bank structures used in this study. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: chresten@chem.ku.dk (C.R.S.), jhjensen@chem.ku.dk (J.H.J.).

■ ACKNOWLEDGMENT

Atomic coordinates for trypsin in complex with the ligands 1bMe, 1c, and 2 were kindly supplied by Paul Czodrowski and Gerhard Klebe. Atomic coordinates for thrombin in complex with the ligand 2 were kindly supplied by David Banner. This work was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 202167 and the Danish Council for Strategic Research through a research grant from the Program Commission on Strategic Growth Technologies (2106-07-0030).

■ REFERENCES

- (1) Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H.; Sastry, G. N. *Curr. Protein Pept. Sci.* **2007**, *8*, 329–351.
- (2) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.
- (3) Czodrowski, P.; Dramburg, I.; Sotriffer, C. A.; Klebe, G. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 424–437.
- (4) Lamotte-Brasseur, J.; Lounnas, V.; Raquet, X.; Wade, R. C. *Protein Sci.* **1999**, *8*, 404–409.
- (5) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 765–783.
- (6) Aguilar, B.; Anandakrishnan, R.; Ruscio, J. Z.; Onufriev, A. V. *Biophys. J.* **2010**, *98*, 872–880.
- (7) Olsson, M. H.; Søndergard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (8) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704–721.
- (9) Klingen, A. R.; Bombarda, E.; Ullmann, G. M. *Photochem. Photobiol. Sci.* **2006**, *5*, 588–596.
- (10) Søndergaard, C. R.; McIntosh, L. P.; Pollastri, G.; Nielsen, J. E. *J. Mol. Biol.* **2008**, *376*, 269–287.
- (11) McIntosh, L. P.; Hand, G.; Johnson, P. E.; Joshi, M. D.; Korner, M.; Plesniak, L. A.; Ziser, L.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **1996**, *35*, 9958–9966.
- (12) Qin, J.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1996**, *35*, 7–13.
- (13) Chivers, P. T.; Prehoda, K. E.; Volkman, B. F.; Kim, B.-M.; Markley, J. L.; Raines, R. T. *Biochemistry* **1997**, *36*, 14985–14991.
- (14) Jeng, M.; Holmgren, A.; Dyson, H. *Biochemistry* **1995**, *34*, 10101–10105.
- (15) van Rossum, G. The Python Language Reference (accessed Feb 14, 2009); <http://docs.python.org/py3k/reference/index.html>.
- (16) Williams, R. pK_a Data Compiled by R. Williams; research.chem.psu.edu/brpgrp/pKa_compilation.pdf.
- (17) Smith, M. Functional Groups. *Organic Chemistry: An Acid-Base Approach*; CRC Press: Boca Raton, FL, 2011; p 142.
- (18) Warshel, A. *Biochemistry* **1981**, *20*, 3167–3177.
- (19) ChemAxon, MarvinSketch 5.2.6 (accessed Jan 15, 2010), 2009; <http://www.chemaxon.com/products/marvin/marvinsketch/>.
- (20) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681–691.
- (21) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (22) Alexov, E. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 572–584.
- (23) Vriend, G. *J. Mol. Graphics* **1990**, *8*, 52–56.
- (24) Dullweber, F.; Stubbs, M. T.; Musil, D.; Stürzebecher, J.; Klebe, G. *J. Mol. Biol.* **2001**, *313*, 593–614.
- (25) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (26) Rostkowski, M.; Olsson, M.; Søndergaard, C.; Jensen, J. H. *BMC Struct. Biol.* **2011**, *11*, 6.
- (27) Velazquez-Campoy, A.; Luque, I.; Todd, M. J.; Milutinovich, M.; Kiso, Y.; Freire, E. *Protein Sci.* **2000**, *9*, 1801–1809.
- (28) Joule, J. A.; Mills, K. The Diazines: Pyridazine, Pyrimidine, and Pyrazine: Reactions and Synthesis. *Heterocyclic Chemistry*, 5th ed.; Wiley-Blackwell: Chichester, UK, 2010; p 254.
- (29) Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. *J. Med. Chem.* **2005**, *48*, 403–413.
- (30) Tari, L.; Matte, A.; Goldie, H.; Delbaere, L. *Nat. Struct. Biol.* **1997**, *4*, 990–994.
- (31) Czodrowski, P.; Sotriffer, C. A.; Klebe, G. *J. Mol. Biol.* **2007**, *367*, 1347–1356.
- (32) Fokkens, J.; Klebe, G. *Angew. Chem., Int. Ed.* **2006**, *45*, 985–989.
- (33) Bernstein, N. K.; Cherney, M. M.; Loetscher, H.; Ridley, R. G.; James, M. N. *Nat. Struct. Mol. Biol.* **1999**, *6*, 32–37.

- (34) Silva, A. M.; Lee, A. Y.; Gulnik, S. V.; Maier, P.; Collins, J.; Bhat, T. N.; Collins, P. J.; Cachau, R. E.; Luker, K. E.; Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Goldberg, D. E.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034–10039.
- (35) Lee, A. Y.; Gulnik, S. V.; Erickson, J. W. *Nat. Struct. Mol. Biol.* **1998**, *5*, 866–871.
- (36) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Hosur, M. V.; Sowder, R. C.; Cachau, R. E.; Collins, J.; Silva, A. M.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6796–6800.
- (37) Pearl, L.; Blundell, T. *FEBS Lett.* **1984**, *174*, 96–101.
- (38) Fitzgerald, P. M.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C. T.; Herber, W. K.; Dixon, R. A.; Darke, P. L. *J. Biol. Chem.* **1990**, *265*, 14209–14219.
- (39) Spinelli, S.; Liu, Q.; Alzari, P.; Hirel, P.; Poljak, R. *Biochimie* **1991**, *73*, 1391–1396.
- (40) Wlodawer, A.; Miller, M.; Jaskólski, M.; Sathyanarayana, B.; Baldwin, E.; Weber, I.; Selk, L.; Clawson, L.; Schneider, J.; Kent, S. *Science* **1989**, *245*, 616–21.
- (41) Baldwin, E.; Bhat, T.; Gulnik, S.; Liu, B.; Topol, I.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. *Structure* **1995**, *3*, 581–590.
- (42) Lam, P. Y. S.; et al. *J. Med. Chem.* **1996**, *39*, 3514–3525.
- (43) Brady, K.; Wei, A.; Ringe, D.; Abeles, R. *Biochemistry* **1990**, *29*, 7600–7607.
- (44) Sidhu, G.; Withers, S.; Nguyen, N.; McIntosh, L.; Ziser, L.; Brayer, G. *Biochemistry* **1999**, *38*, 5346–5354.
- (45) Sun, Z.; Dotsch, V.; Kim, M.; Li, J.; Reinherz, E.; Wagner, G. *EMBO J.* **1999**, *18*, 2941–2949.
- (46) Zuegg, J.; Gruber, K.; Gugganig, M.; Wagner, U.; Kratky, C. *Protein Sci.* **1999**, *8*, 1990–2000.
- (47) Bolin, J.; Filman, D.; Matthews, D.; Hamlin, R.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650–13662.
- (48) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. *Nat. Struct. Mol. Biol.* **1996**, *3*, 946–950.
- (49) Yamazaki, T.; Nicholson, L. K.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Torchia, D. A. *J. Am. Chem. Soc.* **1994**, *116*, 10791–10792.
- (50) Stranzl, G. R.; Gruber, K.; Steinkellner, G.; Zangger, K.; Schwab, H.; Kratky, C. *J. Biol. Chem.* **2004**, *279*, 3699–3707.
- (51) Cocco, L.; Groff, J. P.; Temple, C.; Montgomery, J. A.; London, R. E.; Matwiyoff, N. A.; Blakley, R. L. *Biochemistry* **1981**, *20*, 3972–3978.
- (52) Cocco, L.; Roth, B.; Temple, C.; Montgomery, J. A.; London, R. E.; Blakley, R. L. *Arch. Biochem. Biophys.* **1983**, *226*, 567–577.
- (53) Stone, S.; Morrison, J. *Biochim. Biophys. Acta* **1983**, *745*, 247–258.
- (54) London, R.; Howell, E.; Warren, M.; Kraut, J.; Blakley, R. *Biochemistry* **1986**, *25*, 7229–7235.

A Direct Comparison of the MM-GB/SA Scoring Procedure and Free-Energy Perturbation Calculations Using Carbonic Anhydrase as a Test Case: Strengths and Pitfalls of Each Approach

Cristiano R. W. Guimarães*

Worldwide Medicinal Chemistry Department, Pfizer Inc., 558 Eastern Point Rd, Groton, Connecticut 06340, United States

ABSTRACT: MM-GB/SA scoring and free energy perturbation (FEP) calculations have emerged as reliable methodologies to understand structural and energetic relationships to binding. In spite of successful applications to elucidate the structure–activity relationships for few pairs of ligands, the reality is that the performance of FEP calculations has rarely been tested for more than a handful of compounds. In this work, a series of 13 benzene sulfonamide inhibitors of carbonic anhydrase with binding free energies determined by isothermal titration calorimetry was selected as a test case. R^2 values of 0.70, 0.71, and 0.49 with the experiment were obtained with MM-GB/SA and FEP simulations run with MCPRO+ and Desmond, respectively. All methods work well, but the results obtained with Desmond are inferior to MM-GB/SA and MCPRO+. The main contrast between the methods is the level of sampling, ranging from full to restricted flexibility to single conformation for the complexes in Desmond, MCPRO+, and MM-GB/SA, respectively. The current and historical results obtained with MM-GB/SA qualify this approach as a more attractive alternative for rank-ordering; it can achieve equivalent or superior predictive accuracy and handle more structurally dissimilar ligands at a fraction of the computational cost of the rigorous free-energy methods. As for the large theoretical dynamic range for the binding energies, that seems to be a direct result of the degree of sampling in the simulations since MCPRO+ as well as MM-GB/SA are plagued by this. Van't Hoff analysis for selected pairs of ligands suggests that the wider scoring spread is not only affected by missing entropic contributions due to restricted sampling but also exaggerated enthalpic separation between the weak and potent compounds caused by diminished shielding of electrostatic interactions, thermal effects, and protein relaxation/strain.

INTRODUCTION

Computational methodologies to understand structural and energetic relationships to binding vary in speed and accuracy. Molecular dynamics (MD) and Monte Carlo (MC) simulations coupled with free-energy perturbation (FEP) or thermodynamic integration (TI) calculations are considered the reference computational approaches when it comes to estimating relative binding affinities.^{1–5} Although the FEP and TI equations are exact, the results are plagued by (i) force fields that incompletely or incorrectly describe the protein, the ligand, and their interactions; (ii) limited sampling of the phase space; and (iii) challenges in accurately taking into account changes in hydration.^{6–8} In spite of successful applications to elucidate the structure–activity relationships for a few pairs of ligands in a congeneric series, the reality is that the performance of FEP and TI calculations has rarely been tested for more than a handful of compounds due to the high computational cost, with a few exceptions that are noteworthy.⁹

The combination of molecular mechanics and the Poisson–Boltzmann continuum solvation to compute binding free energies was pioneered by Kollman and Kuhn.¹⁰ The encouraging results obtained with this methodology inspired several authors to use molecular-mechanics-based scoring functions with the generalized Born model¹¹ as the implicit solvent (MM-GB/SA) in the rescoring of docking poses. When compared to docking scoring functions, the MM-GB/SA procedure provides improved enrichment in the virtual screening of databases and better correlation between calculated binding affinities and experimental data.¹² The MM-GB/SA rescoring method, however, performs poorly in some cases, suggesting that success through its application may be system- and/or protocol-dependent.¹³

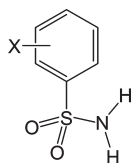
We investigated the performance of our own flavor of MM-GB/SA when rescoring docking poses of congeneric series for the pharmaceutically relevant targets CDK2, Factor Xa, Thrombin, and HIV-RT.¹⁴ The correlation with experimental results obtained with the physics-based scoring was far superior to the one obtained with the Glide XP scoring function.¹⁵ More recently, we addressed the poor estimation of protein desolvation provided by the GB/SA solvation model;¹⁶ replacing this term with the free energy associated with displacing binding-site waters upon ligand binding estimated by the WaterMap method,¹⁷ which treats the solvent explicitly, provides superior results. However, the improvement is modest over results obtained with the MM-GB/SA version that excludes the protein GB/SA desolvation term. This is apparently due to the high correlation between the free energy liberation of the displaced solvent and the protein–ligand van der Waals interactions, which in turn may be interpretable as estimates of the hydrophobic effect and hydrophobic-like interactions, respectively.

The remarkable results obtained with MM-GB/SA^{14,16} warrant a head-to-head comparison with the more rigorous but computationally intensive FEP and TI methodologies to address whether the approximate free-energy method provides a more attractive alternative for rank-ordering. Although the MM-GB/SA approach handles more structurally dissimilar ligands and provides results at a fraction of the computational cost than FEP and TI, it is not clear if it can achieve equivalent predictive accuracy. To accomplish that, a subset with 13 ligands of a series of benzene sulfonamide (BSA) inhibitors of carbonic anhydrase

Received: April 8, 2011

Published: June 02, 2011

Table 1. Isothermal Titration Calorimetry (ITC) Data for the Binding between the Human Carbonic Anhydrase (hCAII) and a Series of Benzene Sulfonamide (BSA) Inhibitors^a



R	ΔG_{bind}^b	$\text{p}K_a^b$	IP^c	$c\Delta G_{\text{bind}}^d$
H	-8.3	10.1	3.5	-11.8
2-F	-8.8	9.6	2.9	-11.7
3-F	-9.4	9.7	3.0	-12.4
4-F	-8.4	10.0	3.4	-11.8
2-Cl	-8.1	9.5	2.7	-10.8
3-Cl	-8.5	9.5	2.7	-11.2
4-Cl	-9.3	9.9	3.3	-12.6
2-CH ₃	-7.9	10.0	3.4	-11.3
3-CH ₃	-8.1	10.2	3.7	-11.8
4-CH ₃	-8.7	10.3	3.8	-12.5
2-NH ₂	-8.2	10.0	3.4	-11.6
3-NH ₂	-8.1	10.0	3.4	-11.5
4-NH ₂	-7.8	10.5	4.1	-11.9

^a Energy values are in kcal/mol. ^b Ref 18. ^c Ionization penalty $\rightarrow \text{IP} = RT \ln(1 + 10^{(\text{p}K_a - \text{pH})})$. ^d $c\Delta G_{\text{bind}} = \Delta G_{\text{bind}} - RT \ln(1 + 10^{(\text{p}K_a - \text{pH})})$.

with binding free energies determined by isothermal titration calorimetry (ITC) was selected (Table 1).¹⁸ This subset is attractive since the degree of similarity is ideal for the FEP and TI techniques; it involves “walks” of typical one-heavy atom substituents around a benzene ring. A negative aspect of this series is the narrow dynamic range for the experimental binding free energies (Table 1).

In this work, relative binding free energies ($\Delta\Delta G_{\text{bind}}$) for the BSA series were calculated by MM-GB/SA and FEP and compared to the free energies obtained with ITC; the original MM-GB/SA implementation¹⁴ and the implementation augmented by Watermap¹⁶ were used. In addition, the large dynamic range observed in the MM-GB/SA scoring compared to the experimental range was revisited. To accomplish that, FEP simulations with different degrees of sampling were performed since this problem resides in the use of a single, relaxed structure for the complex in MM-GB/SA.¹⁶ Finally, a computational Van't Hoff analysis was conducted for selected pairs of ligands to further dissect the origin of the large dynamic range, e.g., enthalpic, entropic, or a combination of both.

METHODS

System. The crystal structure between the human carbonic anhydrase (hCAII) and the unsubstituted BSA analog was used (PDB ID: 2WEJ; Figure 1).¹⁸ The starting structure was refined through a series of restrained, partial minimizations using the OPLS_2005 force field.^{19,20} A view of the binding site reveals that the zinc ion (Zn^{2+}) is coordinated to three histidine residues and the deprotonated nitrogen of the sulfonamide group. This group is not particularly acidic, as shown by the measured $\text{p}K_a$'s in Table 1, so the ITC ΔG_{bind} values have to be corrected by the application of eq 1. The second term represents the ionization penalty (IP) to charge a molecule whose $\text{p}K_a$ value is greater than the pH of the

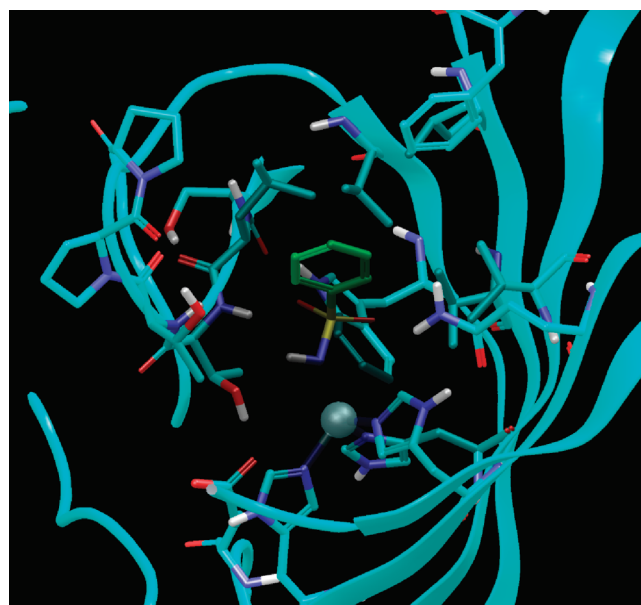


Figure 1. The crystal structure between the human carbonic anhydrase (hCAII) and the unsubstituted benzene sulfonamide inhibitor (PDB ID: 2WEJ).

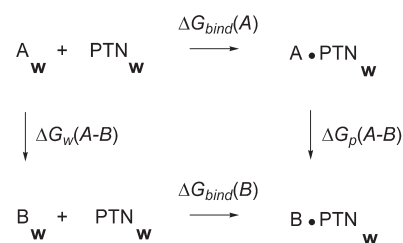


Figure 2. Thermodynamic cycle used for the calculation of relative binding free energies ($\Delta\Delta G_{\text{bind}}$). ΔG_{bind} is the absolute binding free energy. ΔG_w and ΔG_p are the free energy changes for the transformation of ligand A into B in water and in the solvated complex.

experiment, which was set to 7.5. This correction is necessary since the simulations were carried out for the inhibitors in their deprotonated state.

$$c\Delta G_{\text{bind}} = \Delta G_{\text{bind}} - RT \ln(1 + 10^{(\text{p}K_a - \text{pH})}) \quad (1)$$

FEP Simulations. Figure 2 illustrates the thermodynamic cycle used to calculate free energy changes.²¹ Since free energy is a thermodynamic state function, eq 2 can be derived from the cycle and used to calculate $\Delta\Delta G_{\text{bind}}$, where A and B are any two analogs, ΔG_{bind} is the absolute binding free energy, and ΔG_w ($A \rightarrow B$) and ΔG_p ($A \rightarrow B$) are the free energies associated with the transformation of A into B in water and in the solvated complex, computed by the FEP method.

$$\begin{aligned} \Delta\Delta G_{\text{bind}} &= \Delta G_{\text{bind}}(\text{B}) - \Delta G_{\text{bind}}(\text{A}) \\ &= \Delta G_p(\text{A} \rightarrow \text{B}) - \Delta G_w(\text{A} \rightarrow \text{B}) \end{aligned} \quad (2)$$

FEP simulations were performed for the compounds in Table 1 using the unsubstituted analog as the reference state. The MC and

Table 2. λ Schedules Used in the FEP Simulations Run with Desmond

state	1	2	3	4	5	6	7	8	9	10	11	12
VDW A	1.00	1.00	1.00	1.00	1.00	0.67	0.46	0.33	0.25	0.19	0.12	0.00
VDW B	0.00	0.12	0.19	0.25	0.33	0.46	0.67	1.00	1.00	1.00	1.00	1.00
Coulomb A	1.00	0.75	0.50	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Coulomb B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.50	0.75	1.00
Bonded A	1.00	0.91	0.82	0.73	0.64	0.55	0.45	0.36	0.27	0.18	0.09	0.00
Bonded B	0.00	0.09	0.18	0.27	0.36	0.45	0.55	0.64	0.73	0.82	0.91	1.00

MD sampling techniques as implemented in MCPRO+ and Desmond were used.^{22,23} No effort has been made to make the two codes as comparable as possible; the recommended options for each piece of software were used. The goal was to evaluate the performance of each FEP method from a nonexpert user's point of view. However, it should be noted that the default settings have been carefully determined by the experts in the field. These settings were obtained with not only accuracy but also computational time in mind, so that a $\Delta\Delta G_{\text{bind}}$ prediction between two compounds can be provided within 1 to 2 days when multi-CPU processing is available.

The degrees of freedom for the protein backbone atoms were not sampled in the MC simulations. The only protein degrees of freedom allowed to vary during the simulation were the bond angles and dihedral angles for the side chains of residues with any atom within 10 Å from the ligands. The ligands, however, are fully flexible in the MC simulations. In the MD simulations, all degrees of freedom are sampled, except for the bonds that are formed by a heavy atom and a hydrogen atom; those were constrained by the Shake algorithm.²⁴

Charge neutrality for the protein systems in MCPRO+ was imposed by assigning normal protonation states at physiological pH to basic and acidic residues near the active site and making the adjustments for neutrality to the most distant residues. The complexes and the ligands in solution were solvated with a 22-Å-radius water cap. A half-harmonic potential with a force constant of 1.5 kcal/mol·Å² was applied to water molecules at distances greater than 22 Å from the center of the system to discourage evaporation. Residue-based nonbonded cutoffs of 10 Å were employed. In Desmond, charge neutrality was achieved by the addition of a minimal concentration of ions. The complexes and the ligands in solution were placed in a cubic box using periodic boundary conditions with buffer regions of 5 Å and 10 Å, respectively. Short-range van der Waals (VDW) and near electrostatic nonbonded interactions were computed by summing them over all pairs within 9 Å of each other. Far electrostatic interactions were computed using the smooth particle mesh Ewald method.²⁵

The A→B transformations in MCPRO+ were performed using the single topology approach by melding the force field parameters for bond lengths, bond angles, torsions, and nonbonded interactions. In order to keep the number of atoms constant, dummy atoms were introduced for hydrogens that exist in one state and have no counterpart in the other. Desmond performs the transformations using the dual topology approach, where both ligands A and B are simultaneously simulated, but they do not interact with each other.⁷ The free energy changes in MCPRO+ and Desmond employed the double wide sampling and the Bennett acceptance ratio²⁶ methods, respectively. Other important differences between MCPRO+ and Desmond reside in the use by the latter of a softcore potential for VDW interactions and a decoupled λ schedule. The softcore potential is introduced to avoid the so-called van der Waals end point problems when an atom is being created or

annihilated.²⁷ As for the λ schedule, the force field parameters are mutated concomitantly in MCPRO+ using 20 windows with λ values evenly distributed between 0 and 1. In Desmond, the bonded, VDW, and Coulomb interactions for ligands A and B have independent λ schedules, as shown in Table 2.

The A→B transformations were executed at 298 K. The Van't Hoff plots for selected pairs of ligands were obtained by also running the transformations at 298 ± 30 K, the reference temperature. In MCPRO+, initial relaxation of the solvent was performed for 5 × 10⁶ configurations, followed by 10 × 10⁶ configurations of full equilibration and 20 × 10⁶ configurations of data collection for each window in water or in the complex. Established procedures including Metropolis and preferential sampling were employed. Attempted moves of the ligands in water occurred every 60 configurations, while in the complex attempted moves of the protein systems and ligand analogs occurred every 10 and 60 configurations, respectively. In Desmond, the complex and the ligand went through a relaxation process that includes two energy minimizations, with and without solute restraints, followed by four short MD simulations, the first at 10 K in the NVT ensemble, the second also at 10 K but in the NPT ensemble, and the last two at 298 K in the NPT ensemble. The production simulation, also in the NPT ensemble, was run for 0.6 ns for each λ window. The temperature and pressure in the MD simulations were controlled by the Berendsen thermostat and barostat algorithms.²⁸ The TIP4P²⁹ and SPC water models³⁰ were used in MCPRO+ and Desmond, respectively, while the OPLS_2005 force field^{19,20} was employed for both.

To account for the two possible orientations in the binding site for the unsymmetrical ligands, where substituents at the 2 and 3 positions of the phenyl ring may be oriented to the left (L) or right (R) using the pose in Figure 1 as a reference, the free energies of binding of each orientation, $\Delta\Delta G_{\text{bind}}^{\text{L}}$ and $\Delta\Delta G_{\text{bind}}^{\text{R}}$, were combined to produce an overall $\Delta\Delta G_{\text{bind}}$ using eq 3, where R is the ideal gas constant and T is the temperature. The second term in eq 3 penalizes the computed free energies of binding of the unsymmetrical ligands by $RT \ln 2$ because they are relative to the unsubstituted, symmetrical ligand. Thus, when the relative free energies of binding of the two orientations differ by greater than *ca.* 2 kcal/mol, the free energy of binding is essentially that of the more favorable orientation plus $RT \ln 2$. Alternatively, if the relative free energies of binding of the two orientations are the same, the $RT \ln 2$ penalty is removed.^{19b}

$$\Delta\Delta G_{\text{bind}} = -RT \ln \left[\exp \left(-\frac{\Delta\Delta G_{\text{bind}}^{\text{L}}}{RT} \right) + \exp \left(-\frac{\Delta\Delta G_{\text{bind}}^{\text{R}}}{RT} \right) \right] + RT \ln 2 \quad (3)$$

MM-GB/SA Rescoring. In our implementation of the MM-GB/SA rescoring (Figure 3), described in detail in refs 14 and 16,

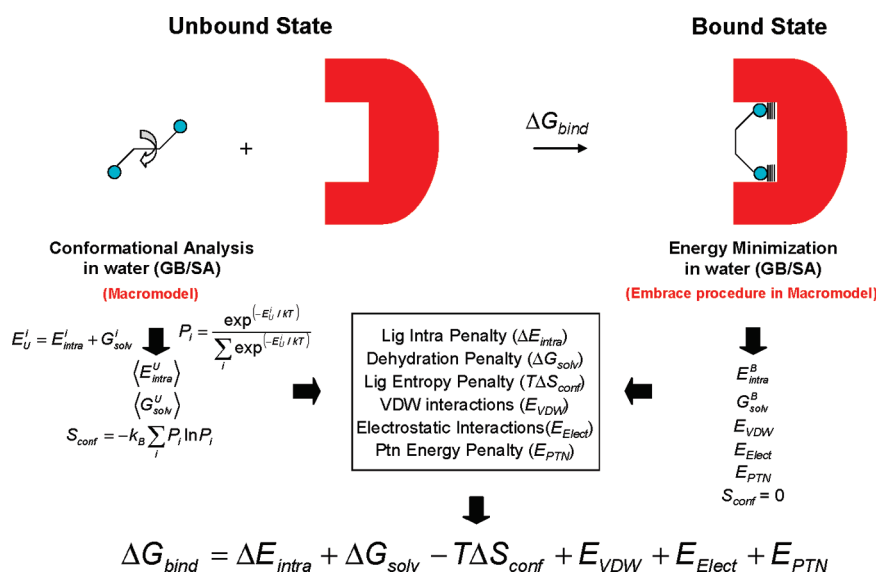


Figure 3. Schematic representation of the MM-GB/SA rescoring procedure with no protein desolvation term.

a conformational search for the inhibitors in the unbound state and energy minimization for the complexes using OPLS_2005 and GB/SA within MacroModel³¹ are performed. All conformers within 5.0 kcal/mol from the lowest energy conformer were retained. Assuming a Boltzmann distribution, the probability for each conformer (P_i) was calculated and used to compute the Boltzmann-averaged intramolecular energy and solvation free energy in the unbound state for each compound. The conformational entropies (S_{conf}) were computed from the probabilities using eq 4, where k_B is the Boltzmann constant.

$$S_{conf} = -k_B \sum_{i=1}^n p_i \ln p_i \quad (4)$$

To better account for the protein flexibility, each inhibitor was energy-minimized in the bound state. The preferred orientations for the unsymmetrical ligands were determined by the FEP simulations. In the energy minimization, no constraints were applied to residues within 5 Å from the center of the system. A second shell of 3 Å around the first shell was defined and constraints of 50 kcal/mol·Å² applied to the residues therein. The remaining residues were held fixed. After the energy minimization step, the protein energy (E_{PTN}) values for all complexes were extracted. This term describes the protein deformation or strain imposed by each ligand. Besides E_{PTN} , the energy-minimized structures for the complexes provided the intramolecular energies and solvation free energies for the ligands in the protein environment and the protein–ligand intermolecular van der Waals (E_{VDW}) and electrostatic (E_{Elect}) interaction energies. In the bound state, it was assumed that there was only one conformation accessible to each ligand; its conformational entropy is therefore zero. In this manner, ΔG_{bind} estimated by MM-GB/SA was calculated as shown in eq 5.

$$\Delta G_{bind} = \Delta E_{intra} + \Delta G_{solv} - T\Delta S_{conf} + E_{VDW} + E_{Elect} + E_{PTN} \quad (5)$$

In eq 5, ΔE_{intra} and ΔG_{solv} are the intramolecular strain and desolvation penalty for each ligand upon binding. Similarly, $-T\Delta S_{conf}$ is the ligand conformational entropy penalty, multiplied by the temperature to convert it into energy. The final ranking was obtained by

calculating relative binding energies, $\Delta\Delta G_{bind}$, using the top-scoring inhibitor as a reference.

