J. Chem. Theory Comput. **2009,** *5,* 439–442 **439**

JCTC Journal of Chemical Theory and Computation

Theoretical Modeling on the Reaction Mechanism of *p*-Nitrophenylmethylphosphate Alkaline Hydrolysis and its Kinetic Isotope Effects

Violeta López-Canut,[†] Javier Ruiz-Pernía,[‡]
Iñaki Tuñón,*[,†] Silvia Ferrer,[§] and Vicent Moliner*[,§]

Departamento de Química Física, Universidad de Valencia, 46100 Burjassot, Spain, Department of Chemistry, University of Bath, Bath BA2 7AY, United Kingdom, and Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castellón, Spain

Received November 4, 2008

Abstract: We have studied the alkaline hydrolysis of *p*-nitrophenylmethylphosphate (*p*-NPmP) in aqueous solution by means of polarizable continuum models and by hybrid quantum-mechanical/molecular-mechanical (QM/MM) methods. The theoretical predictions of kinetic isotope effects (KIEs) are in very good agreement with the experimental data, confirming a concerted asynchronous molecular mechanism. In addition, comparison of high level DFT theory with semiempirical AM1/d Hamiltonian has allowed checking the reliability of the later to be used in modeling very large molecular models containing phosphorus atoms.

Phosphate esters are fundamental molecules in cellular chemistry.[1] Hydrolysis of the phosphorus−oxygen bond of phosphate esters occurs in biochemical processes including energy storage, biosynthesis or replication of genetic material.[2] The kinetic stability of the phosphorus−oxygen bond in aqueous solution imposes the use of enzymes to reach these chemical rates compatible with life.[3] Thus, to know how enzymes that catalyze these processes (such as kinases, ATPases, and phosphatases) work, it would be interesting to understand the reaction in solution, which is the reference reaction to estimate the performance of these catalysts.

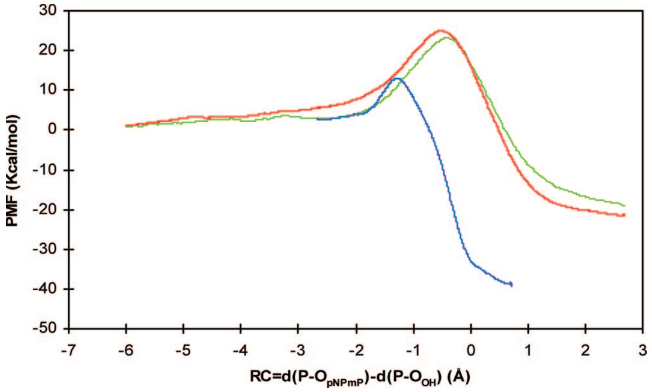


Figure 1. PMF obtained for the hydrolysis of *p*-nitrophenylmethylphosphate (*p*-NPmP) obtained at AM1/MM level at 25 °C (blue line) and at AM1d/MM level at 25 and 95 °C (green and red lines, respectively).

Phosphate esters alkaline hydrolysis may proceed, in principle, with different reaction mechanisms:[2] (i) a S_N1 dissociative mechanism in which departure of the leaving group originates a metaphosphate intermediate, followed by nucleophilic attack ($D_N + A_N$), (ii) an associative mechanism in which nucleophilic attack preceeds the leaving group departure, originating a pentacoordinate intermediate ($A_N + D_N$), or (iii) through a concerted path in which bond breaking and bond forming take place in a single chemical step (A_ND_N). Experimental studies based on kinetic isotope effects (KIEs) on phosphate diesters hydrolysis in alkaline solution suggest that this process takes place through a A_ND_N mechanism.[4,5] Theoretical studies demonstrated that the mechanism becomes more dissociative as the pK_a value of the leaving group is reduced, both for mono- and diesters.[6,7]

Molecular simulations, compared to experimental measurements of KIEs, provide the adequate framework to elucidate the reaction mechanisms in different environments. In this work, the hydrolyses of *p*-nitrophenylmethylphosphate (*p*-NPmP) (see Scheme 1) in aqueous solution has been studied by means of the polarizable continuum model (PCM)[8] and by means of hybrid quantum mechanics/molecular mechanics (QM/MM) methods.[9] In the former, the solvent molecules are described by means of a dielectric continuum and polarizable medium, characterized by a dielectric constant using the Gaussian 03 package of programs.[10] In QM/MM methods, solvent molecules are described explicitly with a box of 55.8 Å side water molecules treated by TIP3P[11] potentials as implemented in DYNAMO library.[12]

Keeping in mind that PCM models describe quite accurately homogeneous environment effects, such as aqueous solutions,[13]

---

* To whom correspondence should be addressed. E-mail: tunon@uv.es (I.K); moliner@uji.es (V.M.).

[†] Departamento de Química Física, Universidad de Valencia.

[‡] Department of Chemistry, University of Bath.

[§] Departament de Química Física i Analítica, Universitat Jaume I.

10.1021/ct800470f CCC: $40.75

Scheme 1

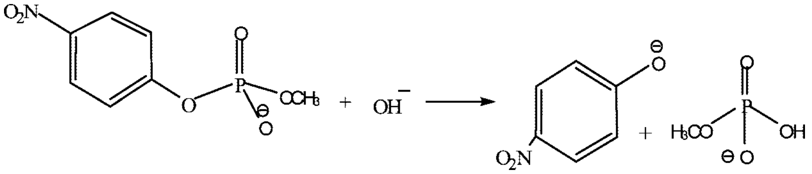


Table 1. Distances (in Å) Defining Breaking and Forming Bonds at the TS Obtained at Two Different Temperatures with the PCM Model and with the Explicit QM/MM (Values Are Reported As Average)[a]

	B3LYP(PCM)		AM1/MM	AM1d/MM	
	25 °C	95 °C	25 °C	25 °C	95 °C
$d(P-O_{OH})$	2.516	2.517	2.913 ± 0.058	2.233 ± 0.057	2.403 ± 0.076
$d(P-O_{pNPmP})$	1.851	1.853	1.663 ± 0.036	1.813 ± 0.047	1.858 ± 0.068
$\Delta G_c^{0,\ddagger}$	23.2	25.1	10.3	20.5	21.6
$\Delta G_c^{0,\ddagger}$ (exptl)	25.9[b]				

[a] Free energies of activation are reported in kcal mol^{-1}. [b] Experimental value was obtained at 42 °C from ref 20.

Table 2. B3LYP(PCM) and AM1d/MM Primary ^{18}O KIE, Secondary ^{18}O KIE, and Secondary ^{15}N KIE for the Hydrolysis of *p*-NPmP in Solution at 95 °C

	B3LYP (PCM)	AM1d/MM	experimental
$(k_{^{16}O}/k_{^{18}O})_{p\text{-NPmP}}$	1.0047	1.0044 ± 0.0033	1.0059 ± 0.0005[a]
$(k_{^{16}O}/k_{^{18}O})_{OH}$	1.0238	1.0125 ± 0.0054	1.0227 ± 0.0100[b]
$(k_{^{16}O}/k_{^{18}O})_{ne,p\text{-NPmP}}$	0.9977	0.9966 ± 0.0032	0.9949 ± 0.0006[a]
$(k_{^{16}O}/k_{^{18}O})_{p\text{-NPmP}}$	1.0029	1.0003 ± 0.0020	1.0016 ± 0.0002[a]

[a] Values obtained from ref 4, measured at 95 °C. [b] Value obtained from ref 5. See ref 22.

B3LYP density functional theory (DFT) functional, within the 6−31+G* basis set, was used for the PCM optimizations and frequency calculations, while electronic energy was further refined by means of single point calculations using the 6−311++G** basis set. These results will be used as the reference for comparison with QM/MM calculations with low-level QM methods. QM region on these calculations were described with the AM1 Hamiltonian,[14] which is known to not be a sufficiently accurate to model reactions involving phosphorus atoms because it does not incorporate d orbitals, as demonstrated by Marcos et al.[15] in a deep comparison study between different semiempirical methods on pentacordinated phosphorus TSs involving reactions in gas phase. For this reason, the new AM1/d-PhoT Hamiltonian (hereafter simply named as AM1d)[16] was also used to describe the QM region. Thus, reactants, transition state (TS) and productlike structures have been localized in the PCM approach at 25 °C, as the reference temperature, and 95 °C, the temperature at which the experimental KIEs for this reaction had been measured.[4] These calculations have been done at the B3LYP(PCM) level by changing the dielectric constant of the solvent from 78.5 to 56.1.[17] Once the TS was localized in the continuum medium, this was used as initial structure to equilibrate the solvent in the QM/MM simulations (both using the AM1 and AM1d methods). The system was relaxed by means of 250 ps of Langevin Dynamics using the NVT ensemble at the two temperatures. A total of 5816 water molecules were present in the 55.8 Å side cubic box at 298 K, while at 368 K only 5576 water molecules were included to have the correct density. The time step employed in all the simulations was 1 fs.

Starting from these structures, the free energy profile of the reaction was then obtained in terms of the potential of mean force (PMF) with the antisymmetric combination of distances defining forming and breaking bond, RC=d(P−OpNPmP)−d(P−OOH), as the distinguished reaction coordinate (negative values correspond to reactants and positive values to products). The different values of the reaction coordinate sampled during the simulations were pieced together by means of the weighted histogram analysis method (WHAM)[18] to construct the full distribution function from which the PMFs were obtained. The value of the force constant used for the harmonic umbrella sampling was 2500 kJ mol^{-1} Å^{-2}. Each window consisted of 10 ps of equilibration, followed by 15 ps of production, using two reference temperatures of 25 and 95 °C. Resulting PMFs are depicted in Figure 1, while activation free energies and distances defining breaking and forming bonds at the TS are reported in Table 1. It must be pointed out that the free energy barriers were computed as free energy difference between the TS and the fully solvated separated reactants plus a correction at standard concentration 1M,[19] while averaged bonds lengths were obtained from the 15 ps of MD production.

Geometrical and energetic results obtained at 25 °C employing the B3LYP(PCM) and the AM1d/MM methods are in reasonable agreement (see Table 1). Both computational levels describe a concerted but asynchronous mechanism for the hydroxide attack to the phosphate diester. As a reference, we also computed the free energy change associated to the complete breaking of the $P-O_{pNPmP}$ bond at the B3LYP(PCM) level. The resulting free energy at 25 °C is 30.4 kcal/mol, substantially higher than the free energy barriers obtained for the concerted mechanism. We also tried to explore an $A_N + D_N$ mechanism through a pentacoordinated phosphorus species. However it was not possible to locate such a structure either as a minimum or as TS. We have estimated that the free energy cost associated to reaching this quadratic region is about 34 kcal/mol, also higher than the free energy barrier obtained in our proposed concerted mechanism. Thus the A_ND_N mechanism seems to be more likely for this reaction than the $D_N + A_N$ or the $A_N + D_N$ one. This finding is in agreement with a previous work.[6] The AM1/MM method describes a too early TS, a larger forming bond and a shorter breaking bond. As a consequence, and in accordance with the Hammond postulate, the free energy barrier appears dramatically smaller, as observed in Figure 1 and in values reported in Table 1. Comparison with experimental barrier obtained by Zalatan et al. at 42 °C,[20] seems to indicate that AM1d/MM calculations render much more reliable data than using the standard AM1 semiempirical method. This result is in accordance with inherent limitations on the later to properly describe phosphorus atoms.

The results obtained at 95 °C computed at B3LYP(PCM) and AM1d/MM levels were also in good agreement, confirming the reliability of the AM1d semiempirical Hamiltonian. Both methods describe also a similar reaction mechanism, although the transition structures are slightly more dissociative than at 25 °C. The increase of the activation free energy with the temperature is caused by a diminution of the dielectric effect, which shields the electrostatic repulsion between the reacting fragments and to the negative sign of the activation entropy characteristic of association processes.

To confirm the nature of the molecular mechanism of the reaction, the effects of isotopic substitutions upon the kinetics (KIE) of the hydrolysis of *p*-NPmP in solution have been computed and compared with available experimental data reported in the literature.[4,5] In particular, primary ^{18}O KIE for substitution at oxygen atom of nucleophile and leaving group, $(k_{^{16}O}/k_{^{18}O})_{OH}$ and $(k_{^{16}O}/k_{^{18}O})_{p\text{- NPmP}}$, respectively, secondary ^{18}O KIE for substitution at the two nonether phosphorus oxygen atoms, $(k_{^{16}O}/k_{^{18}O})_{ne,p\text{-NPmP}}$, and secondary ^{15}N KIE for substitution at nitrogen atom of nitro group of leaving group, $(k_{^{14}N}/k_{^{15}N})_{p\text{-NPmP}}$, have been computed. Calculations have been done within the B3LYP(PCM) and AM1d/MM methods at 95 °C with the CAMVIB/CAMISO programs.[21] The results, reported in Table 2, were obtained with the rigid-rotor/harmonic-oscillator approximation by computing the Hessian of the atoms corresponding to the full solute as depicted in Scheme 1. For the AM1d/MM values, ten different structures of the transition and reactants states were optimized as first-order saddle points and minima on the QM/MM potential energy surface (PES). These were obtained starting from different configurations of the respective MD productions. These QM/MM calculations were carried out with DYNAMO keeping frozen all atoms far from a sphere of 22 Å centered on the solute. Computed KIEs were obtained as an average over these structures.

The analysis of data reported in Table 2 shows that both B3LYP(PCM) and AM1d/MM methods predict the same kind of KIEs: normal primary ^{18}O KIEs, slightly smaller for the labeling at the leaving group position, and inverse secondary ^{18}O KIEs, which is in accordance with other measurements of secondary KIEs for S_N2 reactions. Both methods also render small secondary ^{15}N KIE on the nitro group at para position of the ring. Thus, the AM1d/MM method seems to be a promising approach to study chemical reactions involving phosphorus atoms in complex environments, such as solutions or enzymes, with the evident advantage of being much less computing demanding than high level DFT methods.

Finally, it can be observed that theoretical predictions are in very good agreement with the experimental data, thus giving credit to the proposed molecular mechanism, a concerted but asynchronous associative hydrolysis of the pNPP in solution.

Acknowledgment. This work was supported by DGI project CTQ2006-15447-C02/BQU. V.L.-C. thanks Ministerio Ciencia y Tecnología (Spain) for a doctoral grant.

References

(1) Westheimer, F. H. *Science* **1987**, *235*, 1173–1178.

(2) Cleland, W. W.; Hengge, A. C. *Chem. Rev.* **2006**, *106*, 3252–3278.

(3) Lad, C.; Williams, N. H.; Wolfenden, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5607–5610.

(4) Hengge, A. C.; Tobin, A. E.; Cleland, W. W. *J. Am. Chem. Soc.* **1995**, *117*, 5919–5926.

(5) Cassano, A. G.; Anderson, V. E.; Harris, M. E. *J. Am. Chem. Soc.* **2002**, *124*, 10964–10965.

(6) Klahn, M.; Rosta, E.; Warshel, A. *J. Am. Chem. Soc.* **2006**, *128*, 15310–15323.

(7) Rosta, E.; Kamerlin, S. C. L.; Warshel, A. *Biochemistry* **2008**, *47*, 3725–3735.

(8) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094.

(9) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

(10) 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.

(11) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(12) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*, 1st ed.; Cambridge University Press: Cambridge, U.K., 1999.

(13) (a) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187–4225. (b) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3094. (c) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760–768.

(14) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(15) Marcos, E.; Anglada, J. M.; Crehuet, R. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2442–2450.

(16) Nam, K.; Cui, Q.; Gao, J. L.; York, D. M. *J. Chem. Theory Comput.* **2007**, *3*, 486–504.

(17) Lide, D. R. *Handbook of Chemistry and Physics*, 77th ed; CRC Press: Boca Raton, FL, 1996.

(18) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(19) PMFs were traced up to a reaction coordinate of −6 Å. The free energy contribution resulting from complete separation was

evaluated using the PCM model as a result of the average of 10 independent evaluations for structures obtained at RC = −6 Å.

(20) Zalatan, J. G.; Herschlag, D. *J. Am. Chem. Soc.* **2006**, *128*, 1293–1303.

(21) Williams, I. H. *Chem. Phys. Lett.* **1982**, *88*, 462; Williams, I. H. *THEOCHEM* **1983**,, *11*, 275.

(22) Computed from data measured by Cassano et al.[5] for thymidine-5′*p*-NPP at 37 °C using the following approximation:

$$\ln(\mathrm{KIE}_{95°\mathrm{C}}) = \left(\frac{310K}{368K}\right)\ln(\mathrm{KIE}_{37°\mathrm{C}})$$

CT800470F

J. Chem. Theory Comput. **2009,** *5,* 443–451

JCTC Journal of Chemical Theory and Computation

Computational Investigation of Wave Packet Scattering in the Complex Plane: Propagation on a Grid

Robert E. Wyatt and Brad A. Rowland*

Department of Chemistry and Biochemistry, University of Texas, Austin, Texas 78712

Received June 26, 2008

Abstract: The time-dependent scattering of a wave packet from a Gaussian barrier is investigated computationally in the complex *z*-plane. The initial wave packet and the potential energy are obtained through analytic continuation from functions specified on the real-axis. The wave packet is then propagated on the two-dimensional grid. For a low initial wave packet energy, the time evolution is followed by plotting the following functions: $|\psi(z,t)|$, $real(\psi(z,t))$, and the quantum momentum function (QMF), $p(z,t)$. In the reflected packet, an important role is played by ripples (quasi-nodes) forming above the real axis. As these quasi-nodes move down across the real axis, they are 'detected' as 'interference oscillations' in the density. In contrast, the component of the packet below the real axis makes a significant contribution to the transmitted packet. Vector maps of the QMF show hyperbolic flow around quasi-nodes and counterclockwise circular flow around transient stagnation points, where the QMF vanishes. However, when the Pólya vector field (defined by $P(z,t) = p^*(z,t)$) is plotted, circular counterclockwise flow is obtained near the quasi-nodes. The real and imaginary parts of the quantum action function $S(z,t)$ are plotted and the vorticity, defined by the curl of the Pólya field, is used to pinpoint regions of nonanalyticity in the QMF.

1. Introduction

The scattering of a wave packet from a barrier in one-dimension has been investigated computationally since the mid-1960s. Excellent examples are shown in the animations accompanying the book *Visual Quantum Mechanics*.[1] In studies of this type, the single coordinate (x) is invariably real-valued. In what may seem a bizarre extension, in the present study this coordinate will be replaced by the complex coordinate z, and analysis will be performed on the 'extended' 2D scattering problem in the complex plane. The outcome will be that an observer confined to the real axis will detect only a fraction of the rich dynamics ensuing in the complex plane. However, before getting to this, we need to address the following question: Why study quantum dynamics in the complex plane.

The current investigation was motivated by recent studies using approximate *complex-valued quantum trajectories* to solve the time-dependent Schrodinger equation for barrier scattering problems. Complex-valued *classical trajectories* were introduced in the early years of quantum mechanics to deal with the turning point connection problem in WKB theory,[2,3] and in the early 1970s, trajectories of this type were used for classically forbidden processes.[4–8] In 1987, Huber and Heller[9] generalized the real-space version of Gaussian wave packet dynamics[10] to allow for the propagation of complex-valued classical trajectories. An alternative WKB-type formulation employing complex classical trajectories has also been described.[11]

When it comes to complex *quantum* trajectories, the developments are much more recent. These trajectories provide a method for solving the quantum Hamilton-Jacobi equation (QHJE), which is obtained from the Schrodinger equation through use of the ansatz, $\psi(x,t) = \exp[iS(x,t)/\hbar]$ (see the text by Tannor[12]). The complex action function, $S(x,t)$, is the solution to the QHJE. With the guidance condition, $p(x,t) = \partial S(x,t)/\partial x$, and the dynamical equation $dx(t)/dt = p(x,t)/m$, we are led to trajectories with *complex values* for both x and p (the time remains real-valued).

With the restriction that the wave function is known in advance, it is very informative to analyze and plot complex

---

* Corresponding author e-mail: barowland@gmail.com.

10.1021/ct800248w CCC: $40.75

quantum trajectories. Over the past six years, for the *stationary state* case, quantum trajectories have been analyzed and plotted for bound state and scattering problems.[13−19] A detailed analysis of complex quantum trajectories for several one-dimensional stationary state scattering problems has recently been presented.[20,21] In addition, quantum trajectories have recently been analyzed for several *nonstationary* problems.[20,22]

For *nonstationary problems* where the wave function is *not known* in advance, there have been significant developments[23−31] on the use of *real-valued quantum trajectories* for solving the QHJE. In contrast to these developments concerning real-valued quantum trajectories, the use of *complex valued quantum trajectories* for solving the QHJE is at an earlier stage of development. In 2006, Tannor and co-workers derived equations of motion for approximate quantum trajectories evolving in complex phase space.[32] (Interesting commentary on this work has been presented.[33,34]) This method was applied to the one-dimensional scattering of a wave packet from an Eckart barrier,[32,35] and it has been used to describe interference oscillations in the reflected wave packet and to deal with the node problem.[36,37]

In order to enhance our understanding of quantum dynamics in the complex plane, we will study the scattering of an initial Gaussian wave packet from a complex potential surface on a two-dimensional computational grid. Prior to this work, the exact dynamics of wave packet barrier scattering has not been studied in the complex plane. Some new physical results will arise in this and the following paper, which deals with the dynamics of exact quantum trajectories. Some of these new results include the formation of transport of a string of quasi-nodes away from the barrier region, the twisting of trajectories launched from different initial positions around stagnation curves that form between the quasi-nodes, and the twisting of Pólya vectors around these quasi-nodes. Understanding the exact dynamics may lead to improved methods for propagating approximate quantum trajectories, especially for the reflected wave packet.

The computational strategy mentioned above appears problematic because the amplitude of the wave function becomes very large even for small departures from the real axis. In order to mitigate the effect of these large values, it is necessary to periodically damp the wave packet near the edges of the grid. This approach leads to stable and reproducible results within a subspace of the computational grid termed the ‘viewing window’. The resulting dynamics is then followed by plotting, for a number of time steps, the real part of the wave function and vector maps of the quantum momentum function (QMF). This function is defined by

$$p(z,t)=\frac{\hbar}{i}\frac{1}{\psi(z,t)}\frac{\partial\psi(z,t)}{\partial z}\tag{1}$$

Interesting dynamical features are revealed when this complex function is plotted as a set of vectors, $\vec{p}(z,t) = [p_r(z,t),p_i(z,t)]$, emanating from points in the complex plane. The imaginary component of this vector is responsible for ‘vertical’ transport orthogonal to the real axis. Two special points are of interest: poles in the QMF occur at *nodes* where $\psi(z,t) = 0$, and the QMF becomes zero at *stagnation points* where $d\psi(z,t)/dz = 0$ (and $\psi(z,t) \neq 0$). In addition, for the barrier scattering problem described in this study, the QMF exhibits *hyperbolic flow* around quasi-nodes in the density[21] (these are local minima where the amplitude becomes small but does not reach zero). In order to analyze this flow, the Pólya vector field is introduced,[38−40] and maps of this field show circular counterclockwise flow around quasi-nodes and hyperbolic flow around stagnation points. In addition to these fields, the vorticity of the Pólya field will be shown for the reflected wave packet.

The remainder of this study is organized as follows. The scattering problem and the computational grid are set up in section 2, and computational aspects are discussed in section 3. Plots illustrating wave packet evolution in the complex plane are presented in section 4. Maps showing the quantum momentum field are described in section 5, and properties of the Pólya vector field (including the vorticity) are presented in section 6. The real and imaginary parts of the quantum action function are described in section 7. Finally, a summary appears in section 8.

2. Scattering Problem and Computational Grid

The initial wave packet in the complex plane is obtained through analytic continuation of a Gaussian defined on the real axis. The resulting complex function is

$$\psi(z)=Ne^{-\beta(z-z_0)^2}e^{ik_0(z-z_0)}\tag{2}$$

where the initial translational energy is $E = \hbar^2k_0^2/(2m)$. This function is centered at the point $z_0 = (x_0,0)$ and is normalized *only* for integration along the real axis. The normalization factor is $N = (2\beta/\pi)^{1/4}$. The complex-valued Gaussian barrier, centered on the real axis at the point $z_b = (x_b,0)$, is given by

$$V_G(z)=V_0e^{-\gamma(z-z_b)^2}\tag{3}$$

Both $\psi(z)$and $V_G(z)$ possess essential singularities when $y \rightarrow \pm\infty$.

Both $\psi(z)$and $V_G(z)$ have very large magnitudes for points off the real axis, and this presents a significant challenge for the computational investigation of wave packet scattering in the complex plane. (To illustrate how bad the situation is, let $\beta = 6$ in eq 2. Then, for points fairly close to the real axis, for example $z = (x_0, \pm 3)$, the Gaussian function exp $(-\beta z^2)$ has about the same magnitude as Avogadro’s number!) We will assume that the dynamics relatively close to the real axis is of paramount concern. So, in order to propagate the wave packet on a relatively small two-dimensional computational grid, the packet will be damped or absorbed near the boundaries. However, even if the *initial* packet is damped, the propagated packet will eventually scatter from the boundaries at later times. For this reason, at later times, damping functions and absorbing potentials will be utilized.