In our MM-GB/SA implementation, although solvent effects are included in the protein–ligand complex geometry optimization using GB/SA, the protein desolvation term calculated by the continuum model (ΔG_{solv}^{ptn}), defined as illustrated in Figure 4, is generally excluded from the scoring since it deteriorates the correlation with experimental data.^{14,16} The solvent shielding of protein–ligand electrostatic interactions estimated by the GB model (E_{GB}) as well as rotational, translational, and vibrational entropy changes for the ligand upon binding using the rigid-rotor harmonic oscillator (RRHO) approximation are also excluded, as they have no significant impact on the MM-GB/SA results. Here, we decided to investigate the impact of these contributions on the results for the BSA series in hCAII since the degree of similarity for the ligands (Table 1) is greater than the typical congeneric series previously evaluated.^{14,16}

WaterMap. In WaterMap, the reference state consists of an assumed-to-pre-exist cavity in solution formed to accommodate the ligand. In order to accommodate ligand binding, the binding site waters get displaced to the cavity in solution, leaving a cavity of identical size and shape in the protein (Figure 5). The displaced-solvent functional in WaterMap represents an attempt to estimate the free energy liberation (ΔG_{WM}) for the binding site waters into bulk solution upon cavity transfer between the two environments. This functional depends on the degree of overlap between the ligand heavy atoms and the hydration sites and the energetics of the waters that are displaced. Specifically, the functional considers that a water molecule is completely displaced, and therefore its full energy is liberated when the distance between the hydration site center and the ligand heavy atom approaches zero. The energy of hydration site displacement then decreases linearly to a value of zero when the distance between the two atoms is equal to 80% of the sum of their VDW radii, beyond which there is no displacement. Multiple ligand atoms may contribute to the displacement of a given hydration site; however, these contributions cease once total displacement is achieved. The *ab initio* form of the displaced-solvent functional as described by Abel and co-workers was employed in this

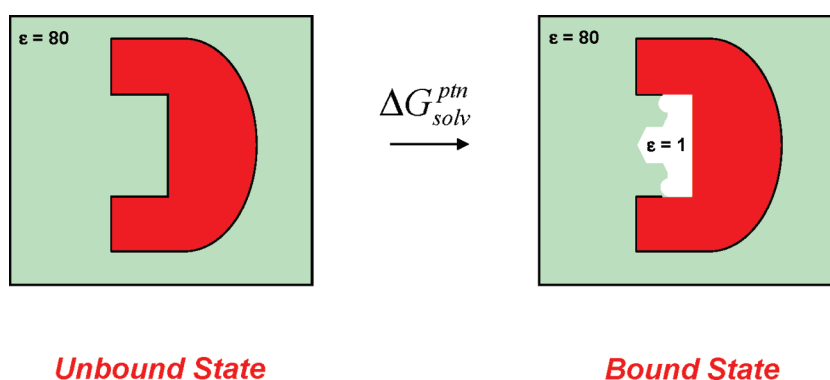


Figure 4. Schematic representation of the protein desolvation term ($\Delta G_{\text{solv}}^{\text{ptn}}$) described by GB/SA. The white area represents the vacuum region in the protein binding site that will be occupied by a ligand of that shape and size.

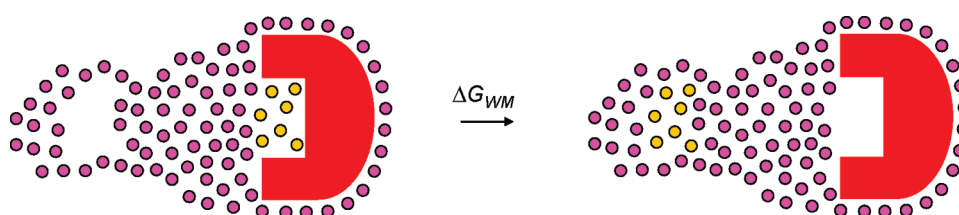


Figure 5. Schematic representation of the process simulated by WaterMap. The white area represents the cavity in the bulk that is transferred to the protein binding site. The orange dots represent the binding site waters that get displaced into the bulk solution. WaterMap estimates the free energy liberation (ΔG_{WM}) for the displaced waters.

work.¹⁷ The thermodynamic and structural properties of theoretical water molecules in the binding sites of hCAII were obtained from MD simulations of 10 ns each using WaterMap.³² The protein heavy atoms were harmonically restrained to their starting coordinates. The ligands were scored using the poses extracted from the energy-minimized complexes.

RESULTS AND DISCUSSION

MM-GB/SA Scoring. The orientations for the unsymmetrical ligands were determined by the FEP simulations. MCPRO+ and Desmond agree in all cases for the 2- and 3-substituted analogs. Specifically, the 2-F, 2-NH₂, and 3-NH₂ analogs adopt the L orientation, where they face a more polar region of the binding site and hydrogen bonds may be established. The 3-F, 2-Cl, 3-Cl, 2-CH₃, and 3-CH₃ analogs adopt the R orientation; in this orientation, they point the more apolar substituents toward the hydrophobic wall of the deep conical-cleft-binding site of hCAII. The FEP predictions for 2-F, 3-F, and 2-Cl agree with the orientations observed in their crystal structures.¹⁸

Table 3 summarizes the impact of different contributions on the results obtained with the MM-GB/SA scoring procedure. The correlation (R^2) between the $\Delta\Delta G_{\text{bind}}$ values calculated by eq 5 and the experimental data, corrected as shown in eq 1, is 0.51. The addition of rotational, translational, and vibrational entropy changes for the ligand upon binding using the RRHO approximation plus the E_{GB} term, the solvent shielding of protein–ligand electrostatic interactions, marginally improves the results. Interestingly in this case, the protein desolvation term, $\Delta G_{\text{solv}}^{\text{ptn}}$ improves the correlation with experimental results, as can be seen when rows 2 and 3 in Table 3 are compared to 4 and 6, respectively. This term is generally excluded from our

Table 3. Correlation (R^2) between the Experimental Data, Corrected As Shown in eq 1, and the MM-GB/SA Scoring Approach Including Different Contributions

	R^2
MM-GB/SA ^a	0.51
MM-GB/SA + RRHO ^b	0.53
MM-GB/SA + RRHO + E_{GB} ^c	0.54
MM-GB/SA + RRHO + $\Delta G_{\text{solv}}^{\text{ptn}}$ ^d	0.60
MM-GB/SA + RRHO + ΔG_{WM} ^e	0.59
MM-GB/SA + RRHO + E_{GB} + $\Delta G_{\text{solv}}^{\text{ptn}}$	0.60
MM-GB/SA + RRHO + E_{GB} + ΔG_{WM}	0.53

^a MM-GB/SA scoring using eq 5. ^b Including rotational, translational, and vibrational entropy changes for the ligand upon binding as estimated by the rigid-rotor harmonic oscillator (RRHO) approximation. ^c Including the solvent shielding of the protein–ligand electrostatic interactions (E_{GB}). ^d Including the protein GB/SA desolvation penalty upon binding ($\Delta G_{\text{solv}}^{\text{ptn}}$). ^e Including the free energy liberation for the binding site waters into bulk solution (ΔG_{WM}).

scoring procedure, as it deteriorates the results.^{14,16} One possible explanation is that errors in the continuum solvation model are more likely to be canceled when there is a higher degree of similarity within the series being scored, which is case with the BSA series.

The WaterMap method alone does not perform particularly well with an R^2 value of 0.28. This contrasts with results obtained for FactorXa and CDK2, where R^2 values of 0.71 and 0.68 were obtained.¹⁶ This is not surprising as the ligands in the BSA series are much more similar in size, indicating that the affinity differences are driven by contributions other than the hydrophobic effect. For example, WaterMap makes no differentiation between

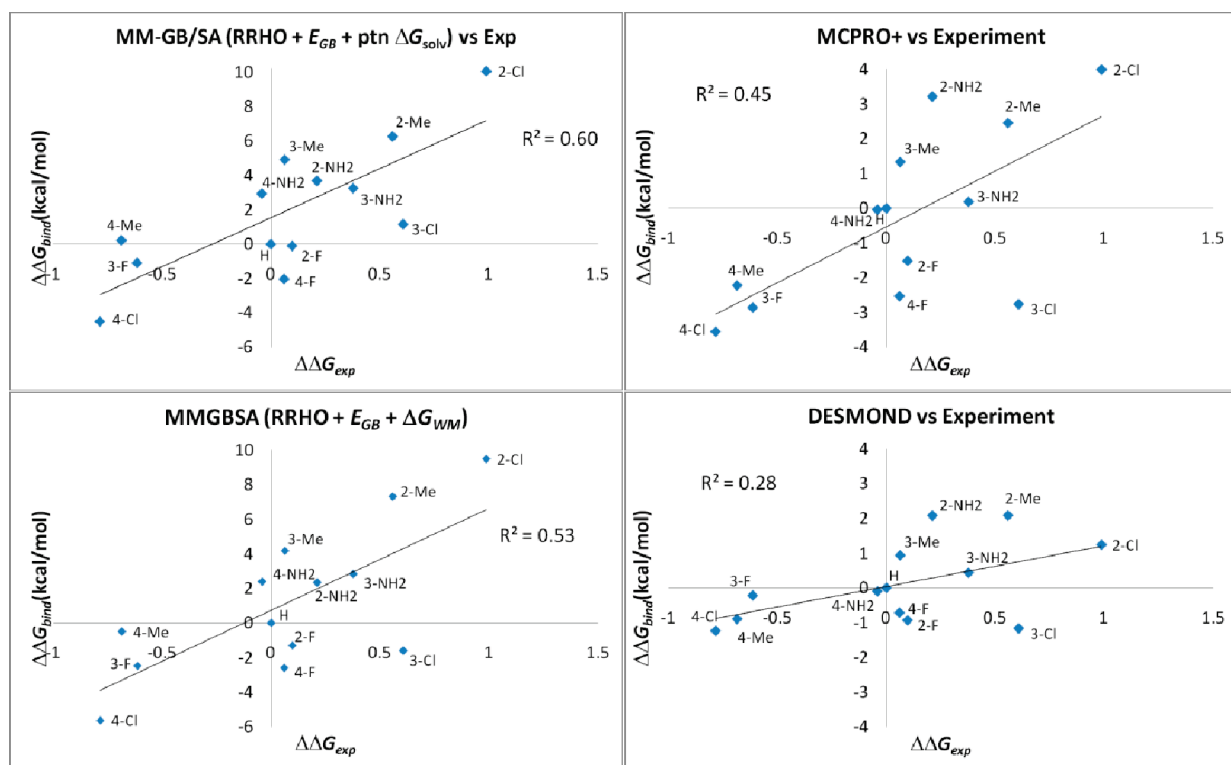


Figure 6. Correlation between isothermal titration calorimetry $\Delta\Delta G_{\text{bind}}$ values for the human carbonic anhydrase inhibitors, corrected as shown in eq 1, and MM-GB/SA (MM-GB/SA + RRHO + E_{GB} + $\Delta G_{\text{solv}}^{\text{ptn}}$), MM-GB/SA combined with WaterMap (MM-GB/SA + RRHO + E_{GB} + ΔG_{WM}), and FEP simulations run with MCPRO+ and Desmond.

the 2-Cl and 2-CH₃ derivatives, as they have identical binding modes and displace essentially the same water molecules upon binding; the $c\Delta G_{\text{bind}}$ values in Table 1 for these analogs, however, are -10.8 and -11.3 kcal/mol, respectively. The same is true for the 4-substituted subset with $c\Delta G_{\text{bind}}$ values of -11.3 , -12.6 , -12.5 , and -11.9 kcal/mol for F, Cl, CH₃, and NH₂; WaterMap provides identical scores to all of them. In spite of that, adding the free energy liberation for the displaced solvent, ΔG_{WM} , to the MM-GB/SA score is either beneficial, when the E_{GB} term is not included (rows 2 and 5), or neutral, in the case that this term is included (rows 3 and 7). Overall, our previous accounts^{14,16} and the results in Table 3 suggest that the terms in eq 5 contain most of the information needed to describe the binding process for congeneric series and that the addition of other contributions provides only incremental improvements to the scoring equation.

MM-GB/SA Scoring versus FEP Simulations. Figure 6 illustrates the correlations with the experimental results obtained with MCPRO+, Desmond, and MM-GB/SA. In the case of MM-GB/SA, the analysis is focused on the equation that includes the RRHO, E_{GB} , and $\Delta G_{\text{solv}}^{\text{ptn}}$ contributions, as it provides the best correlation with the ITC data. The scoring version where $\Delta G_{\text{solv}}^{\text{ptn}}$ is substituted by ΔG_{WM} is also included for comparison.

Superior correlation is obtained with the MM-GB/SA methods, followed by MCPRO+ and Desmond FEP simulations. This cannot necessarily be attributed to a lack of convergence in the MCPRO+ and Desmond simulations, as the calculated values obtained with different levels of sampling agree with each other (Figure 7). Desmond and MCPRO+ $\Delta\Delta G_{\text{bind}}$ values are highly correlated with an R^2 of 0.85. Both FEP methods, in particular MCPRO+, are also correlated with the results provided by the MM-GB/SA method, in spite of the particularities of each approach,

such as explicit versus implicit solvent treatments, which could lead to discrepancies in the free-energy values; R^2 values of 0.81 and 0.61 are observed when the MM-GB/SA scores are plotted against MCPRO+ and Desmond $\Delta\Delta G_{\text{bind}}$ values, respectively. It is interesting to see that MCPRO+, with restricted sampling, behaves almost like an interpolation between two methodological extremes, characterized by MM-GB/SA where a single, energy-minimized structure for each complex is used and Desmond where full sampling is performed.

The fact that the calculated $\Delta\Delta G_{\text{bind}}$ values obtained with the different methods are in closer agreement with each other than they are with the experimental data points to the existence of common outliers. A close inspection of Figure 6 reveals that the 3-Cl derivative is an outlier for all methods, in particular the ones where explicit solvation is used, indicating that the problem with this compound might lie in inaccuracies when simulating its change in hydration upon binding or a more fundamental problem with the energy function or with the experimental data. When the 3-Cl analog is removed, the correlation improves for all methods (Figure 8); R^2 values change from 0.60, 0.53, 0.45, and 0.28 to 0.70, 0.72, 0.71, and 0.49 for MM-GB/SA, MM-GB/SA combined with WaterMap, and FEP simulations run with MCPRO+ and Desmond, respectively. All methods work well, especially considering the narrow experimental dynamic range of 1.8 kcal/mol, but the results obtained with Desmond are inferior to MM-GB/SA and MCPRO+. It is possible that the full flexibility for the systems in Desmond actually adds more noise than information due to the sampling of nonrelevant conformations and/or sampling of relevant conformations that are not adequately weighted in the simulations. The restricted sampling in MCPRO+, besides reducing computational cost, aims at

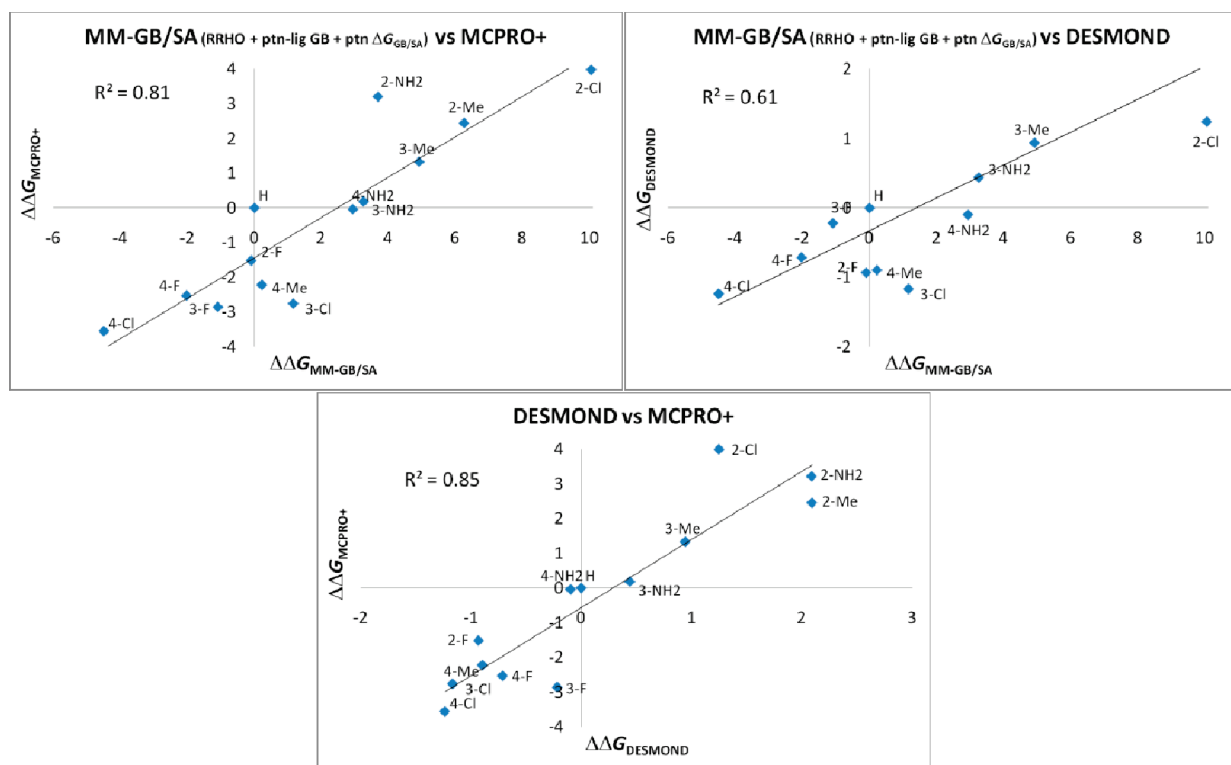


Figure 7. Correlation among MM-GB/SA (MM-GB/SA + RRHO + E_{GB} + ΔG_{solv}^{ptn}) and FEP simulations run with MCPRO+ and Desmond.

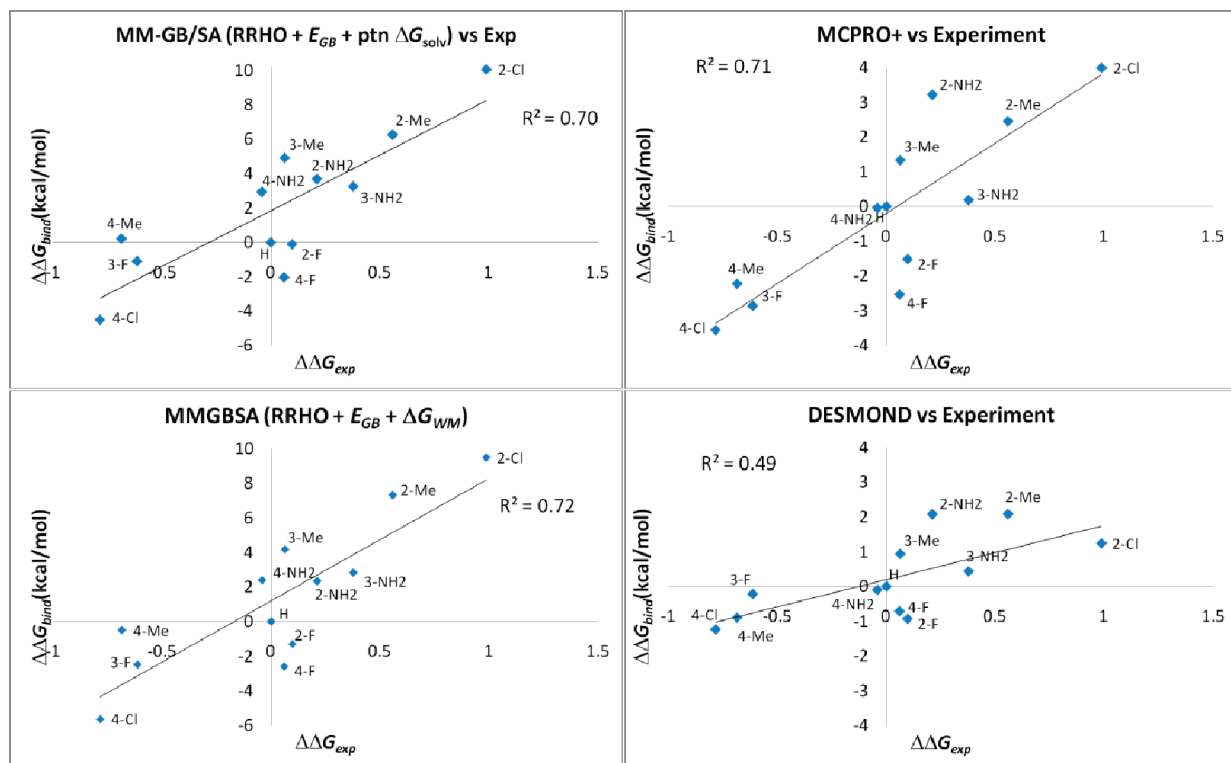


Figure 8. Correlation between isothermal titration calorimetry $\Delta\Delta G_{bind}$ values for the human carbonic anhydrase inhibitors, corrected as shown in eq 1, and MM-GB/SA (MM-GB/SA + RRHO + E_{GB} + ΔG_{solv}^{ptn}), MM-GB/SA combined with WaterMap (MM-GB/SA + RRHO + E_{GB} + ΔG_{WM}), and FEP simulations run with MCPRO+ and Desmond. The 3-Cl analog is removed from the plots.

focusing the MC simulations on crystal-structure-like conformations for the complexes as the protein backbone atoms are

untouched during the MC simulations. The assumption in MM-GB/SA, that one conformation captures the essence of

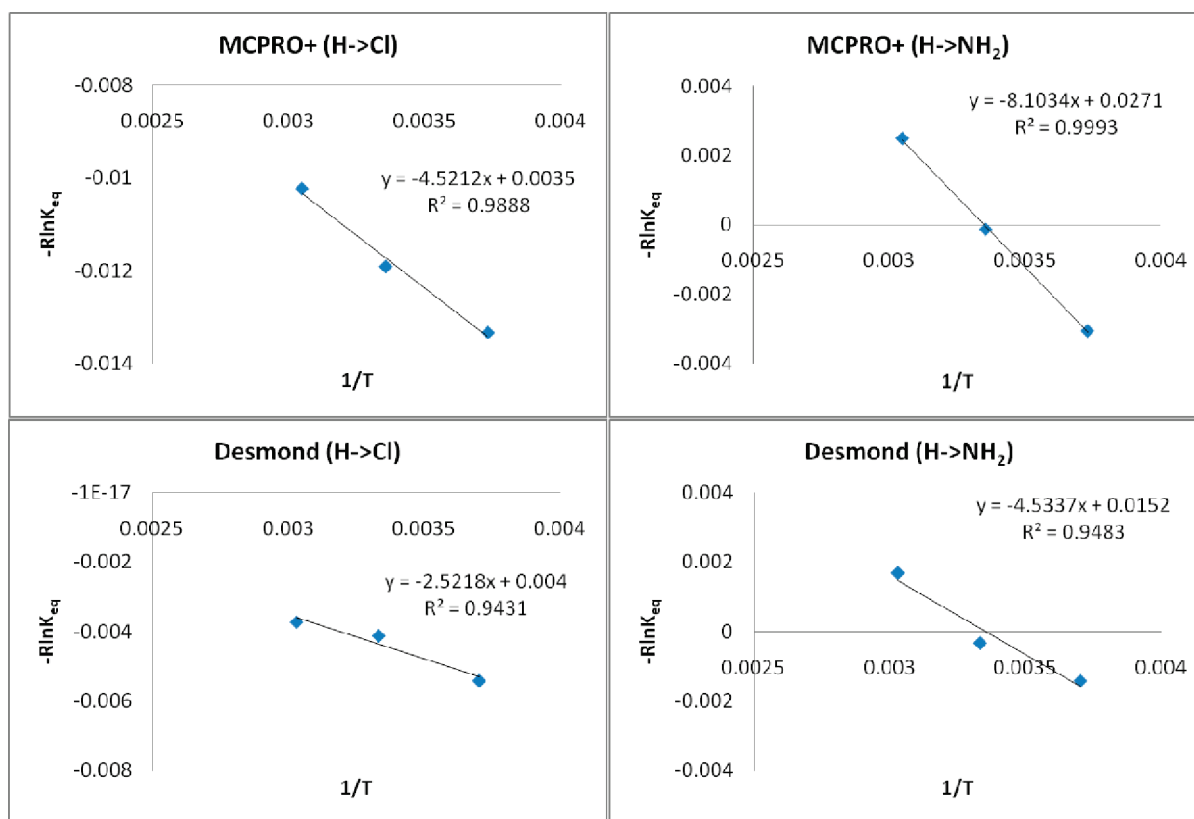


Figure 9. Van't Hoff plots obtained with MCPRO+ (top) and Desmond (bottom) for the H→4-Cl and H→4-NH₂ pairs.

the binding process, is an even more drastic restriction. However, on the basis of the present and historical results,^{14,16} it seems to be reasonable, at least for congeneric series. The results obtained with MM-GB/SA even qualify this approach as a more attractive alternative for rank-ordering; it can achieve equivalent or superior predictive accuracy and handle more structurally dissimilar ligands at a fraction of the computational cost of the rigorous free-energy methods.

Large Dynamic Range in MM-GB/SA: Comparison with FEP Results. Figure 6 shows a large dynamic range for the MM-GB/SA method (~15 kcal/mol) compared to the experimental range (~1.8 kcal/mol). This is a phenomenon typically observed in MM-GB/SA calculations,^{14,16} and it is not particular of the case studied in this work. As discussed previously,¹⁶ the large dynamic range observed in the MM-GB/SA scoring could have its origin in several effects, some enthalpic and others entropic in nature. In the former, it is possible that the wider scoring spread is due to the application of a protein dielectric constant of 1 in a model where protein motions and polarization are not taken into account. Therefore, electrostatic interactions are not shielded enough and protein–ligand electrostatic attractions and repulsions are overestimated, causing the large separation of potent and weak compounds. Other possible enthalpic contributions are related to the lack of thermal effects and protein relaxation/strain as only one structure for each complex is used.

Another potential explanation is associated with the incomplete description of enthalpy–entropy compensation; only the translational, rotational, conformational, and some vibrational entropy changes for the ligand upon binding are included. The important vibrational entropy change associated with the narrowing of the torsional energy wells for the ligands when in the protein

environment is ignored due to its high computational cost. More flexible ligands, which have the opportunity to maximize their interactions with the protein, should pay a more significant torsional entropy penalty due to restriction of their torsional motions.³³ Also ignored are all entropic contributions associated with the protein due to the complexity in computing them.

To evaluate the origin of the large dynamic range in MM-GB/SA, a computational Van't Hoff analysis was conducted for the H→4-Cl and H→4-NH₂ pairs of ligands by running FEP simulations with MCPRO+ and Desmond at 298 K and at ±30 K. In the analysis, it was assumed that the heat capacity remains unchanged over the temperature interval.³⁴ The analogs were chosen for two reasons: (1) to save computational time since they are symmetrical like the unsubstituted analog and (2) because the difference in polarity between Cl and NH₂ could provide a more complete picture of the impact of sampling on the theoretical dynamic range. The assumption here is that the results obtained with MCPRO+, with its restricted sampling, can be extrapolated to MM-GB/SA. As can be seen in Figure 6, MCPRO+ is also plagued by a large dynamic range, although to a lesser extent than MM-GB/SA; the value for MCPRO+ is ~7.5 kcal/mol, which is more than double the value for Desmond (~3.5 kcal/mol). In MCPRO+, only the side chains of residues with any atom within 10 Å from the ligands are varied, and the protein backbone is fixed. Although the ligand is free to move in the unbound state, one should not expect that all accessible conformations and the full range of dihedral angle values for each torsional well are visited, even for very long simulations. This implies that the description of entropic contributions, shielding of electrostatic interactions, thermal effects, and protein relaxation/strain will be somewhat incomplete in the FEP results obtained with MCPRO+.

Table 4. Enthalpic and Entropic Contributions Extracted from Van't Hoff Plots Obtained with MCPRO+ and Desmond (values in kcal/mol)

	H→4-Cl			H→4-NH ₂		
	$\Delta\Delta H_{\text{bind}}$	$-T\Delta\Delta S_{\text{bind}}$	$\Delta\Delta G_{\text{bind}}$	$\Delta\Delta H_{\text{bind}}$	$-T\Delta\Delta S_{\text{bind}}$	$\Delta\Delta G_{\text{bind}}$
MM-GB/SA ^a			-4.50			+2.93
MCPRO+	-4.52 ± 0.19	+0.97 ± 0.24	-3.55 ± 0.12	-8.10 ± 0.14	+8.06 ± 0.17	-0.04 ± 0.09
Desmond	-2.52 ± 0.26	+1.28 ± 0.30	-1.24 ± 0.15	-4.53 ± 0.20	+4.43 ± 0.23	-0.10 ± 0.12
ITC ^b	-0.29	-0.76	-1.05	+0.99	-0.49	+0.50
			(-0.78) ^c			(-0.05) ^c

^a MM-GB/SA scoring using eq 5 + RRHO + $E_{\text{GB}} + \Delta G_{\text{solv}}^{\text{ptn}}$. ^b ITC $\Delta\Delta H_{\text{bind}}$ and $-T\Delta\Delta S_{\text{bind}}$ values may be found in ref 18. ^c Values in parentheses were corrected by the equation: $c\Delta G_{\text{bind}} = \Delta G_{\text{bind}} - RT \ln(1 + 10^{(\text{pK}_a - \text{pH})})$.

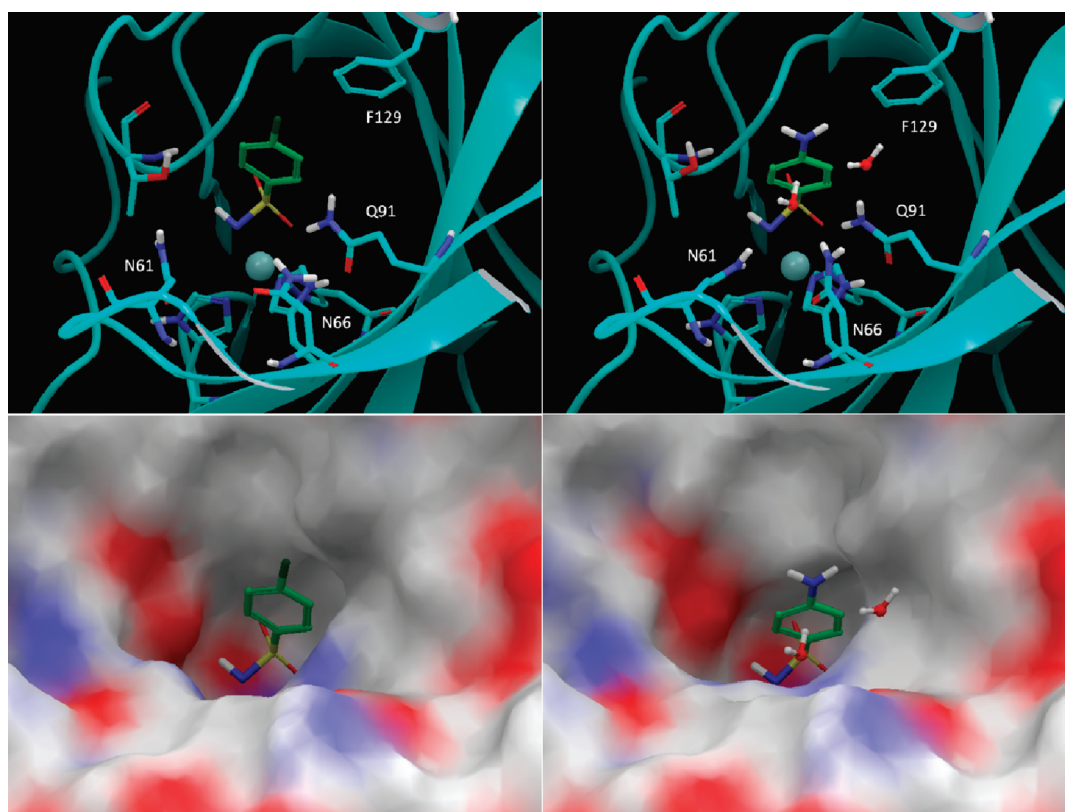
**Figure 10.** Representative theoretical structures of the complexes between human carbonic anhydrase (hCAII) and the 4-Cl (left) and 4-NH₂ (right) substituted benzene sulfonamide inhibitors. Molecular surface representations are also shown at the bottom.

Figure 9 compares the Van't Hoff plots for the H→4-Cl and H→4-NH₂ pairs obtained with Desmond and MCPRO+. The experimental and calculated enthalpic and entropic contributions to the relative binding free energies are shown in Table 4. Although the $\Delta\Delta G_{\text{bind}}$ values obtained with MCPRO+ and Desmond match the experimental values reasonably well, the enthalpic and entropic contributions do not agree with the ITC data. A possible explanation is that, of the parameters measured directly by ITC, the free energy (ΔG^0) has the lowest signal-to-noise ratio, as measured values fall in a relatively narrow range. The enthalpic contribution (ΔH^0) has a higher signal-to-noise ratio, with the entropic contribution (ΔS^0) being even less precise since errors are compounded; ΔS^0 is calculated as the difference between ΔH^0 and ΔG^0 . Alternatively, it is possible that the lack of agreement is due to the fact that the experimental

$\Delta\Delta H_{\text{bind}}$ and $-T\Delta\Delta S_{\text{bind}}$ values still contain the ionization step for the inhibitor before binding to hCAII.

If the analysis is focused on the difference between the FEP methods, it is clear that MCPRO+ provides enthalpic contributions that are exaggerated compared to those from Desmond. In the case of H→4-Cl, it is unlikely that this is due to diminished shielding of electrostatic interaction between the 4-Cl analog and the protein; the substituent is in close contact with a Phe residue (F129; Figure 10). More plausible explanations are either associated with reduced thermal effects that might lead to too favorable VDW interactions and/or limited protein deformation that does not offset the gain in interactions for 4-Cl due to restriction of degrees of freedom in the MC simulations.

It is interesting to see that the MCPRO+ $\Delta\Delta H_{\text{bind}}$ between H and 4-Cl matches their MM-GB/SA relative score of -4.50 kcal/mol,

which despite a few entropic contributions included, is mostly dominated by enthalpic terms. A more significant entropy loss of +0.97 kcal/mol for the 4-Cl derivative compared to the unsubstituted analog results in a $\Delta\Delta G_{\text{bind}}$ of -3.55 kcal/mol obtained with MCPRO+. The increased sampling in Desmond generates a $-T\Delta\Delta S_{\text{bind}}$ value of +1.28 kcal/mol, just slightly more positive than the value reported by MCPRO+ (Table 4). Thus, the larger free energy gap for the H \rightarrow 4-Cl transformation caused by restricted sampling is almost purely enthalpic.

A somewhat different scenario is seen for the H \rightarrow 4-NH₂ pair. Table 4 shows that $\Delta\Delta H_{\text{bind}}$ obtained with MCPRO+ is also exaggerated when compared to Desmond, but differently from the H \rightarrow 4-Cl case, the $-T\Delta\Delta S_{\text{bind}}$ contribution completely offsets that. $-T\Delta\Delta S_{\text{bind}}$ values of +8.06 and +4.43 kcal/mol obtained with MCPRO+ and Desmond, respectively, lead to almost identical $\Delta\Delta G_{\text{bind}}$'s. Figure 10 sheds some light on the enthalpy–entropy compensation observed for the 4-NH₂ analog. It shows that the NH₂ group has the ability to interact with Asn61, Asn66, and Gln91 via water-mediated hydrogen bonds. It is then plausible that in this case the enthalpic and entropic contributions are affected by the different structural and thermodynamic properties of the water molecules in the binding site as a result of restricted versus full protein sampling. A less mobile protein and, consequently, less mobile water molecules in MCPRO+ lead to overly favorable electrostatic interactions for the 4-NH₂ analog due to reduced shielding and thermal effects. The less fluid hydrogen bond network for the 4-NH₂ analog would also result in a greater entropy loss upon binding estimated by MCPRO+. The results obtained with the MC and MD simulations in explicit solvent also explain why the MM-GB/SA relative score for the H \rightarrow 4-NH₂ pair underpredicts the binding energy of the latter; the continuum solvation model is unable to describe the water-mediated hydrogen bonds with the protein that contribute to the binding of the 4-NH₂ analog.

If the implications of MCPRO+ restricted sampling on enthalpic and entropic contributions to binding are extrapolated to MM-GB/SA, a reasonable hypothesis for the large dynamic range in MM-GB/SA is obtained. The scoring spread is not only affected by ignored entropic contributions, i.e., all for the protein and the torsional entropy changes for the ligand, but also exaggerated enthalpic separation between the weak and potent compounds due to the lack of sampling. The resolution of whether this is caused by diminished shielding of electrostatic interactions, thermal effects, or protein relaxation/strain is complex, especially because it seems to be dependent on the case.

CONCLUSIONS

In this work, a series of 13 benzene sulfonamide inhibitors of carbonic anhydrase with binding free energies determined by isothermal titration calorimetry was selected to address whether the MM-GB/SA scoring procedure provides a more attractive alternative for rank-ordering than the more rigorous free energy perturbation (FEP) methodology. In spite of successful applications to elucidate the structure–activity relationships for few pairs of ligands, FEP calculations have rarely been evaluated for more than a handful of compounds. R^2 values of 0.70, 0.71, and 0.49 with the experiment were obtained with MM-GB/SA and FEP simulations run with MCPRO+ and Desmond, respectively, when one outlier (3-Cl derivative) was removed. The main contrast between the methods is the level of sampling, ranging from full to restricted flexibility to single conformation for the

complexes in Desmond, MCPRO+, and MM-GB/SA. All methods work well, but the results obtained with Desmond are inferior to MM-GB/SA and MCPRO+, suggesting that the full flexibility for the complexes in the former leads to additional noise, possibly due to the sampling of nonrelevant conformations and/or relevant conformations that are not adequately weighted in the simulations. In other words, when ranking congeneric series, it is safer to assume that some contributions to binding are roughly constant, and cancel when computing $\Delta\Delta G_{\text{bind}}$, than to actually try to include all of them through full sampling in simulations that are not extremely long.

Regarding the large theoretical dynamic range for the binding energies, that seems to be a direct result of the degree of sampling in the simulations since MCPRO+ as well as MM-GB/SA are plagued by this. Van't Hoff analysis for selected pairs of ligands suggests that the wider scoring spread is not only affected by missing entropic contributions due to restricted sampling but also exaggerated enthalpic separation between the weak and potent compounds caused by diminished shielding of electrostatic interactions, thermal effects, and protein relaxation/strain. Finally, the current and historical results obtained with MM-GB/SA qualify this approach as a more attractive alternative for rank-ordering than the FEP methodology; it can achieve equivalent or superior predictive accuracy and handle more structurally dissimilar ligands at a fraction of the computational cost.

AUTHOR INFORMATION

Corresponding Author

*Phone: (860) 686-2915. E-mail: cristiano.guimaraes@pfizer.com.

ACKNOWLEDGMENT

The author thanks Alan Mathiowetz from Pfizer, Inc. and Robert Abel from Schrodinger, Inc. for helpful discussions.

REFERENCES

- (1) Jorgensen, W. L. Free energy changes in solution. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; Wiley: New York, 1998; Vol. 2, pp 1061–1070.
- (2) Kollman, P. A. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (3) Jorgensen, W. L. Free energy calculations: A breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.
- (4) Simonson, T.; Georgios, A.; Karplus, M. Free energy simulations come of age: Protein-ligand recognition. *Acc. Chem. Res.* **2002**, *35*, 430–437.
- (5) Guimarães, C. R. W.; Boger, D. L.; Jorgensen, W. L. Elucidation of fatty acid amide hydrolase inhibition by potent α -ketoheterocycle derivatives from Monte Carlo simulations. *J. Am. Chem. Soc.* **2005**, *127*, 17377–17384.
- (6) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Energetics of displacing water molecules from protein binding sites: Consequences for ligand optimization. *J. Am. Chem. Soc.* **2009**, *131*, 15403–15411.
- (7) Michel, J.; Foloppe, N.; Essex, J. W. Rigorous free energy calculations in structure-based drug design. *Mol. Inf.* **2010**, *29*, 570–578.
- (8) Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. Understanding noncovalent interactions: Ligand binding energy and catalytic efficiency from ligand-induced reductions in motion with receptors and enzymes. *Angew. Chem., Int. Ed.* **2004**, *43*, 6596–6616.
- (9) (a) Pearlman, D. A.; Charifson, P. S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38

- MAP kinase protein system. *J. Med. Chem.* **2001**, *44*, 3417–3423. (b) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Effects of water placement on predictions of binding affinities for p38 α MAP kinase. *J. Chem. Theory Comput.* **2010**, *6*, 3850–3856. (c) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **2007**, *371*, 1118–1134. (d) Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273. (e) Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J. Mol. Biol.* **2009**, *394*, 747–763. (f) Michel, J.; Verdonk, M. L.; Essex, J. W. Protein-ligand binding affinity predictions by implicit solvent simulations: A tool for lead optimization?. *J. Med. Chem.* **2006**, *49*, 7427–7439. (g) Michel, J.; Essex, J. W. Hit identification and binding mode predictions by rigorous free energy simulations. *J. Med. Chem.* **2008**, *51*, 6654–6664.
- (10) Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to Avidin and Streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43*, 3786–3791.
- (11) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (12) (a) Bernacki, K.; Kalyanaraman, C.; Jacobson, M. P. Virtual ligand screening against *Escherichia coli* dihydrofolate reductase: Improving docking enrichment physics-based methods. *J. Biomol. Screening* **2005**, *10*, 675–681. (b) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-based scoring of protein-ligand complexes: Enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model* **2006**, *46*, 243–253. (c) Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. Molecular mechanics methods for predicting protein–ligand binding. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5166–5177. (d) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* **2006**, *49*, 4805–4808. (e) Lee, M. R.; Sun, Y. Improving docking accuracy through molecular mechanics generalized Born optimization and scoring. *J. Chem. Theory Comput.* **2007**, *3*, 1106–1119. (f) Huang, N.; Jacobson, M. P. Physics-based methods for studying protein-ligand interactions. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 325–331. (g) Foloppe, N.; Hubbard, R. Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* **2006**, *13*, 3583–3608.
- (13) (a) Barril, X.; Gelpí, J. L.; López, J. M.; Orozco, M.; Luque, F. J. How accurate can molecular dynamics/linear response and Poisson–Boltzmann/solvent accessible surface calculations be for predicting relative binding affinities? Acetylcholinesterase Huprine inhibitors as a test case. *Theor. Chem. Acc.* **2001**, *106*, 2–9. (b) Pearlman, D. A. Evaluating the molecular mechanics Poisson–Boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48*, 7796–807. (c) Kawatkar, S.; Wang, H.; Czerminski, R.; Joseph-McCarthy, D. Virtual fragment screening: An exploration of various docking and scoring protocols for fragments using Glide. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 527–539.
- (14) Guimarães, C. R. W.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* **2008**, *48*, 958–970.
- (15) (a) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. (b) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (16) Guimarães, C. R. W.; Mathiowetz, A. M. Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free-energy perturbation calculations. *J. Chem. Inf. Model.* **2010**, *50*, 547–559.
- (17) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the active-site solvent in the thermodynamics of Factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831.
- (18) Scott, A. D.; Phillips, C.; Alex, A.; Flocco, M.; Bent, A.; Randall, A.; O'Brien, R.; Damian, L.; Jones, L. H. Thermodynamic optimization in drug discovery: A case study using Carbonic Anhydrase inhibitors. *Chem. Med. Chem.* **2009**, *4*, 1985–1989.
- (19) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11235.
- (20) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. J. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (21) Wong, C. F.; McCammon, J. A. Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc.* **1986**, *108*, 3830–3832.
- (22) MCPRO+, version 1.36; Schrödinger, LLC: New York, 2008.
- (23) Desmond, version 2.2; Schrödinger, LLC: New York, 2009.
- (24) Anderson, H. C. Rattle: A 'velocity' version of the Shake algorithm for molecular dynamics simulation. *J. Comput. Phys.* **1983**, *52*, 24–34.
- (25) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (26) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (27) Beutler, T.; Mark, A.; van Schaik, R.; Gerber, P.; van Gunsteren, W. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (28) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W.; Dinola, A.; Haak, J. R. Molecular-Dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (30) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (31) MacroModel, version 9.0; Schrödinger, LLC: New York, 2005.
- (32) WaterMap, version 1.0; Schrödinger, LLC: New York, 2008.
- (33) Chang, C. A.; Chen, W.; Gilson, M. K. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1534–1539.
- (34) Prabhur, N. V.; Sharp, K. A. Heat capacity in proteins. *Annu. Rev. Phys. Chem.* **2005**, *56*, 521–548.

Addressing Methionine in Molecular Design through Directed Sulfur–Halogen Bonds

Rainer Wilcken,[†] Markus O. Zimmermann,[†] Andreas Lange,[†] Stefan Zahn,[‡] Barbara Kirchner,[‡] and Frank M. Boeckler^{†,*}

[†]Laboratory for Molecular Design and Pharmaceutical Biophysics, Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmacy, Eberhard-Karls-University Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany

[‡]Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie, Universität Leipzig, Linnéstr. 2, 04103 Leipzig, Germany

 Supporting Information

ABSTRACT: Halogen bonds are directional interactions involving an electron donor as binding partner. Employing quantum chemical calculations, we explore how they can be used in molecular design to address the sulfur atom in a methionine residue in a previously neglected, directed manner. We characterize energetics and directionality of these halogen bonds and elucidate their spatial variability in suboptimal geometries that are expected to occur in protein–ligand complexes featuring a multitude of concomitant interactions. We derive simple rules allowing medicinal chemists and chemical biologists to easily determine preferred areas of interaction within a binding site and to exploit them for scaffold decoration and design. Our work shows that sulfur–halogen bonds may be used to expand the patentable medicinal chemistry space. We demonstrate their potential to increase binding affinities and suggest that they can significantly contribute to inducing and tuning subtype selectivities.

INTRODUCTION

A profound understanding of molecular recognition is the essential basis of structure-based molecular design. Established protein–ligand interactions such as hydrogen bonds and hydrophobic contacts have been increasingly complemented by nonclassical interaction types such as π – π or cation– π contacts which may play important roles and contribute substantially to total binding affinity. However, very few of these nonclassical interaction types have so far been included in common molecular design tools, as emphasized in a recent review.¹

Among these often neglected interactions is halogen bonding, an attractive interaction of the type $R-X \cdots D-R'$, where X represents chlorine, bromine, or iodine and D can be any kind of Lewis base. This interaction has been broadly recognized in materials sciences since the 1970s,^{2,3} but its occurrence in biological systems has been studied only recently, e.g., through statistical PDB evaluations.⁴ The driving force of the interaction can be explained within the σ -hole concept: the halogen atoms possess a characteristic crown of positive charge due to a deficiency in electron density in the outer lobe of the p_z orbital (where z is chosen as the $R-X$ axis).^{5–9} The σ -hole concept has been extended and applied to group VI atoms, which can be relevant in biological systems.^{10,11} Several quantum chemical studies using small model systems have been performed to characterize the nature and strength of halogen bonds.^{12,13} QM/MM calculations on protein–ligand complexes have also been conducted^{14,15} and similarities as well as differences between halogen bonding and hydrogen bonding have been discussed.^{16–19} Halogen bonds involving backbone carbonyl oxygen atoms have been implicated as favorable interactions by some very recent experimental studies.^{20,21}

When searching the PDB for ligand–protein halogen-bond contacts, as done in a recent study,¹⁵ mainly two types of halogen bonds are observed—contacts to backbone carbonyl moieties

(53%) and halogen– π contacts (33%)—while those involving sulfur or nitrogen are found only sparsely (5% S, 9% N). However, this does not mean that sulfur–halogen bonds are unfavorable. Most ligand–protein halogen bonds observed in the PDB were not rationally designed but found serendipitously due to the nonclassical nature of this contact. Thus, it is likely that halogen–sulfur contacts are merely underrepresented given the number of carbonyl oxygen moieties in a protein. We demonstrate that addressing the sulfur-containing side chain of methionine through halogen bonds has distinct potential for improving affinity and selectivity of a compound.

In this work, we use quantum chemical calculations at the DFT-D, MP2, and CCSD(T) level and large basis sets to characterize the interaction strengths between ligand model systems and three different molecular representations of methionine. Where possible, we make sure to obtain interaction geometries and energies that solely represent the halogen-bond contact, avoiding additional secondary interactions. From our model calculations, we deduce simple rules on how to best address methionine through halogen bonding. Additionally, we focus on elucidating the impact of deviations from ideal binding geometries on overall complex formation energies. Finally, we show that our model calculations agree well with an existing crystal structure of IL-2²² and propose a way to visualize favorable areas of interaction within a binding site, making our results immediately applicable for use in molecular design.

RESULTS

Energetics of the Halogen Bond. We represent methionine as a molecular system of three different sizes: (i) as the simple model

Received: April 8, 2011

Published: June 02, 2011

system dimethyl sulfide (MET1); (ii) as 1-(methylsulfanyl)propane (MET2), representing the complete methionine side chain including C α ; and (iii) as full amino acid including backbone atoms, capped by an acetyl group at the N-terminus and a N-methyl amide at the C-terminus (MET3) (Figure 1). Input geometries for MET2 and MET3 were extracted from 3DN4, a PDB structure featuring a short iodine–methionine contact.²³ The backbone conformation of MET3 is α -helical with $\varphi = -59.9^\circ$ and $\psi = -47.5^\circ$, and the side chain conformation is gauche(–)–trans–gauche(+). We have restricted the φ and ψ angles in all optimizations involving MET3 to avoid biologically unreasonable backbone conformations. We employ iodobenzene, bromobenzene, and chlorobenzene as ligand models because substituted aromatic or heteroaromatic scaffolds are common cores in medicinal chemistry and their halogenation may be exploited for scaffold decoration.

Using the small methionine model MET1, we obtain optimized halogen-bonded complexes for all three halobenzenes at the MP2/QZVPP level. Interaction energies are calculated as adduct formation of the halogen-bonded complex from isolated halobenzene and methionine (Table 1). We observe the strongest interaction (-19.3 kJ/mol) for the iodobenzene complex, while bromobenzene and chlorobenzene complexes show weaker interaction energies (-13.0 and -10.1 kJ/mol, respectively). This order of halogen-bond strengths (Cl < Br < I) has been reported numerous times in the literature, regardless of the Lewis base

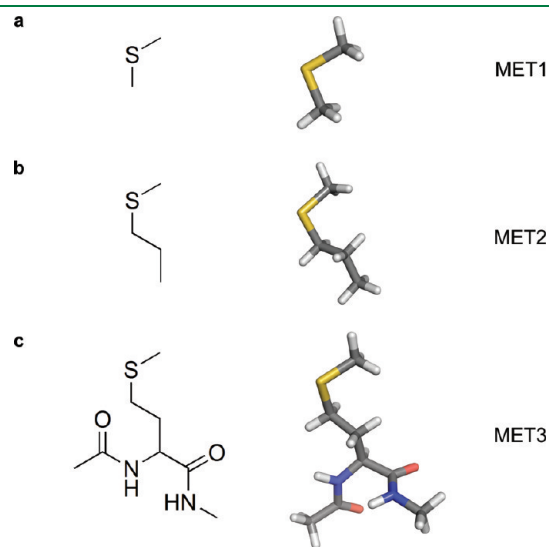


Figure 1. Methionine model systems MET1 (a), MET2 (b), and MET3 (c), represented as structural formulas and stick models (optimized geometries).

interaction partner.^{9,19} In our calculations, all halogen bonds show X \cdots S distances below the sum of van der Waals radii of the two atoms. Despite the size of iodine, the I \cdots S bond is the shortest, implying a very favorable interaction. Interestingly, the equilibrium distances for all X \cdots S bonds are quite similar (around 3.4 Å), indicating that a substitution of chlorine by one of the other two halogens should be feasible despite the differences in size. Employing widely used density functionals augmented by an empirical dispersion correction (DFT-D), we find that some of these quite closely mimic the results obtained with MP2 (Supporting Information, Table S1). To describe the strength of an iodine–sulfur halogen bond with chemical accuracy, we perform CCSD(T) calculations by employing a basis set extrapolation scheme (see Methods, eqs 1 and 2 and Supporting Information Table S2) and obtain an interaction energy of -17.3 kJ/mol. This is reasonably close to our MP2 results.

To put the halogen-bond energies into context, we model two H \cdots S interactions formed by benzene and MET1 (weak hydrogen bond) as well as phenol and MET1 (moderately strong hydrogen bond) at MP2/QZVPP level. Benzene forms a complex with MET1 at a much reduced bond distance compared to the halobenzenes (279 vs 336 pm, H vs I), with a complex formation energy of -11.0 kJ/mol. The phenol–MET1 complex shows the largest interaction energy of all investigated model systems (-29.1 kJ/mol).