In order to implement this procedure, damping functions are introduced near all four boundaries of the 2D computational grid. First, the damping kernel along the x coordinate, for example, is defined by

$$d(a,b;x) = [1 + \tanh(a(x - b))]/2 \tag{4}$$

This function takes the following limits:

$$d \to 0 \text{ for } x \ll b$$
$$d \to 1 \text{ for } x \gg b$$
$$\text{and } d \to 1/2 \text{ for } x = b$$

In addition, with a sign change for a, the function $d(-a,b;x)$ takes the following limits:

$$d \to 1 \text{ for } x \ll b$$
$$d \to 0 \text{ for } x \gg b$$

The product of the two functions $d(a,b_1;x)$ and $d(-a,b_2;x)$ (for $b_1 < b_2$) gives a box function with rounded edges for which the value is essentially zero except when $b_1 < x < b_2$. In the latter region, this function takes the value unity. Raising this function to a (positive) power then gives a function which damps to zero very rapidly outside of the 'internal' region between b_1 and b_2. This function is defined by

$$D(a,n,b_1,b_2;x) = [d(a,b_1;x)d(-a,b_2;x)]^n \tag{5}$$

For brevity, this function will be denoted $D(x)$. The parameters a and n can be used to control how rapidly $D(x)$ drops to zero.

Since both $\psi(z)$ and $V_G(z)$ become very large away from the real axis, each function was multiplied by damping functions along the y coordinate prior to the time propagation. As a result, the damped initial wave function and potential are given by

$$\begin{aligned} \psi_D(z) &= D(y) \cdot \psi(z) \\ V_D(z) &= D(y) \cdot V_G(z) \end{aligned} \tag{6}$$

Damping was not applied in the x direction since these two functions are relatively localized in this direction.

As the wave packet advances in time, parts of it will eventually scatter from the edges of the computational grid. In order to confine the wave packet and prevent edge reflection at later times, two approaches were used. The first method is to repeatedly damp the wave function near the four boundaries of the grid. For this purpose, a two-dimensional damping function was applied every M time steps. This damping function is defined as product of the one-dimensional damping functions, $D(x)D(y)$.

Another approach may also be used to confine the spreading wave packet and prevent edge reflections. For this purpose, negative imaginary absorbing potentials[41,42] are introduced. In this case, most of the amplitude entering boundary regions near the edges of the grid is absorbed, although some edge reflection is hard to avoid. The damping functions defined earlier can be used to construct the absorbing potentials. We first note that the function $(1 - D(x))$ gives a 'well' with rounded edges. Within the internal region $(b_1 < x < b_2)$, this function is zero, and outside of this region it takes on the value one. Along the x coordinate, the absorbing potential, a negative imaginary function, is then

$$V_{abs}(x) = -iV_{ax}(1 - D(x)) \tag{7}$$

where the parameter V_{ax} controls the strength of the absorber. The total potential on the computational grid is then the sum of the damped Gaussian potential in eq 6 added to absorbing potentials along the x and y directions:

$$V(z) = V_D(z) - iV_{ax}(1 - D(x)) - iV_{ay}(1 - D(y)) \tag{8}$$

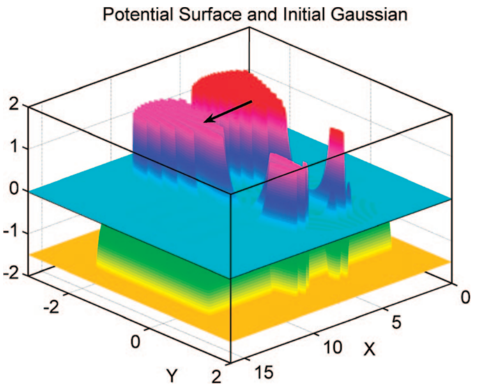


Figure 1. The real part of the (damped) initial Gaussian wave packet (marked with an arrow directed toward the barrier) and the imaginary part of the total potential energy. The initial wave packet energy is $E = V_0/4$. The total potential includes both the Gaussian barrier (damped along the y-coordinate) and the absorbing potential, which is evident near the edges of the computational grid. For viewing purposes, the negative imaginary absorbing potential has been cut off at the lower limit $-1.5i$. (The scaled Gaussian potential is plotted, $V_{sc}(z) = (25/V_0)V_D(z)$).

Parameter values will now be specified (all values are in atomic units). First, the computational grid extends from $x_{\min} = 0$ to $x_{\max} = 16$ and from $y_{\min} = -3$ to $y_{\max} = 2$. There are $N_x = 351$ and $N_y = 251$ grid points along the two coordinates. The initial wave packet is centered at $z_0 = (6,0)$, the width parameter is $\beta = 6$, and the mass is $m = 2000$. The damping parameters used for this function are $a = 10$, $b_1 = -1.7$, $b_2 = 1.4$, and $n = 6$. The barrier is centered at $z_b = (9,0)$, the width parameter is $\gamma = 4$, and $V_0 = 0.035$. The damping parameters used for the potential are $a = 6$, $b_1 = -2.2$, $b_2 = 1.4$, and $n = 6$. The absorbing potential parameters are $V_{ax} = 5$ and $V_{ay} = 1$. The parameters for the $D(x)$ damping function are $a = 10$, $b_1 = 1$, $b_2 = 15$, and $n = 6$. Parameter values (except those specifying the initial wave packet and the Gaussian potential) were determined through experimentation, the goal being to obtain reproducible and accurate values for the scattering wave function at relatively late times ($t \leq 1400$) within a *viewing window* (see below).

Figure 1 shows the real part of the (damped) initial Gaussian wave packet (marked with an arrow directed toward the barrier) and the imaginary part of the total potential energy. The initial wave packet energy is $E = V_0/4$. The pure imaginary absorbing potential is evident from the deep wells near all four edges of the computational grid. For viewing purposes, upper and lower cutoff values were applied to both the wave packet and the potential energy. In order to provide more resolution in the following figures, the results (wave function and quantum momentum) were plotted in a viewing window which is smaller than the computational grid. The size of this window will be evident from the range of the axes in the figures.

3. Computational Aspects

Three explicit time-integration algorithms were utilized for propagating the wave packet in the complex space. The first of these, the forward Euler method (first-order accuracy in Δt, the time step), is given by[43]

$$\psi(z, t+\Delta t) = \psi(z, t) - \Delta t \cdot (i/\hbar)\hat{H}\psi(z, t) \tag{9}$$

where $\hat{H}$ is the Hamilton operator. For this algorithm, the absorbing potential in eq (8) was used to deal with reflection from the edges of the grid. The second propagation algorithm, the leapfrog method (second-order accuracy), is given by[43]

$$\psi(z, t+\Delta t) = \psi(z, t-\Delta t) - 2\Delta t \cdot (i/\hbar)\hat{H}\psi(z, t) \tag{10}$$

For this algorithm, the damping potential $D(x)D(y)$ was applied to the current wave function after every $M = 20$ time steps. Finally, a version of the Verlet method[44] (fourth-order accuracy) for wave function propagation was used. This algorithm is given by

$$\psi(z, t+\Delta t) = 2\psi(z, t) - \psi(z, t-\Delta t) - (\Delta t/\hbar)^2\hat{H}(\hat{H}\psi(z, t)) \tag{11}$$

For this case, in common with the leapfrog method, the damping potential was applied to the current wave function every $M = 20$ time steps.

For some of the computations described later, the leapfrog method was used, with the time step in the range $\Delta t = 0.5-1$. Other calculations were run using the Euler method, with a smaller time step in the range $\Delta t = 0.1-0.5$. Calculations were also run using the Verlet method, with a time step in the range $\Delta t = 1-3$. Within the viewing window mentioned earlier, results obtained using these three algorithms agreed very well with each other.

In order to operate with the Hamiltonian on the wave function, we need to evaluate $\partial^2\psi(z)/\partial z^2$. Assuming analyticity of the wave function on the computational grid, this was done by taking the derivative with respect to x: $\partial^2\psi(z)/\partial x^2$. The quantity $\partial^2\psi(z)/\partial z^2$ can also be obtained by taking the y derivative: $-\partial^2\psi(z)/\partial y^2$. As a result, it can be verified numerically that Laplace's equation[39] is satisfied: $\partial^2\psi(z)/\partial x^2 + \partial^2\psi(z)/\partial y^2 = 0$.

In order to compute the quantum momentum from the wave function according to eq 1, we need to evaluate $\partial\psi(z)/\partial z$. Again assuming analyticity, this was done by taking the derivative with respect to x: $\partial\psi(z)/\partial x$. The same result for $\partial\psi(z)/\partial z$ can also be obtained by taking the y derivative: $-i\partial\psi(z)/\partial y$. As a result, it can be verified numerically that the Cauchy-Riemann equation[39] is satisfied: $\partial\psi(z)/\partial x = -i\partial\psi(z)/\partial y$.

4. Wave Packet Evolution

The time dependence of the wave packet was studied by plotting *real*($\psi(z,t)$) within the viewing window. For six time steps, Figure 2 shows surface plots of this function along with mesh plots of *imag*($V(z)$) (see eq 8). In common with maps showing the quantum momentum (section 5), the initial wave packet energy is $E = V_0/4$. Note that the viewing angle is slightly different for each part of this figure.

Part (a) shows the initial wave packet, labeled with an arrow heading toward the barrier. The component of the wave packet at negative values of y at later times makes the dominant contribution to the transmitted wave packet. However, the component of the wave packet at positive values of y makes an important contribution to oscillatory structure in the reflected wave packet. In part (b), for $t = 300$, the leading edge of the wave packet encounters the barrier potential below the real axis. In part (c), for $t = 500$, the leading edge of the packet has penetrated the barrier region for $y < 0.1$. Note that the de Broglie wavelength for the transmitted packet on the left side of the figure is much smaller than for the reflected packet on the opposite side of the figure. Starting with part (d), for $t = 700$, and continuing with part (e), for $t = 1000$, additional amplitude penetrates to the product side of the barrier. In parts (e) and (f), the latter for $t = 1200$, the component of the wave packet above the real axis undergoes reflection from the barrier. It will be seen in the next section that undulations created *above* the real axis are deflected toward the region below the real axis, and as they cross this axis an observer positioned there would record ripples in the density. In the following section, these features will be linked to the time development of the quantum momentum function.

4. Quantum Momentum Maps

The time dependence of the quantum momentum field, $p(z,t)$, can be studied by superimposing vector maps of the QMF on contour plots of the amplitude of the wave function. (If $\psi(x,y,t) = R(x,y,t)\exp(i\phi(x,y,t))$, where R and ϕ are real-valued, then the amplitude is $|\psi(z,t)| = R(z,t)$.) It was decided not to display flux maps (even though they were computed), where $\vec{j} = \rho\vec{v}$, because very large values for the flux away from the real axis obscured the more interesting (and much smaller) values near this axis. For the same six time steps used in Figure 2, Figure 3 shows the momentum field superimposed upon contour maps of $\log_{10}|\psi|$. (In order not to lose detail near the real axis, *scaled momentum vectors* were plotted: if the magnitude of the momentum vector was greater than 1.5 times the average momentum for vectors within the viewing window, then the length of the momentum vector was reset to 1.5 times the average value).

The initial momentum distribution and amplitude are shown in part (a). These momentum vectors swirl counterclockwise around the point $z_{pp} = (6,0.493)$ (which lies just above the top edge of the figure). This *pinch point*[20] is the saddle point for the initial density in the complex plane and occurs at the distance $y_{pp} = k_0/(2\beta)$ above the center of the initial wave packet on the real axis. In addition, the length of the momentum vectors increases with the radial distance from this point. In part (b), for $t = 300$, the leading edge of the wave packet encounters the lower part of the barrier potential near $y = -0.3$. Momentum vectors in this region (and in the entire leading edge of the wave packet) are directed upward. In contrast, momentum vectors on the trailing edge of the wave packet are directed down toward negative values of y. An interesting feature shown by the amplitude contours is the tilt of the vertical axis, with the result that amplitude below

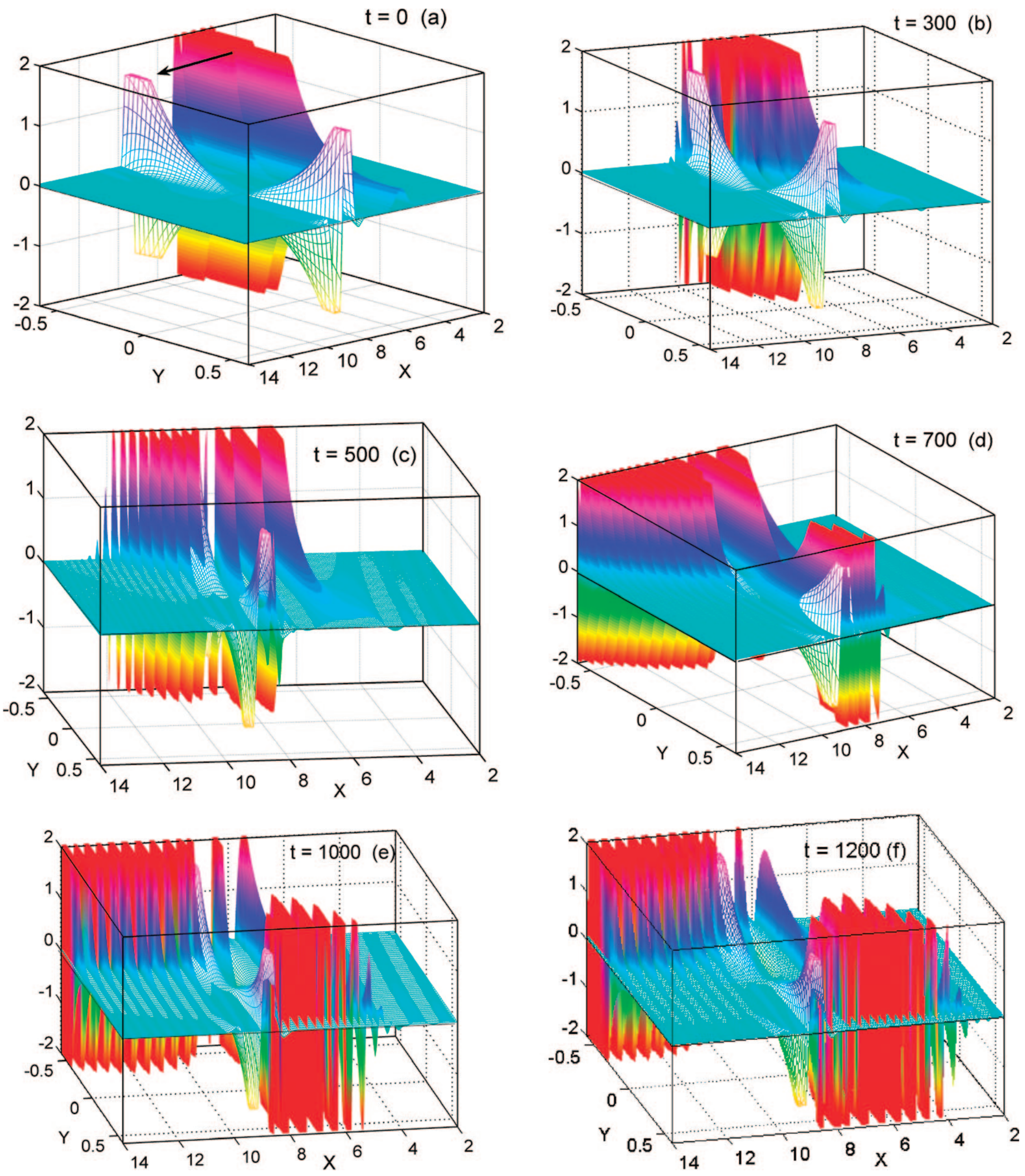


Figure 2. Time dependence of the real part of the scattering wave function. The initial wave packet energy is $E = V_0/4$. In each figure, a mesh plot of the imaginary part of the scaled complex Gaussian potential is also shown (see the caption to Figure 1). For these viewing angles, the initial wave packet moves from right to left, and the transmitted region is on the left-side of the figure.

the real axis encounters the barrier region before the upper component. In part (c), for $t = 500$, the leading edge of the packet for $y < 0.1$ has penetrated the lower part of the barrier region. In addition, the first of a series of ripples or undulations is developing near $y = 0.35$ and $x = 8.2$ on the reflected side of the barrier. These ripples form (for $400 < t < 700$) when amplitude that is thrust upward between $x = 8$ and $x = 9$ interferes with amplitude moving forward for $y > 0.3$. Starting with part (d), for $t = 700$, several ripples are noted above the real axis between $y = 0.1$ and $y = 0.3$. As seen in parts (e) and (f), for $t = 1000$ and $t = 1200$, respectively, these ripples (featuring five local maxima) gradually move down toward negative values of y, and the momentum vectors deflect (see the discussion below) around quasi-nodes in the amplitude. When these ripples pass near the real axis, five relatively large oscillations are observed in the density along this axis. However, an observer confined to this axis would not sense the *vertical component* of the flow around the local minima as they form and decay. An additional feature shown in parts (e) and (f) is the ‘fracture’ of the wave packet near the barrier region. Although the momentum vectors for the transmitted packet have large components parallel to the real axis, those on the reflected side of the barrier form a much more complex pattern characterized by alternating regions with the flow directed either up or down in the y direction.

A blowup of the central portion of Figure 3(e), shown in Figure 4(a), displays counterclockwise circulation of the momentum vectors around four local maxima in the ampli-

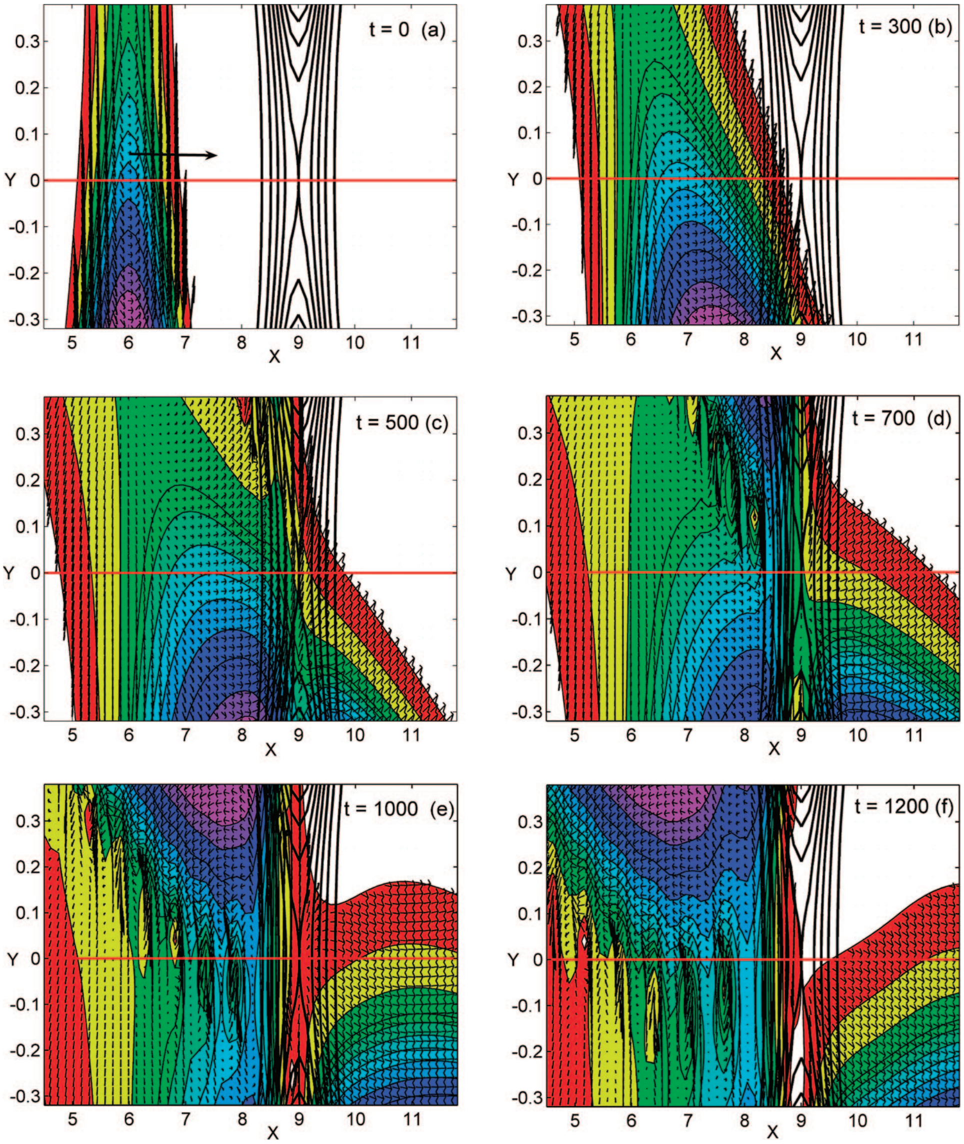


Figure 3. Time dependence of the quantum momentum (arrows) superimposed upon contour maps of log $_{10}|\psi|$. The initial wave packet energy is $E = V_0/4$. The *lower component* of the initial wave packet in (a) moves toward the right. The real axis is shown by the horizontal red line. Contours of the complex Gaussian potential energy are shown by the black contours. The 11 contour values for log $_{10}|\psi|$ are given by the following: −2, −1.3, −0.8, −0.4, −0.25, −0.1, 0.05, 0.25, 0.45, 0.7, 0.9.

tude. The momentum vectors have very small magnitudes near the local maxima, which are transient *stagnation points*. As time proceeds, these maxima gradually move toward the lower left of the figure, i.e., toward smaller values of x and negative values of y. On the real axis, local peaks in the amplitude for the reflected wave packet are approximately equally spaced and occur near the positions $x = 6.6$, 7.1, 7.65, and 8.2. At each of these positions, the momentum vectors have large components toward the left, in accord with the wave packet undergoing reflection from the barrier.

It is shown elsewhere[21] that the QMF near a stagnation point (z_s) may be approximately written $p(z,t) = p'(z_s,t)(z-z_s)$, where the spatial derivative is $p'(z,t) = \partial p(z,t)/\partial z$. If a test particle of mass m is launched from the position z_0, then it will follow the trajectory $z(t) = z_s + [z_0 - z_s] \exp[(p'(z_s,t)/m)t]$. It is useful to express $p'(z,t)$ in terms of real and imaginary parts: $p' = \alpha + i\beta$. For the case where $\alpha = 0$ and $\beta > 0$, the particle will trace out a counterclockwise orbit in the complex plane. This is the situation pertaining to the momentum field shown in Figure 4(a). Other categories of orbits (for various values of α and β) near a stagnation point have been described.[21]

Between the four local maxima in Figure 4(a) are three quasi-nodes (the approximately circular yellow regions). In two or more *real* coordinates, quantum momentum vectors form approximately *circular paths* around nodes in the wave function. The circulation integral around these closed paths is quantized in integer multiples of $2\pi\hbar$. Bearing this in mind, it may be surprising that the momentum vectors in Figure 4(a) do *not* show circular flow around these local minima. Rather, the flow follows four *hyperbolic paths*: directed up on the left-side, down on the right-side, horizontal toward the left above the minimum, and horizontal toward the right below the minimum. In the following section, this flow will be analyzed in terms of another vector field.

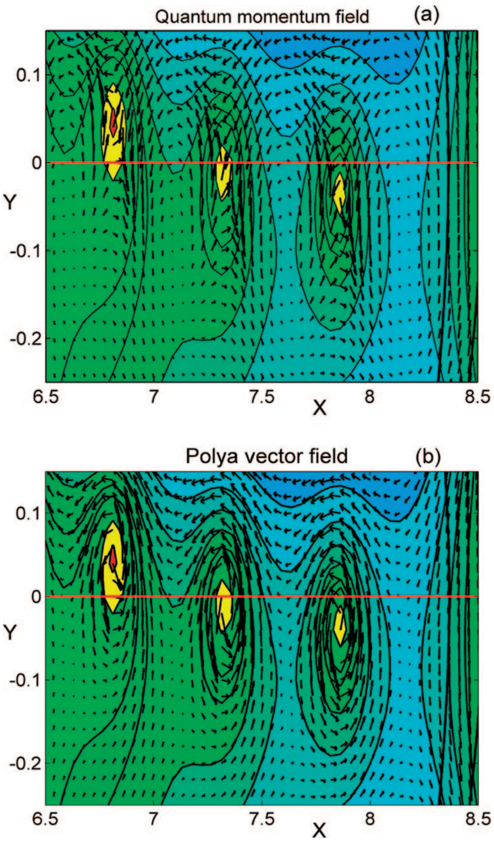


Figure 4. (a) Quantum momentum field (arrows) superimposed upon a contour map of log $_{10}|\psi|$ for $t = 1000$. This figure, which shows hyperbolic flow around three quasi-nodes, is a blowup of the central region of Figure 3(e). The real axis is shown by the red line. Several contours (in black) of the complex Gaussian potential energy are shown on the right side of the figure. (b) Pólya vector field associated with this momentum field, showing counterclockwise circular flow around the three quasi-nodes.