However, adding just one water molecule to the adduct formation reaction for the sake of introducing ligand desolvation energies changes the overall picture dramatically. If we require the ligand to be desolvated upon binding, i.e., if we exchange ligand-bound water by ligand-bound methionine (Figure 2), the reaction energy for phenol is reduced to -1.4 kJ/mol. In contrast, the iodobenzene and benzene reaction energies are only reduced to -11.3 and -6.0 kJ/mol, respectively. This one-molecule approach to solvation is rather simplistic but helps illustrate how desolvation energies can differ between nonpolar and polar ligands. To complete the energetic description of the methionine–halogen bond, we perform geometry optimizations on the larger methionine model systems MET2 and MET3. Switching from the quadruple- ζ basis set (QZVPP) to the smaller triple- ζ basis TZVPP due to computational considerations for the larger models, we obtain smaller interaction energies for all halogen-bonded systems, implying that the MP2 method heavily depends on basis set size. Expanding the methionine in size from MET1 to MET2 and MET3, we observe a moderate rise in interaction energies while retaining virtually unchanged geometries (Supporting Information, Figure S1, Table 2). For the treatment of the chlorobenzene \cdots MET3 adduct, we had to additionally restrict the bond angle $\alpha_{S5-C16-C7}$ during optimization to obtain a halogen-bonded geometry.

Table 1. Interaction Energies (in kJ/mol) for Halogen-Bonded Model Complexes (small methionine model system MET1)^a

complex	ΔE	method	$d_{S5-X6/H6^c}$ pm (% vdW ^b)	$\alpha_{C4-S5-X6/H6^c}$ deg	$\delta_{C3-C4-S5-X6/H6^c}$ deg	$\alpha_{S5-X6/H6-C7/O7^c}$ deg
C ₆ H ₅ I \cdots MET1	-19.3	MP2/QZVPP	336 (88.9)	88.8	-88.6	169.8
C ₆ H ₅ Br \cdots MET1	-13.0	MP2/QZVPP	343 (94.0)	85.9	-85.2	170.1
C ₆ H ₅ Cl \cdots MET1	-10.1	MP2/QZVPP	348 (98.0)	80.9	-79.6	162.7
C ₆ H ₆ \cdots MET1	-11.0	MP2/QZVPP	279 (96.2)	79.3	77.3	156.9
C ₆ H ₅ OH \cdots MET1	-29.1	MP2/QZVPP	226 (77.9)	86.8	-86.1	153.8
C ₆ H ₅ I \cdots MET1	-17.3	CCSD(T)/CBS ^c	345 (91.3)	88.8	-88.5	168.4

^a Interaction energies for complex formation with benzene and phenol (H \cdots S contacts) are given for comparison. Energies were corrected for BSSE using the counterpoise correction. ^b Percentage of the sum of the van der Waals radii of the two atoms directly involved in bonding. ^c CCSD(T) calculations were performed on the SCS-MP2/QZVPP minimum geometry and a basis set extrapolation scheme was employed (see Methods).

In summary, we find that methionine can be favorably addressed through halogen bonds, especially by those involving $\text{Br}\cdots\text{S}$ or $\text{I}\cdots\text{S}$ contacts. Taking into account desolvation penalties implies that these interactions might be superior to weak and moderately strong hydrogen bonds. Interestingly, while most halogen bonds in the PDB are formed by ligands with chloro substituents, our data reconfirm earlier studies, demonstrating that exchange of these chlorines by bromine or iodine is feasible and energetically favorable.

Distance Dependencies of Halogen Bonds. In existing crystal structures, the optimal geometries described above are not always observed due to the interplay of multiple competing primary and secondary interactions involved in ligand binding. For this reason, we investigate the influence of nonideal bond distances on complex formation energy. Starting from MP2/TZVPP minimum structures, we perform distance scans along

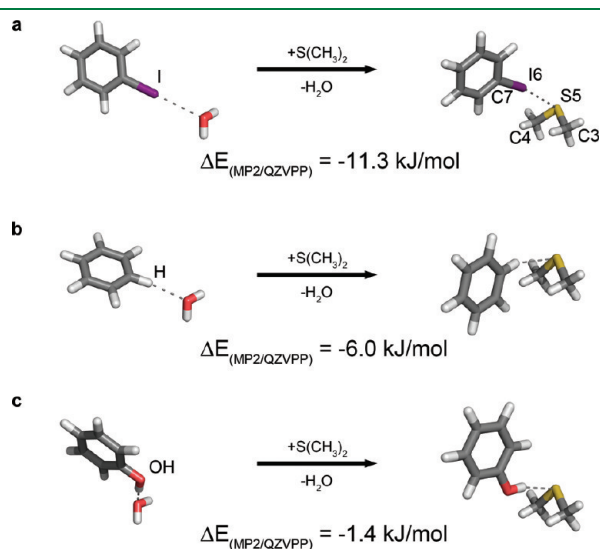


Figure 2. Reactions representing ligand desolvation: Exchange of water from iodobenzene \cdots water (a), benzene \cdots water (b), and phenol \cdots water (c) complexes by dimethyl sulfide (MET1) and corresponding reaction energies (kJ/mol) at the MP2/QZVPP level. These three model systems represent (a) halogen bonds and (b) weak and (c) moderately strong hydrogen bonds. Relevant atom labels, which are used for defining distances, angles, and dihedrals, are shown in the iodobenzene \cdots MET1 complex (a).

the $\text{X}\cdots\text{S}$ bond. The scans are carried out as single point calculations using methionine model MET1 as all three methionine systems show very similar behavior in terms of energetics and distance dependence (Supporting Information, Figure S2).

In rational drug design, introducing an additional moiety like a halogen atom into a lead structure makes most sense if it leads to an affinity increase. We have demonstrated that halogen bonds are favorable interactions. However, we have to keep the “background affinity” in mind that is already provided by an $\text{H}\cdots\text{S}$ contact in an ideal halogen-bond distance, representing an undecorated benzene scaffold. We thus have to compare the complex formation energy of the halobenzene–methionine complexes with the benzene–methionine complex at the ideal halogen bond distance. This distance is very similar for all three halogenated complexes (339, 343, and 347 pm for I, Br, and Cl at the MP2/TZVPP level). Figure 3 shows the distance dependence of the complex formation energies. At 339 pm, the complex formation energy for benzene is $\Delta E(\text{MP2/TZVPP}) = -8.7$ kJ/mol. We use this energy as a threshold: The introduction of a halogen into the ligand structure at any distance $d_{\text{X}\cdots\text{S}}$ needs to be energetically more favorable than 8.7 kJ/mol in order to improve binding energy. This definition (horizontal line in Figure 3b) leads to areas of tolerance for all three halobenzene

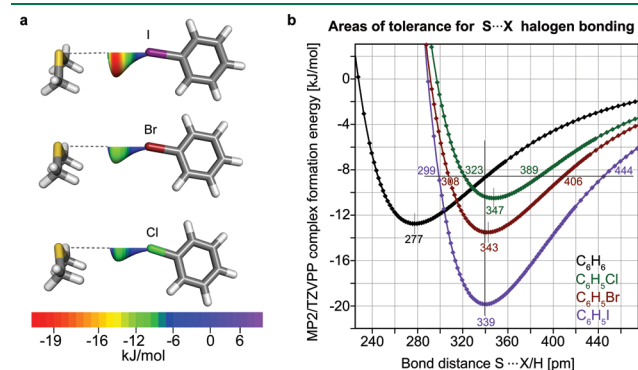


Figure 3. (a) Distance scan plots for the three halobenzene \cdots MET1 complexes. The curves depicted in the model systems highlight preferred regions of halogen placement by their coloring and size. (b) Plot of $\text{X}\cdots\text{S}$ ($\text{H}\cdots\text{S}$) distance vs complex formation energy for iodobenzene, bromobenzene, chlorobenzene, and benzene complexes with MET1 (MP2/TZVPP) and definition of areas of tolerance for the halobenzene \cdots methionine complexes.

Table 2. Interaction Energies (in kJ/mol) for Adduct Formation Yielding Halogen-Bonded Complexes with the Three Methionine Models at MP2/TZVPP Level^a

complex	$\Delta E_{\text{MP2/TZVPP}}$	$d_{\text{S}\cdots\text{X6}}$, pm	$\alpha_{\text{C4}\cdots\text{S5}\cdots\text{X6}}$, deg	$\delta_{\text{C3}\cdots\text{C4}\cdots\text{S5}\cdots\text{X6}}$, deg	$\alpha_{\text{S5}\cdots\text{X6}\cdots\text{C7}}$, deg
$\text{C}_6\text{H}_5\text{I}\cdots\text{MET1}$	-15.2 (-19.8)	339	88.3	-88.0	169.5
$\text{C}_6\text{H}_5\text{I}\cdots\text{MET2}$	-16.2 (-21.2)	339	88.3	-87.0	170.3
$\text{C}_6\text{H}_5\text{I}\cdots\text{MET3}$	-16.6 (-21.7) ^b	342	88.9	-84.2	168.1
$\text{C}_6\text{H}_5\text{Br}\cdots\text{MET1}$	-10.7 (-13.5)	343	85.8	-85.1	170.1
$\text{C}_6\text{H}_5\text{Br}\cdots\text{MET2}$	-11.2 (-14.1)	342	85.6	-85.2	172.0
$\text{C}_6\text{H}_5\text{Br}\cdots\text{MET3}$	-12.6 (-15.6) ^b	347	84.7	-80.2	163.7
$\text{C}_6\text{H}_5\text{Cl}\cdots\text{MET1}$	-7.8 (-10.5)	347	80.9	-79.5	162.7
$\text{C}_6\text{H}_5\text{Cl}\cdots\text{MET2}$	-8.2 (-11.1)	348	80.1	-79.6	163.6
$\text{C}_6\text{H}_5\text{Cl}\cdots\text{MET3}$	-10.0 (-13.1) ^{b,c}	349	80.6	-75.9	162.4

^a Energies are counterpoise-corrected; uncorrected energies are given in brackets. ^b Methionine model system MET3 was extracted from PDB structure 3DN4, and backbone φ and ψ dihedrals were frozen in the original α -helical geometry. ^c For the chlorobenzene \cdots MET3 complex, the bond angle $\alpha_{\text{S5}\cdots\text{Cl6}\cdots\text{C7}}$ was constrained to 162.4° to obtain a halogen-bonded minimum structure.

Table 3. Comparison of the MP2/TZVPP Complex Formation Energies (in kJ/mol) at Optimal Distances and at the Equilibrium Distance of Iodobenzene...MET1 Yields Distance Areas of Tolerance^a

complex	$d_{S5-X6/H6}$ pm	$\Delta E_{MP2/TZVPP}$		
		at optimal distances	at 339 pm	areas of tolerance for $d_{S5-X6/H6}$, pm
C ₆ H ₅ I...MET1	339	-19.8	-19.8	299–444
C ₆ H ₅ Br...MET1	343	-13.5	-13.5	308–406
C ₆ H ₅ Cl...MET1	347	-10.5	-10.4	323–389
C ₆ H ₆ ...MET1	277	-12.7	-8.7	N/A

^aAll halogen-bonded complexes with S...X distances within tolerance have interaction energies greater than 8.7 kJ/mol, which is the complex formation energy of benzene...MET1 at 339 pm.

ligand systems (presented in detail in Table 3). For chlorobenzene, the area of tolerance reaches from 323 to 389 pm, and the energy gain is at most 1.8 kJ/mol compared to benzene (at the ideal bond distance of 343 pm). While this explains the occurrence of some Cl...S halogen bonds in the PDB, it also illustrates that there is no big affinity increase to be expected by simply exchanging H for Cl at one position on a benzene-based scaffold, unless this interaction is supplemented by other interactions that the chlorine may participate in. For bromobenzene, the area of tolerance becomes larger (308–406 pm) and the energy gain more pronounced. For the iodobenzene complex, the distance tolerance is especially large (299–444 pm), with a favorable complex formation energy of less than $\Delta E(MP2/TZVPP) = -18$ kJ/mol within a range of about 320–365 pm.

In addition to the MP2 method, we also carried out the distance scans employing four widely used density functionals (BP86-D, BLYP-D, B3LYP-D, and TPSS-D, Supporting Information, Figure S3). TPSS-D shows the best agreement with the MP2 data while being much less computationally expensive.

The agreement of our distance scans with experimental data is exemplified in two crystal structures from the PDB containing iodine–sulfur halogen bonds. For 3DN4,²³ the authors report a bond distance of 3.3 Å, while for 2PIW, the thyroid hormone 3,3',5-triiodo-L-thyronine is bound to the androgen receptor²⁴ with an S...I distance of 3.7 Å. Both distances are well within our areas of tolerance.

Energetic Impact of σ -Hole Bond Directionality. It is generally agreed that the driving force of halogen bonding is the σ -hole, a crown of positive charge on chlorine, bromine, and iodine residues. The σ -hole concept explains halogen bonding as an electrostatic interaction similar to hydrogen bonding.^{5,19} One study has suggested that halogen bonds may be driven by both electrostatic and dispersive forces, using symmetry-adapted perturbation theory.¹²

A purely electrostatics-driven interaction would translate into a very directional bond, meaning that deviations from the ideal bond angle $\alpha_{S5-X6-C7}$ would be heavily penalized. If the interaction was mainly driven by dispersion, one would expect more undirected, “greasy” behavior, and deviations from the ideal angle would be tolerated to a large extent. These two scenarios are decisively different and have great impact in molecular design approaches, e.g., in scaffold placement. In the dispersion-driven case, the scaffold and the halogen solely need to be placed in the right distance from a methionine residue, while in the

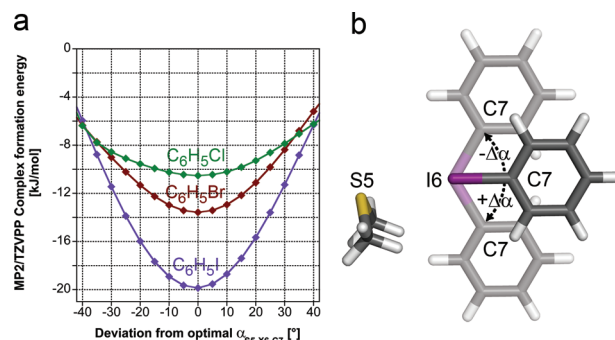


Figure 4. (a) Impact of deviations from optimal σ -hole angles $\alpha_{S5-X6-C7}$; energies are calculated using MP2/TZVPP. (b) Illustration of scan setup: the ligand model system is rotated around the axis perpendicular to the paper plane passing through I6.

electrostatics-driven case, both the distance $d_{X6...S5}$ and the angle $\alpha_{S5-X6-C7}$ are very important and need to be matched as closely as possible.

To evaluate the energetic penalties of nonideal σ -hole bond angles, we first optimize the halobenzene...MET1 dimer constrained to C_s symmetry at the MP2/TZVPP level in order to rule out any secondary interactions of the aromatic ring with the methyl moieties of the MET1 system. The resulting complexes possess $\alpha_{S5-X6-C7}$ angles of 168.5° (I), 166.9° (Br), and 162.6° (Cl). We then successively change the bond angle $\alpha_{S5-X6-C7}$ in 5° steps and perform MP2/TZVPP single point calculations. The results of the scans are shown in Figure 4a, with a schematic representation of positive ($+\Delta\alpha$) and negative ($-\Delta\alpha$) deviations from ideal (C_s) geometry given in Figure 4b. Interestingly, even though the strengths of the halogen bonds for the three halobenzenes are very different, as discussed earlier, the derived tolerance areas are similar at about $\pm 30^\circ$ deviation from ideal angles. It is important to note that the potential curves are fairly symmetric, even though the “optimized” angle is not 180°. Another surprise is that the curves cross at about $\pm 40^\circ$, which means that at this angle, there is no advantage of substituting one halogen for another. From these scans, we derive that halogen bonds are strongly directional and that large deviations from the ideal $\alpha_{S5-X6-C7}$ angle have the potential to ruin binding affinities, especially in the case of the favorable I...S interaction. This confirms the electrostatics-driven nature of the interaction and is in accordance with studies on other model systems.¹⁸ Deviations from the ideal angle amounting to more than 20°–30° should be avoided at all costs, and scaffolds should be selected (e.g., in a scaffold-hopping approach) or designed de novo to allow for close to optimal ($\approx 170^\circ$) $\alpha_{S5-X6-C7}$ angles.

Dihedral Angle Variations—Spherical Scans. Taking one step toward integrating halogen bonds in molecular design tools, we investigate degrees of freedom in the spherical orientation of the halogen with respect to the methionine sulfur. Starting from an optimized input geometry and using 5° steps in varying the angles (see Methods), we generate 2664 variations of the interaction geometry, keeping the halogen at the equilibrium distance on a sphere around the sulfur atom as its center. We prepare the input geometry by conducting a constrained optimization (MP2/TZVPP) in C_s symmetry with the bond angle $\alpha_{S5-X6-C7}$ restricted to 180° and the dihedral angle $\delta_{C3-C4-S5-X6}$ restricted to 90° to avoid unsymmetric artifacts due to bias in the input structure. This bias arises predominantly from the deviation of the σ -hole angle $\alpha_{S5-X6-C7}$ from 180° in the optimized structure. The constraints

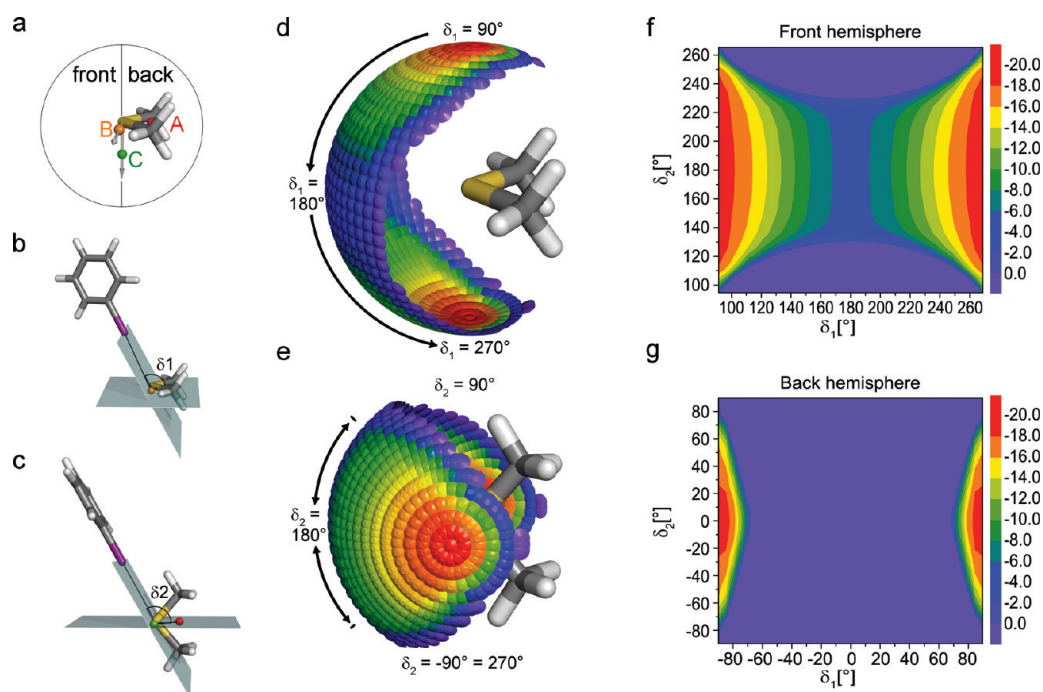


Figure 5. Spherical scans for iodobenzene. (a) Definition of front and back hemisphere for MET1 and position of dummy points A–C used to define dihedral angles. Positions of dummy points B and C are highlighted by vectors centered at S5. (b) Illustration of dihedral angle δ_1 (out-of-plane rotation). δ_1 represents the rotation around an axis through dummy point B and atom S5, which is part of the C–S–C plane and is oriented perpendicularly to the mirror plane of the methionine fragment. (c) Illustration of dihedral angle δ_2 (in-plane rotation). δ_2 represents the rotation around an axis from dummy point C to atom S5, which is oriented perpendicularly to the C–S–C plane. (d) Interaction energy sphere for the iodobenzene...MET1 complex, oriented to exemplify variations in δ_1 . (e) Interaction energy sphere for iodobenzene...MET1 complex, oriented to exemplify variations in δ_2 . (f, g) 2D plots (δ_1 vs δ_2) of interaction energies on the front hemisphere (f) and back hemisphere (g).

have a mild impact on energies compared to the freely optimized complexes (1.1 kJ/mol for iodobenzene...MET1, 1.2 kJ/mol for bromobenzene...MET1, 2.0 kJ/mol for chlorobenzene...MET1; overlays of constrained and freely optimized structures are shown in the Supporting Information (Figure S4). The cutoff for the areas of tolerance changes to -6.4 kJ/mol as the same constraints are applied to the benzene...MET1 complex. We conduct the spherical scans with the more efficient DFT-D method TPSS-D/TZVPP, which delivers interaction energies comparable to the more expensive MP2/TZVPP method as exemplified in the distance scans (Supporting Information, Figure S3).

The obtained interaction sphere for iodobenzene...MET1 is presented in Figure 5d,e from two different perspectives. These perspectives are chosen to focus on variations of the two dihedral angles δ_1 ($=\delta_{C_4-B-S_5-X_6}$) and δ_2 ($=\delta_{A-C-S_5-X_6}$) as defined in Figure 5, parts b and c, respectively. While Figure 5b,d highlights the variation of the “out-of-plane” δ_1 angle, changes in the “in-plane” δ_2 angle are more easily perceived from Figure 5c,e. The plane is defined by the three heavy atoms (C–S–C) of the methionine fragment. To emphasize the areas of tolerance regarding the dihedrals δ_1 and δ_2 , we present 2D plots for the interaction areas on the front and back hemisphere in Figure 5f,g. A definition of front and back hemisphere is given in Figure 5a.

In order to deduct simple rules, we discuss particular points on the sphere indicative of general trends. Starting from the input geometry of the sphere, we first focus on the variation of the δ_1 angle from 90° out-of-plane toward in-plane ($\delta_1 = 180^\circ$) and further toward 270° , while keeping $\delta_2 = 180^\circ$ (i.e., placing the iodine right in front of the sulfur atom). In Figure 5d, this corresponds to starting from the red area above MET1 and

proceeding down along the circular path to the red area below MET1 (black arrow). In Figure 5f, this circular path becomes a straight line from the left to the right at $\delta_2 = 180^\circ$. We observe a significant loss of interaction energy toward only -4 to -6 kJ/mol remaining for the in-plane geometry ($\delta_1 = 180^\circ$, $\delta_2 = 180^\circ$). Thus, there is a strong energy dependence on changes in δ_1 , with out-of-plane angles close to $\pm 90^\circ$ being highly favored over in-plane geometries. It becomes much more pronounced when we look at the back of the sphere (Figure 5g). As this hemisphere includes both methyl groups of the methionine, the decrease in energy is much more rapid and results in numerous sterical clashes, leaving just a small tolerance area between $\delta_1 = 80^\circ$ and 90° or -80° and -90° .

In contrast to the large impact of changes in δ_1 , it is easily possible to vary δ_2 within as much as $\pm 40^\circ$ from the 180° geometry. This can be observed particularly well in Figure 5e, where the symmetry with respect to $\delta_2 = 180^\circ$ becomes obvious. As we proceed down the sphere from the red area (corresponding to changes in δ_1), the coloring at different δ_2 angles (proceeding circularly along each “ δ_1 level”) is identical to a large extent.

For the lighter halogens chlorine and bromine, the interaction spheres show strongly attenuated energies of complex formation. Thus, Figure 6a,d (as well as corresponding 2D plots in Figure 6b,c, e,f) gives a quantitatively different, but qualitatively similar, impression when compared to the iodine figures. As the halogen bond strength to the sulfur atom of methionine is much weaker for the lighter halogens, the preferred interaction areas are much smaller than for iodine. While there is still a reasonably large interaction area for bromobenzene...MET1, the area for chlorobenzene shrinks to only those data points close to the optimal interaction geometry.

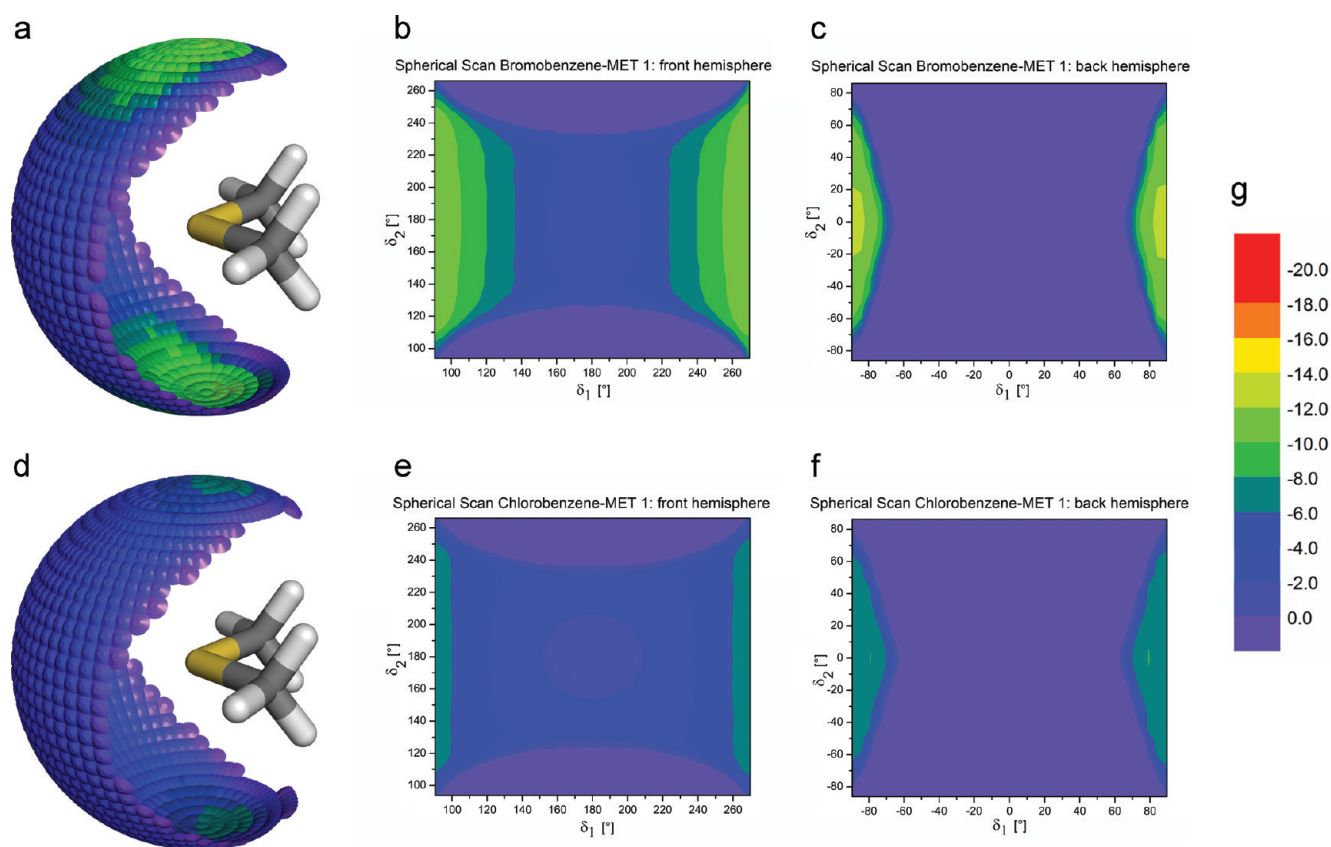


Figure 6. Spherical scans for bromobenzene and chlorobenzene. (a) Interaction energy sphere for the bromobenzene...MET1 complex (b, c) 2D plots (δ_1 vs δ_2) of interaction energies on the front hemisphere (b) and back hemisphere (c) for the bromobenzene...MET1 complex. (d) Interaction energy sphere for the chlorobenzene...MET1 complex. (e, f) 2D plots (δ_1 vs δ_2) of interaction energies on the front hemisphere (e) and back hemisphere (f) for the chlorobenzene...MET1 complex. (g) Color scale.

Finally, we also perform spherical scans starting from the unconstrained $X \cdots \text{MET}1$, $X \cdots \text{MET}2$, and $X \cdots \text{MET}3$ geometries. These are shown in the Supporting Information (Figure S5). For the MET1 interaction spheres, we observe a slight loss of symmetry of the energy profiles which can be explained by taking into account that the optimal σ -hole angle $\alpha_{S5-X6-C7}$ is not 180° . This means that two points placed on opposite sides of the sphere are not identical in terms of the σ -hole angle and thus do not lead to the same energies. Increasing the methionine model in size from MET1 to MET2 and MET3, we observe a steady increase of secondary interactions. An obvious example is Figure S5c (Supporting Information), where the areas of favorable interaction are extended toward $C\beta$ and $C\gamma$ of the aliphatic side chain. These secondary interactions (mostly $\text{CH}\cdots\pi$ contacts) even outweigh the halogen bond in strength in the $\text{Cl} \cdots \text{MET}3$ complex (Supporting Information, Figure S5i). It thus appears that the smallest model system (MET1) is best-suited to describe an isolated halogen bond void of secondary interactions.

DISCUSSION

We have proposed that halogen bonds addressing the sulfur atom of a methionine side chain are favorable interactions that have so far been neglected in molecular and drug design. We have shown that all halobenzene model systems form complexes featuring very similar equilibrium distances. This leads to the assumption that chloro, bromo, and iodo moieties are easily

interchangeable as halogen bond partners of the sulfur atom. Our calculations show that the $\text{I} \cdots \text{S}$ bond is rather strong [CCSD(T) energies of -17.9 kJ/mol or -4.3 kcal/mol], strongest among the halogens, and is superior to weak and moderately strong hydrogen bonds ($\text{H} \cdots \text{S}$ contacts) when taking ligand desolvation into account. We have elucidated the spatial behavior of halogen bonds by performing scans varying distances, angles, and dihedral angles, deducing simple rules of thumb useful for computational biologists and medicinal chemists alike. In Figure 7 we apply our results by projecting the interaction sphere onto Met39 in the crystal structure of IL-2 with the bound ligand SP4160.²² The actual geometry of the halogen on the sphere closely matches the optimal interaction area predicted by the calculations (Figure 7a). A comparison with the iodobenzene...MET1 interaction sphere exemplifies how a putative gain of binding energy could be achieved by exchanging chlorine by iodine (Figure 7b,c). It is possible to map these interaction spheres on methionine residues in any crystal structure, enabling medicinal chemists to determine favorable areas of interaction within a binding site and exploit them for scaffold decoration or de novo design.

This work demonstrating the strength of sulfur-halogen bonds has implications not only for gaining compound affinity, but particularly for tuning selectivity. So far, there is no way to address the sulfur atom in a methionine residue selectively and with a directed interaction, as can be achieved through halogen bonding. This allows for novel scaffolds and unorthodox

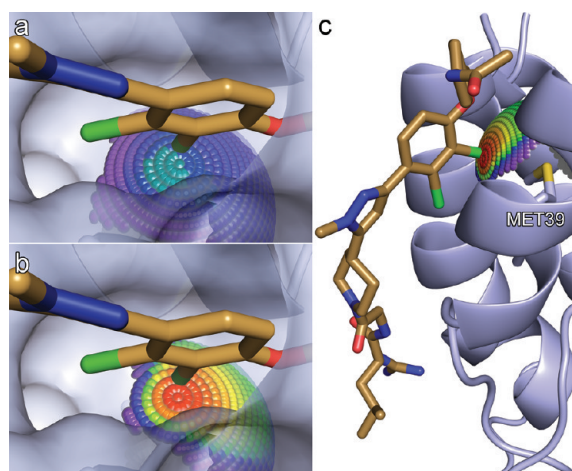


Figure 7. (a) Interaction sphere of chlorobenzene...MET1 mapped onto Met39 in PDB structure 1QVN, illustrating agreement of calculations with an experimentally determined crystal structure. (b) Mapping of the iodobenzene...MET1 interaction sphere onto Met39 exemplifies putative gain of binding energy by exchanging chlorine by iodine. (c) Embedding of halogen-bonding interaction spheres in protein–ligand environments facilitates recognition of preferential binding geometries in molecular design.

approaches in medicinal chemistry. It facilitates exploiting new, patentable chemistry and avoiding already patented chemical space. Finally, our results provide blueprints for integrating sulfur-halogen bonds into molecular design tools.

METHODS

DFT-D and MP2 Structure Optimizations. All DFT and MP2 calculations in this work were performed using the Turbomole 6.2 suite of programs.^{25,26} The basis sets used were of triple- ζ (def2-TZVPP)²⁷ and quadruple- ζ (def2-QZVPP)²⁷ quality. Relativistic effects for iodine were considered by an effective core potential (ECP).²⁸ MP2 calculations were done in combination with the resolution of identity (RI) technique^{29–32} and the frozen core approximation. The frozen core orbitals were attributed by the default setting in Turbomole, by which all orbitals possessing energies below 3.0 au are considered core orbitals. The SCF convergence criterion was increased to 10^{-8} Hartree for all calculations. Interaction energies were counterpoise-corrected using the procedure of Boys and Bernardi³³ where indicated (Tables 1 and 2) in order to correct for basis set superposition errors (BSSEs). DFT calculations were performed using the RI approximation^{34–36} with the BP86,^{37,38} BLYP,^{37,39} and TPSS⁴⁰ functionals. The RI approximation is not efficient for hybrid functionals and hence was not used in B3LYP^{37,39,41} calculations. All functionals were augmented with an empirical dispersion correction as proposed by Grimme,⁴² which we have indicated by appending “-D” to their names (i.e., BP86-D, BLYP-D, B3LYP-D, TPSS-D). We used the def2-TZVPP²⁷ basis set throughout, employing a relativistic ECP for iodine.²⁸

CCSD(T) Calculations. Structure optimizations for iodobenzene, MET1, and the adduct system were performed with the Turbomole suite employing the SCS-MP2⁴³ method and the def2-QZVPP²⁷ basis set. Additionally, we performed MP2 single point calculations with the cc-pVTZ and cc-pVQZ^{44–46} basis sets. These MP2 calculations were done in combination with the

RI technique^{29–32,47} and the frozen core approximation as described above. Relativistic effects for iodine were considered by an effective core potential (ECP).⁴⁶

CCSD(T) calculations were carried out with the cc-pVTZ^{44–46} basis set employing the MOLPRO⁴⁸ program package. Relativistic effects for iodine were again considered by a relativistic pseudopotential. All CCSD(T) and MP2 energies were counterpoise-corrected using the procedure of Boys and Bernardi.³³ The contribution of higher order correlation energy was determined by the following scheme:^{49,50}

$$\Delta E_{\text{CBS}}^{\text{CCSD(T)}} = \Delta E_{\text{CBS}}^{\text{MP2}} + (\Delta E_{\text{cc-pVTZ}}^{\text{CCSD(T)}} - \Delta E_{\text{cc-pVTZ}}^{\text{MP2}}) \quad (1)$$

This is based on the assumption that the difference between the CCSD(T) and MP2 interaction energies ($\Delta E^{\text{CCSD(T)}} - \Delta E^{\text{MP2}}$) depends only slightly on basis set size and therefore can be determined with small- or medium-sized basis sets like cc-pVTZ.

$\Delta E_{\text{CBS}}^{\text{MP2}}$ is the MP2 energy at the complete basis set limit which was obtained by the extrapolation procedure proposed by Halkier et al.:⁵¹

$$\Delta E_{\text{CBS}}^{\text{MP2}} = \frac{\Delta E_X^{\text{MP2}} X^3 - \Delta E_Y^{\text{MP2}} Y^3}{X^3 - Y^3} \quad (2)$$

where X and Y are the cardinal numbers of the cc-pVTZ and cc-pVQZ basis set, respectively.

Distance Scans. All distance scans were performed using the optimized MP2/TZVPP geometries as starting points. The bond distances $d_{\text{X}\dots\text{S}}$ ($d_{\text{H}\dots\text{S}}$ for benzene) were then elongated or shortened in 2 pm steps (5 pm further away from the minimum), with the rest of the ligand structure (halobenzene or benzene) transformed accordingly. For each step, single-point (SCF) calculations were performed using MP2/TZVPP.

Spherical Scans. Input files were generated from the optimized MP2/TZVPP geometry in C_s symmetry with the angle $\alpha_{\text{S5-X6-C7}}$ restricted to 180° and the dihedral angle $\delta_{\text{C3-C4-S5-X6}}$ restricted to 90° (comparison of minimum geometry and constrained geometry given in Supporting Information, Figure S4). In order to generate a full sphere of input geometries for subsequent calculations, the optimized structure was transformed as follows: The sulfur atom was placed on the origin of the coordinate system and the entire complex was rotated until the halogen atom was positioned on the positive x -axis. Let α denote the angle of rotation counterclockwise around the z -axis and β denote the angle of rotation counterclockwise around the x -axis. α was gradually increased from 0° to 180° in steps of 5° . For each α -value, β was varied from 0° to 355° in steps of 5° , leading to a total number of 2664 halogen positions distributed on a sphere. The structure of the ligand was not altered during this transformation process. Calculations were done as single point (SCF) calculations using the TPSS-D(RI)/TZVPP method.

For visualization purposes, several Python scripts were written and executed in PyMOL.⁵² Interaction energies were partitioned into bins and spectrum colors (red to blue to purple) were assigned. At the positions of the halogens small CONE objects with the appropriate coloring were generated. For the 2D plots, 36×36 ($\delta_1 \times \delta_2$) matrices were generated from a worksheet of δ_1 , δ_2 , and complex formation energy values using the XYZ gridding tool in Origin 8.5 (OriginLab, Northampton, MA). We used “weighted average” as the gridding method with a search radius of 15. Final figures were generated as 2D contour plots.

■ ASSOCIATED CONTENT

S Supporting Information. A comparison of MP2 and DFT methods for the complex formation energies and interaction geometries is given in Table S1. The results of the basis set extrapolation procedure to obtain CCSD(T)/CBS energies are given in Table S2. A comparison of the freely optimized geometries for iodobenzene with all three methionine model systems is given in Figure S1. Distance scans with methionine model systems of different sizes are given in Figure S2. A comparison of common DFT-D functionals with MP2 exemplified in the halobenzene–MET1 distance scans is presented in Figure S3. A comparison of the constrained input structures for the spherical scans and freely optimized structures is given in Figure S4, and depictions of interaction spheres from nonconstrained spherical scans are given in Figure S5. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +4970712974567. Fax: +497071295637. E-mail: frank.boeckler@uni-tuebingen.de.

■ ACKNOWLEDGMENT

We acknowledge the help of bachelor students Leo Rosenkranz and Alwin Balhuber in the initial stages of this project. High-performance computing resources of the LRZ (Leibniz Rechenzentrum) Munich and of the BW-Grid were kindly made available by the federal states of Bayern and Baden-Wuerttemberg.

■ ABBREVIATIONS:

PDB, Brookhaven Protein Data Bank; QM/MM, quantum mechanics/molecular mechanics; DFT, density functional theory; DFT-D, density functional theory augmented by an empirical dispersion correction; MP2, second-order Møller–Plesset perturbation theory; SCS-MP2, spin-component scaled second-order Møller–Plesset perturbation theory; CCSD(T), coupled-cluster method with single, double and perturbative triple excitations; IL-2, interleukin-2; MET1–MET3, the various models for methionine; Met, methionine; SCF, self-consistent field calculation; ECP, effective core potential; BSSE, basis set superposition error

■ REFERENCES

- (1) Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010**, *53*, 5061–5084.
- (2) Hassel, O. Structural Aspects of Interatomic Charge-Transfer Bonding. *Science* **1970**, *170*, 497–502.
- (3) Pierangelo, M.; Franck, M.; Tullio, P.; Giuseppe, R.; Giancarlo, T. Halogen Bonding in Supramolecular Chemistry. *Angew. Chem., Int. Ed.* **2008**, *47*, 6114–6127.
- (4) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16789–16794.
- (5) Clark, T.; Hennemann, M.; Murray, J.; Politzer, P. Halogen Bonding: The σ -Hole. *J. Mol. Model.* **2007**, *13*, 291–296.
- (6) Murray, J.; Lane, P.; Politzer, P. Expansion of the σ -Hole Concept. *J. Mol. Model.* **2009**, *15*, 723–729.
- (7) Murray, J. S.; Lane, P.; Politzer, P. A Predicted New Type of Directional Noncovalent Interaction. *Int. J. Quantum Chem.* **2007**, *107*, 2286–2292.

- (8) Murray, J. S.; Riley, K. E.; Politzer, P.; Clark, T. Directional Weak Intermolecular Interactions: σ -Hole Bonding. *Aust. J. Chem.* **2010**, *63*, 1598–1607.
- (9) Politzer, P.; Lane, P.; Concha, M.; Ma, Y.; Murray, J. An Overview of Halogen Bonding. *J. Mol. Model.* **2007**, *13*, 305–311.
- (10) Murray, J.; Lane, P.; Clark, T.; Politzer, P. σ -Hole Bonding: Molecules Containing Group VI Atoms. *J. Mol. Model.* **2007**, *13*, 1033–1038.
- (11) Murray, J. S.; Lane, P.; Politzer, P. Simultaneous Sigma-Hole and Hydrogen Bonding by Sulfur- and Selenium-Containing Heterocycles. *Int. J. Quantum Chem.* **2008**, *108*, 2770–2781.
- (12) Riley, K. E.; Hobza, P. Investigations into the Nature of Halogen Bonding Including Symmetry Adapted Perturbation Theory Analyses. *J. Chem. Theory Comput.* **2008**, *4*, 232–242.
- (13) Riley, K. E.; Murray, J. S.; Politzer, P.; Concha, M. C.; Hobza, P. Br---O Complexes as Probes of Factors Affecting Halogen Bonding: Interactions of Bromobenzenes and Bromopyrimidines with Acetone. *J. Chem. Theory Comput.* **2009**, *5*, 155–163.
- (14) Lu, Y.; Shi, T.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. Halogen Bonding, a Novel Interaction for Rational Drug Design?. *J. Med. Chem.* **2009**, *52*, 2854–2862.
- (15) Lu, Y.; Wang, Y.; Zhu, W. Nonbonding Interactions of Organic Molecules in Biological Systems: Implications for Drug Discovery and Biomolecular Design. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4543–4551.
- (16) Voth, A. R.; Khoo, P.; Oishi, K.; Ho, P. S. Halogen Bonds As Orthogonal Molecular Interactions to Hydrogen Bonds. *Nat. Chem.* **2009**, *1*, 74–79.
- (17) Politzer, P.; Murray, J. S.; Lane, P. Sigma-Hole Bonding and Hydrogen Bonding: Competitive Interactions. *Int. J. Quantum Chem.* **2007**, *107*, 3046–3052.
- (18) Shields, Z. P.; Murray, J. S.; Politzer, P. Directional Tendencies of Halogen and Hydrogen Bonds. *Int. J. Quantum Chem.* **2010**, *110*, 2823–2832.
- (19) Politzer, P.; Murray, J. S.; Clark, T. Halogen Bonding: An Electrostatically-Driven Highly Directional Noncovalent Interaction. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7748–7757.
- (20) Fedorov, O.; Huber, K.; Eisenreich, A.; Filippakopoulos, P.; King, O.; Bullock, A. N.; Szklarczyk, D.; Jensen, L. J.; Fabbro, D.; Trappe, J.; Rauch, U.; Bracher, F.; Knapp, S. Specific CLK Inhibitors from a Novel Chemotype for Regulation of Alternative Splicing. *Chem. Biol.* **2011**, *18*, 67–76.
- (21) Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; Plancher, J.-M.; Hartmann, G.; Banner, D. W.; Haap, W.; Diederich, F. Systematic Investigation of Halogen Bonding in Protein–Ligand Interactions. *Angew. Chem., Int. Ed.* **2011**, *50*, 314–318.
- (22) Thanos, C. D.; DeLano, W. L.; Wells, J. A. Hot-Spot Mimicry of a Cytokine Receptor by a Small Molecule. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15422–15427.
- (23) Liu, L.; Baase, W. A.; Matthews, B. W. Halogenated Benzenes Bound within a Non-Polar Cavity in T4 Lysozyme Provide Examples of I...S and I...Se Halogen-Bonding. *J. Mol. Biol.* **2009**, *385*, 595–605.
- (24) Estébanez-Perpiñá, E.; Arnold, L. A.; Nguyen, P.; Rodrigues, E. D.; Mar, E.; Bateman, R.; Pallai, P.; Shokat, K. M.; Baxter, J. D.; Guy, R. K.; Webb, P.; Fletterick, R. J. A surface on the androgen receptor that allosterically regulates coactivator binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16074–16079.
- (25) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
- (26) Turbomole v6.2 (June 2010), available from <http://www.turbomole.com>. 2010.
- (27) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (28) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically Convergent Basis Sets with Relativistic Pseudopotentials. II. Small-Core Pseudopotentials and Correlation Consistent Basis Sets

for the Post-d Group 16–18 Elements. *J. Chem. Phys.* **2003**, *119*, 11113–11123.

(29) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of Approximate Integrals in Ab Initio Theory. An Application in MP2 Energy Calculations. *Chem. Phys. Lett.* **1993**, *208*, 359–363.

(30) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: Optimized Auxiliary Basis Sets and Demonstration of Efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143–152.

(31) Hättig, C. Optimization of Auxiliary Basis Sets for RI-MP2 and RI-CC2 Calculations: Core-Valence and Quintuple- ζ Basis Sets for H to Ar and QZVPP Basis Sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.

(32) Hellweg, A.; Hättig, C.; Höfener, S.; Klopper, W. Optimized Accurate Auxiliary Basis Sets for RI-MP2 and RI-CC2 Calculations for the Atoms Rb to Rn. *Theor. Chem. Acc.* **2007**, *117*, 587–597.

(33) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.

(34) Baerends, E. J.; Ellis, D. E.; Ros, P. Self-Consistent Molecular Hartree–Fock–Slater Calculations I. The Computational Procedure. *Chem. Phys.* **1973**, *2*, 41–51.

(35) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On Some Approximations in Applications of X Alpha Theory. *J. Chem. Phys.* **1979**, *71*, 3396–3402.

(36) Weigend, F. Accurate Coulomb-Fitting Basis Sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

(37) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098.

(38) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **1986**, *33*, 8822.

(39) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle–Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785.

(40) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative Assessment of a New Nonempirical Density Functional: Molecules and Hydrogen-Bonded Complexes. *J. Chem. Phys.* **2003**, *119*, 12129–12137.

(41) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(42) Grimme, S. Semiempirical GGA-type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(43) Grimme, S. Improved Second-Order Møller–Plesset Perturbation Theory by Separate Scaling of Parallel- and Antiparallel-Spin Pair Correlation Energies. *J. Chem. Phys.* **2003**, *118*, 9095–9102.

(44) Dunning, J. T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(45) Woon, D. E.; Dunning, J. T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum through Argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(46) Peterson, K. A.; Shepler, B. C.; Figgen, D.; Stoll, H. On the Spectroscopic and Thermochemical Properties of ClO, BrO, IO, and Their Anions. *J. Phys. Chem. A* **2006**, *110*, 13877–13883.

(47) Weigend, F.; Kohn, A.; Hättig, C. Efficient Use of the Correlation Consistent Basis Sets in Resolution of the Identity MP2 Calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.

(48) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. MOLPRO, version 2006.1, a package of ab initio programs; <http://www.molpro.net>. 2006.

(49) Hobza, P.; Šponer, J. Toward True DNA Base-Stacking Energies: MP2, CCSD(T), and Complete Basis Set Calculations. *J. Am. Chem. Soc.* **2002**, *124*, 11802–11808.

(50) Jurecka, P.; Šponer, J.; Cerny, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) complete basis set limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, And Amino Acid Pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(51) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-Set Convergence in Correlated Calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(52) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific LLC: Palo Alto, CA, USA. 2008.

Using the Wimley–White Hydrophobicity Scale as a Direct Quantitative Test of Force Fields: The MARTINI Coarse-Grained Model

Gurpreet Singh[†] and D. Peter Tieleman^{*,†,‡}

[†]Department of Biological Sciences, and [‡]Institute for Biocomplexity and Informatics, 2500 University Drive N.W., University of Calgary, Calgary, Alberta, Canada T2N 1N4

ABSTRACT: The partitioning of proteins and peptides at the membrane/water interface is a key step in many processes, including the action of antimicrobial peptides, cell-penetrating peptides, and toxins, as well as signaling. To develop a computational model that can be used to accurately represent such systems, the underlying model must be able to quantitatively represent the partitioning preferences of amino acids in the lipid membrane. The MARTINI model provides a consistent set of parameters for building coarse-grained models of systems involving lipids and proteins. Even though MARTINI is parametrized to reproduce the partitioning behavior of small molecules, its ability to reproduce partitioning preferences of amino acids at lipid/water interfaces has never been tested. In this study, we measured the partitioning free energies of side chains of amino acids using alchemical simulations and umbrella sampling. The pentapeptides of sequence Ac-WLXLL were simulated at the POPC/water and cyclohexane/water interfaces using MARTINI, and the computed free energies were compared with the Wimley–White hydrophobicity scale. The free energy values obtained using the free energy perturbation, thermodynamic integration, and umbrella sampling methods were compared to gain insight into the most efficient method and the degree of sampling required to obtain statistically accurate free energies for use with atomistic force fields in future work. With the standard MARTINI water model, the amino acids D, E, K, and R were found to be significantly too favorable in hydrophobic environments, whereas with the polarizable water model, the amino acids D, E, K, and R were found to give correct free energies of partitioning. The amino acids P and F showed significant deviations from the experimental values. This model system will be used in future improvements to the MARTINI model.

1. INTRODUCTION

Over the past few decades, computer simulations of biomolecules have become a useful tool in understanding the dynamics and mechanism of function with atomistic details.¹ New sampling methods coupled with increased computational resources have made it possible to simulate the folding of small peptides and proteins.^{2–4} The underlying parameters or force fields used in atomistic representations of proteins have matured to the point that free energies calculated using computer simulation methods are in quantitative agreements with experiments.⁵

At the current state of simulation algorithms and computer hardware, the time scales of typical atomistic simulations are limited to (sub)microseconds, and the system size is limited to hundreds of thousands of atoms. Sampling all relevant degrees of freedom still remains a major challenge in biomolecular simulations, considering that many interesting biological processes occur on time scales beyond the reach of atomistic simulations.

To simulate processes such as vesicle fusion, pore formation in membranes, and the formation of protein complexes, several simplified models have been developed.^{6–8} The use of these coarse models allows for simulations of larger systems, for longer time scales, while still providing significant structural detail. Marrink and co-workers developed one such coarse-grained model for simulations of lipids and surfactant, coined MARTINI.^{9,10} In the MARTINI model, several atoms are grouped together in one bead that interacts with other beads through an effective potential. This model was successfully applied to study long-time-scale phenomena such as lipid monolayer collapse, domain formation in vesicles, and vesicle fusion.^{11–13}

Consistent with the original design philosophy, the interaction parameters for the beads representing amino acids were chosen to reproduce the experimental partitioning free energies of amino acid analogues between water and cyclohexane.¹⁴ The experimental data on partitioning free energies of amino acid analogues have also been used to parametrize recent versions of the GROMOS force field.¹⁵ However, as free energy calculations are computationally intensive, small-molecule analogues rather than complete amino acids or peptides are generally used and compared with experimental data to obtain insight into the performance of various classical force fields.^{16–19}

To test our ability to accurately model lipid–protein interactions, critical experimental data are essential, but such data are rare.²⁰ In addition, some accurate data are difficult to reproduce by simulations because of the time scales or system complexity that is required for an accurate comparison.^{21,22} Wimley and White created a whole-residue hydrophobicity scale by measuring the partitioning free energies of model peptides of the sequence Ac-WLXLL, where X can be any of the 20 amino acids, at a phospholipid (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine, POPC)/water interface and in a water/octanol mixture.^{23,24} This scale is routinely used in hydrophobicity plots for the prediction of transmembrane regions in membrane proteins.^{25,26} These peptides, because of their small size, can be simulated relatively easily and can be used to gain further insight into the thermodynamics of lipid–protein interactions. Some of these peptides have previously been studied using atomistic simulations at

Received: April 15, 2011

Published: June 06, 2011

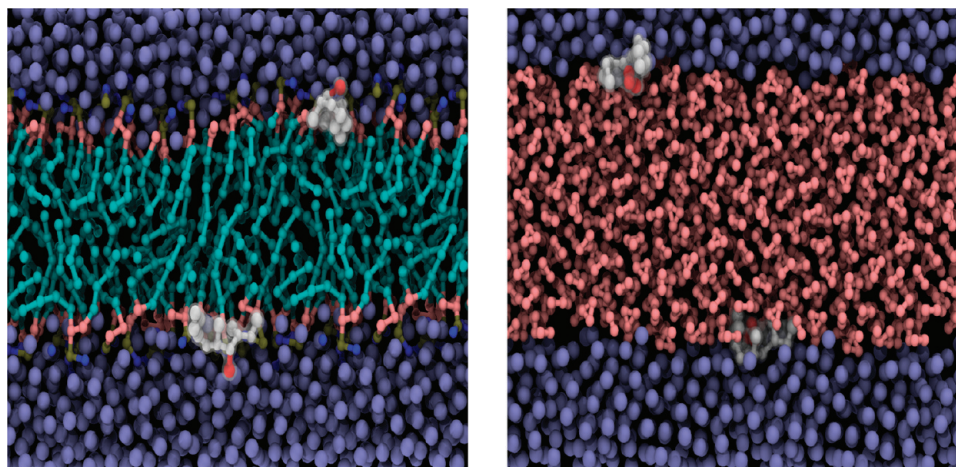


Figure 1. Snapshots of simulations depicting WLELL peptides at the POPC/water interface (left) and WLFL peptides at the cyclohexane/water interface (right). Water beads are colored ice blue, cyclohexane is colored pink, peptide beads are colored white, and X3 beads are colored red.

cyclohexane/water, octanol/water, and 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC)/water interfaces and were found to adsorb readily to the interface.^{27,28}

The potential of mean force (PMF) profiles of amino acid analogues across the DOPC bilayer were studied by Macallum et al.^{29,30} In that study, the computed bulk to interface transfer free energies were compared with free energies of partitioning of Wimley–White peptides at octanol/water and POPC/water interfaces, and moderate correlations (correlation coefficients of 0.84 and 0.61, respectively) were observed. These results were attributed to the fact that the Wimley–White scale represents the free energy of partitioning of side chains in the context of the pentapeptide rather than partitioning of small-molecule analogues.

In this study, we computed the partitioning free energies of side chains of all 20 amino acids as part of the WLXLL peptide at the cyclohexane/water and POPC/water interfaces using the MARTINI model. The latter set of values can be directly compared with the experimental results of the Wimley–White scale. The main motivations of this study were to investigate the feasibility of a direct calculation of the Wimley–White scale for use in the development and testing of force fields, to test and possibly improve the MARTINI model's ability to describe lipid–protein interactions, and to establish a well-defined experimental test system and simulation protocol that will be useful in the future development of the MARTINI model.