6. The Pólya Vector Field and Vorticity

In order to analyze hyperbolic flow, the QMF for a model wave function will be plotted. For the function $\phi(z) = (z - z_0)\exp(-\beta(z - z_0)^2)$, which has a node at z_0, the QMF is given by

$$p(z) = \frac{\hbar}{i}\left[\frac{1}{z - z_0} - 2\beta(z - z_0)\right] \tag{12}$$

This function has a first order pole at z_0. The vector map for $p(z)$ is shown in Figure 5(a). (Similar plots appear in *Visual Complex Analysis*:[39] see Figures 1 and 5 in Chapter 10.) Near the node, two diagonal separatrices confine the flow within four quadrants. Because of this confinement, momentum vectors cannot circulate completely around the node. Within each quadrant, the flow is directed inward near one separatrix and outward along the other. Quantum trajectories (which are tangent to the streamlines) are forced away from the nodal point by the quantum force, the gradient of the quantum potential. Using the first term in eq 12, the quantum force near the node is given by[40]

$$F_q(z) = -\frac{\hbar}{2mi}\frac{\partial^2 p}{\partial z^2} = \frac{\hbar^2}{m}\frac{1}{(z - z_0)^3} = \frac{\hbar^2}{m}\frac{1}{r^3}e^{-3i\theta} \tag{13}$$

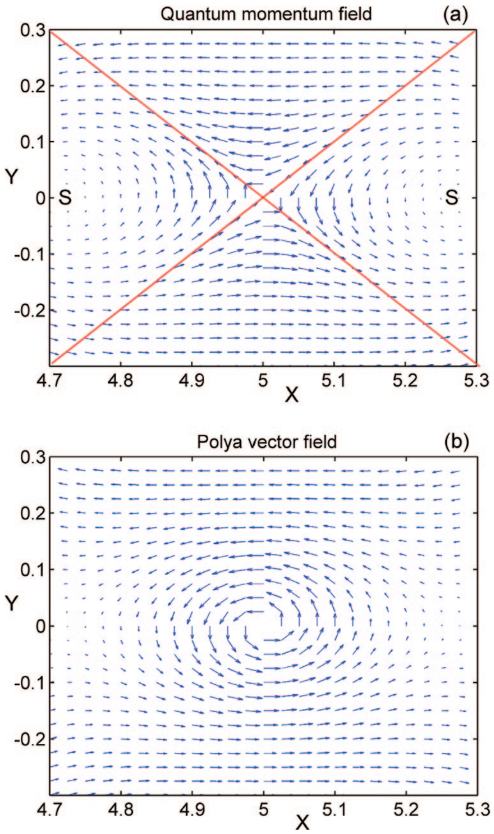


Figure 5. Quantum momentum field near a node in a model wave function is shown in part (a). This wave function has a node at the middle of the figure. In-flow and out-flow asymptotes (separatrices) are shown by the two (red) diagonal lines. Two stagnation points (S) are shown on the left and right edges of the figure. Part (b) shows the associated Pólya vector field. This field shows counterclockwise circular flow around the nodal point.

where $z = z_0 + r\exp(i\theta)$ was used in the last step. The vector map of this function (not shown here) clearly shows the r^{-3} repulsive character near the node (except along the asymptote at $\theta = \pi/4$.)

An alternative way of describing and visualizing flow in the complex plane will now be introduced. This function was utilized previously[40] in an analysis of quantum vorticity in the complex plane. We begin by considering the circulation integral: the line integral of the QMF around a simple closed curve in the complex plane. According to the Cauchy integral theorem,[38] this integral vanishes if $p(z)$ is analytic on and within the bounding contour. If both the line element and the QMF in the circulation integral are decomposed into real and imaginary parts, we obtain[38,39] (the momentum vector has the components $\vec{p}(z) = [p_r, p_i]$)

$$\begin{aligned}\oint p(z)dz &= \oint (p_r + ip_i)(dx + idy)\\ &= \oint (p_r dx - p_i dy) + i\oint (p_r dy + p_i dx)\\ &= \oint \vec{P}\cdot d\vec{t} + i\oint \vec{P}\cdot d\vec{n}\end{aligned} \tag{14}$$

In the last step, the Pólya vector field has been introduced;[38,39] this vector is defined by $\vec{p} = [p_r, -p_i]$ or as a complex function, $P(z,t) = p^*(z,t)$. In addition, the tangent along the integration contour and the outward normal vector along the

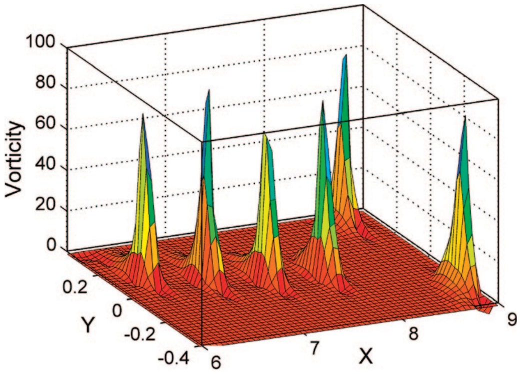


Figure 6. Vorticity field for time $t = 1000$. The four peaks for $x < 8$ form at the positions of the four peaks in *imag*($S(z,t)$) shown in Figure 7(a).

integration contour are $d\vec{t} = [dx,dy]$ and $d\vec{n} = [dy,-dx]$, respectively.

Although the Pólya field contains the same information as the momentum field, this information is displayed differently. To continue with the example mentioned previously, Figure 5(b) shows the Pólya field corresponding to the QMF plotted in part (a) of this figure. The vectors plotted in this figure clearly show counterclockwise circular flow around the nodal point. In this case, if the integration contour encloses the pole in $p(z)$, the imaginary term in the last step of eq 14 vanishes and the circulation integral is real-valued (and takes the value $2\pi\hbar$, corresponding to a quantized vortex in the 'ground state'). Returning to the reflected wave packet for the barrier scattering problem, Figure 4(b) shows the Pólya field corresponding to the QMF in part (a). The Póla field, as expected, shows counterclockwise circular flow around each of the three quasi-nodes. There are two additional features evident from this figure: (a) the Pólya vectors are always tangent to contours of the wave function amplitude (or the density); (b) the Pólya vectors display hyperbolic flow around stagnation points, such as the local maxima in the amplitude shown in this figure. The first property is generally valid for any wave function.[40] It has also been shown that Pólya vectors are parallel to contours of *imag*(S) and orthogonal to contours of *real*(S), where $S(z,t)$ is the complex action function.[40] The complex action function is considered further in the next section.

The local circulation of a vector field is frequently measured by the vorticity, the curl of the field. Accordingly, the vorticity of the Pólya field is given by

$$\vec{\Omega}(z,t) = \vec{\nabla} \times \vec{P}(z,t) = \frac{\partial P_i}{\partial x} - \frac{\partial P_r}{\partial y} \tag{15}$$

In regions where the field is analytic, the vorticity is zero, so that the magnitude of the vorticity vector may be viewed as a local measure of nonanalyticity. The vorticity is plotted in Figure 6 for $t = 1000$. The four sharp peaks for $x < 8$ form at the positions of the four quasi-nodes in the wave function (and at the four peaks in *imag*($S(z,t)$) shown in Figure 7(a)). These peaks gradually move to the left toward small values of x as the packet reflects from the barrier. There are also two peaks close to the barrier maximum, at $x_b = 9$, and these gradually decay as the wave packet bifurcates.

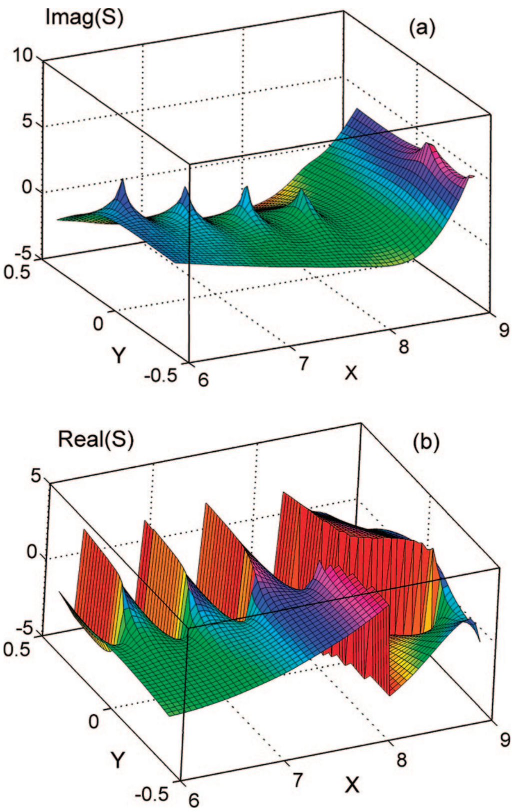


Figure 7. The imaginary (a) and real (b) parts of the complex quantum action function for the time $t = 1000$. In part (b), the principal value of the multivalued function is shown. These figures should be compared with the amplitude and QMF plotted in Figure 3(e) and the QMF and Pólya vector fields plotted in Figure 4.

7. The Complex Action

The complex action function, $S(z,t)$, not only determines the wave function through the exponential ansatz, $\psi(z,t) = \exp[iS(z,t)/\hbar]$, but also displays interesting features that will be illustrated in this section. If this function is decomposed into real and imaginary parts, $S = S_r + iS_i$, the wave function then becomes $\psi(z,t) = \exp[-S_i(z,t)/\hbar]\exp[iS_r(z,t)/\hbar]$. From this expression, it is evident that the imaginary part of the action determines the amplitude of the wave function. Large positive values of S_i imply small values for the amplitude. In addition, the real part of the action, the 'pure phase', also has interesting properties, which will now be considered.

For the barrier scattering example, Figure 7(a) shows the imaginary part of $S(z,t)$, while part (b) shows the real part of this function. In part (a), the four peaks in *imag*($S(z,t)$) between $x = 6$ and $x = 8$ are associated with the four quasi-nodes evident in Figures 3(e) and 4. Part (b) shows the principal zone of the phase, with values in the range $-\pi \le$ *real*($S(z,t)$) $\le \pi$. Starting near the position of each quasi-node and progressing in the positive y direction are sharp discontinuities of 2π in the phase. Discontinuities of this type are expected near nodes or quasi-nodes. For example, the model wave function $\phi(z) = z\exp(-\beta z^2 - iKz)$ has a QMF displaying hyperbolic flow near $z = 0$ and a phase discontinuity in *real*($S(z,t)$) along the positive y direction. For $\phi(z)$, the action function is given by $S(z) = -i\hbar[\ln(z) - \beta z^2 - iKz]$. It is important to note that the complex function $\ln(z)$

is multivalued: $\ln(z) = \ln|z| + iArg(z) + 2m\pi i$, where the principal zone of the argument is $\pi \leq Arg(z) \leq \pi$, and $m = 0, \pm1, \pm2,...$ As a result, the action function is also multivalued. When the principal sheet, corresponding to $m = 0$, and the sheets for $m = \pm1, \pm2,...$ are plotted, the Riemann surface for this function is obtained.

8. Summary

The time-dependent scattering of a wave packet from a Gaussian barrier was computationally investigated in the complex plane. Time evolution of the wave packet was followed by plotting the real part of the wave function and the quantum momentum field. In the reflected packet, an important role was played by ripples forming *above* the real axis. When these ripples pass near the real axis, 'interference fringes' are observed in the density. However, an observer confined to this axis would be ignorant of the significant vertical component of the flow (along the imaginary axis) near the local maxima and minima. In contrast, the component of the packet *below* the real axis makes a significant contribution to the transmitted packet. The quantum momentum vector maps show hyperbolic flow forming four 'petals' around transient quasi-nodes in the amplitude and counterclockwise circular flow around stagnation points, including local maxima in the amplitude. When the Pólya vector field associated with this momentum field was plotted, circular counterclockwise flow was obtained around the quasi-nodes. The vorticity, the curl of the Pólya field, is used to pinpoint regions of nonanalyticity in the QMF. These are also the positions where peaks occur in the imaginary part of the quantum action function.

Wave packet dynamics in the complex plane has been studied through an alternate approach: *numerical analytic continuation* from the wave function computed directly on the real axis.[45] Fortunately, within the viewing window, $\psi(z,t)$ turns out to be in excellent agreement with the results described in the current study.

Acknowledgment. This work was supported in part by the Robert Welch Foundation. We thank Julianne David and Chia-Chun Chou for helpful comments.

References

(1) Thaler, B. *Visual Quantum Mechanics*; Springer-Verlag: New York, 2000.

(2) Kemble, E. C. *Phys. Rev.* **1935**, *48*, 549.

(3) Berry, M. V.; Mount, K. E. *Rep. Prog. Phys.* **1972**, *35*, 315.

(4) Stine, J.; Marcus, R. A. *Chem. Phys. Lett.* **1972**, *15*, 536.

(5) Miller, W. H.; George, T. F. *J. Chem. Phys.* **1972**, *5668*, 56.

(6) George, T. F.; Miller, W. H. *J. Chem. Phys.* **1972**, *57*, 2458.

(7) Doll, J. D.; George, T. F.; Miller, W. H. *J. Chem. Phys.* **1973**, *58*, 1343.

(8) Miller, W. H. *Adv. Chem. Phys.* **1974**, *25*, 69. 1974.

(9) Huber, D.; Heller, E. J. *J. Chem. Phys.* **1987**, *87*, 5302.

(10) Heller, E. J. *J. Chem. Phys.* **1975**, *62*, 1544.

(11) Boiron, M.; Lombardi, M. *J. Chem. Phys.* **1998**, *108*, 3431.

(12) Tannor, D. J. *Introduction to Quantum Mechanics: A Time Dependent Perspective*; University Science Books: Sausalito, CA, 2007.

(13) John, M. V. *Found. Phys. Lett.* **2002**, *15*, 329.

(14) Yang, C. D. *Chaos, Solitons Fractals* **2006**, *30*, 342.

(15) Yang, C. D. *Chaos, Solitons Fractals* **2006**, *32*, 312.

(16) Yang, C. D. *Annal. Phys.* **2005**, *319*, 399.

(17) Yang, C. D. *Int. J. Quantum Chem.* **2006**, *106*, 1620.

(18) Yang, C. D. *Annal. Phys.* **2005**, *319*, 444.

(19) Yang, C. D. *Chaos, Solitons Fractals* **2006**, *30*, 41.

(20) Chou, C. C.; Wyatt, R. E. *Phys. Rev. A* **2006**, *76*, 012115.

(21) Chou, C. C.; Wyatt, R. E. *J. Chem. Phys.* **2008**, *129*, 124113.

(22) Sanz, A. S.; Miret-Artés, S. *Chem. Phys. Lett.* **2008**, *458*, 239.

(23) Lopreore, C. L.; Wyatt, R. E. *Phys. Rev. Lett.* **1999**, *82*, 5190.

(24) Wyatt, R. E. *Quantum Dynamics with Trajectories*; Springer: New York, 2005.

(25) Trahan, C. J.; Hughes, K.; Wyatt, R. E. *J. Chem. Phys.* **2003**, *118*, 9911.

(26) Trahan, C. J.; Wyatt, R. E. *J. Chem. Phys.* **2003**, *119*, 7017.

(27) Trahan, C. J.; Wyatt, R. E. *Chem. Phys. Lett.* **2004**, *385*, 280.

(28) Hughes, K. *J. Chem. Phys.* **2005**, *122*, 11.

(29) Bittner, E. R. *J. Chem. Phys.* **2004**, *119*, 1358.

(30) Trahan, C. J.; Wyatt, R. E.; Poirier, B. *J. Chem. Phys.* **2005**, *122*, 164104.

(31) Rowland, B. A.; Wyatt, R. E. *Chem. Phys. Lett.* **2006**, *426*, 209.

(32) Goldfarb, Y.; Degani, I.; Tannor, D. J. *J. Chem. Phys.* **2006**, *125*, 231103.

(33) Rowland, B. A.; Wyatt, R. E. *J. Phys. Chem. A* **2007**, *111*, 10234.

(34) Sanz, A. S.; Miret-Artés, S. *J. Chem. Phys.* **2007**, *127*, 197101.

(35) Goldfarb, Y.; Degani, I.; Tannor, D. J. *J. Chem. Phys.* **2007**, *127*, 197102.

(36) Goldfarb, Y.; Schiff, J.; Tannor, D. J. *J. Phys. Chem. A* **2007**, *111*, 10416.

(37) Goldfarb, Y.; Tannor, D. J. *J. Chem. Phys.* **2007**, *127*, 161101.

(38) Pólya, G.; Latta, G. *Complex Variables*; Wiley: New York, 1974.

(39) Needham, T. *Visual Complex Analysis*; Oxford University Press: Oxford, 1997.

(40) Chou, C. C.; Wyatt, R. E. To be published.

(41) Zhang, J. Z. H. *Theory and Application of Quantum Molecular Dynamics*; World Scientific: Singapore, 1999.

(42) Vibok, A.; Balint Kurti, G. G. *J. Phys. Chem.* **1992**, *96*, 8712.

(43) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes;* Cambridge University Press: New York, 1992.

(44) Verlet, L. *Phys. Rev.* **1967**, *159*, 98.

(45) Rowland, B. A.; Wyatt, R. E. To be published.

CT800248W

JCTC Journal of Chemical Theory and Computation

Computational Investigation of Wave Packet Scattering in the Complex Plane: Dynamics of Exact Quantum Trajectories

Robert E. Wyatt and Brad A. Rowland*

Department of Chemistry and Biochemistry, University of Texas, Austin, Texas 78712

Received June 26, 2008

Abstract: In two previous studies, the time-dependent scattering of a wave packet from a Gaussian barrier was investigated computationally in the complex *z*-plane. One of these involved the 'direct' propagation of the wave packet in the complex space, and the other used numerical analytic continuation techniques to generate the dynamics in the complex plane from the wave function computed on the real-axis. In the current study, the dynamics of exact quantum trajectories are analyzed for the same barrier scattering problem. Thousands of quantum trajectories were launched from positions near the center of the initial wave packet. These trajectories were computed by integrating equations-of-motion involving the quantum momentum function, which was obtained from the time-dependent wave function and its derivative. In order to analyze the dynamics, many trajectories were plotted on space-time diagrams. Particular emphasis was placed upon trajectories undergoing reflection in the barrier region. Some groups of strongly correlated trajectories form long-lived highly organized patterns, including helical wrappings around a series of stagnation filaments. These curves alternate with quasi-nodes where the amplitude of the wave function reaches low values. In addition, other trajectories for short times follow hyperbolic paths as they propagate near vorticity tubes surrounding these quasi-nodes.

1. Introduction

The current study is an extension of our previous investigations on wave packet scattering from a barrier potential in the complex plane.[1,2] This 'extended' 2D scattering problem was approached in two quite different ways: 'direct' propagation of a wave packet in the 2D complex *z*-space[1] and numerical continuation into the complex plane of time-dependent information computed on the real-axis.[2] The rich dynamics ensuing in the complex plane were analyzed by plotting the amplitude of the wave function and several other functions. In the introductory sections of these studies, we reviewed the use of complex-valued classical trajectories, analytical quantum trajectories computed for systems where the wave function is known, and for synthetic approximate quantum trajectories wherein information (namely, the quantum action function) is computed along each trajectory so that the wave function may be synthesized on-the-fly. Our previous papers should be consulted for references to these earlier studies.

The current study, dealing with 'exact' quantum trajectories for a nonstationary scattering problem, complements the detailed analysis of complex-valued quantum trajectories for one-dimensional *stationary state* scattering problems (the Eckart barrier, the soft potential step, and a downhill potential with a barrier).[3,4] The only previous studies of complex quantum trajectories for *nonstationary* problems have dealt with barrier-free systems for which the time-dependent wave function is known (or can be found) analytically; these include the free Gaussian wave packet[3] and the frontal collision of two Gaussian wave packets.[5]

In the current study, thousands of quantum trajectories were evolved in the complex space for the barrier scattering problem. In order to do this, the quantum momentum function along each trajectory was computed from the time-dependent wave function which was simultaneously computed on a large computational grid. Time histories of the

---

* Corresponding author e-mail: barowland@gmail.com.

10.1021/ct8002496 CCC: $40.75

Table 1. Categories of Quantum Trajectories

type	description
transmitted	transmit beyond barrier maximum ($x > x_b$)
reflected trajectories:	
upper	propagate toward upper part of barrier region
down-up	move down initially and deflect toward larger values of y without loops or spirals
loop-simple	move in toward barrier and deflect toward larger values of y without spirals
loop-down	move in toward barrier and deflect toward smaller values of y without spirals
loop-complex	move upward as propagation proceeds in toward barrier and reflect toward smaller values of y with some undulations
twist	move upward as propagation proceeds in toward barrier and then reflect toward smaller values of y with several loops or spirals
spiral	move upward as propagation proceeds in toward barrier and then reflect toward smaller values of y with complex spiraling

reflected trajectories were studied in detail, because many of them have much more complicated features than the transmitted trajectories. In addition, strong correlation between these trajectories leads to the formation of highly organized patterns. In space-time diagrams for the reflected wave packet, we will identify *vorticity tubes* (to be defined shortly). Groups of quantum trajectories coming from widely separated initial locations end up wrapping around stagnation curves which form in pairs adjacent to these tubes. Other quantum trajectories experience 'hyperbolic indentations' when they pass near the vorticity tubes. The helical coils of trajectories around the stagnation curves and the hyperbolic flow near vorticity tubes exemplify the most complicated types of trajectories that form during the reflection of the wave packet.

In an analysis of quantum vorticies that form around nodes in the complex plane for *stationary states*,[6] vector maps of the QMF were shown to exhibit hyperbolic flow close to these nodes. However, when the Pólya vector field,[7,8] defined as the conjugate of the QMF ($P(z) = p^*(z)$), is plotted, circular counterclockwise flow is obtained near the nodes. The vorticity, the curl of the Pólya field, $\vec{\Omega}(z) = \vec{\nabla} \times \vec{P}(z)$, provides a quantitative measure of the circulation of this field. For *nonstationary states*, as time proceeds, the nodes generally translate along curves, and this leads to time-dependence of the vorticity field. At a particular time, if a closed curve encloses a nodal region such that the magnitude of the vorticity is constant on this curve, then for later times a vorticity tube is formed by the isosurface surrounding this nodal curve. (As a historical note, Helmholtz introduced vorticity tubes[9] into hydrodynamics in 1858). For purposes of the current study, the vorticity tube will be defined as an

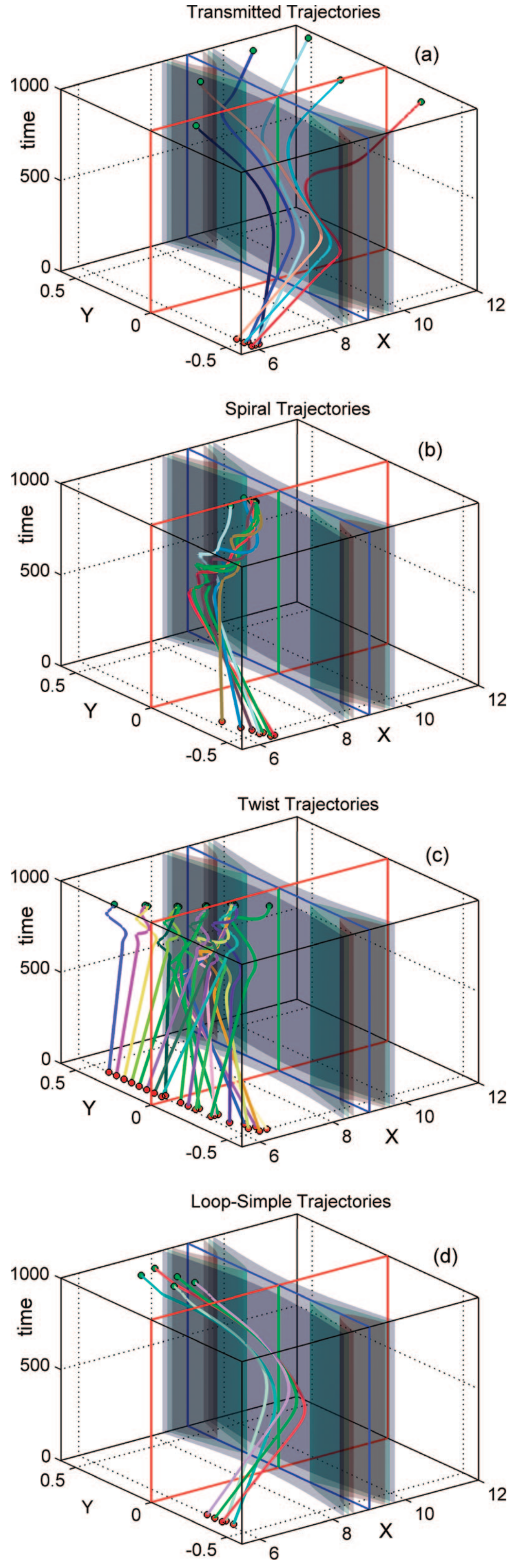


Figure 1. Time-dependence for four groups of quantum trajectories: (a) transmitted trajectories; (b) spiral trajectories; (c) twist trajectories; and (d) loop-simple trajectories. Propagation time is plotted on the vertical axis above the complex plane. In each part, the vertical slabs are isosurfaces of the absolute value of the potential energy.