2. METHODS

2.1. Coarse-Grained Simulations. All simulations were performed with the GROMACS software (version 4.0 and above).³¹ In all simulations, a leapfrog integrator was used. All systems were simulated at a temperature of 300 K and a pressure of 1 bar using periodic boundary conditions. Peptides with the sequence WLXLL, where X can be any of the 20 amino acids, were simulated at the cyclohexane/water interface, the POPC/water interface, and in bulk water, using MARTINI force field v2.1¹⁴ for protein parameters and v2.0¹⁰ for lipid parameters. In the MARTINI model, four heavy atoms are generally represented by a single interaction site. Polar (P), nonpolar (N), apolar (C), and charged (Q) are the four main types of interaction sites. These primary interaction types are further divided into subtypes based on either hydrogen-bonding capabilities or degree of

polarity. To model the N-terminal cap of the Wimley–White peptides, the default N-terminal bead was replaced with the P₅ bead.

Amino acids were represented in their default protonations state for pH 8; that is, the D and E side chains were negatively charged, the K and R side chains were positively charged, and the rest were neutral. The C terminus was modeled with a negative charge to mimic the deprotonated state at pH 8, except for X = K or R, which were represented by a neutral C-terminus bead (P₅) to mimic pH 2. Additional simulations for the system with X = A were carried out with a neutral C terminus.

For the systems containing either the cyclohexane/water or POPC/water interface, the initial configuration was prepared by inserting two peptides in the water phase and deleting the overlapping water molecules. In the case of the system containing the POPC/water interface, the initial adsorption of the peptides at opposite leaflets of the bilayer was facilitated by subjecting the peptides to a constant acceleration of 0.05 nm ps⁻² in opposite directions, for a duration of 30 ns. The system was further simulated without any acceleration for 500 ns, during which time the temperature was maintained by coupling the system to a Nosé–Hoover^{32,33} thermostat with a coupling constant of 1.5 ps and the pressure was maintained by coupling the system semi-isotropically to a Parrinello–Rahman barostat³⁴ with a coupling constant of 2.5 ps. The systems containing cyclohexane/water were coupled to a Berendsen pressure bath,³⁵ in only the Z dimension (perpendicular to the cyclohexane/water interface), with a coupling constant of 1 ps and were simulated for an additional 2 μs. Snapshots of the cyclohexane/water and POPC/water systems with peptides are shown in Figure 1 and were prepared using VMD software.³⁶

2.1.1. Free Energy Calculations. The free energies of partitioning of the side-chain residues can be calculated using the thermodynamic cycle shown in Figure 2. Simulations in which a chemical species is transformed into another through an unphysical pathway are referred to as alchemical simulations.³⁷ Alternatively, a system can be transformed from one state into another by performing simple displacements. Thus, the free energies of partitioning can be obtained by either following the vertical lines using alchemical simulations or by following the horizontal lines and computing the potential of mean force as a function of distance from the interface, for example, by using

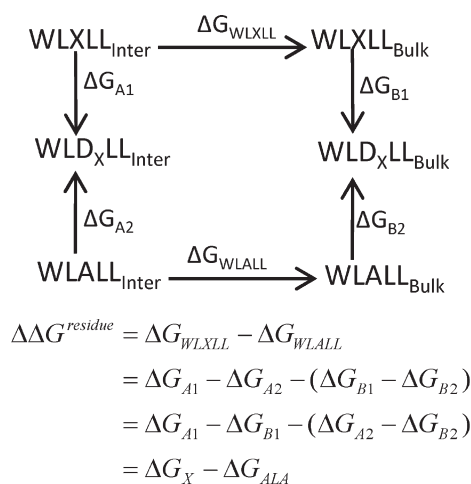


Figure 2. Thermodynamic cycle for the calculation of free energies of partitioning of amino acids, $\Delta\Delta G^{\text{residue}}$, at interfaces. $\text{WLXLL}_{\text{Inter}}$ and $\text{WLXLL}_{\text{Bulk}}$ represent the peptide at the interface and in the bulk, respectively. WLD_XLL represents a peptide in which the side-chain atoms of residue X have been converted to dummy atoms that have no Lennard-Jones or Coulombic interactions and the backbone bead is converted to P_S .

umbrella sampling (US).³⁸ Free energies from alchemical simulations can be computed using the thermodynamic integration (TI) method, in which the derivative of the Hamiltonian is computed along a reaction coordinate describing the alchemical transformation, which is then integrated numerically to obtain the free energy difference,³⁹ or using the free energy perturbation (FEP) method, in which the free energy difference is estimated from the exponential average of the difference in energy between the states.⁴⁰ To calculate the free energy estimates from the data obtained using either the FEP or US approach, methods such as the weighted histogram analysis method (WHAM) or the multiple Bennett acceptance ratio (MBAR) can be employed.^{41,42}

The partitioning free energies for amino acids were calculated at the cyclohexane/water interface and POPC/water interface using the thermodynamic cycle shown in Figure 2. At the cyclohexane/water interface, free energies were computed using both alchemical simulations and umbrella sampling. For alchemical simulations, the free energies were computed using both FEP and TI. The MBAR method implemented in the pyMBAR⁴¹ program was used to estimate free energies and uncertainties from FEP data, and for TI, the uncertainties were calculated using the method described by Hess, which involves fitting an analytic function to a standard error estimate of a measured observable ($\partial H/\partial\lambda$, in this case) as a function of block size.⁴³

The initial configurations of production runs were used as the starting configurations for each intermediate state, and the peptides were accelerated toward the interface for 200 ps. Nineteen equally spaced intermediate states were defined using the coupling parameter λ , which linearly switches off Lennard-Jones (LJ) and Coulomb interactions of the perturbed side chain as λ goes from 0 to 1. To avoid singularities, soft-core interactions were used for LJ interactions.⁴⁴ All simulations included 10 ns of equilibration, followed by production runs of 0.5 μs for each λ value. For TI, the numerical integration of $\partial H/\partial\lambda$ was carried out using Simpson's rule. The potential energy differences between all intermediate states are required in MBAR and were obtained

by recalculating the potential energies from the trajectories using the -rerun option of the GROMACS mdrun program. For peptides containing charged residues, free energy calculations were also performed using the polarizable MARTINI water model, and the charges were decoupled separately.⁴⁵

The PMF profiles for pulling the entire peptide from an interface to the bulk were computed using the distance between the center of mass (COM) of cyclohexane and the COM of the peptide as the reaction coordinate. Thirty-one equally spaced windows (0.1 nm) were used. The COM of the peptide was held at its position by applying a harmonic potential with a force constant of 1000 kJ nm^{-2} . After initial equilibration, data were collected for 250 ns for each window, and the entire simulation was repeated twice. The data were analyzed using the weighed histogram method (WHAM) as implemented in GROMACS.⁴⁶ The free energy of adsorption at the interface from the bulk was calculated as $\Delta G_X = -RT \ln[\int_{z_f}^{z_s} e^{-\Delta G(z)/RT} dz]$, where $z_s = 1.5$ nm and $z_f = 4.0$ nm represent the coordinates of the cyclohexane/water interface and bulk water, respectively. The integrations were carried out with numerical integration methods using Simpson's rule. The averages and standard errors were computed using the two data sets and are reported.

All peptides were modeled without any dihedral potential on consecutive backbone beads. However, to determine the effects of different pseudodihedral parameters defining the backbone geometry, two peptides (with X = I or E) were chosen, and the free energy of partitioning at the cyclohexane/water interface was computed with different backbone parameters representing helix, extended, turn, and bend structures and with a pseudodihedral potential derived from an atomistic simulation (only WLELL). In all cases, the backbone was represented by a P_S bead.

2.2. All-Atom Simulations. All atomistic simulations were performed using the OPLS force field and SPCE water model.^{47,48} The N terminus of the peptide was capped with an acetyl group, and the C terminus was deprotonated. To obtain the backbone dihedral potential for the coarse-grained WLELL peptide, 40 atomistic simulations of 40-ns duration were carried out. The probability densities of pseudodihedrals between four consecutive C- α carbons were measured, and the dihedral potential was derived by the Boltzmann inversion of the average probability density.

3. RESULTS AND DISCUSSION

Experimental studies of these peptides in POPC/water have shown that the peptides partition at the interface, without deep penetration into the hydrocarbon core.²⁴ This behavior was attributed to the lack of secondary structure in the peptides, resulting in a high energy cost of partitioning of the non-hydrogen-bonded peptide backbone into the hydrocarbon core. In agreement with the experimental observations, in our simulations at both interfaces, the peptides were adsorbed at the interface, with the charged C terminus oriented toward water.

Because we used two peptides (one at each interface), we looked at the convergence of the simulations by comparing the time evolution of the average probability density of the center of mass (COM) of side chains at the interface. Figure 3 shows the time evolution of side chain X for a few selected peptides. The converged distributions were obtained within 300 ns of simulation, and further sampling did not change the shape of the distributions. Also, similar distributions were obtained from the simulations of peptides at the cyclohexane/water interface, with

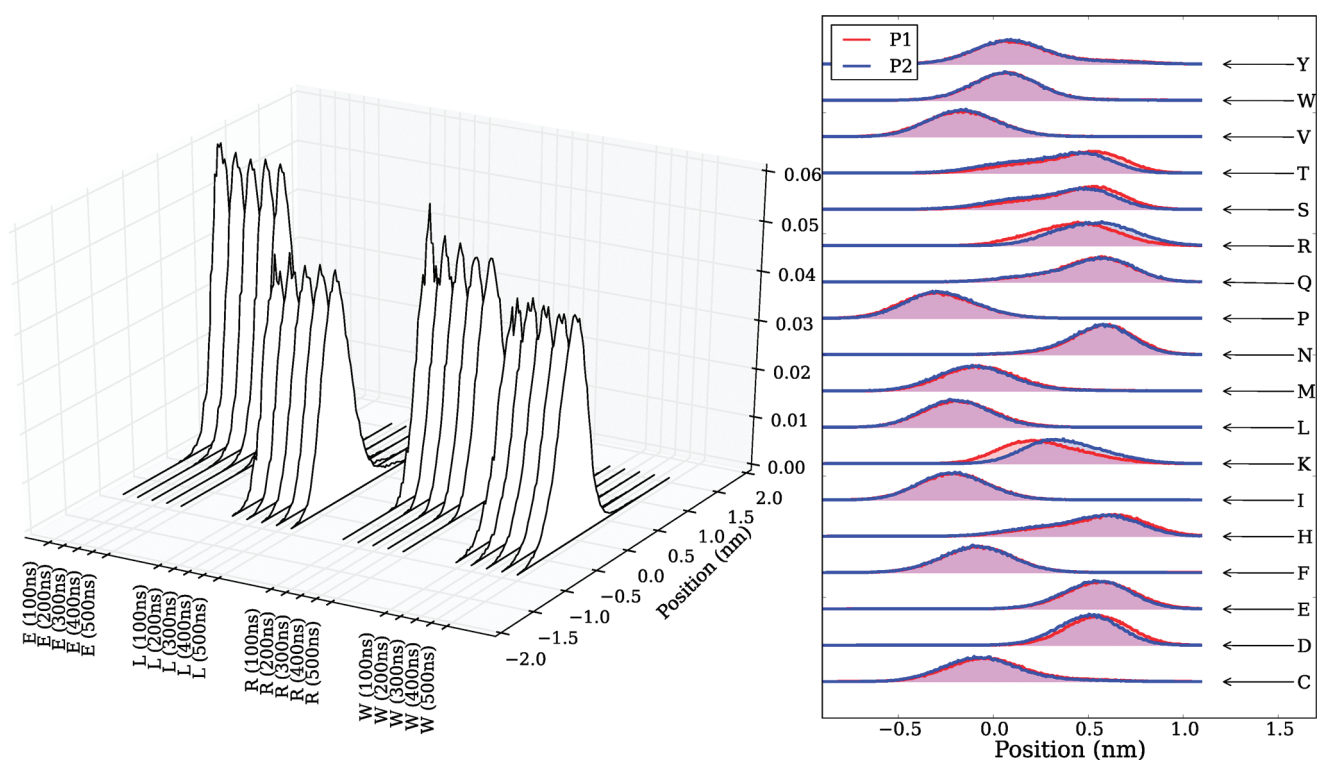


Figure 3. (Left) Probability density of the side-chain COM as a function of the position at the interface and the simulation length at the POPC/water interface for a few selected peptides. (Right) Distributions of the side-chain COM of amino acids at the cyclohexane/water interface for simulations with (P2) and without (P1) an angle potential. For comparison of the distributions of different amino acids, the x axis was adjusted such that the intersection of the densities of cyclohexane and water was always at 0. In the case of the POPC/water interface, the center between Q_0 and COM of the first two C_1 positions was set as 0.

or without angle potential parameters. Figure 4 shows the distributions of side chain X for all peptides. At the cyclohexane/water interface, the density of water drops from its bulk value to almost 0 within ~ 1 nm. The change in density of water is more gradual near the membrane interface, with the decrease in bulk water density marked by an increase in density of the choline group (Q_0), followed by that of the phosphate group (Q_a), both of which interact most favorably with the water bead (P_4). The region of the glycerol ester group (N_a beads) is chemically most diverse, as all of the beads have a presence in that region. The decrease in density of water from its bulk value to almost 0 occurs in ~ 1.7 nm near POPC. At the POPC/water interface, the density distributions of D and E coincide with those of beads Q_0 and Q_a , whereas the F, I, L, and P density distributions coincide with those of the C_1 and N_a beads. Even though the POPC/water interface is far more complex and chemically diverse than the cyclohexane/water interface, certain similarities between the side-chain distributions at the two interfaces can be seen. All of the distributions have a single maximum, and the position of this maximum correlates with the hydrophobicity of the amino acids, with polar side chains preferring water and apolar side chains oriented away from bulk water. The similarity in side-chain distributions at the cyclohexane/water and POPC/water interfaces indicates that the partial density of water at the interface has the strongest effect on the side-chain positions.

The side-chain distributions of the rest of the amino acids in the pentapeptide (W1, L2, L4, and L5) for a few selected peptides are shown in Figure 5. The type of amino acid at X does not significantly affect the positions of the rest of the amino

acids, indicating that the peptides have enough conformational flexibility to allow almost independent orientation of the individual X amino acid. The distributions obtained from the two peptides in the system are very similar, indicating that the amount of sampling was adequate.

3.1. Free Energies. The partitioning free energies of side chains at the cyclohexane/water interface obtained using the FEP, TI, and US methods are reported in Table 1. Although similar free energy estimations were obtained for all three methods, the estimated errors were smallest for the free energy estimates from FEP using the MBAR method. As both the TI and US methods involve numerical integration of the primary results ($\partial H/\partial \lambda$ and an average force as a function of distance, respectively), the resulting values have comparatively larger uncertainties. However, PMF profiles along physically relevant reaction coordinates can provide additional insight; for example, the PMF profiles at the cyclohexane/water interface of one representative amino acid from each group are shown in Figure 6. All of these profiles have a single minimum at the interface, indicating that the adsorption of the peptide from bulk water is barrierless, and the free energy increases rapidly while moving away from the interface in either direction. Very similar PMF profiles for Q, E, and A also indicate that the standard MARTINI model does not sufficiently distinguish between water and hydrophobic environments for charged particles.

The free energies of partitioning of side-chain residues at the POPC/water interface, $\Delta \Delta G^{\text{residue}}$, as computed using the FEP method with the thermodynamic cycle in Figure 2, are shown in Figure 7. Based on experimental free energies of partitioning, the amino acids can be roughly divided into five groups, as shown in

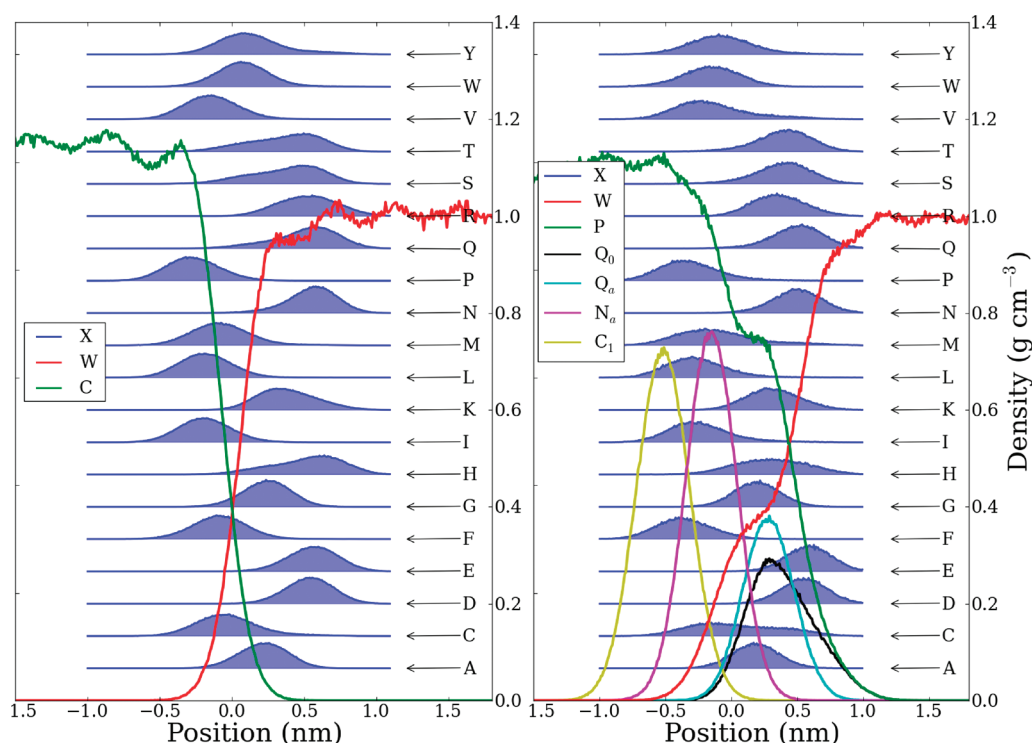


Figure 4. Distributions of the side-chain COM of amino acids at the cyclohexane/water (left) and POPC/water (right) interfaces. The side-chain distributions (legend label X) are filled. The partial densities of water, cyclohexane, and POPC are labeled as W, C, and P respectively. In the case of the POPC/water interface, the density distributions of Q_0 , Q_a , N_a , and C_1 beads are also shown. For the N_a and C_1 distributions, the COMs of the two N_a beads and the first two C_1 beads, respectively, were used. For comparison of the distributions of different amino acids, the x axis was adjusted such that the intersection of the densities of cyclohexane and water was always at 0. In the case of the POPC/water interface, the center between Q_0 and the COM of the first two C_1 positions was set as 0. The average densities of water, cyclohexane, and POPC as functions of the Z dimension are also plotted.

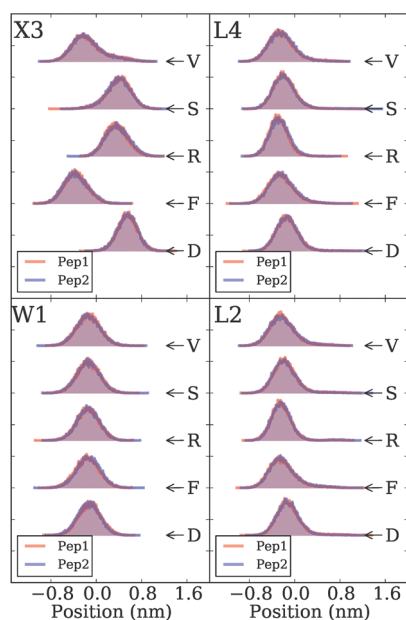


Figure 5. Distributions of the side-chain COM of amino acids at the POPC/water interface. Pep1 and Pep2 are the two peptides. The distributions obtained from the two peptides (one at each interface) are shown for W1 (lower left panel), L2 (lower right panel), X3 (top left panel), and L4 (top right panel) side chains.

Figure 7. Group I contains the amino acids that are most hydrophobic on the Wimley–White scale, in particular those

with aromatic side chains; group II consists of amino acids with aliphatic side chains such as L and I and side chains containing sulfur (C and M); group III consists of amino acids whose partitioning free energies are very similar to that of alanine (A, S, T, V, H, and G); group IV includes side chains containing amide groups and proline (N, Q_a , and P); and group V consists of the charged amino acids (D, E, K, and R). Considering the coarse-grained description of the system, the experimental and computed values are in good agreement for groups II and III. F, P, and the amino acids in group V have the largest deviations from the experimental values.

The free energies obtained at the POPC/water and cyclohexane/water interfaces are compared in Figure 8. The side chains of amino acids P and L are the most hydrophobic at the cyclohexane/water interface, whereas W is the most hydrophobic at the POPC/water interface. Also, the cyclohexane-to-water transfer free energies are generally more positive than the transfer free energies for POPC/water, except for W, Y, and H. In membrane proteins, the W and Y residues are known to be preferentially located at the regions corresponding to the membrane interfaces. This preference is generally attributed to several factors, including their flat rigid shape, cation– π interactions, and hydrogen bonding.^{49–51}

3.1.1. Group V. For side chains containing a net charge, Figure 4 shows that at the cyclohexane/water interface, the peptides orient themselves to position these charged side chains toward the bulk water, thus correctly preferring the water phase over cyclohexane. However, the calculated free energies of partitioning are close to 0, as the free energies of annihilation of charged

Table 1. Values of $\Delta\Delta G$ (kJ mol^{-1}) at the Cyclohexane/Water Interface Calculated Using the FEP, TI, and US Methods

AA	$\Delta\Delta G_{\text{FEP}}^{\text{res}}$	$\Delta\Delta G_{\text{TI}}^{\text{res}}$	$\Delta\Delta G_{\text{US}}^{\text{res}}$
C	6.09 ± 0.03	6.1 ± 0.1	6.2 ± 0.1
D	0.68 ± 0.04	0.7 ± 0.1	-0.6 ± 0.1
E	0.34 ± 0.04	0.3 ± 0.1	-1.4 ± 0.2
F	10.45 ± 0.05	10.5 ± 0.1	10.4 ± 0.2
G	-1.011 ± 0.001	-1.02 ± 0.02	-1.5 ± 0.2
H	-2.20 ± 0.05	-2.2 ± 0.2	-1.9 ± 0.2
I	14.24 ± 0.03	14.2 ± 0.1	13.8 ± 0.3
K	1.76 ± 0.05	1.8 ± 0.1	2.6 ± 0.2
L	14.68 ± 0.03	14.7 ± 0.1	14.5 ± 0.2
M	7.83 ± 0.04	7.8 ± 0.1	7.9 ± 0.1
N	-1.38 ± 0.04	-1.4 ± 0.1	-1.2 ± 0.4
P	20.24 ± 0.04	20.3 ± 0.1	20.0 ± 0.1
Q	-1.37 ± 0.04	-1.4 ± 0.1	-1.3 ± 0.3
R	-1.12 ± 0.05	-1.2 ± 0.1	0.4 ± 0.5
S	-0.27 ± 0.03	-0.3 ± 0.1	-0.4 ± 0.1
T	-0.25 ± 0.04	-0.2 ± 0.1	-0.1 ± 0.3
V	11.52 ± 0.03	11.5 ± 0.1	11.1 ± 0.3
W	5.43 ± 0.06	5.4 ± 0.2	5.6 ± 0.1
Y	2.15 ± 0.05	2.2 ± 0.1	2.0 ± 0.3

side chains in the bulk and at the interface are very similar, suggesting that the charged side chains are effectively in bulk water. At the POPC/water interface, side chains D and E are close to positively charged choline beads, whereas K and R have maxima close to the negatively charged phosphate bead of POPC. Even in this case, the calculated free energy values are close to 0. The water molecule in the standard MARTINI model is represented by a Lennard-Jones bead, and the interaction level of charged beads with the water bead is the same as for other polar beads, resulting in underestimation of the hydration free energies of charged side chains. This problem was recently fixed by introducing point charges into the water model to make the water molecules polarizable and by modifying the interactions of Q-type particles.⁴⁵ Using this polarizable water model in the free energy calculations of charged side chains resulted in significant improvements (see Figure 7) at the POPC/water interface. In Figure 9, the side-chain COM distributions obtained using the polarizable water model are compared with those obtained from standard MARTINI. The side-chain COM distributions with respect to the interface in the two cases are similar.

In comparison to POPC, cyclohexane is far more apolar than the lipid head groups. Therefore, it can be argued that the free energy of partitioning of charged residues should be more negative; alternatively, the cyclohexane/water interface is narrower than the POPC/water interface, and reorientation of peptides at cyclohexane/water interface could place the charged side chains effectively in bulk water, resulting in low partitioning free energies for the charged residues as compared to those at the POPC/water interface. Because of the unavailability of experimental data for these peptides at the cyclohexane/water interface, it is not possible to validate the computed values.

3.1.2. F and P. In the MARTINI force field, the aromatic side chains are modeled using ring particle beads that have lower masses than other beads, with ring–ring interactions that are scaled ($\sigma = 0.43$ and ϵ is scaled to 75% of its original value) compared to interactions with the rest of the beads. From Figure 4, it can be seen that the side chain of amino acid F predominantly interacts with the tail beads (bead type C_1) of POPC molecules. The computed free energy of partitioning of F

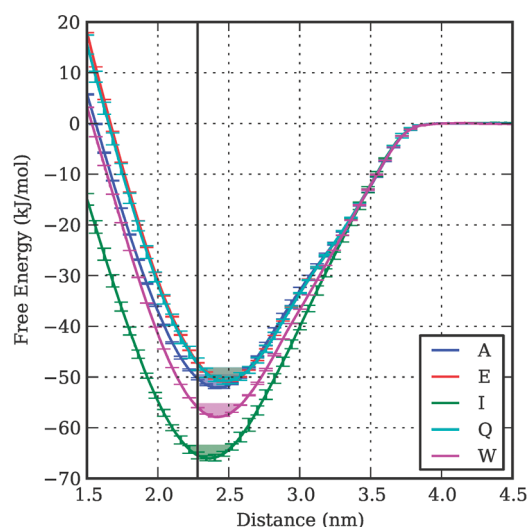


Figure 6. PMF profiles for peptides with X = A, E, I, Q, or W at the cyclohexane/water interface as a function of the distance between the COMs of cyclohexane and the peptide. The vertical black line represents the interface, that is, where the densities of water and cyclohexane are equal. The minimum for each profile is filled up to $1RT$. The error bars represent the variance multiplied by 10 so that they can be distinguished from the lines.

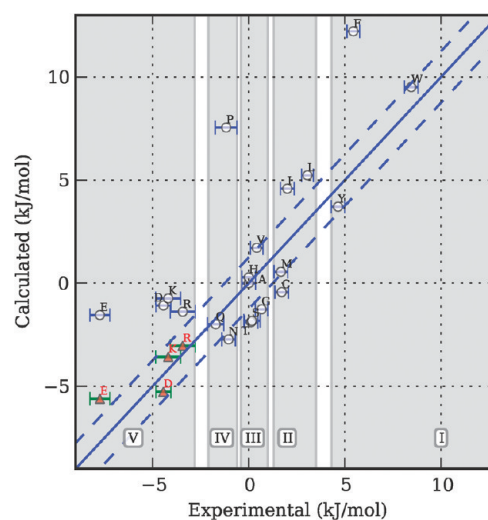


Figure 7. Comparison of experimental (horizontal axis) and calculated (vertical axis) partitioning free energies at the POPC/water interface as obtained using the standard MARTINI FF (circles). Residues are indicated by a corresponding single-letter amino acid code. The partitioning free energies calculated using the polarizable water model for side chains E, D, K, and R are represented by triangles. The solid blue line indicates perfect agreement with the experimental data, and the dashed lines indicate $\pm 0.5RT$.

indicates that the C_1 – SC_4 interactions are strongly attractive, resulting in large positive values as compared to the experimental data. Proline is modeled using C_2 and N_a beads, where the N_a bead represents the backbone. This mapping leads to a higher free energy of partitioning. Using beads that have more attractive interactions with water molecules could improve the free energy of partitioning compared to experimental values; however, because of the lack of experimental solvation data for the proline

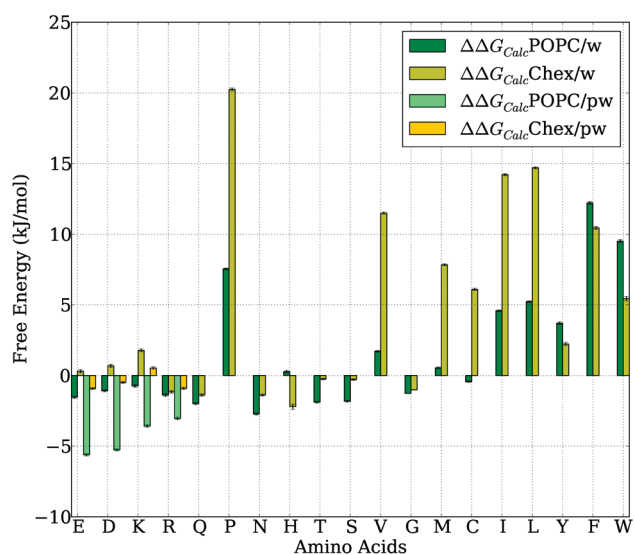


Figure 8. Calculated side-chain partitioning free energies at the POPC/water and cyclohexane/water interfaces as obtained using standard MARTINI are labeled as $\Delta\Delta G_{\text{calc}}^{\text{POPC/w}}$ and $\Delta\Delta G_{\text{calc}}^{\text{Chex/w}}$, respectively. Free energies calculated using MARTINI with the polarizable water model (for side chains D, E, R and K) are labeled as $\Delta\Delta G_{\text{calc}}^{\text{POPC/pw}}$ and $\Delta\Delta G_{\text{calc}}^{\text{Chex/pw}}$, respectively.

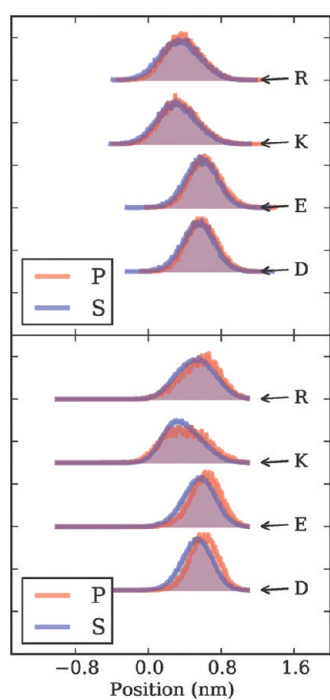


Figure 9. Distributions of the side-chain COM of amino acids at the cyclohexane/water (bottom) and POPC/water (top) interfaces, as obtained using the polarizable water model (labeled as P and colored red) or standard MARTINI (labeled as S and colored blue).

residue, it is difficult to justify such mappings for proline. The side-chain mappings for these two residues will be addressed in a future version of MARTINI.

3.2. Effects of Secondary Structure Parameters. In the current MARTINI model, a protein can be modeled as having a helix, coil, extended, turn, or bend secondary structure. The

Table 2. $\Delta\Delta G^{\text{s}}$ for Various Secondary Structures

secondary structure	$\Delta\Delta G^{\text{s}}$ (kJ mol ⁻¹)	
	E	I
helix	0.9	18.7
extended	1.3	1.4
turn	1.1	9.4
bend	1.0	10.1
dih ^a	0.7	

^a Dih represents simulations carried out using the dihedral potential extracted from atomistic simulations.

choice of secondary structure affects the type of backbone bead and associated bonded parameters. The current MARTINI version does not model changes in secondary structure. Even though the pentapeptides in this study do not have any preferred secondary structure, the conformations adopted by a peptide in the bulk water phase and at the interface could be significantly different. At the interface, the peptide is more likely to adopt conformations in which all of the hydrophobic side chains are oriented toward the cyclohexane and the charged side chains are oriented toward water, even when these conformations cause some strain in the backbone structure. As well, the addition of a dihedral potential will affect the conformational space explored by the peptide in the bulk solvent. These effects would be highly dependent on the peptide sequence, thus affecting partitioning free energies to different degrees. We investigated the effect of imposing a secondary structure on the free energy of partitioning by carrying out simulations with various secondary-structure parameters and computing the relative free energy of transfer of the peptide, with given secondary-structure constraints, to the bulk water as $\Delta\Delta G^{\text{s}} = (\Delta G_{\text{inter}}^{\text{s}} - \Delta G_{\text{bulk}}^{\text{s}}) - (\Delta G_{\text{inter}}^{\text{c}} - \Delta G_{\text{bulk}}^{\text{c}})$, where $\Delta G_{\text{inter}}^{\text{c}}$ and $\Delta G_{\text{bulk}}^{\text{c}}$ are annihilation free energies of side-chain beads of E or I at the cyclohexane/water interface and in bulk water, respectively, for the peptide modeled as a coil and $\Delta G_{\text{inter}}^{\text{s}}$ and $\Delta G_{\text{bulk}}^{\text{s}}$ represent the corresponding free energies for the peptide represented with helix, turn, extended, or bend secondary structures. Furthermore, the dihedral potentials derived from bulk atomistic simulations of the same peptide were also employed. The resulting free energy values are reported in Table 2. The WLILL peptide favors the interface by 18.7 kJ mol⁻¹ when modeled as a helix as compared to the coil structure, whereas for the WLELL peptide, the differences between various secondary structure representations are less than 1.5 kJ mol⁻¹. As all of the side chains in the WLILL peptide are hydrophobic, the helical conformations shield the polar backbone beads from cyclohexane beads, allowing the peptide to penetrate deeper into the cyclohexane region, resulting in better interactions between the side chains and the cyclohexane beads. On the other hand, the side-chain bead of E is always positioned toward the bulk water, regardless of the secondary-structure constraints.

Enforcing a particular secondary structure can have a significant impact on the partitioning free energy of a peptide at the interface, particularly for extreme cases such as enforcing an α helix for a hydrophobic peptide. This finding highlights the importance of carefully considering the treatment of secondary structure when using MARTINI to model peptides or proteins and the need for further improvements in MARTINI such that secondary-structure transitions can be adequately incorporated.

4. CONCLUSIONS

We have calculated the partitioning free energy of amino acid side chains using the MARTINI model at cyclohexane/water and POPC/water interfaces for the Wimley–White hydrophobicity scale peptides WLXLL. The free energies obtained using TI, FEP, and umbrella sampling were found to be in good agreement with each other. Among the three methods employed in this study, the free energies obtained using the FEP method coupled with MBAR for free energy estimation had the smallest statistical uncertainties.

Comparison of the experimental free energies at the POPC/water interface with computed values revealed that the amino acids F and P and the charged amino acids have the largest deviation. Hydrophobic residues such as L, I, and V and those with a net charge on the side chain are biased toward the hydrophobic phase, whereas most of the polar amino acids have a slightly higher preference for bulk water. In the case of amino acids with a net charge, significant improvements in calculated free energies were obtained by using the polarizable MARTINI water model. P and F can be improved in future versions of MARTINI with different bead mappings. For peptides at the cyclohexane/water interface, the polarizable water model did not have a significant effect on the free energy of partitioning of charged residues. This could be due to limitations of the model, but no experimental data are available for this system.

The choice of secondary-structure constraints imposed on the peptides affects the partitioning free energies, highlighting the importance of improving the representation of secondary structure in MARTINI, because, for simulations of peptide adsorption at an interface, it is important that the coarse model of the peptide be able to model the secondary-structure changes that could occur during peptide adsorption.

Because the coarse-grained representation was used, simulations of 0.2–0.5 μ s per λ point were feasible, resulting in good convergence. It can be estimated that, if all-atom models were to be employed, sampling of an order of magnitude higher would likely be required to obtain free energy values with reasonable uncertainty. Although this is currently a significant computational challenge, such simulations will be in easy reach in the near future and allow access to key thermodynamic data on lipid/peptide interactions to further improve simulations of biological membranes.

AUTHOR INFORMATION

Corresponding Author

*E-mail: tieleman@ucalgary.ca.

ACKNOWLEDGMENT

This work was supported by the Natural Sciences and Engineering Research Council (Canada). G.S. was supported by a postdoctoral fellowship from the Alberta Heritage Foundation for Medical Research (AHFMR). D.P.T. is an AHFMR Scientist. Calculations were performed in part on WestGrid/Compute Canada facilities.

REFERENCES

- (1) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57–83.
- (2) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168–174.

- (3) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102–106.
- (4) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258–11259.
- (5) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. *J. Chem. Phys.* **2005**, *123*, 084108.
- (6) Smit, B.; Hilbers, P. A. J.; Esselink, K.; Rupert, L. A. M.; van Os, N. M.; Schlijper, A. G. *Nature* **1990**, *348*, 624–625.
- (7) Nielsen, S.; Lopez, C.; Srinivas, G.; Klein, M. J. *Phys.: Condens. Matter* **2004**, *6*, R481–R512.
- (8) *Coarse-Graining of Condensed Phase and Biomolecular Systems*, 1st ed.; Voth, G. A., Ed.; CRC Press: Boca Raton, FL, 2009.
- (9) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.
- (10) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (11) Baoukina, S.; Monticelli, L.; Risselada, H. J.; Marrink, S. J.; Tieleman, D. P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10803–10808.
- (12) Risselada, H. J.; Marrink, S. J. *Soft Matter* **2009**, *5*, 4531–4541.
- (13) Baoukina, S.; Tieleman, D. P. *Biophys. J.* **2010**, *99*, 2134–2142.
- (14) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (15) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- (16) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740–5761.
- (17) MacCallum, J. L.; Tieleman, D. P. *J. Comput. Chem.* **2003**, *24*, 1930–1935.
- (18) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.
- (19) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616–17626.
- (20) Tieleman, D. P.; MacCallum, J. L.; Ash, W. L.; Kandt, C.; Xu, Z.; Monticelli, L. *J. Phys.: Condens. Matter* **2006**, *18*, S1221.
- (21) Ozdirekcan, S.; Etchebest, C.; Killian, J. A.; Fuchs, P. F. *J. Am. Chem. Soc.* **2007**, *129*, 15174–15181.
- (22) Monticelli, L.; Tieleman, D. P.; Fuchs, P. F. *J. Biophys. J.* **2010**, *99*, 1455–1464.
- (23) Wimley, W. C.; Creamer, T. P.; White, S. H. *Biochemistry* **1996**, *35*, 5109–5124.
- (24) Wimley, W. C.; White, S. H. *Nat. Struct. Biol.* **1996**, *3*, 842–848.
- (25) White, S. H.; Wimley, W. C. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319–365.
- (26) Snider, C.; Jayasinghe, S.; Hristova, K.; White, S. H. *Protein Sci.* **2009**, *18*, 2624–2628.
- (27) Aliste, M. P.; MacCallum, J. L.; Tieleman, D. P. *Biochemistry* **2003**, *42*, 8976–8987.
- (28) Aliste, M. P.; Tieleman, D. P. *BMC Biochem.* **2005**, *6*, 30.
- (29) MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. *J. Gen. Physiol.* **2007**, *129*, 371–377.
- (30) MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. *Biophys. J.* **2008**, *94*, 3393–3404.
- (31) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (32) Nose, S. *Mol. Phys.* **1984**, *52*, 255–268.
- (33) Hoover, W. *Phys. Rev. A.* **1985**, *31*, 1695–1697.
- (34) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (35) Berendsen, H. J. C.; Postma, J.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (36) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14* (33–8), 27–8.
- (37) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. *Annu. Rep. Comput. Chem.* **2007**, *3*, 41–59.
- (38) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578–581.
- (39) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175–1188.
- (40) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

- (41) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.
- (42) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (43) Hess, B. *J. Chem. Phys.* **2002**, *116*, 209–217.
- (44) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
- (45) Yesylevskyy, S. O.; Schafer, L. V.; Sengupta, D.; Marrink, S. J. *PLoS Comput. Biol.* **2010**, *6*, e1000810.
- (46) Hub, J. S.; de Groot, B. L.; van der Spoel, D. *J. Chem. Theory Comput.* **2010**, *6*, 3713–3720.
- (47) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- (48) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (49) Yau, W. M.; Wimley, W. C.; Gawrisch, K.; White, S. H. *Biochemistry* **1998**, *37*, 14713–14718.
- (50) Killian, J. A.; von Heijne, G. *Trends Biochem. Sci.* **2000**, *25*, 429–434.
- (51) Sun, H.; Greathouse, D. V.; Andersen, O. S.; Koeppe, R. E. *J. Biol. Chem.* **2008**, *283*, 22233–22243.

Performance of Density Functional Theory and Møller–Plesset Second-Order Perturbation Theory for Structural Parameters in Complexes of Ru

Anant D. Kulkarni and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, Minnesota 55455-0431, United States

S Supporting Information

ABSTRACT: We assess the performance of density functional theory (DFT) and Møller–Plesset second-order perturbation theory (MP2) for predicting structural parameters in Ru complexes, in particular, a Ru(IV) allyl dicationic complex with the formula $[\text{Ru}(\eta^5\text{-Cp}^*)(\eta^3\text{-CH}_2\text{CHCHC}_6\text{H}_5)(\text{NCCH}_3)_2]^{2+}$ and the molecules RuO_4 and $\text{Ru}(\text{C}_2\text{O}_4)_2(\text{H}_2\text{O})_2^-$, where Cp^* denotes C_5Me_5 and Me denotes methyl. The density functionals studied are B3LYP, B3PW91, M05, M06, M06-L, MOHLYP, MPW3LYP, PBE0, PW6B95, SOGGA, $\tau\text{HCTHhyb}$, ωB97X , and $\omega\text{B97X-D}$, in combination with three different basis sets, namely, LANL2DZ, def2-SVP, and def2-TZVP. The theoretically computed Ru–C distances corresponding to the phenylallyl complex are especially well predicted by the SOGGA (pure DFT) and $\omega\text{B97X-D}$ (DFT plus an empirical molecular mechanics term) methods. This contrasts with an article in this Journal [Calhorda, M. J.; Pregosin, P. S.; Veiros, L. F. J. *Chem. Theory Comput.* **2007**, *3*, 665–670] in which it was found that DFT cannot account for these Ru–C distances. Averaging over four Ru–C distances in the allyl complex and three unique Ru–O distances in RuO_4 and $\text{Ru}(\text{C}_2\text{O}_4)_2(\text{H}_2\text{O})_2^-$, the SOGGA and $\omega\text{B97X-D}$ methods have both a smaller mean unsigned error than MP2 and the same maximum error. The M06, PW6B95, PBE0, M06-L, and ωB97X density functionals also have a smaller or the same mean unsigned error as MP2.

INTRODUCTION

The organometallic chemistry of ruthenium is very interesting from both synthetic and theoretical points of view. Ruthenium catalysts are used in a wide range of applications including hydroamination,¹ metathesis,² hydrogenation,³ selective transformation of alkynes,⁴ and allylic alkylation.⁵ Quantum chemical modeling of such catalytic systems is a key step in the rational design and understanding of the detailed mechanisms of catalytic processes. The large size of catalytic systems makes their study with reliable wave function theory (WFT) difficult. Density functional theory (DFT) offers an affordable and efficient alternative if one uses an accurate enough exchange–correlation functional and a suitable large basis set.

Fomine et al.⁶ performed DFT studies on ruthenium-alkylidene-mediated metathesis of various olefins using a variety of functionals, in particular, B3LYP,⁷ M05,⁸ and PBE0,⁹ and they discussed stereoselectivity and the steric effect on activation energy. Mathew et al.¹⁰ used B3LYP to study the structural and energetic features of decomposition pathways of a Grubbs second-generation olefin metathesis catalyst. Occhipinti et al.¹¹ studied various Ru–L bonds in $\text{L}(\text{PCy}_3)_2\text{Ru}=\text{CH}_2$ complexes, where L is a dative ligand. Noncovalent interactions in a Ru-triphenyl phosphine complex were studied by Sieffert and Büehl¹² using the B3LYP, BP86,¹³ B3LYP-D,¹⁴ B97-D,¹⁵ M05,⁸ M05-2X,¹⁶ M06-L,¹⁷ M06-HF,¹⁸ and M06-2X¹⁹ functionals. According to their study, B97-D, M06-HF, M06-L, M06, and M06-2X produce the best agreement with the experimental²⁰ binding enthalpies. Piacenza et al.²¹ assessed the performance of the B3LYP, BP86, B97-D, TPSSH,²² and B2-PLYP²³ functionals

for ruthenium-catalyzed olefin metathesis. The popular B3LYP has the largest overall error. Tsipis et al.²⁴ and Zhao and one of the authors^{25,26} also applied DFT to ruthenium olefin catalysts, and M06-L and M06 were found to perform better than older functionals such as BP86, B3LYP, and TPSSH. Pandian et al.²⁷ found the M06 functional to be useful for studying ruthenium-catalyzed ring-closing metathesis. Diesendruck et al.²⁸ showed how BP86 and M06-L calculations can predict the configuration of chelated ruthenium complexes. Sliwa and Handzlik²⁹ evaluated 31 different density functionals for reaction energies of model Grubbs-type ruthenium complexes and found the best performance with M06 and $\omega\text{B97X-D}$. They also examined 20 density functionals for reproduction of the PCy_3 dissociation energy of a Grubbs catalyst and found the best performance with M06-L, M06, and M06-2X. Dutta et al.³⁰ studied reactions of alkynes with cyclopentadienyl-ruthenium half-sandwich compounds, and M06 calculations were in excellent agreement with the bond lengths and bond angles and with the orientation of the alkyne ligand; they could rationalize steric effects and relative stabilities of mechanistic intermediates.

In light of these successes, we became interested in a paper in this Journal entitled “Geometry Optimization of a Ru(IV) Allyl Dicationic Complex: A DFT Failure?”³¹ That paper concluded that popular density functionals, namely B3LYP, BLYP,^{13a,32} BPW91,^{13a,33} B3PW91,^{7a} BP86, PW91,³³ PBE,³⁴ PBE0,^{34,35} and mPW1PW91,³⁵ even with reasonably large basis sets, greatly

Received: March 19, 2011

Published: May 27, 2011

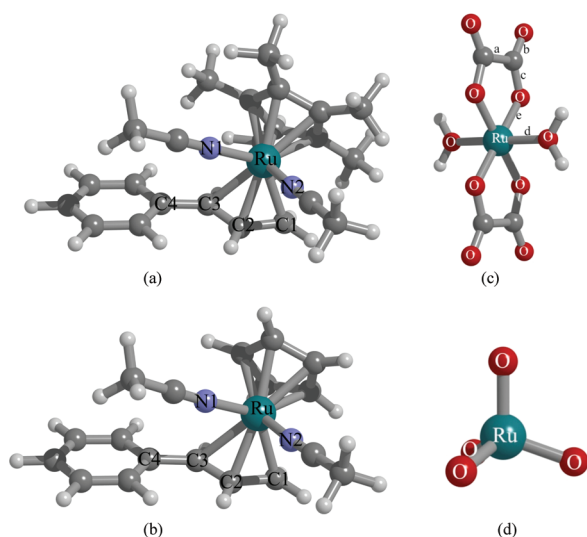


Figure 1. Structures of (a) Complex-1, (b) Complex-2, (c) $\text{Ru}(\text{C}_2\text{O}_4)_2(\text{H}_2\text{O})_2^-$, and (d) RuO_4 . Color conventions: hydrogen, light gray; carbon, black; nitrogen, blue; oxygen, red; and ruthenium, green.

overestimate the weak Ru–C bond length involving a phenylallyl ligand in the catalytic dication $[\text{Ru}(\eta^5\text{-Cp}^*)(\eta^3\text{-CH}_2\text{CHC-CHC}_6\text{H}_5)(\text{CH}_3\text{CN})_2]^{2+}$, abbreviated here as Complex-1. Here, the acetonitrile ligands are coordinated at the nitrogen, and Cp^* , as usual, denotes C_5Me_5 where Me is a methyl group. The experimental structure was obtained by X-ray diffraction from the salt crystal. The DFT overestimates that they found for Ru–C bond distances are 0.38–0.46 Å, and Hartree–Fock theory overestimates the key Ru–C bond length (called Ru–C3 in Figure 1) by 0.60 Å. The authors also concluded that WFT in the form of Møller–Plesset second-order perturbation theory (MP2) is more accurate than DFT and yields results that agree with the experimental ones within 0.07 Å. However, they did not test more recently developed density functionals such as M05, M06, M06-L, SOGGA,³⁶ ωB97X ,³⁷ and $\omega\text{B97X-D}$.³⁸ Thus, it was the initial goal of the present study to test these functionals on the problem studied by Calhorda et al.³¹ To make the process more complete, we also tested seven other density functionals and Hartree–Fock theory.

In order to test the generality of our conclusions, we also consider bond lengths in two other Ru complexes, RuO_4 and $\text{Ru}(\text{C}_2\text{O}_4)_2(\text{H}_2\text{O})_2^-$.

METHODOLOGY

We have employed 15 levels of electronic structure theory, in particular, Hartree–Fock theory, MP2 theory, and 13 density functionals, namely, B3LYP, B3PW91, M05, M06-L, M06-2X, MOHLYP,³⁹ MPW3LYP,⁴⁰ PW6B95,⁴¹ PBE0, SOGGA, $\tau\text{HCTHhyb}$,⁴² ωB97X , and $\omega\text{B97X-D}$.

The SOGGA and M06-L functionals are particularly interesting because previous work³⁶ has shown that they have good accuracy for geometries, and furthermore, because they do not involve any Hartree–Fock exchange, they are particularly efficient for geometry optimization of large systems.

We combine these levels with a variety of basis sets, namely, LANL2DZ,⁴³ def2-SVP,⁴⁴ and def2-TZVP,⁴⁴ for a total of 45 model chemistries. Use of the LANL2DZ basis set also implies usage of its associated effective core potential (ECP) to replace

the 28-electron inner core of Ru, whereas calculations with the def2-class basis sets replace the 28-electron inner core of Ru with the Stuttgart ECPs.⁴⁵ Both the LANL2DZ and Stuttgart ECPs account for scalar relativistic effects (they are therefore sometimes called relativistic ECPs, i.e., RECPs). Gaussian 09⁴⁶ and a locally modified version, MNGFM4.1,⁴⁷ were used for the geometry optimization calculations with the option “scf=tight”. The integration grid employed is “ultrafine” as defined in Gaussian 09.⁴⁶ The calculations on the odd-electron anion were carried out with spin unrestricted methods (UHF and spin-polarized DFT). The atomic charges were computed employing ChelpG,⁴⁸ Merz–Kollamn,⁴⁹ and natural bond orbital (NBO) analysis⁵⁰ schemes. For ChelpG and Merz–Kollman methods, we set the radius of Ru to 2.07 Å.⁵¹

STRUCTURES

We will consider two Ru dication structures. The experimental structure⁵² that forms the basis for the comparison of Calhorda et al.³¹ is $[\text{Ru}(\eta^5\text{-Cp}^*)(\eta^3(\text{CH}_2\text{CHCHC}_6\text{H}_5)(\text{CH}_3\text{CN})_2)]^{2+}$, abbreviated here as Complex-1. However, Calhorda et al. modeled this as $[\text{Ru}(\eta^5\text{-Cp})(\eta^3\text{-CH}_2\text{CHCHC}_6\text{H}_5)(\text{CH}_3\text{CN})_2]^{2+}$ (which we call Complex-2), where Cp denotes C_5H_5 . We will compare the calculated structural parameters of Complex-1 to experimental results and discuss unsigned errors and mean unsigned errors (MUEs). Coordinates of calculated structures are in the Supporting Information. For Complex-2, we will compare the calculated structure to the calculated structure for Complex-1. In this case, we discuss the results as unsigned deviations and mean unsigned deviations (MUDs). The comparisons for Complex-1 and Complex-2 are focused on Ru–C distances and C–C–C bond angles in the allyl ligand.

To test the performance of the theoretical models used in this study beyond Ru–C bonding, we have also selected two structures containing Ru–O bonds, namely, RuO_4 ⁵³ and the molecular fragment $\text{Ru}(\text{C}_2\text{O}_4)_2(\text{H}_2\text{O})_2^-$ from the *trans*-dimethylammonium bis(oxalato) diaquaruthenate (III) tetrahydrate⁵⁴ complex. The coordinates of these systems were taken from their experimental crystal structures available at ICSD.⁵⁵ Figure 1 depicts the structures of Complex-1, Complex-2, RuO_4 ,⁵³ and $\text{Ru}(\text{C}_2\text{O}_4)_2(\text{H}_2\text{O})_2^-$.⁵⁴

RESULTS AND DISCUSSION

Figure 1 shows that Complex-1 and Complex-2 involve coordination of a phenylallyl ligand to ruthenium. For a structural analysis of Complex-1 and Complex-2, we select all of the Ru–C bond lengths, namely, Ru–C1, Ru–C2, and Ru–C3, where C1, C2, and C3 are identified in Figure 1, and Ru–CR, where CR denotes one of the carbons of the cyclopentadienyl ligand. In all cases, we use the average of the five Ru–CR values and treat this average as a single variable. We also consider the C1–C2–C3 bond angles in the allyl portion of the phenylallyl ligand, denoted by $\angle\text{C1–C2–C3}$. In order to test the ability of the model chemistries to reproduce the geometrical parameters, we compute the errors and mean unsigned errors (MUEs) for these parameters with respect to the experimental parameters.⁵²

In assessing the qualities of HF, MP2, and the density functionals, we consider only the largest basis set, namely, def2-TZVP. We will, however, also discuss basis set effects.

Tables 1 and 2 show that the new-generation functionals, namely, SOGGA, M05, M06, M06-L, PW6B95, ωB97X , and $\omega\text{B97X-D}$, when used with the def2-TZVP basis set, yield small

Table 1. Unsigned Error and Mean Unsigned Error (both in Å) Computed for Complex-1 Employing Different Model Chemistries with Respect to the Experimental Values

bond distance	basis set	B3LYP	B3PW91	HF	M05	M06	M06-L	MOHLYP	MP2
Ru–C1	LANL2DZ	0.044	0.025	0.092	0.011	0.017	0.007	0.040	0.067
	def2-SVP	0.007	0.086	0.005	0.016	0.007	0.012	0.015	0.010
	def2-TZVP	0.007	0.015	0.001	0.019	0.007	0.012	0.013	0.020
Ru–C2	LANL2DZ	0.114	0.098	0.154	0.084	0.068	0.043	0.179	0.032
	def2-SVP	0.078	0.034	0.101	0.035	0.032	0.025	0.125	0.063
	def2-TZVP	0.088	0.044	0.106	0.042	0.039	0.028	0.142	0.074
Ru–C3	LANL2DZ	0.360	0.335	0.555	0.311	0.234	0.203	0.618	0.011
	def2-SVP	0.306	0.194	0.409	0.193	0.138	0.148	0.523	0.078
	def2-TZVP	0.337	0.227	0.424	0.216	0.159	0.160	0.563	0.073
Ru–CR ^a	LANL2DZ	0.088	0.040	0.088	0.040	0.052	0.043	0.102	0.076
	def2-SVP	0.035	0.011	0.035	0.011	0.003	0.016	0.045	0.027
	def2-TZVP	0.029	0.015	0.029	0.015	0.000	0.021	0.039	0.040
MUE ^b	LANL2DZ	0.15	0.12	0.22	0.11	0.09	0.07	0.23	0.05
	def2-SVP	0.11	0.08	0.14	0.06	0.05	0.05	0.18	0.04
	def2-TZVP	0.12	0.08	0.14	0.07	0.05	0.06	0.19	0.05

^a Average over five Ru–C distances for C atoms in the cyclopentadienyl ring. ^b Mean unsigned error: average over four previous values.