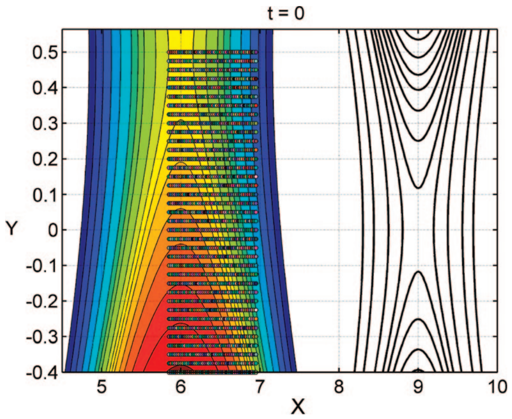


Figure 2. Contours (color) of log $_{10}|\psi|$ for the initial wave packet in the complex plane. Contours (black) of the magnitude of the potential, $|V(z)|$, are shown on the right side of the figure. In addition, a 45 × 43 grid of quantum trajectories at their initial locations are shown (colored dots).

isosurface corresponding to a constant (but small) value of $|\psi(z,t)|$ enclosing the path followed by a moving quasi-node. Within these tubes, $|\vec{\Omega}(z,t)|$ takes on large values, but the vorticity vanishes outside of these tubes (because $\psi(z,t)$ is an analytic function in these regions).

The remainder of this study is organized as follows. Computation of quantum trajectories is described in section 2, and eight types of reflected and transmitted trajectories are described in section 3. A series of time slices showing trajectory locations are presented in section 4, and vertical tubes and two types of reflected trajectories are described in more detail in section 5. Finally, a summary appears in section 6.

2. Propagation of Quantum Trajectories

Before presenting the method used to propagate quantum trajectories, the scattering problem will be briefly described.[1,2] The initial wave packet in z-space is given by

$$\psi(z) = (2\beta/\pi)^{1/4} e^{-\beta(z - z_0)^2} e^{ik_0(z-z_0)} \qquad (1)$$

where the translational energy is $E = \hbar^2 k_0^2/(2m)$. This function is centered at the position $z_0 = (x_0,0) = (6,0)$, the width parameter is $\beta = 6$, and the mass is $m = 2000$ (all parameters are given in atomic units). The complex-valued Gaussian potential is given by

$$V_G(z) = V_0 e^{-\gamma(z - z_b)^2} \qquad (2)$$

This barrier is centered at $z_b = (x_b,0) = (9,0)$, the width parameter is $\gamma = 4$, and the barrier height on the real-axis is $V_0 = 0.035$. The initial wave packet energy is $E = V_0/4$.

Quantum trajectories in the complex plane are obtained by integrating the equation of motion

$$\frac{dz(t)}{dt} = \frac{p(z,t)}{m} \qquad (3)$$

where the quantum momentum function is obtained from the wave function through the relation

$$p(z,t) = \frac{\hbar}{i}\frac{1}{\psi(z,t)}\frac{\partial \psi(z,t)}{\partial z} \qquad (4)$$

The real and imaginary components of the momentum vector, $\vec{p}(z,t) = [p_r(z,t),p_i(z,t)]$, are responsible for 'horizontal' and 'vertical' components of motion in the complex plane. This function possesses two special points of interest: poles occur at *nodes* where $\psi(z,t) = 0$, and $p(z,t)$ becomes zero at *stagnation points* where $d\psi(z,t)/dz = 0$ (and $\psi(z,t) \neq 0$). In addition, as described in detail in our earlier studies, the quantum momentum field exhibits *hyperbolic flow* around quasi-nodes in the density (these are local minima where the amplitude becomes small but does not reach zero). In order to analyze this flow, the Pólya vector field was introduced, and maps of this field show circular counterclockwise flow around quasi-nodes and hyperbolic flow around stagnation points.

The following procedure was used to propagate trajectories across the computational grid. At each time step, it is necessary to find the indices of the corner grid points for the cell in the computational grid in which the trajectory is located. These indices for the four grid points at the corners of this cell are denoted (j_0,k_0), (j_0+1,k_0), (j_0,k_0+1), and (j_0+1,k_0+1). Since the wave function is known at each of these points, $d\psi/dz$ may be computed at each corner point using wave function values known at neighboring grid points. Equation 4 was then used to compute the quantum momentum function at each corner grid point. Finally, linear interpolation was used to find the quantum momentum at the position of the trajectory within this cell. This procedure was then repeated for each trajectory at each time step.

3. Types of Quantum Trajectories

Starting from a Cartesian grid near the center of the initial wave packet, hundreds of quantum trajectories were launched toward the barrier region. Although it is trivial to categorize these trajectories as either 'transmitted' or 'reflected', it is more subjective to group the reflected trajectories into the seven categories listed in Table 1. It is useful to note that some of the reflected trajectories display simple behavior, while others display very complicated motions. Some propagate directly toward the barrier region, while others tend to move either up or down relative to the vertical (imaginary) axis. The complicated reflected trajectories arise because of quantum vortices that develop above the real-axis. Some of the quantum trajectories become trapped near transient stagnation points located between the vortices, while others execute complex loops or spirals as they escape from or avoid these regions.

Figure 1 shows plots for four of the eight trajectory categories listed in Table 1. The vertical coordinate is the propagation time, and these trajectories were propagated for 1000 atomic time units. In each figure, isosurfaces having constant values of the potential energy are shown. Of course, the potential is independent of time, so that each isosurface appears as a curved vertical slab. The real-axis, when displaced vertically in time, is outlined by the red box. Part (a) of this figure shows six *transmitted trajectories*, of which two barely make it to the transmitted side of the barrier. In part (b), seven *spiral trajectories* are shown. These trajectories initially move to larger values of y, but after about $t = 600$, they spiral back toward the real axis. In part (c), 22

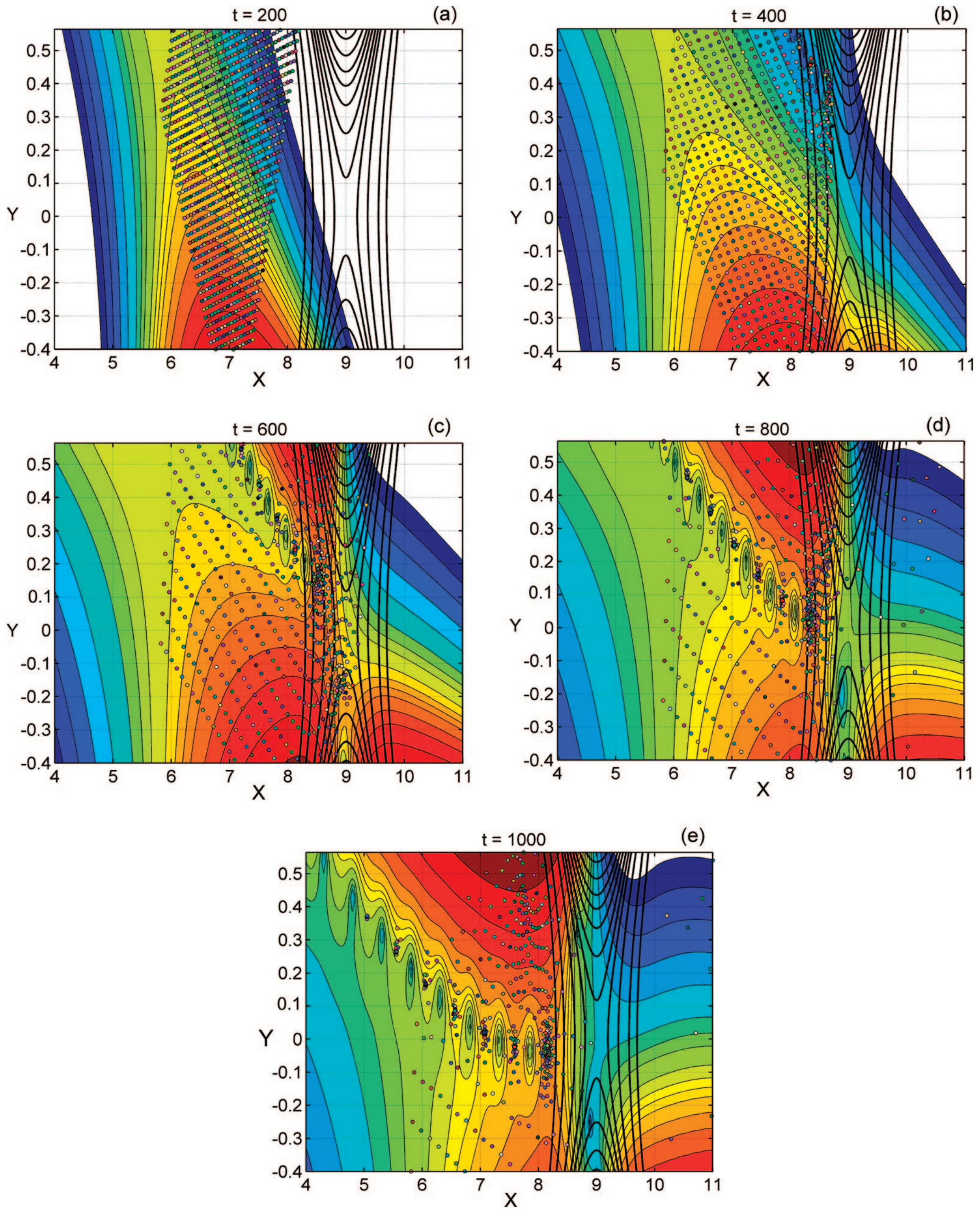


Figure 3. Contours (color) of log $_{10}|\psi|$ showing propagation of the wave packet in the complex plane. Contours (black) of the magnitude of the potential, $|V(z)|$, are shown on the right side of the figure. In addition, locations the quantum trajectories are shown (colored dots).

twist trajectories are plotted. After reflection from the barrier, these trajectories execute at least one twist as they move toward smaller values of y. Finally, in part (d), 5 *loop-simple trajectories* are shown. As these trajectories approach the barrier region, they move upward toward more positive values of y, and then reflect down toward smaller y values, without looping or spiraling. In the next section, we will observe that twist and spiral trajectories acquire their complicated motions by passing close to quasi-nodes, where they are strongly influenced by the repulsive quantum potential. Of the seven types of reflected trajectory listed in Table 1, only the *loop-simple* and *upper* trajectories are *not* strongly influenced by the quasi-nodes that form on the reflected side of the barrier. Trajectory dynamics near the vortical regions are described in more detail in section 5.

4. Quantum Trajectory Dynamics

Quantum trajectory evolution was studied by launching a set of trajectories from a rectangular grid located near the starting position of the initial wave packet. In Figure 2, the colored dots show the initial positions of 1935 trajectories superimposed upon a color-fill contour map of the amplitude of the initial wave function. Note that the amplitude has significantly higher values below the real-axis. The trajectories were propagated for the time $t = 1000$, and during this interval some trajectories transmitted to the right side

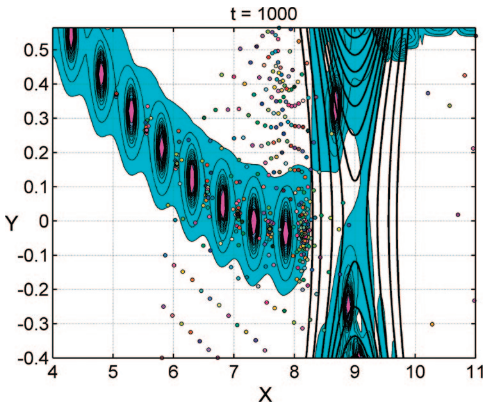


Figure 4. Absolute value of the quantum potential at time *t* = 1000. The pink regions between *x* = 4 and *x* = 8 locate eight quasi-nodes that gradually move toward the lower left. The quantum trajectories (colored dots) are repelled from regions of high quantum potential. Clusters of trajectories are noted around transient stagnation points near the quasi-nodes.

of the barrier, but most moved back to the left during the formation of the reflected wave packet.

Snapshots of the quantum trajectories and the amplitude of the wave function are shown at five time steps in Figure 3. Each trajectory always carries the same color, so it is possible to follow some of the trajectories as they interact with the potential. In part (a), the trajectory grid undergoes a shearing deformation when trajectories on the front edge of the wave packet are thrust upward with high speeds, while those on the upper left move down toward the real-axis with low speeds. In part (b), trajectories on the upper right are starting to form clusters when the wave packet begins to reflect from the upper part of the barrier potential. In part (c), a series of five quasi-nodes is forming above the real-axis from about $x = 6.5$ to $x = 8.5$. Some of the quantum trajectories are clustering near stagnation points located to the lower-right or upper-left of each quasi-node. Also, a few trajectories have crossed over to the transmitted side of the barrier. In parts (d) and (e), these quasi-nodes move further toward the lower-left of the figure as the reflected wave packet recedes from the barrier region. Some of the quasi-nodes in part (e) are moving across the real-axis, where an observer would detect 'interference oscillations' in the amplitude and density.

We will see in section 5 that trajectories trapped near stagnation points (which themselves are time-dependent) move away from the barrier region in correlated groups. In addition, trajectories become 'indented' when they temporarily follow hyperbolic paths near quasi-nodes. The quasi-nodes modify the local trajectory dynamics through the intense quantum potential and the resulting strong quantum forces in there regions. The absolute value of the quantum potential at time $t = 1000$ is shown in Figure 4. The pink regions of high quantum potential between $x = 4$ and $x = 8$ locate eight quasi-nodes that continue to move toward the lower left. The quantum trajectories (colored dots) are repelled from regions of high quantum potential. Clusters of trajectories are noted around pairs of transient stagnation points which form close to each quasi-node (especially near $x = 8$).

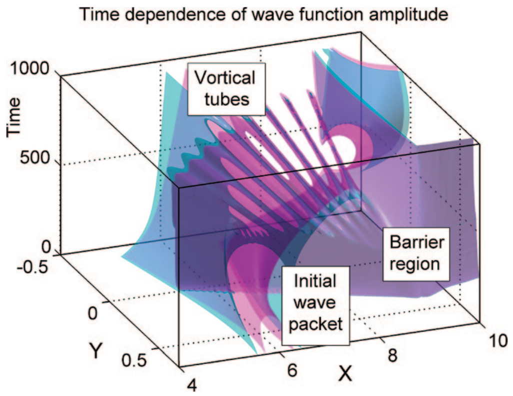


Figure 5. Two isosurfaces of log $_{10}|\psi(z,t)|$ are drawn for the values °0.6 (inner pink surface) and °0.9 (outer cyan surface). The initial wave packet is centered near $z_0 = (6,0)$, and the barrier is centered near $z_b = (9,0)$. Starting around $t = 350$, several vorticity tubes develop in the region above the real-axis. As time proceeds, these tubes bend back toward the real axis.

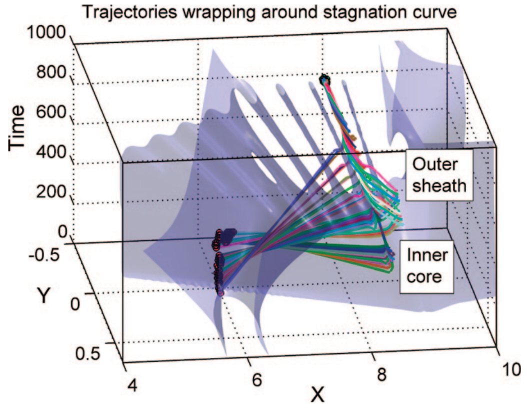


Figure 6. Quantum trajectories (colored lines) wrapping around the stagnation curve between vorticity tubes 1 and 2. One isosurface of log $_{10}|\psi(z,t)|$ is drawn for the value °0.7. Trajectories launched from positions relatively far from the real-axis form the tight inner core of the bundle, while trajectories launched closer to the real-axis form the loosely bound outer sheath.

5. Trajectory Dynamics near Vorticity Tubes

Trajectory dynamics within the reflected wave packet is strongly influenced by multiple quasi-nodes that form above the real-axis and which subsequently move toward the region below this axis as the reflected packet recedes from the barrier region. In order to explore features of quantum trajectories near these quasi-nodes, the trajectories will be plotted on the type of three-dimensional diagram shown in Figure 5, where propagation time appears on the vertical axis above the complex plane. (In Figures 5−8, note that the positive *y*-axis is pointing *toward* the viewer.) In this figure, isosurfaces have been plotted for two constant values of the wave function amplitude. As the initial wave packet scatters

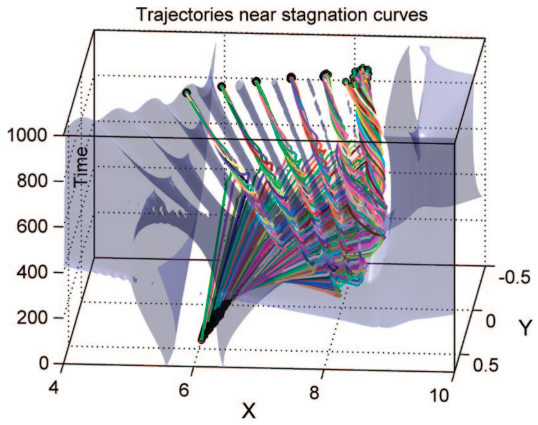


Figure 7. Quantum trajectories wrapping counterclockwise abound six stagnation curves which form near the vorticity tubes (487 trajectories are plotted). As we progress from stagnation curve 1 on the right to curve 6 on the left, the trajectories become more tightly wrapped. One isosurface of log $_{10}|\psi(z,t)|$ is drawn for the value °0.7.

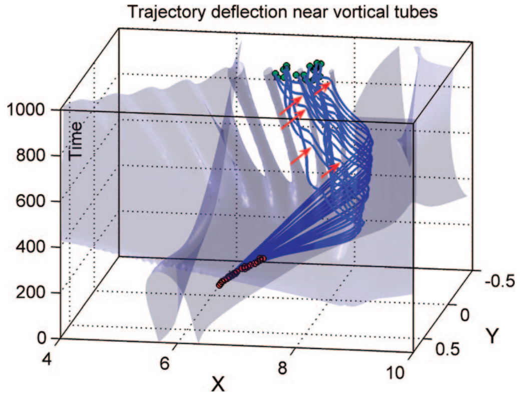


Figure 8. Trajectories deflecting around two vorticity tubes (22 trajectories were plotted). When each trajectory passes close to a tube number 1 or number 2, it follows a hyperbolic path. The five red arrows show hyperbolic indentations as the trajectory passes near a tube.

from the barrier region on the left side of the figure, a vorticity tube surrounds each of the quasi-nodes that forms above the real-axis. Each tube encloses a region of low amplitude, one of the quasi-nodes. As time proceeds, the reflected wave packet moves away from the barrier region, which causes these tubes to bend toward (and extend below) the real-axis.

In order to elucidate the dynamics, a large number of quantum trajectories (1521) were launched from a 39 × 39 grid located near the center of the initial wave packet. This grid covered the range [0.4,−0.55] along the y-axis and [6,6.95] along the x-axis. (This grid is slightly different from the one shown earlier in Figure 5.) All of these trajectories were plotted, but only those that eventually propagated near the vorticity tubes will be described in this section. In Figures 5− 8, sets of quantum trajectories are plotted, along with one or more isosurfaces of the wave function amplitude.

In Figure 6, 57 quantum trajectories (colored lines) are shown wrapping counterclockwise (when viewed from the $-y$ direction) around the stagnation curve that forms between vortical tubes 1 and 2. (These tubes are numbered consecutively from the barrier region out toward the location of the initial wave packet.) There are three types of trajectories that eventually wrap around this stagnation curve. Trajectories launched from positions relatively far from the real-axis (between $y = -0.45$ and -0.55) gather in the time interval $t = 300-350$ to form the tight inner core of the bundle. After that, trajectories launched slightly closer to the real axis (between $y = -0.40$ and 0.48) gather in the time interval $t = 450-750$ to form the middle segment of the bundle. Finally, trajectories launched closer to the real-axis (between $y = 0.05$ and 0.35) gather in the time interval $t = 550-750$ to form the more loosely bound outer sheath. Even though trajectories from these three sets travel different distances with different velocities, they eventually become correlated and twist in unison around the stagnation curve.

In Figure 7, 487 quantum trajectories are shown wrapping counterclockwise abound six stagnation curves which form near vorticity tubes 1−7. As we progress from right to left, the trajectories become more tightly wrapped around the stagnation curves. The trajectories that wrap around a particular stagnation curve are launched from a specific curve in the $x-y$ plane. Finally, in Figure 8, 22 trajectories are shown deflecting around vorticity tubes 1 and 2. When each trajectory passes near one of these tubes, it is first drawn in toward the tube before repelling away (along a hyperbolic path). The red arrows show indentations in five trajectories as they pass near these tubes.

6. Summary

In this study, exact quantum trajectories were analyzed for the barrier scattering problem that was computationally investigated in our previous papers.[1,2] These exact quantum trajectories were computed by integrating equations-of-motion involving $p(z,t)$, the quantum momentum function, which in turn was obtained from the time-dependent wave function. Thousands of quantum trajectories were launched from positions near the center of the initial wave packet, and trajectory evolution was illustrated in space-time diagrams. These trajectories were divided into several categories, depending upon whether they were transmitted or reflected during the scattering process and upon the complexity of the dynamics in the reflected region. Special emphasis was placed upon selected sets of reflected trajectories. Among the most interesting are those that follow counterclockwise helical paths as they wind, like cooked spaghetti, around stagnation curves that form adjacent to quasi-nodes (where the amplitude reaches a local minimum) in the reflected wave packet. The trajectories that wrap around the stagnation curves originate from different regions near the center of the initial wave packet. In addition, other reflected trajectories follow hyperbolic paths when propagating near vorticity tubes which surround these quasi-nodes. As the reflected wave packet recedes from the barrier region, these tubes move away from the barrier region and eventually cross the real-axis, where an observer would detect ‘interference oscillations’ in the density. However, this observer would be clueless regarding the rich dynamics ensuing in the complex space that preceded these observations.

Acknowledgment. This research was supported in part by funding from the Robert Welch Foundation. We thank Julianne David and Chia-Chun Chou for helpful comments.

References

(1) Wyatt, R. E.; Rowland, B. A. To be published.

(2) Rowland, B. A.; Wyatt, R. E. To be published.

(3) Chou, C. C.; Wyatt, R. E. *Phys. Rev. A* **2006**, *76*, 012115.

(4) Chou, R. E.; Wyatt, R. E. *J. Chem. Phys.* **2008**, *129*, 124113.

(5) Sanz, A. S.; Miret-Artés, S. *Chem. Phys. Lett.* **2008**, *458*, 239.

(6) Chou, C. C.; Wyatt, R. E. *J. Chem. Phys.* **2008**, *128*, 234106.

(7) Pólya, G.; Latta, G. *Complex Variables*; Wiley: New York, 1974.

(8) Needham, T. *Visual Complex Analysis*; Oxford University Press: Oxford, 1997.

(9) von Helmholtz, H. J. *Für Reine Angew. Mathematik* **1858**, *55*, 25.

CT8002496

J. Chem. Theory Comput. **2009,** *5,* 459–467 **459**

Journal of Chemical Theory and Computation

X-Pol Potential: An Electronic Structure-Based Force Field for Molecular Dynamics Simulation of a Solvated Protein in Water

Wangshen Xie,[†] Modesto Orozco,[*,‡] Donald G. Truhlar,[*,†] and Jiali Gao[*,†,‡]

Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis Minnesota 55455-0431, and Barcelona SuperComputer Center—Institute for Research in Biomedicina Joint Program on Computational Biology, C/ Jordi Girona 29, 08034 Barcelona, Spain and Josep Samitier 1-5, 08028 Barcelona, Spain

Received June 21, 2008

Abstract: A recently proposed electronic structure-based force field called the explicit polarization (X-Pol) potential is used to study many-body electronic polarization effects in a protein, in particular by carrying out a molecular dynamics (MD) simulation of bovine pancreatic trypsin inhibitor (BPTI) in water with periodic boundary conditions. The primary unit cell is cubic with dimensions $\sim 54 \times 54 \times 54$ Å^3, and the total number of atoms in this cell is 14281. An approximate electronic wave function, consisting of 29026 basis functions for the entire system, is variationally optimized to give the minimum Born−Oppenheimer energy at every MD step; this allows the efficient evaluation of the required analytic forces for the dynamics. Intramolecular and intermolecular polarization and intramolecular charge transfer effects are examined and are found to be significant; for example, 17 out of 58 backbone carbonyls differ from neutrality on average by more than 0.1 electron, and the average charge on the six alanines varies from −0.05 to +0.09. The instantaneous excess charges vary even more widely; the backbone carbonyls have standard deviations in their fluctuating net charges from 0.03 to 0.05, and more than half of the residues have excess charges whose standard deviation exceeds 0.05. We conclude that the new-generation X-Pol force field permits the inclusion of time-dependent quantum mechanical polarization and charge transfer effects in much larger systems than was previously possible.

1. Introduction

Molecular dynamics simulation has become a powerful tool for studying biochemical properties ranging from protein and nucleic acid dynamics and structural prediction to chemical reactions in enzymes.[1] At the heart of these calculations is the potential energy function that describes intermolecular interactions in the system,[2−8] and often it is the accuracy of the potential energy surface (or its gradient field, called the force filed) that determines the reliability of simulation results. Because of the size of condensed-phase systems and the complexity of biomacromolecules, one typically uses molecular mechanics force fields, in which the potential energy surface for a macromolecular system is approximated by analytical functions describing bond stretches, bond angle bends, torsions, and nonbonded van der Waals and Coulomb interactions.[2,3] The computational efficiency of analytic molecular mechanics force fields allows molecular dynamics simulations of biopolymer systems to be carried out with the extensive sampling required for rare event simulation, and classical simulations may be extended for long time scales and to large molecular systems. There is therefore considerable effort being expended to improve the physical representation and accuracy of such force fields, with special emphasis on including polarization effects to better represent electrostatic forces.[3,4]

The development of molecular mechanics dates back to early studies of steric effects of organic compounds,[5−7] and

---

[*] Corresponding author e-mail: modesto@mmb.pcb.ub.es (M.O.), truhlar@umn.edu (D.G.T.), gao@jialigao.org (J.G.).

[†] University of Minnesota.

[‡] Barcelona SuperComputer Center.

10.1021/ct800239q CCC: $40.75

the basis of the current generation of force fields for biomolecular systems was established in the 1960s.[2,8] The first molecular dynamics simulation of a protein was reported in 1977 by McCammon, Gelin, and Karplus;[9] it lasted 8.8 ps for a small protein, in particular, bovine pancreatic trypsin inhibitor (BPTI) in the gas phase. The authors noted two limitations in that study. The first "is the approximate nature of the potential energy function," which was essentially of the same type as we are using today, and the second "is the neglect of solvent". Although a relatively short simulation was performed on a dry protein, the 1977 article[9] is one of the classic studies in our field because of its vision that paved the way for molecular simulation and modeling as we know it today. Of course, solvation effects are now recognized as an unavoidable essential element in dynamics simulations because they play an inescapeable role in biomolecular function, and tremendous progress has been made in the accuracy of conventional force fields for modeling biopolymers.[3,10−15] Yet, it is sobering to notice that the physical representation and functional forms used in molecular mechanics have hardly changed.

Despite the success of molecular mechanics in biomacromolecular modeling, there are also shortcomings, such as inapplicability to chemical reactions and lack of polarization. Recognition of the latter has motivated many efforts[3,4,16−33] to parametrize nonadditive polarization effects in the force fields. An alternative is to make a fundamental paradigm change in the functional form of the force fields and the representation of biomolecular systems,[31−33] moving beyond the present classical development. With this motivation, we have introduced an explicit polarization (X-Pol) model based on quantum mechanics as a framework for a next-generation force field.[34−36] In what follows, we report a molecular dynamics simulation of a fully solvated BPTI protein that, although short, employs an explicit quantal force field. This force field is currently represented by an available semiempirical method, namely the AM1 approximate molecular orbital theory. Nevertheless, the present study demonstrates the feasibility of an entirely new concept in force field development for large-scale simulation.

Polarization and charge transfer are intrinsic properties of the electronic structure of a molecular system, resulting in polar bonds, nonunit charge on functional groups, and electric response of electron density to an external field. Although molecular polarization is a well-defined property, its incorporation into an MM force field is not unique.[3,4,16−30] Consequently, numerous models have been proposed for the classical treatment of polarization effects, and their validity is a subject of ongoing validation. In the X-Pol potential,[31−36] the internal energy terms and electrostatic potentials used in the force field are described explicitly by a quantum chemical wave function. Since molecular polarization and charge transfer are represented naturally by electronic structure theory, no polarization terms need to be added, and hence there is no ambiguity in the choice of functional form for polarization terms or in the selection of internal degrees of freedom to define these terms. Furthermore, such a method can be used to model chemical reactions.

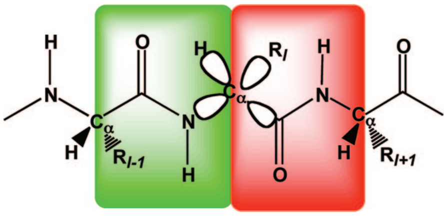


Figure 1. Definition of peptide units and the division of a protein into fragments at the C_α boundary atom. Two quantum mechanical fragments are highlighted in green and red, respectively, corresponding to residues $I - 1$ and I.

In the X-Pol potential, a molecular system is partitioned into fragments, such as an individual solvent molecule or a peptide unit or a group of such entities. The electronic interaction within each fragment is treated using electronic structure theory, while the interfragment electrostatic interactions are treated[31,34−36] using a quantal analog of the combined quantum mechanical and molecular mechanical (QM/MM) approach (hence these interaction terms are sometimes called electrostatic QM/MM terms). Because the wave function of the entire system (which is assumed to be a Hartree product of antisymmetrized fragment wave functions) is variationally optimized,[35] we can take advantage of analytic gradient theory[37] to develop efficient methods for evaluating the contributions of internal energies and electrostatic interactions to forces.[35] Exchange repulsion and dispersion-like attraction between fragments are added by pairwise additive functions.

The X-Pol potential has been tested and applied to the simulation of liquid water[32] and liquid hydrogen fluoride,[33] and it has been recently extended to treat fragments that are covalently bonded to one another.[34−36] Here we employ analytic gradients to carry out the dynamical simulation of a small protein, in particular BPTI, in a water box with a size of about 54 × 54 × 54 Å^3; the water box contains 4461 water molecules and one copy of the protein. We will analyze the polarization of the charge distribution of the protein and its significance for the description of protein−solvent interactions. The stability of MD simulations using the X-Pol potential for macromolecular simulations in water will also be demonstrated. The calculations presented here are designed to set the stage for systematically parametrizing the X-Pol potential to achieve the goal of chemical accuracy in such simulations.

In Section 2 we briefly review the theory of the X-Pol potential, and section 3 gives computational details. Sections 4−6 present results, timings, discussion, and comments on future prospects.

2. Theoretical Background

The design of the X-Pol potential has been described in a series of publications,[31−36] and here we only present the necessary background for the simulation of a solvated protein. We adopt the peptide unit convention as defined by IUPAC,[38] although we note that the residue convention is typically used in other force fields.[3,10−14] As endorsed in the IUPAC rules,[38] we will refer to a peptide unit by its residue name. Figure 1 shows the division of a peptide chain into peptide units at a C_α carbon; each peptide unit is defined as a quantum mechanical (QM) fragment in the present calculation, and the C_α atom is called a boundary atom. Only the valence electrons of the boundary atom are treated explicitly, so the effective nuclear charge is four and the associated number of electrons is also four. Both the nuclear charge and electrons are divided equally into the two neighboring fragments. The boundary atom has four hybrid bonding orbitals,[34,39] such that each fragment has two of them as active orbitals and the other two as auxiliary orbitals. This partition of a polypeptide results in two "pseudo atoms", which have identical coordinates, and each of which is half of a boundary C_α atom.[34−36]

The wave function of the entire system is written as a Hartree product of Slater determinant wave functions of individual fragments.[31,34,35] In the X-Pol potential, the internal energies of the fragments are treated with electronic structure theory, and the interactions between fragments are described by a combined quantum mechanical and molecular mechanical (QM/MM[40−42]) approach. No bond stretching, bending, or torsion terms appear because such interactions are represented by quantum mechanics, and no harmonic assumptions or analytic anharmonicity terms appear. It should be emphasized that at the top of the hierarchy of approximations in the X-Pol model, there is no restriction on the level of molecular orbital theory or density functional theory used to represent each individual fragment. In principle, it is possible to use a high-level quantum model to treat the region of interest and to use a lower level of theory to represent the rest of the system. In the present version of the method, the electronic structure calculations are carried out by valence-only semiempirical molecular orbital theory with the neglect of diatomic differential overlap (NDDO[43]). Thus the electronic wave function includes only valence electrons, and core electrons are combined with nuclei and treated as frozen atomic cores.

The total energy of the system includes the electronic energies of the fragments (each including half of the electronic Coulombic interactions and half of the core interactions to avoid double counting) plus an empirical van der Waals term. Thus,

$$E_{tot} = E_{elec} + E_{vdw} \tag{1}$$

where the van der Waals energy term is required because the electronic structure calculation omits electron correlation and exchange repulsion between electrons in different fragments. The van der Waals term is a sum[32,35] of Lennard-Jones potentials, including both repulsion due to exchange and dispersion-like attraction due to medium-range correlation energy. Note that Lennard-Jones interactions are omitted for atom pairs within the same fragment and for those separated by less than 3 bonds (*i.e.*, 1−2 and 1−3 van der Waals interactions are excluded) as in most of the conventional force fields. In the current study, the atomic Lennard-Jones parameters are taken directly from the CHARMM[11] protein force field without modification, and pair parameters are obtained by the usual combining rules. Furthermore the NDDO parameters for nonboundary atoms are taken from Austin Model 1 (AM1[44]) without modification. The semiempirical parameters for the carbon boundary atom are the same as in AM1 except that the values of U_{ss} and U_{pp} are scaled by 0.99 as in previous studies.[34,36]

In calculating E_{elec}, the electric potential due to fragments sharing a boundary atom with the QM fragment under current consideration is calculated by explicit Coulomb integrals;[36] the electric potential from non-neighboring fragments are approximated by one-electron integrals with partial atomic charges. In the present calculations these charges are obtained by the Mulliken approximation[45] applied to wave functions of each of the other fragments.

The total electronic energy of the system is determined by a double self-consistent-field (DSCF) procedure.[32,34−36] Starting with an initial guess of the one-electron density matrix for each fragment, one cycles over all fragments in the system and performs electronic structure calculations for each fragment (peptide unit or water molecule) in the presence of Mulliken charges of the other fragments until the change in total electronic energy or density matrix satisfies a predefined tolerance.[32,34,36] To facilitate the DSCF convergence, we introduce a quantum mechanical buffer zone for the peptide unit (m) currently being treated quantum mechanically in the inner SCF iteration.[35,36] Thus, in addition to this fragment m, we also include the peptide units prior to and after fragment m in each explicit QM treatment. In turn, fragment m becomes a buffer fragment for peptide units $m - 1$ and $m + 1$, respectively. Note that during the SCF optimization of the wave function for fragment m, the electron densities of the buffer peptide units are kept frozen[36] at values derived from a previous outer SCF iteration. Although it increases the number of two-electron integrals, the use of a buffer zone[36] reduces the time spent on matrix transformations (as compared to the earlier formulations) because no atom needs special treatment to avoid double counting or unphysical interactions with virtual orbitals. Once the wave function is converged, the forces are calculated analytically.[35] When the DSCF process has converged, the chemical potentials of all fragments will have been equalized. This allows mutual polarization of all fragments subject to the constraint that there is no charge transfer between fragments.

A key methodological issue is that the Fock matrix is expressed in a mixed basis consisting of atomic orbitals for nonboundary atoms and hybrid orbitals[34,39] for boundary atoms. The present usage of hybrid orbitals is expected to be more accurate than the original use[39] because the charges in the hybrid orbitals are all determined self-consistently rather than determining some of them from MM parameters.

References 32 and 34−36 contain further details of the method.

3. Computational Details

The initial structure of a BPTI protein molecule solvated in a cubic box of water molecules is constructed using a developmental version of CHARMM (version c34a1),[46] in which the present X-Pol potential has been implemented. The non-hydrogen atomic coordinates are taken from the structure 6PTI in the protein databank (PDB), and all hydrogen atoms are built using the HBUILD function in

CHARMM based on standard equilibrium geometrical parameters.[11] There are two disulfide bonds in BPTI, which are terminated by hydrogen atoms in the X-Pol treatment to allow for convenient partition of the protein into peptide units. We note that a simple extension of the procedure already implemented into CHARMM can be made to handle disulfide bond connecting two amino acids. For the present study, our current treatment is sufficient. We have used a neutral side chain for each histidine residue, while all other titratable residues are assigned a protonation state corresponding to a pH of 7. The BPTI protein is then solvated by a previously equilibrated water box of about 54 × 54 × 54 Å^3, deleting all water molecules within 2.7 Å of any protein atoms, resulting in a total of 4461 water molecules, giving rise to a total of 14281 atoms, including protein, solvent, and counterions. There are 4519 fragments (58 amino acid residues and 4461 waters).

An MD simulation in the *NPT* ensemble at 300 K and 1 atm is carried out for 100 ps using the CHARMM22 force field[11] for protein and the three-point-charge TIP3P model[47] for water to equilibrate the system. The resulting coordinates are used as the initial structure for a 50 ps *NVT* MD simulation at 300 K using the X-Pol potential.[34−36] The box length is set to 53.65 Å in the X-Pol calculations, which is the average value from the MD simulation using the CHARMM22 force field. A Nosé-Hoover thermostat[48] is used for temperature control.

All simulations utilized an integration time-step of 1 fs, and the SHAKE algorithm[49] is used to constrain bond distances involving hydrogen atoms at the equilibrium values defined in the CHARMM22 force field.[11] Electrostatic interactions between fragment pairs whose centers of mass are separated beyond 11 Å are truncated (shifting or switching can be introduced as a refinement in later work). The convergence criterion for average diagonal elements of the density matrix is set to 10^{-6}.

All calculations were carried out using a locally modified version of the CHARMM[46] software package. The X-Pol potential was initially developed based on a new semiempirical code written by Walker et al.,[50] and the current X-Pol software is essentially entirely rewritten.

4. Results and Discussion

A snapshot of the BPTI structure at the end of the 50 ps MD simulation using the X-POL potential is displayed in Figure 2 along with the structure at the end of a 50 ps MD simulation using the CHARMM22 force field and the crystal structure 6PTI from the protein data bank (PDB). These figures show that the secondary structures of BPTI retained their configurations in 50 ps molecular dynamics simulations employing the X-Pol potential (Figure 2a) in comparison with the crystal structure (Figure 2c). It appears that the two β-strand configurations are somewhat weakened from simulations using the classical force field. Compared with the crystal structure, the side chains show significant conformational change from both classical and X-Pol molecular dynamics simulations; charged residues are more exposed to the solvent on the protein surface. One realizes that the semiempirical AM1 model was not developed to treat intermolecular interactions accurately. Thus, we note that the present X-Pol potential with the original AM1 quantum mechanical method is not a quantitatively accurate force field for performing quantitative studies of the dynamics of solvated proteins. To achieve this goal, a reparameterized and well-tested QM model is needed, and the development of such a model is left for future research.

The fluctuation of the total potential energy is displayed in Figure 3 for the entire 50 ps (50000 steps) of simulation, showing that the energy exhibited an initial rise in the first 15 ps and then quickly settled to a stable average throughout the rest of the simulation. At each MD step, about 7 iterations were sufficient to achieve SCF convergence to an accuracy of 10^{-6} in the electronic one-electron density matrix.

The main physical result of the present study is the extent of electronic polarization and intramolecular charge transfer in the solvated protein. The net charge from Mulliken population analysis of the wave function for each carbonyl group (C=O) in the protein backbone is calculated and averaged for the last 30000 MD steps. The average net charge on the backbone carbonyl (C=O) group of each residue along the peptide chain is shown in Figure 4. We found that all carbonyl groups bear a negative net charge, which is reasonable since C=O is a strong electron withdrawing group. The average net charges on the carbonyl groups range from −0.05 to −0.16 (all partial charges are in units of a proton charge), with 17 of them more negative than −0.10. In comparison, the CHARMM22 force field[11] employs fixed partial atomic charges with the convention that the net group charge for each carbonyl unit (C=O) is zero in the protein backbone. Since it is computationally efficient for each group charge to be zero (forcing groups of 2−10 atoms to be perfectly neutral allows not only for more easily transferable charge parameters but also for more efficient truncation of long-range electrostatics[3a]), this can only be remedied in conventional molecular mechanics calculations by using larger units as groups and by parametrizing the groups to allow different charges on carbonyl groups in different environments. However, even if that is done, the charge on each carbonyl group would be independent of time and environment, neither of which is found to be the case in the X-Pol calculations.

To analyze charge transfer effects between different residues, we calculated the net charge of each residue by Mulliken population analysis. Here, we note that formally there is no charge transfer between fragments treated in the X-Pol potential. However, effective charge transfer can be observed through the boundary atom due to intrafragment electronic polarization. Thus, it is possible that the electron density of the two active orbitals in the *(I-1)th* residue (green fragment in Figure 1) is depleted into the rest of the fragment, whereas the two active orbitals in the *Ith* residue (red in Figure 1) attract greater charge density in that fragment. Thus, the net partial atomic charge on the boundary atom has contributions from the charge densities of both neighboring residues. In the following discussion, the term “charge transfer” or excess charge is used to describe the difference of the total Mulliken population charge of a formal residue (not the peptide unit used in the definition of the QM

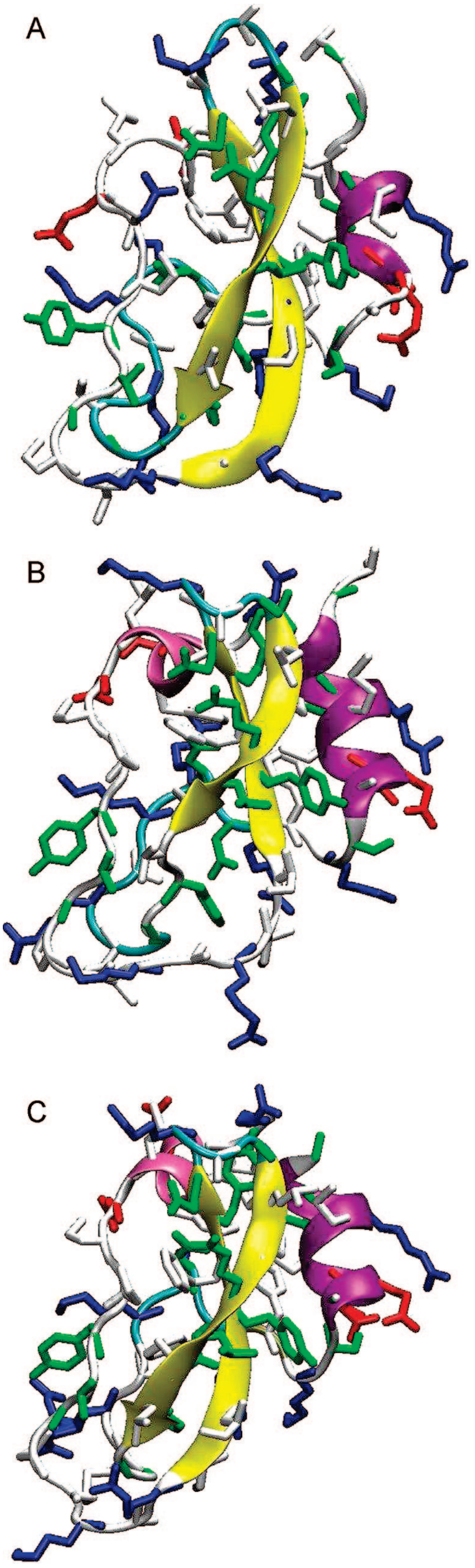


Figure 2. (a) Snapshot of the BPTI structure from MD simulation with the X-Pol potential, (b) snapshot of the BPTI structure from MD simulation with the CHARMM22 force field, and (c) X-ray crystal structure. Secondary structures are shown in yellow for β-strands, in purple for α-helices, and in gray and cyan for loops. Side chains are depicted in gray for hydrophobic residues, in green for polar residues, in blue for cationic reisdues, and in red for anionic residues.

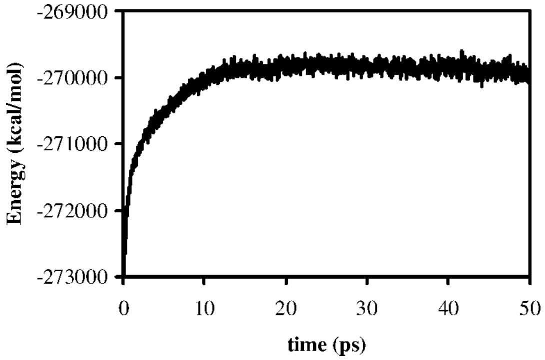


Figure 3. Histogram of the potential energy (kcal/mol) during the 50 ps molecular dynamics simulation of BPTI in water using the X-Pol potential.

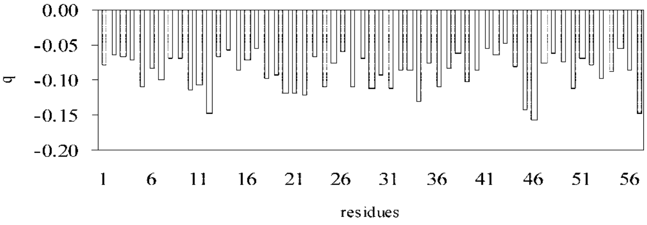


Figure 4. Average net partial charges (in atomic units) on the backbone carbonyl (C=O) group of each amino acid residue in BPTI. The carbonyls are arranged in order of sequence number.

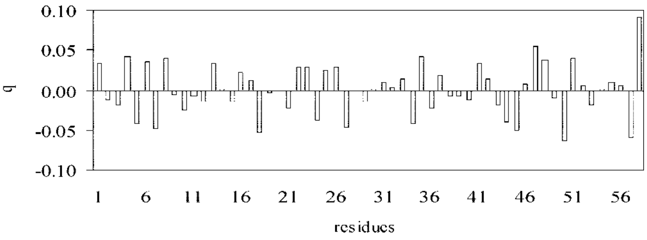


Figure 5. Average excess charge (in atomic units) for each residue of BPTI in water. The residues are arranged in order of sequence number. Note that Figures 5, 6, and 7(b) refer to conventional residues, not to peptide units.

fragment) in the protein and that of an isolated residue. Thus, the dominant contributor to charge transfer in this model is intrafragment polarization in the X-Pol representation. If one is interested in charge transfer between two residues, for example, in an ion pair, salt bridge, or hydrogen bond to a charged residue, the two moieties between which charge transfer is to be allowed should be treated as a single fragment.

Figure 5 shows the excess charge (calculated by subtracting the formal charge associated with the protonation state from the Mulliken charge) for all residues averaged over the last 30000 steps of MD simulations. The average excess charge ranges from -0.06 to $+\ 0.09$. The excess charge of the same type of amino acid at different locations in the protein is displayed in Figure 6. This figure shows that charge transfer can be quite different depending on the specific position of a given residue as well as its environment and the protein sequence, for example, the average excess charge on phenylalanine ranges from -0.05 to $+0.04$, that on cysteine from -0.04 to $+0.04$, and that on alanine from

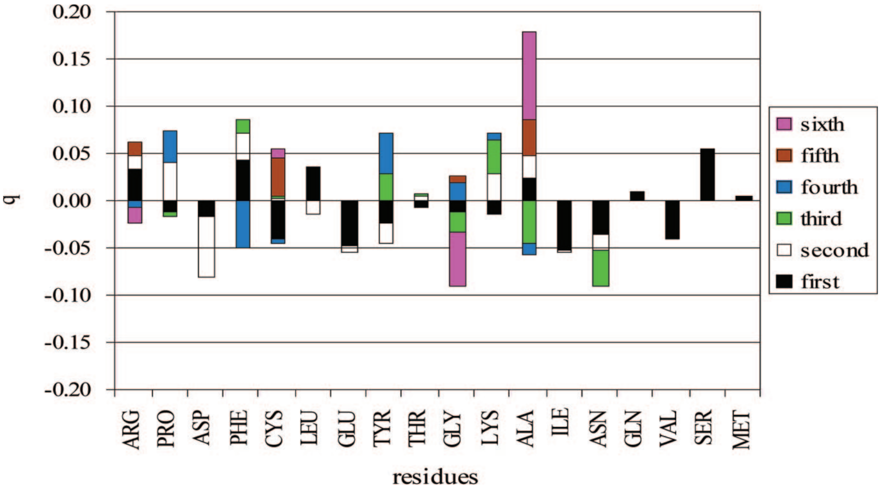


Figure 6. Average excess charge (in atomic units) for each residue of BPTI in water. The results for peptide units of the same type are shown together. The ordering of the residues, to which the colors refer, is by sequence number. Negative excess charges are shown below the baseline, and positive charges above it. The magnitude of the charge associated with each colored segment is from the bottom of a given colored segment to the top, not from the baseline to the top/bottom of each color bar. So the magnitude of charge associated with the residue with the longest colored segment is the largest.