Table 2. Unsigned Error and Mean Unsigned Error (both in Å) Computed for Complex-1 Employing Different Model Chemistries with Respect to the Experimental Values

bond distance	basis set	MPW3LYP	PBE0	PW6B95	SOGGA	τ HCTHhyb	ω B97X	ω B97X-D
Ru–C1	LANL2DZ	0.033	0.010	0.018	0.004	0.025	0.013	0.023
	def2-SVP	0.008	0.018	0.013	0.024	0.002	0.020	0.008
	def2-TZVP	0.008	0.023	0.016	0.029	0.006	0.022	0.011
Ru–C2	LANL2DZ	0.117	0.059	0.059	0.025	0.098	0.066	0.059
	def2-SVP	0.072	0.016	0.020	0.017	0.051	0.018	0.017
	def2-TZVP	0.083	0.022	0.025	0.014	0.059	0.024	0.021
Ru–C3	LANL2DZ	0.384	0.222	0.196	0.137	0.335	0.172	0.148
	def2-SVP	0.280	0.123	0.115	0.052	0.228	0.066	0.056
	def2-TZVP	0.320	0.149	0.129	0.066	0.260	0.088	0.074
Ru–CR ^a	LANL2DZ	0.084	0.053	0.049	0.048	0.076	0.040	0.043
	def2-SVP	0.033	0.002	0.004	0.004	0.024	0.009	0.002
	def2-TZVP	0.031	0.002	0.010	0.007	0.020	0.010	0.004
MUE ^b	LANL2DZ	0.15	0.09	0.08	0.05	0.13	0.07	0.07
	def2-SVP	0.10	0.04	0.04	0.02	0.08	0.03	0.02
	def2-TZVP	0.11	0.05	0.05	0.03	0.09	0.04	0.03

^a Average over five Ru–C distances for C atoms in cyclopentadienyl ring. ^b Mean unsigned error: average over four previous values.

Table 3. Error (in degrees) Computed for \angle C1–C2–C3 for Complex-1

basis set	B3LYP	B3PW91	HF	M05	M06	M06-L	MOHLYP	MP2
LANL2DZ	0.9	0.5	1.4	0.5	0.1	0.8	2.1	3.7
def2-SVP	0.9	0.3	1.3	0.1	0.8	0.8	1.9	4.6
def2-TZVP	1.0	0.2	1.1	0.1	0.7	0.9	1.9	4.9

MUE values typically in the same range as those for MP2 calculations. The older PBE0 functional also does well. For Ru–C bond lengths, the popular functionals B3LYP and B3PW91 yield larger MUEs of ~ 0.12 and 0.08 Å, respectively. It is also encouraging that Tables 3 and 4 show that *all* of the density functionals studied here have much smaller errors in \angle C1–C2–C3 than does MP2.

The key issue in the article by Calhorda et al. was the weak Ru–C3 bond. They defined 0.07 Å as an “acceptable” error for this bond but found that all density functionals tested had an error of at least 0.23 Å. Tables 1 and 2 show eight functionals with an error less than 0.23 Å for this bond. Four of these (PW6B95, SOGGA, ω B97X, and ω B97X-D) have an error of 0.13 Å or less, and two (SOGGA and ω B97X-D) have an acceptable error of 0.07 Å or less.

Table 4. Error (in degrees) Computed for $\angle C1-C2-C3$ for Complex-1

basis set	MPW3LYP	PBE0	PW6B95	SOGGA	τ HCTHhyb	ω B97X	ω B97X-D
LANL2DZ	1.1	0.5	1.2	2.0	0.5	1.0	1.3
def2-SVP	0.7	0.6	1.4	2.5	0.1	1.8	1.9
def2-TZVP	0.9	1.0	1.6	3.4	0.2	1.5	1.8

Table 5. Unsigned Deviations and Mean Unsigned Deviations (both in Å) Computed for Complex-2 Employing Different Model Chemistries with Respect to the Original Complex, Complex-1

bond distance	basis set	B3LYP	B3PW91	HF	M05	M06	M06-L	MOHLYP	MP2
Ru–C1	LANL2DZ	0.044	0.025	0.092	0.011	0.017	0.007	0.040	0.067
	def2-SVP	0.007	0.086	0.005	0.016	0.007	0.012	0.015	0.010
	def2-TZVP	0.007	0.015	0.001	0.019	0.007	0.012	0.013	0.020
Ru–C2	LANL2DZ	0.114	0.098	0.154	0.084	0.068	0.043	0.179	0.032
	def2-SVP	0.078	0.034	0.101	0.035	0.032	0.025	0.125	0.063
	def2-TZVP	0.088	0.044	0.106	0.042	0.039	0.028	0.142	0.074
Ru–C3	LANL2DZ	0.360	0.335	0.555	0.311	0.234	0.203	0.618	0.011
	def2-SVP	0.306	0.194	0.409	0.193	0.138	0.148	0.523	0.078
	def2-TZVP	0.337	0.227	0.424	0.216	0.159	0.160	0.563	0.073
Ru–CR ^a	LANL2DZ	0.103	0.087	0.101	0.048	0.056	0.043	0.114	0.085
	def2-SVP	0.035	0.012	0.035	0.011	0.003	0.016	0.045	0.027
	def2-TZVP	0.033	0.009	0.029	0.015	0.000	0.021	0.039	0.040
MUD ^b	LANL2DZ	0.155	0.136	0.226	0.114	0.094	0.074	0.238	0.049
	def2-SVP	0.106	0.082	0.138	0.064	0.045	0.050	0.177	0.045
	def2-TZVP	0.116	0.074	0.140	0.073	0.051	0.055	0.189	0.052

^a Average over five Ru–C distances for C atoms in the cyclopentadienyl ring. ^b Mean unsigned error: average over four previous values.

Table 6. Unsigned Deviations and Mean Unsigned Deviations (both in Å) Computed for Complex-2 Employing Different Model Chemistries with Respect to the Original Complex, Complex-1

bond distance	basis set	MPW3LYP	PBE0	PW6B95	SOGGA	τ HCTHhyb	ω B97X	ω B97X-D
Ru–C1	LANL2DZ	0.033	0.010	0.018	0.004	0.025	0.013	0.023
	def2-SVP	0.008	0.018	0.013	0.024	0.002	0.020	0.008
	def2-TZVP	0.008	0.023	0.016	0.029	0.006	0.022	0.011
Ru–C2	LANL2DZ	0.117	0.059	0.059	0.025	0.098	0.066	0.059
	def2-SVP	0.072	0.016	0.020	0.017	0.051	0.018	0.017
	def2-TZVP	0.083	0.022	0.025	0.014	0.059	0.024	0.021
Ru–C3	LANL2DZ	0.384	0.222	0.196	0.137	0.335	0.172	0.148
	def2-SVP	0.280	0.123	0.115	0.052	0.228	0.066	0.056
	def2-TZVP	0.320	0.149	0.129	0.066	0.260	0.088	0.074
Ru–CR ^a	LANL2DZ	0.090	0.062	0.055	0.055	0.087	0.050	0.052
	def2-SVP	0.033	0.002	0.004	0.004	0.024	0.009	0.002
	def2-TZVP	0.031	0.002	0.010	0.007	0.020	0.010	0.004
MUD ^b	LANL2DZ	0.16	0.09	0.08	0.06	0.14	0.08	0.07
	def2-SVP	0.10	0.04	0.04	0.02	0.08	0.03	0.27
	def2-TZVP	0.11	0.05	0.05	0.03	0.09	0.04	0.03

^a Average over five Ru–C distances for C atoms in the cyclopentadienyl ring. ^b Mean unsigned error: average over four previous values.

For Ru–C bond distances, Tables 1 and 2 show that the popular LANL2DZ basis set gives larger MUEs than the other two basis sets, def2-SVP and def2-TZVP, and we judge it to be unreliable. The other valence double- ζ basis set studied here, namely, def2-SVP, performs as well as the valence triple- ζ basis set def2-TZVP.

In addition to the fact that they used older density functionals and an inadequate basis set, one reason for the different conclusions here and in ref 31 is that the model system, Complex-2,

selected by Calhorda et al.³¹ is not a good choice for modeling the original system, Complex-1. Tables 5 and 6 list the absolute deviation and MUD computed for Ru–C bond lengths for Complex-2 with respect to the original complex (Complex-1); Tables 7 and 8 list the deviation computed for $\angle C1-C2-C3$ in Complex-2 with respect to Complex-1. It may be seen from these tables that the substituents (five methyl groups) present on the cyclopentadiene ring play an important role in the modification

Table 7. Deviation (in degrees) Computed for $\angle C1-C2-C3$ in Complex-2 with Respect to Complex-1

basis set	B3LYP	B3PW91	HF	M05	M06	M06-L	MOHLYP	MP2
LANL2DZ	0.9	0.5	1.4	0.5	0.1	0.8	2.1	3.7
def2-SVP	0.9	0.3	1.3	0.1	0.8	0.8	1.9	4.6
def2-TZVP	1.0	0.2	1.1	0.1	0.7	0.9	1.9	4.9

Table 8. Deviation (in degrees) Computed for $\angle C1-C2-C3$ in Complex-2 with Respect to Complex-1

basis set	MPW3LYP	PBE0	PW6B95	SOGGA	τ HCTHhyb	ω B97X	ω B97X-D
LANL2DZ	1.1	0.5	1.2	2.00	0.5	1.0	1.3
def2-SVP	0.7	0.6	1.4	2.5	0.1	1.8	1.9
def2-TZVP	0.9	1.0	1.6	3.4	0.2	1.5	1.8

Table 9. MUE^a (Å) Computed for Ru–O Distance in RuO₄ and Ru(C₂O₄)₂(H₂O)₂[−]

basis set	B3LYP	B3PW91	HF	M05	M06	M06-L	MOHLYP	MP2
LANL2DZ	0.056	0.045	0.054	0.059	0.055	0.054	0.074	0.075
def2-SVP	0.042	0.036	0.071	0.055	0.042	0.048	0.064	0.040
def2-TZVP	0.042	0.034	0.066	0.054	0.041	0.048	0.062	0.040

^a Mean unsigned error: average over three Ru–O values.

Table 10. MUE^a (Å) Computed for Ru–O Distance in RuO₄ and Ru(C₂O₄)₂(H₂O)₂[−]

basis set	MPW3LYP	PBE0	PW6B95	SOGGA	τ HCTHhyb	ω B97X	ω B97X-D
LANL2DZ	0.055	0.038	0.041	0.033	0.048	0.046	0.045
def2-SVP	0.041	0.034	0.027	0.014	0.034	0.046	0.044
def2-TZVP	0.040	0.031	0.027	0.012	0.033	0.045	0.043

^a Mean unsigned error: average over three Ru–O values.

Table 11. Overall Mean Unsigned Errors and Largest Errors in Ru–C and Ru–O Bond Lengths (Å) for def2-TZVP Basis Set

functional	MUE (Ru-allyl) ^a	MUE (Ru-X) ^b	largest ^c
SOGGA	0.04	0.02	0.07
ω B97X-D	0.04	0.03	0.07
ω B97X	0.04	0.04	0.09
MP2	0.06	0.05	0.07
PW6B95	0.06	0.04	0.13
PBE0	0.06	0.04	0.15
M06	0.07	0.05	0.16
M06-L	0.07	0.05	0.16
M05	0.09	0.07	0.22
B3PW91	0.10	0.06	0.23
τ HCTHhyb	0.11	0.06	0.26
MPW3LYP	0.14	0.08	0.32
B3LYP	0.14	0.09	0.34
HF	0.18	0.11	0.42
MOHLYP	0.24	0.13	0.56

^a Average over three Ru–C distances in which C is an allyl carbon.

^b Average over seven values for X = C or O; equivalent to a 4:3 weighted average of the bottom rows of Tables 1 and 2 with the bottom rows of Tables 9 and 10. ^c Largest absolute error for any of the four Ru–C or three Ru–O distances.

of all Ru–C bond lengths; the MUDs with def2-TZVP are as large as 0.19 Å. In fact, Calhorda et al. had studied this effect of the methyl groups, but they reported bond length differences smaller than 0.1 Å with the LANL2DZ basis set for Ru. We find larger differences even with LANL2DZ. They found large errors in Ru–C3 distances just as we find in Table 1 for B3LYP/LANL2DZ, where the error is 0.36 Å. These errors are reduced to 0.14 Å by SOGGA/LANL2DZ and to 0.07 Å with SOGGA/def2-TZVP.

For RuO₄ and Ru(C₂O₄)₂(H₂O)₂[−], we focus our attention on the Ru–O bond lengths. The four Ru–O bonds in RuO₄ are equivalent and symmetric: for Ru(C₂O₄)₂(H₂O)₂[−], two Ru–O distances involving the water ligands are identical, and four Ru–O distances involving oxalates are identical. Therefore, we computed the MUEs as unweighted averages over the three unique values. Tables 9 and 10 show the resulting MUE computed for the Ru–O distances in these molecules with various model chemistries. These tables show that the MUE in the DFT bond lengths is typically 0.03–0.07 Å, with the best results obtained by SOGGA (0.01 Å) and PW6B95 (0.03 Å). For Ru–O bond lengths, the def2-SVP basis set is almost as good as def2-TZVP, and LANL2DZ is again significantly less accurate.

Finally, in Table 11, we present an overall assessment of the density functionals for computed bond distances. For this table, we report the results with the most complete basis set, namely, def2-TZVP, and we average over the errors in the seven Ru–C

Table 12. Difference between the ChelpG Charges of Complex-1 and Complex-2 Calculated with the def2-TZVP Basis Set for the Five Model Chemistries with the Smallest Errors in Geometries ($q_{\text{difference}} = q_{\text{Complex-1}} - q_{\text{Complex-2}}$)

atoms	M06	M06-L	PW6B95	PBE0	ω B97X
Ru	0.38	0.33	0.29	0.31	0.42
$C_{\text{Cp}^*} - C_{\text{Cp}}^a$	0.03	0.05	0.07	0.06	0.01
$-C_6H_5$	-0.05	-0.05	-0.05	-0.09	-0.04
$-N1CCH_3$	-0.22	-0.21	-0.18	-0.15	-0.17
$-N2CCH_3$	0.02	0.04	0.03	-0.02	-0.04
Allyl	-0.15	-0.15	-0.15	-0.12	-0.18
C1	-0.18	-0.12	-0.07	-0.07	-0.16
C2	-0.33	-0.37	-0.38	-0.42	-0.38
C3	0.15	0.19	0.18	0.20	0.18

^a Difference of sum over all of the atoms of C_{P}^* and Cp rings.**Table 13.** Difference between the MK Charges of Complex-1 and Complex-2 Calculated with the def2-TZVP Basis Set for the Five Model Chemistries with the Smallest Errors in Geometries ($q_{\text{difference}} = q_{\text{Complex-1}} - q_{\text{Complex-2}}$)

atoms	M06	M06-L	PW6B95	PBE0	ω B97X
Ru	0.33	0.10	0.20	0.33	0.37
$C_{\text{Cp}^*} - C_{\text{Cp}}^a$	0.04	0.14	0.10	0.05	0.02
$-C_6H_5$	-0.08	-0.08	-0.10	-0.07	-0.07
$-N1CCH_3$	-0.23	-0.18	-0.15	-0.16	-0.17
$-N2CCH_3$	0.05	0.10	0.03	-0.02	-0.02
Allyl	-0.11	-0.07	-0.09	-0.14	-0.14
C1	-0.04	-0.05	-0.04	-0.08	-0.05
C2	-0.02	0.00	-0.01	0.03	0.00
C3	0.01	0.02	0.04	-0.03	0.01

^a Difference of sum over all of the atoms of C_{P}^* and Cp rings.**Table 14.** Difference between the NBO Charges of Complex-1 and Complex-2 Calculated with the def2-TZVP Basis Set for the Five Model Chemistries with the Smallest Errors in Geometries ($q_{\text{difference}} = q_{\text{Complex-1}} - q_{\text{Complex-2}}$)

atoms	M06	M06-L	PW6B95	PBE0	ω B97X
Ru	0.02	0.01	0.01	0.01	0.02
$C_{\text{Cp}^*} - C_{\text{Cp}}^a$	0.13	0.14	0.13	0.13	0.12
$-C_6H_5$	-0.07	-0.07	-0.08	-0.08	-0.06
$-N1CCH_3$	-0.01	-0.01	0.00	-0.01	0.00
$-N2CCH_3$	-0.02	-0.03	-0.02	-0.02	-0.02
Allyl	-0.07	-0.06	-0.06	-0.06	-0.07
C1	-0.03	-0.02	-0.02	-0.55	-0.54
C2	0.01	0.01	0.02	-0.49	-0.48
C3	-0.05	-0.04	-0.05	-0.28	-0.27

^a Difference of sum over all of the atoms of C_{P}^* and Cp rings.

and Ru–O bond distances. For every density functional, we tabulate the mean unsigned error and the maximum error; the latter is usually, but not always, the error in the Ru–C3 distance (in a few cases, namely, MP2, SOGGA, ω B97X, and ω B97X-D, it is an error in a Ru–O distance). Table 11 shows that seven

functionals have an average error smaller than or equal to that of MP2, and of these seven functionals, two also have the same maximum error. The two functionals that outperform MP2 in MUE and do not have a larger maximum error than MP2 are SOGGA and ω B97X-D. Although ω B97X-D contains an empirical molecular mechanics term, SOGGA is simply a density functional, as are all other methods in Table 11 except HF and MP2. Because density functional calculations are much less expensive than MP2 for large systems, it is very encouraging that some of the density functionals outperform MP2. As an example of the cost differential, with the def2-TZVP basis set, we found for Complex-1 that a geometry optimization step with the SOGGA, M06-L, or PBE0 density functionals is about 40 times faster than one with MP2, and a geometry optimization step with M06, ω B97X-D, or PW6B95 is about 30 times faster than with MP2.

In order to understand the effect of methyl groups on C_{P}^* , we used three schemes for electrostatic fitting to determine the partial charges on the atoms of Complex-1 and Complex-2. In particular, we used ChelpG,⁴⁸ Merz–Singh–Kollamn,⁴⁹ and natural bond orbital (NBO)⁵⁰ methods. Tables 12–14 show the differences in partial atomic charges between Complex-1 and Complex-2 as determined by these analyses for the five model chemistries with smallest errors in geometries. (The charges themselves are given in the Supporting Information.) It is well-known that C_{P}^* is a better electron donor than Cp, and Tables 12–14 (row 2) confirm this in that the sum of the charges on all of the atoms in C_{P}^* in methylated Complex-1 is more positive than that for Cp from Complex-2. However, the Ru atoms in Complex-1 are also more positive, by 0.02–0.4 au depending on the method. The extra electron density donated by C_{P}^* ends up on the phenyl group, the NCCCH₃ ligand on the phenyl side, and the allyl group (especially C1), which are all less positive or more negative in Complex-1 than in Complex-2.

CONCLUSIONS

In the present study, we employed 45 different model chemistries to assess the performance of density functional methods in predicting the structural parameters of Ru bonds to C and O. The important highlights of our study are as follows:

- 1 Ru–C distances in Complex-2 are quite different from those of Complex-1, contrary to a previous³¹ assumption.
- 2 Selection of the model chemistry is an important step for obtaining accurate geometries in studying catalytic complexes.
- 3 The popular B3LYP functional does not yield accurate Ru–C bond lengths.
- 4 The popular LANL2DZ basis set is found to be inadequate for studying the structural parameters.
- 5 Seven density functionals, namely, SOGGA, ω B97X-D, PW6B95, M06, M06-L, ω B97X-D, and PBE0, have an average error in Ru–X bond lengths, where X is C or O, that is smaller than or equal to that of the more expensive MP2 method. Hence, we recommend these functionals with the def2-SVP or def2-TZVP basis set for calculating Ru–C and Ru–O bond lengths.
- 6 The presence of methyl groups on the Cp ring modifies the electronic environment of the Ru center, making that environment more negative.

The progress of DFT since the Callhorda et al. study of four years ago is very encouraging.

■ ASSOCIATED CONTENT

S Supporting Information. The coordinates computed by SOGGA/def2-TZVP and M06/def2-TZVP model chemistries and tables of the structural parameters (bond lengths and bond angles) of Complex-1, Complex-2, $\text{Ru}(\text{C}_2\text{O}_4)_2(\text{H}_2\text{O})_2^-$, and RuO_4 , computed for all of the model chemistries employed herein, as well as a discussion exploring additional ruthenium complexes. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: truhlar@umn.edu.

■ ACKNOWLEDGMENT

The authors are grateful to Yan Zhao for helpful correspondence. This work was supported in part by the Air Force Office of Scientific Research (AFOSR) and by National Science Foundation under grant no. CHE09-56776.

■ REFERENCES

- (1) Takaya, J.; Hartwig, J. F. *J. Am. Chem. Soc.* **2005**, *127*, 5756.
- (2) Grubbs, R. H. *Tetrahedron* **2004**, *60*, 7117.
- (3) (a) Haack, K. J.; Hashiguchi, S.; Fujii, A.; Ikariya, T.; Noyori, R. *Angew. Chem.* **1997**, *36*, 285. (b) Hashiguchi, S.; Fujii, A.; Haack, K. J.; Matsumura, K.; Ikariya, T.; Noyori, R. *Angew. Chem.* **1997**, *109*, 300.
- (4) Dérien, S.; Dixneuf, P. H. *J. Organomet. Chem.* **2004**, *689*, 1382.
- (5) (a) Trost, B. M.; Crawley, M. L. *Chem. Rev. (Washington, DC, U. S.)* **2003**, *103*, 2921. (b) Trost, B. M.; Fraisse, P. L.; Ball, Z. T. *Angew. Chem.* **2002**, *114*, 1101.
- (6) (a) Fomine, S.; Vargas, S. M.; Tlenkopatchev, M. A. *Organometallics* **2002**, *22*, 93. (b) Fomine, S.; Ortega, J. V.; Tlenkopatchev, M. A. *Organometallics* **2005**, *24*, 5696. (c) Fomine, S.; Tlenkopatchev, M. A. *Organometallics* **2007**, *26*, 4491. (d) Fomine, S.; Tlenkopatchev, M. A. *Organometallics* **2010**, *29*, 1580.
- (7) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648. (b) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
- (8) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103.
- (9) (a) Perdew, J. P.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982. (b) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1998**, *298*, 113.
- (10) Mathew, J.; Koga, N.; Suresh, C. H. *Organometallics* **2008**, *27*, 4666.
- (11) (a) Occhipinti, G.; Bjørsvik, H.-R.; Törnroos, K. W.; Jensen, V. R. *Organometallics* **2007**, *26*, 5803. (b) Minenkov, Y.; Occhipinti, G.; Jensen, V. R. *J. Phys. Chem. A* **2009**, *113*, 11833.
- (12) Sieffert, N.; Büehl, M. *Inorg. Chem.* **2009**, *48*, 4622.
- (13) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098. (b) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.
- (14) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
- (15) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.
- (16) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.
- (17) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.
- (18) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13126.
- (19) Zhao, Y.; Truhlar, D. *Theor. Chem. Acc.* **2008**, *120*, 215.
- (20) Seetharaman, S. K.; Chung, M.-C.; English, U.; Ruhlandt-Senge, K.; Sponsler, M. B. *Inorg. Chem.* **2006**, *46*, 561.
- (21) Piacenza, M.; Hyla-Kryspin, I.; Grimme, S. *J. Comput. Chem.* **2007**, *28*, 2275.
- (22) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129.
- (23) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
- (24) Tsepis, A. C.; Orpen, A. G.; Harvey, J. N. *Dalton Trans.* **2005**, 2849.
- (25) Zhao, Y.; Truhlar, D. G. *Chem. Phys. Lett.* **2011**, *502*, 1.
- (26) (a) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
- (b) Zhao, Y.; Truhlar, D. G. *Org. Lett.* **2007**, *9*, 1967.
- (27) Pandian, S.; Hillier, I. H.; Vincent, M. A.; Burton, N. A.; Ashworth, I. W.; Nelson, D. J.; Percy, J. M.; Rinaudo, G. *Chem. Phys. Lett.* **2009**, *476*, 37.
- (28) Diesendruck, C. E.; Tzur, E.; Ben-Asuly, A.; Goldberg, I.; Straub, B. F.; Lemcoff, N. G. *Inorg. Chem.* **2009**, *48*, 10819.
- (29) Sliwa, P.; Handzlik, J. *Chem. Phys. Lett.* **2010**, *493*, 273.
- (30) Dutta, B. F. E.; Curchod, B.; Campomanes, P.; Solari, E.; Scopelliti, R.; Rothlisberger, U.; Severin, K. *Chem.—Eur. J.* **2010**, *16*, 8400.
- (31) Calhorda, M. J.; Pregosin, P. S.; Veiros, L. F. *J. Chem. Theory Comput.* **2007**, *3*, 665.
- (32) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (33) Perdew, J. P. In *Electronic Structure of Solids*; Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991; p 11.
- (34) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (35) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.
- (36) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2008**, *128*, 184109.
- (37) Chai, J.-D.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 084106.
- (38) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
- (39) Schultz, N. E.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 11127.
- (40) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.
- (41) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656.
- (42) Boese, A. D.; Handy, N. C. *J. Chem. Phys.* **2002**, *116*, 9559.
- (43) (a) Dunning, T. H., Jr.; Hay, P. J. *Modern Theoretical Chemistry*; Plenum: New York, 1976; Vol. 3, p 1. (b) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.
- (44) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
- (45) Andrae, D.; Häussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chem. Acc.* **1990**, *77*, 123.
- (46) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
- (47) Zhao, Y.; Truhlar, D. G. *Minnesota Gaussian Functional Module*, Version 4.1. <http://comp.chem.umn.edu/mn-gfm/> (accessed June 2011).
- (48) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
- (49) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.
- (50) (a) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211. (b) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066. (c) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736. (d) Carpenter, J. E.; Weinhold, F. *THEOCHEM* **1988**, *169*, 41.
- (51) Mantina, M.; Valero, R.; Cramer, C. J.; Truhlar, D. G. Atomic Radii of the Elements. In *CRC Handbook of Chemistry and Physics*, 91st ed.; Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, 2010; pp 9–49.

- (52) Fernández, I.; Hermatschweiler, R.; Breher, F.; Pregosin, P. S.; Veiros, L. F.; Calhorda, M. J. *Angew. Chem., Int. Ed.* **2006**, *45*, 6386.
- (53) Pley, M.; Wickleder, M. S. *J. Solid State Chem.* **2005**, *178*, 3206.
- (54) Murphy, C. A.; Cameron, T. S.; Cooke, M. W.; Aquino, M. A. S. *Inorg. Chim. Acta* **2000**, *305*, 225.
- (55) The Inorganic Crystal Structure Database (ICSD). <http://www.fizinformationsdienste.de/en/DB/icsd/> (accessed April 10, 2010).

Correction to Application of Molecular Dynamics Simulations in Molecular Property Prediction. 1. Density and Heat of Vaporization [*Journal of Chemical Theory and Computation* **2011**, *7*

DOI: 10.1021/ct200142z]. Junmei Wang and Tingjun Hou
Department of Pharmacology, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9050, United States

Institute of Nano & Soft Materials (FUNSOM) and Jiangsu Key Laboratory for Carbon-Based Functional Materials & Devices, Soochow University, Suzhou, Jiangsu 215123, P. R. China

One of the author's last and first names were inverted in the original published article. In the original published article, the author's name was Hou Tingjun. The correct author listing should be Tingjun Hou.

DOI: 10.1021/ct2004287

Published on Web 06/24/2011