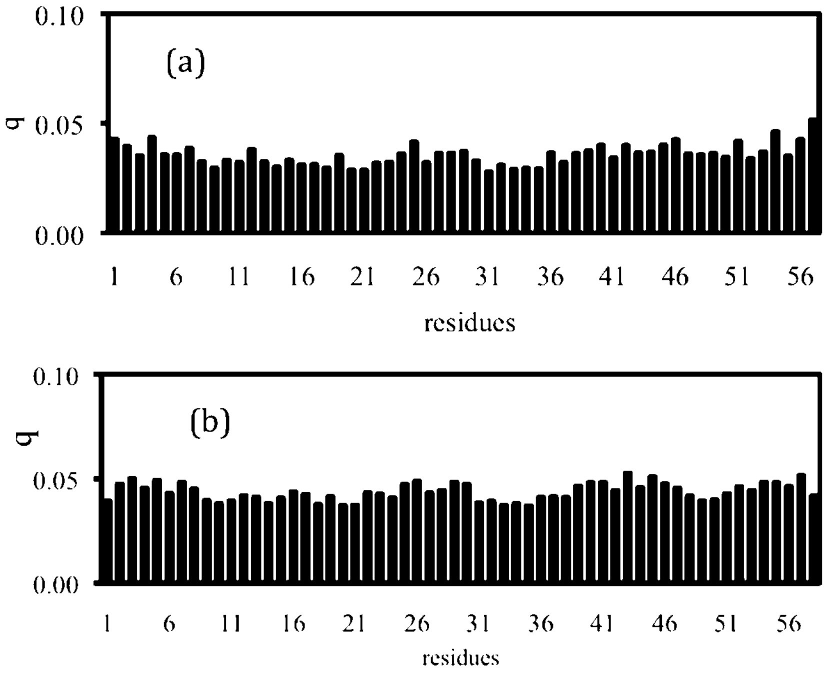


Figure 7. Standard deviations of the charges shown in Figures 4 and 5: (a) partial charges (in atomic units) on the backbone carbonyls and (b) excess charges on the residues.

−0.05 to +0.09. The instantaneous excess charges vary even more widely.

In Figures 4−6 we reported averages for net carbonyl charges and for excess residue charges over 30000 time frames. Using the same data, we also computed standard deviations for these 30000 time frames, and this provides a measure of the variation of the instantaneous charges with time (as the environment changes). These results are given in Figure 7. This variation is different from the sequence dependence of the average (shown in Figure 6), and it is a feature of the true dynamics that a nonpolarizable force field can never reproduce. Figure 7 shows that the standard deviations are in the range of 0.03 to 0.05.

5. Computational Considerations and Future Improvements

The total CPU time for the 50 ps simulation using a single 2.66-GHz SGI Altix XE 1300 Linux Cluster processor is 377 h, whereas it took 62.6 h to run 5 ps on a single 1.5-GHz IBM Power4 processor. During each DSCF calculation, all one- and two-electron integrals are saved in the memory although one has the option to calculate the electron integrals on the fly. The number of one-electron QM/MM integrals scales as N^2 where N is the number of fragments, and the total memory used to store one-electron integrals is about 458 MB; however, the two-electron integrals are only evaluated within each fragment and its buffer fragments that share a boundary atom, so their number scales as N, and the total memory used to store two-electron integrals is merely 13 MB. In principle, by reorganizing the algorithm, the X-Pol potential is highly parallelizable, at least up to a number of processors equal to the number of fragments, since fragments or groups of fragments and the associated QM/MM integrals can be distributed over processors, and the only messages to be communicated for potential energy evaluation are the DSCF updates to auxiliary density matrices, the buffer density matrices, and the partial charges. A parallel version of the X-Pol potential is currently being developed.

In the present X-Pol potential, atom-centered point charges are calculated from class II Mulliken population analysis. In future work, it may be worthwhile to explore other approximations to the MM potential such as charges fitted to electrostatic potentials (class III ESP charges),[51] CM4 class IV charges,[52] or distributed multipoles.[53]

In the present study, charge transfer along the chain occurs through boundary atoms, whereas charge transfer between protein and solvent is not included. Although the present study used small fragments (peptide units and water molecules), the method is very flexible, and one can use larger fragments to include charge transfer between fragments not connected by a chain of bonds or to include charge transfer between connected fragments more self-consistently. For an example of the latter, one can include in a single fragment either neighboring peptide units or peptide units interacting in secondary or tertiary structure via hydrogen bonds or salt bridges. As mentioned above, one can include solute−solvent charge transfer by, for example, treating a peptide unit and a nearby water molecule (or molecules) as a single fragment. Since water molecules exchange in and out of the first solvation shell, such solute−solvent fragments should be treated by an adaptive[54] algorithm that allows such exchanges. For studying enzyme kinetics, one could take the entire substrate and coenzyme to be a single fragment, if desired.

An important objective in future work is to carefully parametrize the quantum mechanical model to achieve the desired accuracy for properties both in the gas phase and in the condensed phase.

6. Concluding Remarks

We have demonstrated the applicability of an explicit polarization (X-Pol) method[31−36] to a solvated protein in a water box with $>10^4$ atoms. The calculation presented here is equivalent to a molecular dynamics simulation of a 14281-atom system consisting of 29026 basis functions with direct dynamics based on an explicit quantum mechanical electronic wave function for the entire system plus a van der Waals term for interfragment exchange repulsion and dispersion forces. The present molecular dynamics simulation generated a trajectory over a 50 ps time interval in an NVT ensemble using an existing semiempirical model. Remarkably, on a single processor without carefully optimizing the quantum mechanical code, it is possible to run more than 3 ps of direct dynamics per day. Whereas all atoms are treated quantum mechanically in this simulation, a typical combined QM/MM simulation treats only 10^1-10^2 atoms quantum mechanically (for example, the quantum-electronic-structure subsystem had 69,[55] 47−60,[56] 53−56,[57] 102,[58] and 71[50] atoms in some recent studies). Analysis of the wave function implies that the polarization and charge transfer effects are significant in the condensed phase and protein. Water molecules display a significant polarization effect in the condensed phase. Carbonyl groups in the protein backbone bear a negative net charge. The net charges of the backbone carbonyl group in different amino acid residues are different by an amount up to about 0.1 e which suggests that in some cases intramolecular charge transfer needs to be considered explicitly. Analysis of the excess charge of each amino acid shows fluctuating charge transfer between amino acid residues in the protein. The same residue may have significantly different charge transfer effects depending on protein sequence.

In closing, it is worthwhile to emphasize the differences between the present linear scaling method and problem decomposition by a divide-and-conquer-type[59] approach. Methodologically, X-Pol and divide-and-conquer are not the same, even when the same quantum mechanical model is used. Divide-and-conquer is a linear scaling method to efficiently obtain a solution of the quantum mechanical model (e.g., Hartree−Fock or Kohn−Sham equations) for a large system. In contrast, X-Pol is a quantum mechanical force field, whose energy is not the Hartree−Fock or Kohn−Sham energy of the system. X-Pol is variational and constitutes an efficient method that can be used to run molecular dynamics simulations for a fully solvated protein, whereas D&C requires many more SCF iterations to obtain an energy that is accurate enough for gradients (forces) in MD simulations. Although AM1 was used in the present study for the purpose of proof of concept, the X-Pol potential is designed as a force field to be parametrized just as “standard” CHARMM, AMBER, or OPLS force fields are parametrized.

Acknowledgment. This work was supported in part by the National Institutes of Health under award no. GM46736, the Office of Naval Research under grant no. N00012-05-01-0538, and the National Science Foundation under grant no. CHE07-04974. J.G. thanks the Ministerio de Educación y Ciencia, España for partial support during his sabbatical leave and the BSC for human and technical resources.

References

(1) (a) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186. (b) Elber, R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 151. (c) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679. (d) Rueda, M.; Ferrer, C.; Meyer, T.; Pérez, A.; Camps, J.; Hospital, A.; Gelpı´, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 796.

(2) Bixon, M.; Lifson, S. *Tetrahedron* **1967**, *23*, 769.

(3) (a) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27. (b) Patel, S.; Brooks, C. L., III. *Mol. Simul.* **2005**, *32*, 231. (c) Senn, H. M.; Thiel, W. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182. (d) MacKerell, A. D., Jr. *Annu. Rep. Comput. Chem.* **2005**, *1*, 91.

(4) (a) Xie, W.; Pu, J.; MacKerell, A. D., Jr.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1878. For a thematic issue devoted to current research on polarization and polarizable force fields, see: (b) Jorgensen, W. L. *J. Chem. Theory Comput.* **2007**, *3*, 1877.

(5) (a) Hill, T. *J. Chem. Phys.* **1946**, *14*, 465. (b) Westheimer, F. H.; Mayer, J. E. *J. Chem. Phys.* **1946**, *14*, 733.

(6) Hendrickson, J. B. *J. Am. Chem. Soc.* **1961**, *83*, 4537.

(7) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.

(8) For a historical account of the development of the Lifson-style force field, see Levitt, M. *Nat. Struct. Mol. Biol.* **2001**, *8*, 392.

(9) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.

(10) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(11) MacKerell, A. D., Jr.; 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., III.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(12) 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.

(13) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F. *Eur. Biophys. J.* **2005**, *34*, 273.

(14) Maple, J. R.; Hwang, M. J.; Jalkanen, K. J.; Stockfisch, T. P.; Hagler, A. T. *J. Comput. Chem.* **1998**, *19*, 430.

(15) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730.

(16) Van Belle, D.; Couplet, I.; Prevost, M.; Wodak, S. J. *J. Mol. Biol.* **1987**, *198*, 721.

(17) Niesar, U.; Corongiu, G.; Clementi, E.; Kneller, G. R.; Bhattacharya, D. K. *J. Phys. Chem.* **1990**, *94*, 7949.

(18) Sprik, M.; Klein, M. L.; Watanabe, K. *J. Phys. Chem.* **1990**, *94*, 6483.

(19) Dang, L. X.; Chang, T. M. *J. Chem. Phys.* **1997**, *106*, 8149.

(20) (a) Rick, S.; Stuart, S.; Berne, B. *J. Chem. Phys.* **1994**, *101*, 6141. (b) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515. (c) Maple, J. R.; Cao, Y.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A. *J. Chem. Theory. Comput.* **2005**, *1*, 694.

(21) Ding, Y.; Bernardo, D. N.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1995**, *99*, 11575.

(22) (a) Gao, J.; Habibollazadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460. (b) Gao, J.; Pavelites, J. J.; Habibollazadeh, D. *J. Phys. Chem.* **1996**, *100*, 2689.

(23) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048.

(24) (a) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933. (b) Grossfield, A.; Ren, P.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 15671. (c) Rasmussen, T. D.; Ren, P.; Ponder, J. W.; Jensen, F. *Int. J. Quantum Chem.* **2007**, *107*, 1390.

(25) Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025.

(26) Cubero, E.; Luque, F. J.; Orozco, M.; Gao, J. *J. Phys. Chem. B* **2003**, *107*, 1664.

(27) (a) Rappé, A.; Goddard, W. A., III. *J. Phys. Chem.* **1991**, *95*, 3358. (b) Ogawa, T.; Kuita, N.; Sekino, H.; Kitao, O.; Tanaka, S. *Chem. Phys. Lett.* **2004**, *397*, 382.

(28) Vorobyov, I. V.; Anisimov, V. M., Jr. *J. Phys. Chem. B* **2005**, *109*, 18988.

(29) Patel, S.; Brooks, C. L., III. *J. Chem. Phys.* **2005**, *123*, 164502.

(30) Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Dehez, F.; Angyan, J. G.; Orozco, M.; Chipot, C.; Luque, J. *J. Chem. Theory Comput.* **2007**, *3*, 1901.

(31) Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657.

(32) Gao, J. *J. Chem. Phys.* **1998**, *109*, 2346.

(33) Wierzchowski, S. J.; Kofke, D. A.; Gao, J. *J. Chem. Phys.* **2003**, *119*, 7365.

(34) Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890.

(35) Xie, W.; Song, L.; Truhlar, D.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108.

(36) Xie, W.; Song, L.; Truhlar, D.; Gao, J. *J. Phys. Chem. B* **2008**, *112*, 14124.

(37) Pulay, P. *Adv. Chem. Phys.* **1987**, *69*, 241.

(38) *Pure Appl. Chem.* **1974** *40* 291 See also: Moss, G. P. *Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains*, IUPAC-IUB Commission on Biochemical Nomenclature (CBN): London, 1974. Available at http://www.chem.qmul.ac.uk/iupac/misc/ppep1.html (accessed Dec 20, 2008).

(39) (a) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714. (b) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. *Theor. Chem. Acc.* **2000**, *104*, 336.

(40) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(41) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

(42) (a) Gao, J.; Xia, X. *Science* **1992**, *258*, 631. (b) Gao, J. *Rev. Comput. Chem.* 1995, *7*, 119.

(43) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.

(44) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(45) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(46) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(48) (a) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511. (b) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

(49) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(50) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019.

(51) (a) Momany, F. A. *J. Phys. Chem.* **1978**, *82*, 592. (b) Gao, J.; Luque, F. J.; Orozco, M. *J. Chem. Phys.* **1993**, *98*, 2975.

(52) (a) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87. (b) Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**,

100, 16385. (c) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820. (d) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.

(53) (a) Stone, A. J.; Price, S. L. *J. Phys. Chem.* **1988**, *92*, 3325. (b) Chipot, C.; Ángyán, J. G. *New J. Chem.* **2005**, *29*, 411.

(54) (a) Heyden, A.; Lin, H.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 2231. (b) Heyden, A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 217.

(55) Garcia-Viloca, M.; Truhar, D. G.; Gao, J. *Biochemistry* **2003**, *42*, 13558.

(56) Spiegel, K.; Rothlisberger, U.; Carloni, P. *J. Phys. Chem. B* **2004**, *108*, 2699.

(57) Rodriguez, A.; Oliva, C.; Gonzalez, M.; Van der Kamp, M.; Mulholland, A. J. *J. Phys. Chem. B* **2007**, *111*, 12909.

(58) Tuttle, T.; Thiel, W. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159.

(59) (a) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438. (b) Stechel, E. B. In *Domain-Based Parallelism and Problem Decomposition Methods in Computational Science and Engineering*; Keyes, D. E., Saad, Y., Truhlar, D. G., Eds.; SIAM: Philadelphia, 1995; pp 217−238. (c) Bowler, D. R.; Aoki, M.; Goringe, C. M.; Horsfield, A. P.; Pettifor, D. G. *Modell. Simul. Mater. Sci. Eng.* **1997**, *5*, 199. (d) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085. (e) Gogonea, V.; Westerhoff, L. M.; Merz, K. M., Jr. *J. Chem. Phys.* **2000**, *113*, 5604. (f) Shimojo, F.; Kalia, R. K.; Nakano, A.; Vashishta, P. *Comput. Phys. Commun.* **2005**, *167*, 151.

CT800239Q

JCTC Journal of Chemical Theory and Computation

Tunneling Dynamics in a Double-Well Model of an H Transfer Reaction

Ashley E. Myers, Matt R. Teague, and Michael Messina*

Department of Chemistry and Biochemistry, University of North Carolina, Wilmington, North Carolina 28403

Received October 15, 2008

Abstract: We study the role that tunneling can play in the reaction dynamics of H atom transfer. The small mass of the H atom offers it another, nonclassical route, from reactants to products, tunneling through an activation barrier. In this work, we carefully define the portion of a reaction rate constant that is caused by tunneling in such reactions. We do this by decomposing an initial H atom wavepacket into above and below the barrier components. We show that for a very particular decomposition, the quantum dynamics of the system can be separated into two events: tunneling and above-the-barrier product production. We show for such decomposition it is possible to determine a rate constant because of tunneling alone. Finally, we demonstrate that from a single experimental observable, the overall decay of reactant concentration, one can extract structural and dynamical information about the H atom transfer reaction.

I. Introduction

Recently there has been great interest in the role tunneling plays in reaction dynamics, particularly enzyme catalyzed H transfer reactions.[1−10] As is well-known, the small mass of the H atom offers it another, nonclassical route, from reactants to products, tunneling through an activation barrier. There have been experimental measures, calculations, and talk in the literature on the role tunneling plays in catalyzed H atom transfer.[11−15] But, we feel, as of yet, tunneling in such reactions has not been carefully defined. Consider the following scenario: An experimentalist follows the dynamics of H atom transfer for a short time and then deduces a first-order rate constant, for example, by fitting the short-time decay of reactant concentration to a decaying exponential. Can this rate constant be considered as the “rate of tunneling”? One might argue if the barrier is large enough, the H atom had no other way to get to products other than tunneling, and yes indeed the measured rate constant is representative of a tunneling rate. There is a problem with this argument. What barrier is “high enough” to preclude all other routes to product production other than tunneling? The usual argument is that a barrier is too high for over the barrier production if its energy is much greater than the average of the thermal distribution of initial H atom energies at the temperature at which the rate is measured. The problem with this argument is that the H atom can be prepared in a nonequilibrium state and is not distributed according to a Boltzmann distribution. There have been arguments in the literature, for example, that claim that enzymes could prepare reactants in the most advantageous way as to approach products.[16−18] For example, an enzyme can cause the transfer of electron density away from bonding or into antibonding orbitals involved in binding the H atom to a donor atom of a substrate. In this case the H-donor bond will elongate and weaken, resulting in an H atom that is no longer in an equilibrium state. This H atom is then described, at $t = 0$, by a wavepacket not at equilibrium. A classical H atom, at zero temperature, prepared with energy, E_0, has this single, definite value of energy. The more correct view is that the H atom is a quantum mechanical particle, and even at zero temperature, is represented by a wavepacket having a distribution of energies. Say the H atom must overcome an activation barrier of energy E_A to become products. The H atom wavepacket at time zero, particularly if it is prepared in a nonequilibrium state, can have energy components above the barrier as well as components below the barrier. The ensuing quantum dynamics of this reactive system contains two events, an above-the-barrier H atom transfer from the components of the wavepacket initially above the barrier at

---

* To whom correspondence should be addressed. E-mail: messinam@uncw.edu.

10.1021/ct800436u CCC: $40.75

time zero and tunneling through the barrier caused by the below the barrier components of the initial wavepacket. In this scenario the initial, short-time, decay of reactants would be dominated by the faster, above the barrier event, and the deduced rate constant would have very little to do with tunneling.

In this work we demonstrate that there is a reasonable definition of a rate constant for tunneling alone in a reactive event involving an atom that can tunnel through an activation barrier. For such a separation of dynamics in a single reactive event there has to be a rigorous definition of "components" of an initial wavepacket above and below a reactive barrier. We take the philosophy here that if we decompose an initial wavepacket into two such components, these components must satisfy two criteria.

The first one is that these two components must evolve independently of one another. For example, if we decompose the initial reactant wavepacket into an above the barrier and a below the barrier component, the quantum dynamics of the below the barrier component, that is, the one that represents tunneling alone, should evolve independently of the other component. If this is the case, then we could indeed view the overall single reactive process as being composed of a tunneling event and an above the barrier route to product production. We will see that there is only one way to decompose an initial wavepacket into these two components where each component evolves independently of the other. This is to expand the initial reactive wavepacket in the eigenstates of the $\hat{H}$ operator governing the time evolution of the initial wavepacket in the time-dependent Schrödinger equation (TDSE).

The second criteria is that we must be able to extract, out of the quantum dynamics of the system, a time-evolving probability density for each event. We show that such a division is possible for the common and ubiquitous double-well reactive system and, thus, are able to extract out a rate constant for tunneling alone.

This paper is organized as follows. In section II, we discuss the theory. We describe how an initial, nonequilibrium, H atom wavepacket can be decomposed into above and below the barrier components. We show that for one, special, decomposition, the ensuing dynamics of the system is separable into a tunneling and an above the barrier component. We also show in section II how this special decomposition of the initial H atom wavepacket yields a way to extract out a rate constant for the transfer event that is due to tunneling alone. In section III, we describe numerical results. In section III, we also make the connection between the theory presented and experimental observable quantities in a kinetics experiment. We show that it may be possible for an experimentalist to extract out a rate constant for tunneling alone in an H transfer reaction, just from experimentally observable data. Finally, in section IV, we discuss the results and conclusions.

II. Theory

(a). Decomposition of an Initial Wavepacket. Consider a quantum mechanical particle, such as an H atom, moving in a single spatial dimension, x. For now we leave the potential energy of the system unspecified. We assume that the particle can undergo a "reaction" and transfer from a reactant-side to a product-side of a reaction profile. We assume that an activation barrier of energy, E_A, spatially divides the reactants and products from one another. The reactants and products are separated by a dividing point at $x = x^*$. Thus, if the particle is at positions $x < x^*$, it is considered reactant, and is considered product if the particle is at positions $x > x^*$.

A classical particle, prepared with an energy E_0 at a temperature of $T = 0$ has a single, definite value of energy of E_0. Thus, if this classical particle has $E_0 > E_A$, it would have enough energy to overcome the activation barrier, allowing a reaction to occur. If $E_0 < E_A$, then the particle would not have enough energy to overcome the activation barrier and no reaction would take place.

The situation is more complicated if we treat the particle correctly, according to quantum mechanics. Say the particle, prepared at time, $t = 0$, is in a nonequilibrium state on the reactant side of the reaction, and is represented by a wave function, $\Psi_0(x,t = 0)$. A particle represented by such a wave function does not have a single value of energy, but rather a distribution of energies, $P(E)$. For example, the probability of the particle having the activation energy is $P(E_A)$. The wavepacket, $\Psi_0(x,t = 0)$, contains some energy values greater than E_A and others that are less than E_A. The components with energy greater than E_A have enough energy to overcome the activation barrier. Thus, the components of the initial wavepacket with $E > E_A$ do not get to products via tunneling but become products much as a classical particle would. If we want to ascribe a rate of reaction caused by tunneling alone, we must be careful to exclude these components of the initial wave function from consideration. The components of the initial reactant wavepacket with energy less than E_A, on the other hand, can only become products via tunneling and are the components of the initial wavepacket that contribute to the portion of the reaction rate caused by tunneling.

The initial wavepacket, $\Psi_0(x,t = 0)$, evolves in time according to the time-dependent Schrödinger equation (TDSE). At time, t, the wavepacket evolves into $\Psi(x,t)$, and the probability of the particle being on the reactant side of the barrier at this time is defined as $P_R(t)$ shown below

$$P_R(t) = \int_{x=-\infty}^{x=x^*} |\Psi(x,t)|^2 \mathrm{d}x \tag{1a}$$

If the reaction follows first-order kinetics $P_R(t)$ should decay exponentially according to

$$P_R(t) = \mathrm{e}^{-\kappa t} \tag{1b}$$

A method for determining the rate constant, κ, is as follows: Evolve $\Psi_0(x,t = 0)$, in time by solving the appropriate TDSE, compute $P_R(t)$ by using eq 1a, and then fit the results to an exponential according to eq 1b and extract out the value of the rate constant κ. The problem with this method is that the rate constant found using this approach would not only include the components of $\Psi_0(x,t = 0)$ that tunnel through the barrier; it would also include the components of $\Psi_0(x,t$

= 0) that can energetically overcome the barrier. Thus, the rate constant determined by such an algorithm would not represent the rate of tunneling alone, but a combination of tunneling and above-the-barrier dynamics. If one wants to extract out a rate constant from the quantum dynamics just from tunneling another algorithm must be used.

What we have in mind is the decomposition of the initial reactant wavepacket into two components; $\Psi_0^-(x,t=0)$, the component of initial wavepacket below, and $\Psi_0^+(x,t=0)$, the component of initial wavepacket above the barrier, at $t = 0$. We then decompose the initial wavepacket as

$$\Psi_0(x,t=0) = \Psi_0^+(x,t=0) + \Psi_0^-(x,t=0) \tag{2}$$

The initial wavepackets $\Psi_0^-(x,t=0)$ and $\Psi_0^+(x,t=0)$ evolve according to the TDSE into $\Psi^-(x,t)$ and $\Psi^+(x,t)$, respectively. We want to be able to assign two rates of product production, κ^- and κ^+, where κ^- is the rate of product production because of tunneling alone, and κ^+ is the rate of product production due to above the barrier product production. The only way for these rates to be well defined is if we could define two probability densities, $P_R^-(t)$ and $P_R^+(t)$ as shown below

$$P_R^-(t) = \int_{x=-\infty}^{x=x^*} |\Psi^-(x,t)|^2 \mathrm{d}x \tag{3a}$$

$$P_R^+(t) = \int_{x=-\infty}^{x=x^*} |\Psi^+(x,t)|^2 \mathrm{d}x \tag{3b}$$

Both $P_R^-(t)$ and $P_R^+(t)$ should be decreasing functions of time because they describe the loss of reactant. Our conjecture is that there should be a wide disparity in the time−decay between $P_R^-(t)$ and $P_R^+(t)$, with the loss of reactant through tunneling, $P_R^-(t)$, being the slower process. For example, if both $P_R^-(t)$ and $P_R^+(t)$ were each described by a single exponential decay then we could write

$$P_R^-(t) \propto e^{-\kappa^- t} \tag{4a}$$

$$P_R^+(t) \propto e^{-\kappa^+ t} \tag{4b}$$

where κ^- is the rate constant for product production via tunneling and κ^+ is the rate constant for over the barrier production of product. If our conjecture about the wide disparity between the decay of $P_R^-(t)$ and $P_R^+(t)$ were true, then we would find $\kappa^- \ll \kappa^+$. In the Numerical Results section, we will say more about the particular functional forms of $P_R^-(t)$ and $P_R^+(t)$.

We now explore if we can indeed decompose the quantum dynamics of the system into two events: a tunneling event and an above the barrier production event.

To show that such a partition of the quantum dynamics, as shown in eqs 3a−4a, is possible, we now specify the Hamiltonian operator, $\hat{H}$, for the system as

$$\hat{H} = \frac{-\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V^*(x) \tag{5}$$

where the first term is the kinetic energy operator and the second term is potential energy of the system. The set of eigenstates, $\{\varphi_n(x)\}$ of any $\hat{H}$ operator forms a complete set of states. Thus, the initial wavepacket, $\Psi_0(x,t=0)$, can be expanded in terms of this complete set according to

$$\Psi_0(x,t=0) = \sum_n c_n |\varphi_n(x)\rangle \tag{6}$$

In the above equation, the set of eigenstates, $\{\varphi_n(x)\}$ can be a set of solutions to *any* time-independent Schrodinger equation, and the basis coefficients, $\{c_n\}$, can be determined by the projections

$$c_n = \langle\varphi_n(x)|\Psi_0(x,t=0)\rangle \tag{7}$$

Thus, the $\{\varphi_n(x)\}$can be taken as the solutions of *any* time-independent Schrödinger equation

$$\hat{H}\varphi_n(x) = E_n\varphi_n(x) \tag{8}$$

where $\hat{H}$ is a Hamiltonian operator containing any potential energy function, $V(x)$ and the $\{E_n\}$ are the set of eigenenergies. Within this spectrum of eigenenergies are two subsets, one, $\{E^-\} = \{E_1, E_2, E_3,..., E_m\}$, is the set of energies below the barrier, and the other, $\{E^+\} = \{E_{m+1}, E_{m+2},...\}$, is the subset containing energies above the barrier. We can expand each of the initial wave functions, $\Psi_0^-(x,t=0)$ and $\Psi_0^+(x,t=0)$in this set of eigenstates as

$$\Psi_0^-(x,t=0) = \sum_{j=1}^{m} c_j|\varphi_j(x)\rangle \tag{9a}$$

$$\Psi_0^+(x,t=0) = \sum_{j=m+1} c_j|\varphi_j(x)\rangle \tag{9b}$$

Now we come to an important juncture. If the $\hat{H}$ operator in eq 8 does not contain the potential energy, $V^*(x)$, of our system, then the energies, $\{E_n\}$, are not the eigenenergies of our system, and we would not be sure that the subset of energies $\{E^-\} = \{E_1, E_2, E_3,..., E_m\}$ truly represent components of the initial reactant wave function that are below the barrier. We now show that if the initial wavepacket was expanded in a set of eigenstates that are not the eigenstates of the $V^*(x)$ potential, then the wavepackets $\Psi^-(x,t)$ and $\Psi^+(x,t)$ do not evolve independently of one another and the separation of the quantum dynamics into separate tunneling and nontunneling events would not make sense.

We now introduce a set of eigenfunctions, $\{\psi_n(x)\}$, which *are* eigenfunctions of the Hamiltonian operator, $\hat{H}$ for our system as described in eq 5. These eigenfunctions have an associated set of eigenenergies,$\{E_n\}$, that are the true eigenenergies of our system. We can then, unambiguously, decompose this spectrum of energies into two subsets: $\{E^-\} = \{E_1, E_2,..., E_m\}$, the set of energies having values less than E_A, and $\{E^+\} = \{E_{m+1}, E_{m+2},...\}$, the set containing above the barrier energies. This allows us to truly decompose the initial wavepacket into above and below the barrier components as

$$\Psi_0^-(x,t=0) = \sum_{j=1}^{m} c_j|\psi_j(x)\rangle \tag{10a}$$

$$\Psi_0^+(x,t=0) = \sum_{j=m+1} c_j|\psi_j(x)\rangle \tag{10b}$$

Now we consider the time-evolution of the initially decomposed wavepacket. We write the wavepacket for all times as

$$\Psi(x,t) = \Psi^-(x,t) + \Psi^+(x,t) \tag{11}$$

This wavepacket evolves according to the TDSE shown below

$$\hat{H}\Psi(x,t) = i\hbar\frac{\partial\Psi(x,t)}{\partial t} \tag{12}$$

Substituting eq 11 into the TDSE in eq 12 results in

$$\hat{H}\{\Psi^-(x,t) + \Psi^+(x,t)\} = i\hbar\frac{\partial\{\Psi^-(x,t) + \Psi^+(x,t)\}}{\partial t} \tag{13}$$

If the time-evolution of our system is truly composed of two separate independent dynamic events, then each event should be governed by it's own TDSE. Thus, there should be a TDSE for the tunneling event and one for the above the barrier dynamics. With this in mind, we decompose eq 13 into the following equations

$$i\hbar\frac{\partial\Psi^-(x,t)}{\partial t} = \hat{H}\Psi^-(x,t) \tag{14a}$$

$$i\hbar\frac{\partial\Psi^+(x,t)}{\partial t} = \hat{H}\Psi^+(x,t) \tag{14b}$$

We now explore the separability of the quantum dynamics of the system into tunneling and above-the-barrier dynamical components for a general expansion of $\Psi_0(x,t = 0)$ in a general set of eigenfunctions, $\{\varphi_n(x)\}$, as in eq 6. We define the set of time-dependent basis coefficients corresponding to eigenfunctions with energies below and above the barrier as $\{c^-(t)\}$ and $\{c^+(t)\}$, respectively. We solve the TDSE by putting eq 6 into eq 13 to get

$$i\hbar\left\{\sum_{j=1}^{m}\dot{c}_j^-(t)|\varphi_j(x)\rangle + \sum_{j=m+1}\dot{c}_j^+(t)|\varphi_j(x)\rangle\right\} = \sum_{j=1}^{m}c_j^-(t)\hat{H}|\varphi_j(x)\rangle + \sum_{j=m+1}c_j^+(t)\hat{H}|\varphi_j(x)\rangle \tag{15}$$

Now we multiply eq 15 on the left by, $\varphi_\ell(x)$, an eigenfunction with an energy below the barrier, and then integrate over all space to get

$$i\hbar\dot{c}_\ell^-(t) = \sum_{j=1}^{m}c_j^-(t)H_{\ell j} + \sum_{j=m+1}c_j^+(t)H_{\ell j} \tag{16}$$

In eq 16, the Hamiltonian matrix elements, $H_{\ell j}$, are defined by

$$H_{\ell j} = \langle\varphi_\ell(x)|\hat{H}|\varphi_j(x)\rangle \tag{17}$$

According to eq 16, the time-evolution of $c_\ell^-(t)$, a basis coefficient governing the dynamical evolution of a wavepacket component below the barrier is not independent of the above the barrier components of the wave function, that is, $\Psi^+(x,t)$. This is because, as shown in eq 16, the differential equation governing below the barrier coefficients, $\{c^-(t)\}$, depends on above the barrier coefficients, $\{c^+(t)\}$. Thus, in a general expansion of the initial wavepacket in a general set of eigenstates, $\{\varphi_n(x)\}$, we cannot separate the dynamics into tunneling dynamics and above the barrier dynamics. Now we ask if such a separation of dynamics is possible if $\Psi_0(x,t = 0)$ is expanded in the special set of complete states, that is, $\{\psi_j(x)\}$, that are solutions of the TDSE with $V = V^*(x)$, that is, the potential energy of the system being studied. If we use these eigenstates as a basis, then the Hamiltonian matrix elements in eq (17) become

$$H_{\ell j} = \langle\psi_\ell(x)|\hat{H}|\psi_j(x)\rangle = \mathrm{E}_j\langle\psi_\ell(x)|\psi_j(x)\rangle = \mathrm{E}_j\cdot\delta_{\ell j} \tag{18}$$

Equation 16 then, in this case, becomes

$$\dot{c}_\ell(t) = -\frac{i}{\hbar}\mathrm{E}_\ell\cdot c_\ell(t) \tag{19}$$

A similar differential equation obtains, for all the basis coefficients, if we expand $\Psi^0(x,t = 0)$ in the set of eigenfunctions, $\{\psi_j(x)\}$, of our potential energy, $V^*(x)$. As shown in eq 19, each below the barrier coefficient, $c_\ell(t)$, evolves independently of all the others, and most importantly, independently of the above the barrier coefficients, $\{c^+(t)\}$. This demonstrates that tunneling and over-the-barrier quantum dynamics can be considered separately if we expand $\Psi_0(x,t = 0)$ in the set of eigenfunctions of the $\hat{H}$ operator, that is, in the set $\{\psi_j(x)\}$. Within this expansion the basis coefficients evolve according to

$$c_\ell(t) = c_{\ell,0}\cdot \mathrm{e}^{-\frac{i}{\hbar}\mathrm{E}_\ell t} \tag{20a}$$

where the initial value of the basis coefficients, $c_{\ell,0}$, are calculated from

$$c_{\ell,0} = \langle\psi_\ell(x)|\Psi^0(x,t=0)\rangle \tag{20b}$$

Thus far we have shown that the quantum dynamics of a "reactive" system can be separated into two dynamic events: tunneling and above the barrier transfer. We have also shown that this separation is only obtained if the initial wavepacket is decomposed in terms of the eigenstates of the system of interest. We have *not* yet demonstrated that the total time evolution of the probability density, $P_{\mathrm{R,T}}(t)$, can be broken up into two pieces as suggested by eqs 3b−4b. This must be the case if we are to be able to ascribe a rate solely for tunneling in a reaction. The total probability density of reactant at a time t is given by

$$P_{\mathrm{R,T}}(t) = \int_{x=-\infty}^{x=x^*}|\Psi(x,t)|^2\mathrm{d}x = \int_{x=-\infty}^{x=x^*}|\Psi^-(x,t) + \Psi^+(x,t)|^2 \tag{21}$$

The second equality comes from the definition of the wavepacket, $\Psi(x,t)$, in eq 11. We then expand out the square in eq 21. Using the definition for the separate probability densities given by eqs 3a, we can then write eq 21 as

$$P_{\mathrm{R,T}}(t) = P_{\mathrm{R}}^-(t) + P_{\mathrm{R}}^+(t) + 2\cdot\mathrm{Re}\left[\int_{x=-\infty}^{x=x^*}\Psi^-(x,t)\cdot\Psi^+(x,t)\right] \tag{22}$$

Thus, the definition of two probability densities, one for tunneling, and one for above the barrier transfer, will only be true if the last term in eq 22 is zero, or nearly so. To show that this is indeed the case we now discuss the potential energy under consideration here in more detail.

(b). Double-Well Potential. The double-well potential, is by far, the most commonly used potential energy surface used to describe the transfer of an atom from one moiety to another.[19,20] We consider a double-well potential as naturally arising from two diabatic states. We consider here a double-well potential built out of two harmonic single-dimensional diabatic states, a reactant state, $V_R(x)$, and a product state, $V_P(x)$. These diabatic states are defined below

$$V_R(x) = \frac{m\omega_R^2}{2}(x - x_R^0)^2 \tag{23a}$$

and

$$V_P(x) = \frac{m\omega_P^2}{2}(x - x_P^0)^2 \tag{23b}$$

Here $\{\omega_R,\omega_P\}$are the frequencies of the reactant and product wells, respectively, and $\{x_R^0,x_P^0\}$ are the minima of the wells. These diabatic states are coupled to one another via a coupling function, $g(x)$, given by

$$g(x) = g_0 \cdot e^{-\alpha(x-x^*)^2} \tag{24}$$

The degree of coupling between the states is set by the coupling strength, g_0, and the maximum coupling occurs at the dividing point between reactants and products, x^*. Thus, we take the dividing line between reactants and products as the value of position x^* lying between the two minima,$\{x_R^0,x_P^0\}$, and making $V_R(x^*) = V_P(x^*)$. The length scale of the coupling is set by α. If α is large the coupling between the states becomes more localized around the crossing point, x^*.

A well-known, and useful approximation, to the diabatic system described above is it is adiabatic, double-well analogue.[20] In this approximation, there is a single wave-packet describing the evolution of the system, $\Psi(x,t)$, which evolves on a single adiabatic potential, $V^*(x)$. This adiabatic potential is taken as the lowest eigenvalue of the potential energy matrix

$$\underset{\approx}{V} = \begin{bmatrix} V_R(x) & g(x) \\ g(x) & V_P(x) \end{bmatrix} \tag{25}$$

This lowest eigenvalue describes a double-well potential and is given by

$$V^*(x) = \frac{V_R(x) + V_P(x)}{2} - \frac{\sqrt{[V_R(x) - V_P(x]^2 + 4 \cdot g(x)^2}}{2} \tag{26}$$

The quantum dynamics of this system is fully described by a single, time-evolving wavepacket, $\Psi(x,t)$, that evolves according to the TDSE

$$i\hbar\frac{\partial\Psi(x,t)}{\partial t} = [\hat{T} + V^*(x)] \cdot \Psi(x,t) \tag{27}$$

The initial adiabatic wavepacket is given by, $\Psi(x,t = 0) = \Psi_0(x,t = 0)$, where $\Psi_0(x,t = 0)$ is a wavepacket localized in the reactant well, that is, having it's maximum near to $x = x_R^0$. This initial wavepacket can be decomposed into above and below the barrier components, as described above, by expanding it in the set of eigenfunctions of the time-independent Schrödinger equation below

$$[\hat{T} + V^*(x)] \cdot \Psi(x) = \mathrm{E} \cdot \Psi(x) \tag{28}$$

(c). Separation of the Probability Densities. We still have to show that the definition of two probability densities, that is, $P_R^-(t)$ for tunneling production and $P_R^+(t)$ for above the barrier production of products makes sense. Above we have shown that we can indeed ascribe two separate probability densities for the dynamics of each event if eq 22 is true, that is, if the last term on the RHS is zero, or nearly so.

It is well-known, that in a double well system as described here, the eigenstates come in pairs. Each pair is composed (approximately) of a plus and minus combination of the eigenstates of $V_R(x)$ and $V_P(x)$. Specifically, we can write the eigenstates of or system as

$$\psi_{2j+1}(x) \approx c_{2j+1}\{|R_j\rangle + |P_j\rangle\} \quad j = 0, 1, 2, ... \tag{29a}$$

and

$$\psi_{2j}(x) \approx c_{2j}\{|R_{j-1}\rangle - |P_{j-1}\rangle\} \quad j = 1, 2, 3, ... \tag{29b}$$

Here $\{|R_j\rangle,|P_j\rangle\}$are the eigenstates of $V_R(x)$ and $V_P(x)$, respectively. [Note these are harmonic oscillator eigenstates, whose index conventionally starts at $j = 0$.] The $|R_j\rangle$ eigenstates are localized on the reactant side of the double-well potential, and the $|P_j\rangle$ eigenstates are localized in the product well and are nearly zero in the region $x = -\infty \rightarrow x^*$. Since the integral in eq 22 runs from $x = -\infty \rightarrow x^*$, that is, the portion of space where the $|P_j\rangle$ components of the$\{\psi_j(x)\}$are nearly zero we can define components of $\Psi^+(x,t)$ and $\Psi^-(x,t)$ that contribute to the integral in eq 22 as $\chi^+(x,t)$ and $\chi^-(x,t)$ as shown below

$$\chi^-(x,t) = \sum_{j=0}^{m}(c_{2j+1} + c_{2j+2}) \cdot |R_j(x)\rangle \tag{30a}$$

and

$$\chi^+(x,t) = \sum_{\ell=m+1}^{m}(c_{2\ell+1} + c_{2\ell+2}) \cdot |R_\ell(x)\rangle \tag{30b}$$

The integral in eq 22 then becomes

$$\int_{x=-\infty}^{x=x^*}[\Psi^-(x,t)]^* \cdot \Psi^+(x,t)\mathrm{d}x \approx \int_{x=-\infty}^{x=x^*}[\chi^-(x,t)]^* \cdot \chi^+(x,t)\mathrm{d}x \tag{31}$$

Substitution of eq 30a into eq 31 shows that the integral in eq 22 is composed of a sum of terms of the form $\int_{x=-\infty}^{x=x^*} R_j(x) \cdot R_l(x)\,\mathrm{d}x$ with $j \neq \ell$ in each case. Since the $|R_j\rangle$ eigenstates are nearly orthogonal over this interval, the integral in eq 31 and, hence, the one in eq 22 are nearly zero. This means that the separation of the evolved probability densities into a tunneling and above the barrier probability density can be achieved. There are two important points to make about this argument. First, the last term on the RHS of eq 22 is only very small if the initial wavepacket, $\Psi_0(x,t = 0)$, is expanded in the exact eigenstates of the double-well system, $\{\psi_j(x)\}$. Second, there is one case where eq 22 will not be even approximately true. This is the case where the plus eigenstate $|R_j\rangle + |P_j\rangle$ is below the barrier and its companion $|R_j\rangle - |P_j\rangle$ is above. In this case, there will be a large contribution to the integral from a term of the form $\int_{x=-\infty}^{x=x^*} R_j(x) \cdot R_l(x)\,\mathrm{d}x$ with $j = \ell$, which is not zero and is in fact approximately unity. But, in all cases studied, we have not found a case where this situation occurs. Thus, we feel that this happens rarely enough to be neglected.

III. Numerical Results

For all calculations we use the lower adiabatic, double-well, potential given by $V^*(x)$ in eq 26. Some parameters are constant throughout the calculations and we discuss these first. Since we plan to focus on tunneling, we choose the mass of the system, m, to be that of an H atom, m_H. We also explore the isotope effect on the tunneling dynamics by setting the system mass to that of a Deuteron in some calculations, in these cases $m = 2m_H$. The minima of the reactant and product wells, that is, x_R^0 and x_P^0, in eqs 25−26 are 3.0 and 4.25 Å, respectively. Thus, in all calculations the distance between the well minima is 1.25 Å. The frequency of the product well, ω_P, is $\omega_P = 2000$ cm^{-1} in all calculations. We take the length scale of the nonadiabatic coupling, α, to be 1.0 Å^{-2} in all calculations. Some parameters are varied to explore their effect on the quantum dynamics of the system. The reactant well frequency, ω_p, varies between 1500 and 2000 cm^{-1}. The nonadiabatic coupling strength, g_0, is varied between 100 and 220 kJ mol^{-1}. These coupling strengths result in barrier heights, E_A, ranging between 45 and 125 kJ mol^{-1}.

The eigenstates, $\{\psi_j\}$, of the double-well potential are calculated by using a particle in a box basis set. The spatial grid runs from $x = 0$ Å → $x = 6$ Å, with 256 equally spaced grid points. In all calculations the results converged with a basis set comprised of 100 basis functions.

The initial wavepacket of the system, $\Psi_0(x,t = 0)$, is localized on the reactant side of the potential and is taken as the Gaussian shown below

$$\Psi_0(x, t=0) = \left(\frac{m\omega_0}{\pi\hbar}\right)^{\frac{1}{4}} \cdot \exp\left\{-\left(\frac{m\omega_0}{2\hbar}\right) \cdot [x - (x_R^0 - \delta x)]^2\right\} \tag{32}$$

The width of the initial wavepacket is set by the frequency, ω_0, which we take as 2000 cm^{-1} in all calculations. The initial displacement of the wavepacket from equilibrium is controlled by the parameter δx in eq 32. If $\delta x = 0$, then the wavepacket is centered at its equilibrium position at $t = 0$ in the reactant well. As δx increases the initial reactant wavepacket is displaced more from equilibrium and, thus, possesses more initial energy.

As described previously, this initial wavepacket is then expressed as a superposition of the eigenstates of the system, that is, $\{\psi_j\}$, and further broken down into below and above the barrier components as described in eqs 2, 9a and 9b. We chose to study a system where there is no chance for reactant that has crossed the dividing point $x = x^*$ to recross and go back toward reactants. To do this we need a mechanism for pulling probability density off of the product well. We cannot do this by evolving the basis coefficients as shown in eq 20a. We carry this out by discretizing the initial wavepackets on a spatial grid and solving the TDSEs in eqs 14a and 14b via the Feit−Fleck split operator technique. We are then able to pull probability density off of the product state by setting $\Psi(x,t) = 0$ for all $x > x_P^0$ during the course of the time-evolution of the wavepackets. For all calculations in which we solve the TDSE, the spatial grid is the same as that for the basis set calculations and the time step is d$t = 0.01$ fs.

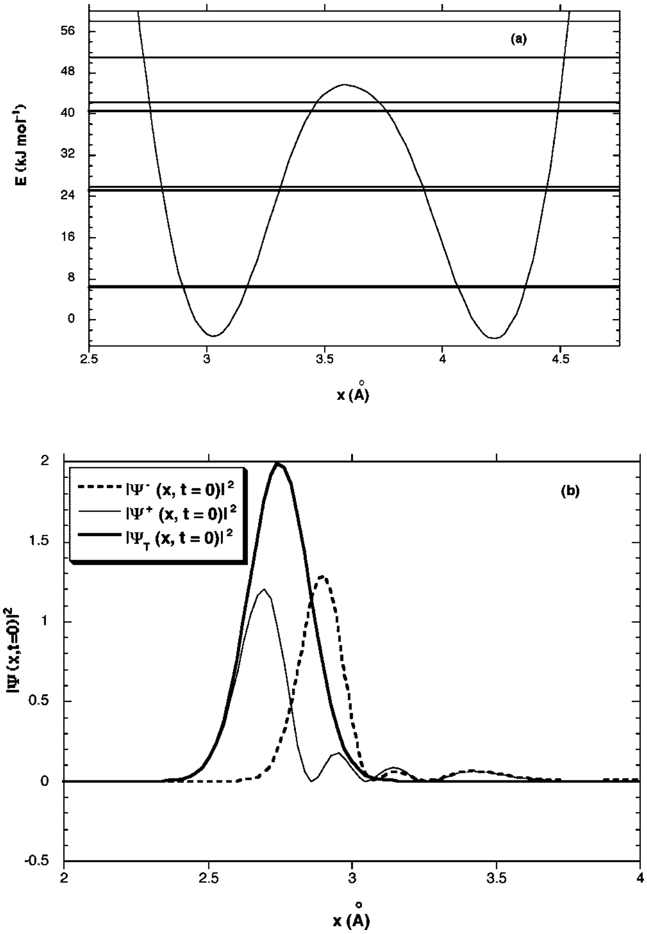


Figure 1. (a) Double-well potential when $\omega_R = \omega_P = 2000$ cm^{-1} with a nonadiabatic coupling strength of $g_0 = 230$ kJ mol^{-1}. The eigenenergies of this system, $\{E\}$, are shown as horizontal lines inside the potential. (b) The squares of the initial wave function and it is below and above the barrier components as described in the text with the thick solid line as the total wave function, the thick dashed line as the below the barrier component and the thin solid as the above the barrier component.

In Figure 1a, we show the double-well potential when $\omega_R = \omega_P = 2000$ cm^{-1} with a nonadiabatic coupling strength of $g_0 = 230$ kJ mol^{-1}. This leads to a double-well potential with a barrier height of $E_A = 45$ kJ mol^{-1}, which is the lowest barrier considered. The eigenenergies of this system, $\{E\}$, are shown as horizontal lines inside the potential. There are 6 eigenstates below the barrier, and as described above, they come in pairs of two. All the other eigenenergies that are above the barrier are no longer paired as those below.

In Figure 1b, we show the initial wave function and its below and above the barrier components when the initial wave function is displaced by an amount $\delta x = 0.25$ Å from equilibrium at $t = 0$. The thick solid line in Figure 1b is the square of the total wave function, $\Psi_0(x,t = 0)$, which is centered at position $x_R^0 - \delta x = 2.75$ Å. The thick dashed line is the square of the below the barrier component of the initial wave function, $\Psi_0^-(x,t = 0)$, of the initial wave function. The thin solid line in Figure 1b is the square of the above the barrier component, $\Psi_0^+(x,t = 0)$, of the initial wave function. Note that the more energetic above the barrier component has more nodes than the below the barrier component, which is consistent with it being the more energetic component. Further, the maximum of the above the barrier wavepacket, $\Psi_0^+(x,t = 0)$, occurs at $x \approx 2.6$ Å, which is 0.4 Å from equilibrium. The maximum of the below the barrier component, on the other hand, occurs at $x \approx 3.0$ Å, which is where the equilibrium of the reactant well occurs. Thus, the component of the wave function that is responsible for product production through tunneling is that component of the initial reactant wavepacket that is nearly at equilibrium on the reactant state.

We now consider the ensuing quantum dynamics of this initially displaced wavepacket. We express this wavepacket as a superposition of the double-well eigenstates and form $\Psi_0^-(x,t = 0)$ and $\Psi_0^+(x,t = 0)$ according to eqs 10a and 10b. We can determine how much of the wavepacket is above and below the barrier at $t = 0$, that is, $P_R^-(t = 0)$ and $P_R^+(t = 0)$, respectively, from

$$P_R^-(t=0) = \int_{-\infty}^{+\infty} |\Psi_0^-(x, t=0)|^2 \cdot dx \qquad (33a)$$

and

$$P_R^+(t=0) = \int_{-\infty}^{+\infty} |\Psi_0^+(x, t=0)|^2 \cdot dx \qquad (33b)$$

With an initial displacement of the wavepacket of $\delta x = 0.25$ Å $P_R^-(t = 0) = 0.5$ and $P_R^+(t = 0) = 0.5$. Thus, in this case half the initial wavepacket contains energy components above the barrier. We then evolve $\Psi_0^-(x,t = 0)$ and $\Psi_0^+(x,t = 0)$ in time according to eqs 14a and 14b using the Feit−Fleck operator algorithm,[21] and in Figure 2a, we show the evolution of the total reactant probability density, $P_{R,T}(t)$. The solid line is approximation to the total reactant probability density obtained when we separate it into it is $P_R^-(t)$ and $P_R^+(t)$ components and set $P_{R,T}(t) \approx P_R^-(t) + P_R^+(t)$. The dashed line in Figure 2a is the exact total reactant probability density. We obtain this probability density by evolving the whole of the initial wavepacket, $\Psi_0(x,t = 0)$, in time according to the TDSE in eq 12. The close agreement between the solid and dashed lines shows that, indeed, in this case the total reactant probability density can be split up into its below and above the barrier components. In fact, we have found that this was the case in all calculations we have performed. Although it may not be obvious from Figure 2a, the time-evolution of the reactant probability density does not follow a single decaying exponential.

Again, we assume that there is a large separation between the time scales of tunneling and over the barrier production. We still need to choose particular functional forms for $P_R^-(t)$ and $P_R^+(t)$ to quantify these time scales. For simplicity, we take both $P_R^-(t)$ and $P_R^+(t)$ as being described by an exponential decay as described in eq 4a. It is important to keep in mind that we have offered no proof that each of $P_R^-(t)$ and $P_R^+(t)$ should be exactly represented by a single exponential decay. In fact, if each below the barrier eigenstate decayed exponentially then the aggregate decay because of tunneling, $P_R^-(t)$, should not be exactly represented by a single decaying exponential but rather by a sum of decaying exponentials. We are assuming that a *comparison* of the two dynamic events, that is, tunneling and over the barrier production, could be well quantified by treating both events as first order events. The following numerical results will demonstrate that this assumption has some validity.

The thicker solid line in Figure 2a shows the best single exponential fit of $P_{R,T}(t)$. In Figure 2b, we show the evolution of the below the barrier component of the wavepacket, that is, $P_R^-(t)$. This probability density represents what we are claiming is the tunneling dynamics of the system. The thicker solid line in Figure 2b shows the best fit to $P_R^-(t)$ to a single, decaying, exponential. The tunneling dynamics alone fits a single exponential decay much more closely than that of the whole time evolution of the probability density. We extract out a rate constant for tunneling from the exponential fit in Figure 2b according to eq 4a and find, in this case, $\kappa^- = 7.5 \times 10^{-6}$ au. In Figure 2c, we show the evolution of the above the barrier probability density, that is, $P_R^+(t)$. The thick solid line in Figure 2c shows the best single decaying exponential fit to $P_R^+(t)$. Although the single exponential fit of $P_R^+(t)$ is not as good as that of $P_R^-(t)$ the fit does follow the overall decay trend pretty well. We extract out a rate constant for above the barrier product production from the exponential fit in Figure 2c according to eq 4b and find, in this case, $\kappa^+ = 5.3 \times 10^{-4}$ au. There is a large separation of time scales between the product production through tunneling, that is, κ^-, and the above the barrier production, that is, κ^+. In this case, in fact, there is a 2 orders of magnitude difference, with the tunneling being the slower process.

Now lets discuss the results in Figure 2 from an experimental viewpoint. The only observable that an experimentalist can measure is the total reactant probability density, that is, $P_{R,T}(t)$. If an experimentalist follows the decay of reactant concentration they would generate a concentration versus time profile like that shown by the thin solid line in Figure 2a. We have shown that this time profile results from two different dynamic routes toward product production: tunneling and above the barrier production. Further, the separation in time scales of these two events is very large. Thus, how can an experimentalist go from a measured concentration versus time profile like that in Figure 2a and extract out a "tunneling rate"? We offer an answer here.

We have shown that the total probability density, at least for a double-well potential, can be written as a sum of two distinct probability densities: $P_{R,T}(t) \approx P_R^-(t) + P_R^-(t)$. Thus, if both of the dynamic events, tunneling and above the barrier production are first-order events then each of $P_R^-(t)$ and $P_R^+(t)$ will decay exponentially and perhaps we can write the overall decay of the reactant concentration in time as

$$P_{R,T}(t) = P_R^-(0) \cdot e^{-\kappa^- t} + P_R^+(0) \cdot e^{-\kappa^+ t} \qquad (34)$$

Here, $\{P_R^-(0), P_R^+(0)\}$ are the below and the above the barrier probability densities contained in the initial wavepacket as described in eqs 33a and 33b. The two rate constants, $\{\kappa^-, \kappa^+\}$, are the rate constants for product production through tunneling and above the barrier transfer, respectively. Thus, we are claiming that the overall decay of the reactant concentration should follow, at least approximately, a biex-

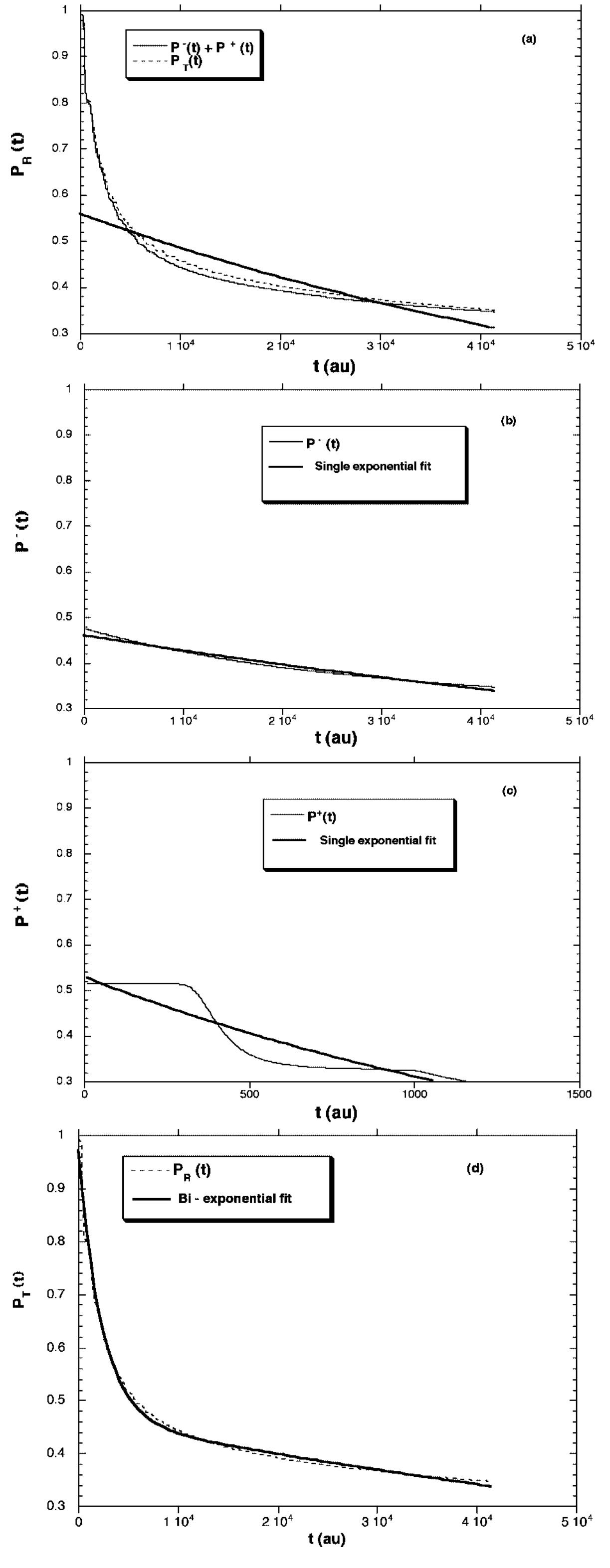


Figure 2. Evolution of the reactant probability density, $P_R(t)$, in the potential energy shown in Figure 1(a) and an initial displacement of $\delta x = 0.25$ Å: (a) the exact total reactant probability density (dashed line) and the approximate probability density (thin solid line). The thicker solid line shows the best single exponential fit of $P_{R,T}(t)$. (b) The evolution of the below the barrier probability density, $P_R^-(t)$ (thin solid line), and the best single exponential fit (thick solid line). (c) Evolution of the above the barrier probability density, that is, $P_R^+(t)$ (thin solid line), and the best exponential fit (thick solid line). (d) The biexponential fit to the reactant probability density, $P_R(t)$ (thick solid line), and the exact total reactant probability density (dashed line).

ponential decay. To explore the validity of this claim we fit the decay of the total reactant probability, $P_{R,T}(t)$, to a biexponential of the form

$$P_{R,T}(t) = a_0 \cdot e^{-a_1 t} + b_0 \cdot e^{-b_1 t} \qquad (35)$$

using the Levenberg−Marquardt algorithm.[22] In performing this fit, we did *not* use initial guesses that reflect the fact that $a_0 = P_R^-(0)$ and $b_0 = P_R^+(0)$. This is because an experimentalist would not know this information; all they know is the observed total decay of the reactant concentration in time. In all calculations we use the initial guesses for the fit parameters of $a_0 = a_1 = b_0 = b_1 = 0.1$. In Figure 2d, we show the biexponential fit to the $P_{R,T}(t)$ in Figure 2a. The solid line is the fit, and the dashed line is the decay of total reactant probability density. As shown by Figure 2d, the time profile $P_{R,T}(t)$ is fit extremely well with a biexponential decay. In this fit, the parameters in eq 35 were as follows: $a_0 = 0.5$ au, $a_1 = 7.7 \times 10^{-6}$ au, $b_0 = 0.5$ au, $b_1 = 4.3 \times 10^{-4}$ au. A comparison of eqs 34 and 35 shows that $\{a_1,b_1\} = \{\kappa^-,\kappa^+\}$, and $\{a_0,b_0\} = \{P_R^-(0),P_R^+(0)\}$. As shown in Figure 2b, the tunneling rate constant was found to be $\kappa^- = 7.5 \times 10^{-6}$ au, which compares very favorably to the biexponential fit parameter, $a_1 = 7.7 \times 10^{-6}$ au. The over the barrier rate constant found from the single exponential fit in Figure 2c was $\kappa^+ = 5.3 \times 10^{-4}$ au, which compares very favorably to the biexponential fit parameter, $b_1 = 4.3 \times 10^{-4}$ au. Finally, as described above, for the case represented by Figures (2) the initial probability densities were $\{P_R^-(0),P_R^+(0)\} = \{0.5,0.5\}$, which compares very favorably to the fit parameter values of $\{a_0,b_0\} = \{0.5,0.5\}$.

In Figure 3, we show results for the same case as Figure 2, but now the initial wavepacket is displaced by an amount $\delta x = 0.10$ Å from equilibrium at $t = 0$. With this initial displacement of the wavepacket, $P_R^-(t = 0) = 0.935$ and $P_R^+(t = 0) = 0.065$. Thus, in this case only ~5% of the initial wavepacket contains energy components above the barrier. In Figure 3a, we show the decay of the total reactant probability density in time with the dashed line. The solid line is the best biexponential fit to $P_{R,T}(t)$. The parameters from the fit lead to the values for the two rate constants of $\kappa^- = 1.5 \times 10^{-6}$ au for the tunneling rate and $\kappa^+ = 2.1 \times 10^{-4}$ au for the above the barrier rate. In Figure 3b, we show the time evolution of the $P_R^-(t)$ probability density with a solid line. The dashed line in Figure 3b is the best single exponential fit which leads to another value for the tunneling rate constant: $\kappa^- = 1.8 \times 10^{-6}$ au. Again, an experimentalist can only observe the data reflected in Figure 3a, that is, the time evolution of $P_{R,T}(t)$. From the observable data along with a biexponential fit the experimentalist would conclude that the tunneling rate is $\kappa^- = 1.5 \times 10^{-6}$ au, which is in very good agreement to the actual rate of tunneling from the unobservable, but theoretically calculated value of $\kappa^- = 1.8 \times 10^{-6}$ au. In Figure 3c, we show the time evolution of the above the barrier probability density, $P_R^+(t)$ (thin line). The thick solid line is the best single exponential fit to $P_R^+(t)$, which leads to another value of κ^+ of $\kappa^+ = 1.6 \times 10^{-4}$ au, which compares favorably to the rate extracted from the experimentally realizable data of $\kappa^+ = 2.1 \times 10^{-4}$ au.

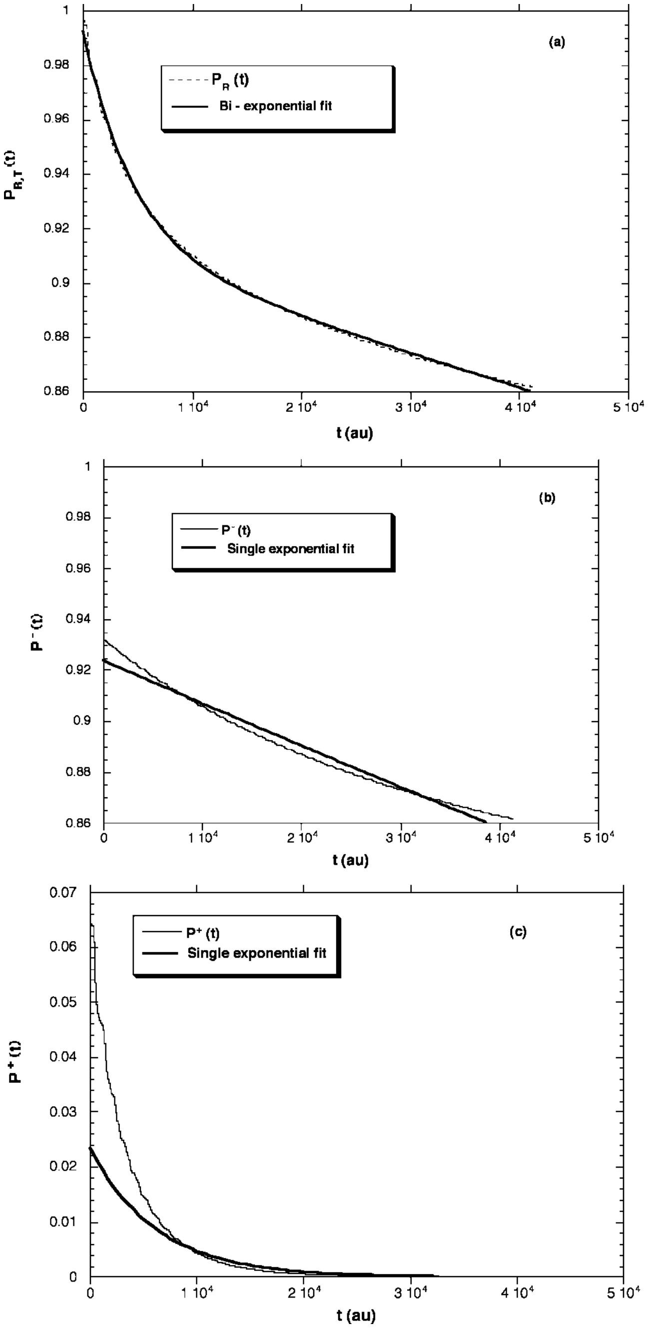


Figure 3. The evolution of the reactant probability density, $P_R(t)$ in the potential energy shown in Figure 1a and an initial displacement of $\delta x = 0.10$ Å: (a) the exact total reactant probability density (dashed line), and the biexponential fit (solid line). (b) The evolution of the below the barrier probability density, $P_R^-(t)$ (thin solid line) and the best single exponential fit (thick solid line). (c) Evolution of the above the barrier probability density, that is, $P_R^+(t)$ (thin solid line) and the best exponential fit (thick solid line).

Next, we explore if the claims above hold for systems with a larger barrier between the reactants and products. In Figures 4 and 5, we show results for a system with potential parameters of $\omega_R = \omega_P = 2000$ cm^{-1} and a nonadiabatic coupling strength of $g_0 = 150$ kJ mol^{-1}. This leads to a double well potential with a barrier height of $E_A = 125$ kJ mol^{-1}. In this case 12 eigenstates are below the barrier, and as described above, they come in pairs of two. In Figure 4,

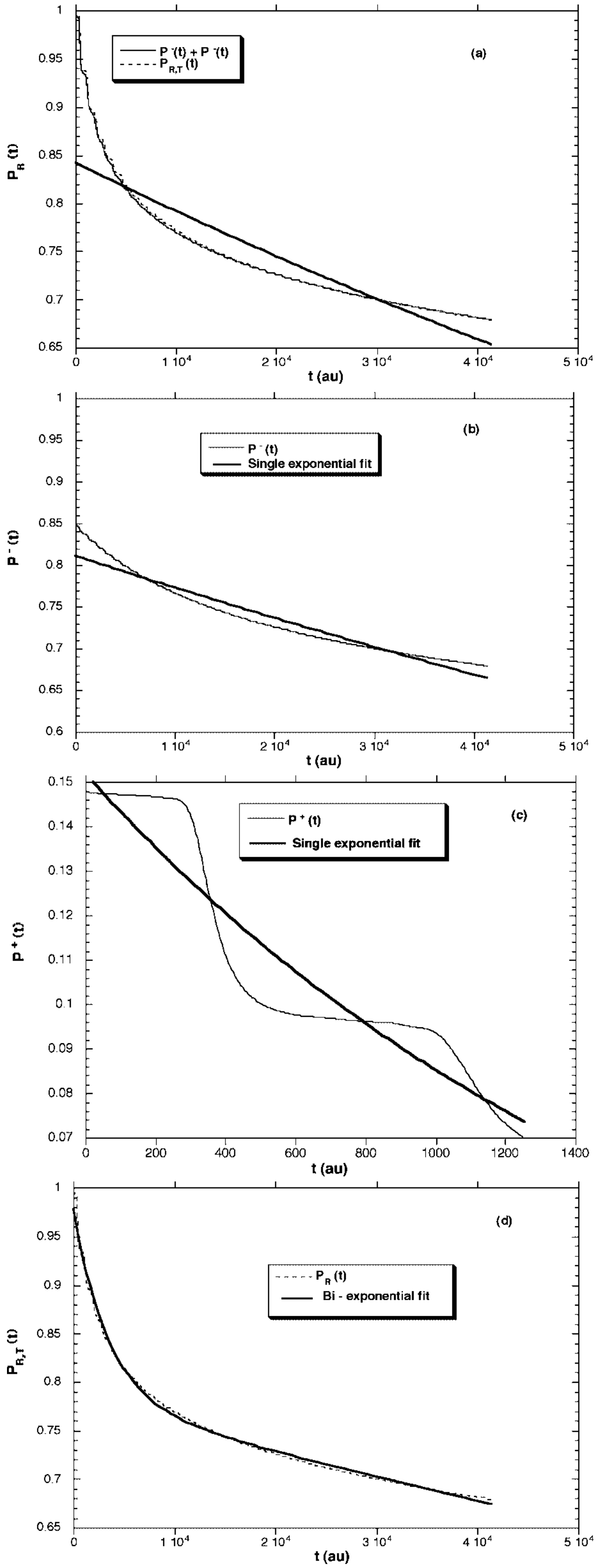


Figure 4. Evolution of the reactant probability density, $P_R(t)$, in a double-well potential with a barrier height of 125 kJ mol^{-1} and an initial displacement of $\delta x = 0.30$ Å: (a) the exact total reactant probability density (dashed line), the approximate probability density (thin solid line). The thicker solid line shows the best single exponential fit of $P_{R,T}(t)$. (b) The evolution of the below the barrier probability density, $P_R^-(t)$ (thin solid line) and the best single exponential fit (thick solid line). (c) Evolution of the above the barrier probability density, that is, $P_R^+(t)$ (thin solid line) and the best exponential fit (thick solid line). (d) The biexponential fit to the reactant probability density, $P_R(t)$, (solid line) and the exact total reactant probability density (dashed line).

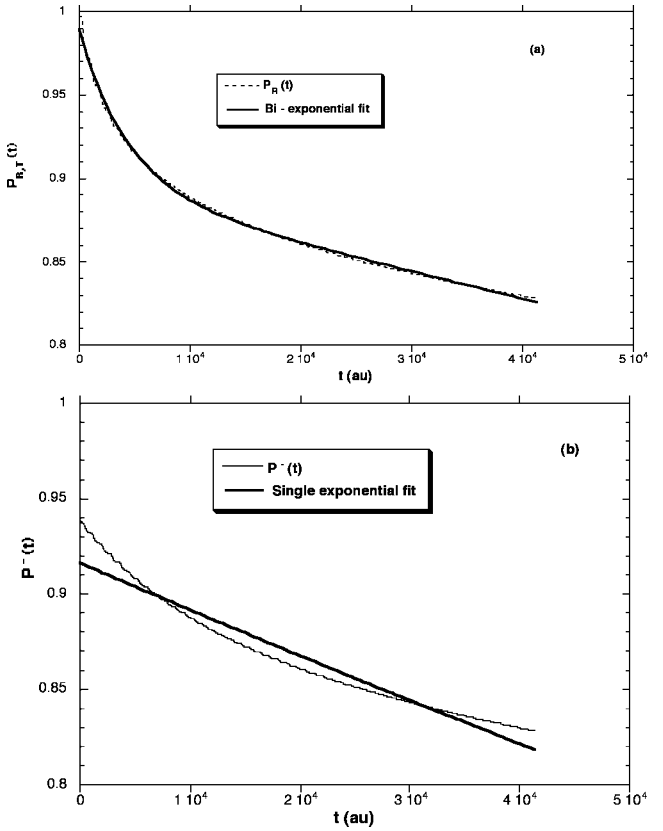


Figure 5. Evolution of the reactant probability density, $P_{\mathrm{R}}(t)$, in a double-well potential with a barrier height of 125 kJ mol^{-1} and an initial displacement of $\delta x = 0.25$ Å: (a) the exact total reactant probability density (dashed line) and the biexponential fit (solid line). (b) The evolution of the below the barrier probability density, $P_{\mathrm{R}}^{-}(t)$ (thin solid line) and the best single exponential fit (thick solid line).

we show results when the initial reactant wavepacket is displaced by an amount $\delta x = 0.3$ Å from equilibrium at $t = 0$. With this displacement and this potential, the initial below and above the barrier probability densities are $P_{\mathrm{R}}^{-}(t = 0) = 0.85$ and $P_{\mathrm{R}}^{+}(t = 0) = 0.15$, respectively. In Figure 4a, we show the evolution of the total reactant probability density, $P_{\mathrm{R,T}}(t)$. The thin solid line is approximation to the total reactant probability density obtained when we separate it into it's $P_{\mathrm{R}}^{-}(t)$ and $P_{\mathrm{R}}^{+}(t)$ components and the dashed line in Figure 4a is the exact total reactant probability density. Because the thin solid and dashed lines are in close agreement, the dynamics of the system is separable into below and above the barrier probability densities even with a larger barrier. The thicker solid line in Figure 4a shows the best single exponential fit to the data, which is poor. In Figure 4b, we show the evolution of the below the barrier component of the wavepacket, that is, $P_{\mathrm{R}}^{-}(t)$ (thin line). The thick line in Figure 4b shows the best fit of $P_{\mathrm{R}}^{-}(t)$ to a single, decaying, exponential. Again, the tunneling dynamics alone fits a single exponential decay much more closely than that of the whole time evolution of the probability density. We extract out a rate constant for tunneling from the exponential fit in Figure 4b and find, in this case, $\kappa^{-} = 4.8 \times 10^{-6}$ au. The thin line in Figure 4c is the evolution of the above the barrier probability density, that is, $P_{\mathrm{R}}^{+}(t)$. The thick line in Figure 4c shows the best single decaying exponential fit to $P_{\mathrm{R}}^{+}(t)$. We extract out a rate constant for above the barrier product production from the exponential fit in Figure 4c and find, in this case, $\kappa^{+} = 4.3 \times 10^{-4}$ au. In Figure 4d, we show the decay of the total reactant probability density in time with the dashed line. The solid line is the best biexponential fit to $P_{\mathrm{R,T}}(t)$. The parameters from the fit lead to the values for the two rate constants of $\kappa^{-} = 3.6 \times 10^{-6}$ au for the tunneling rate and $\kappa^{+} = 3.0 \times 10^{-4}$ au for the above the barrier rate. Again, these values represent information that can be extracted from the experimentally observable total rate of decay of reactant concentration. These values compare favorably with calculated separate tunneling and above the barrier rates of $\kappa^{-} = 4.8 \times 10^{-6}$ au and $\kappa^{+} = 4.3 \times 10^{-4}$ au, respectively.

In Figure 5, we show results when the initial reactant wavepacket is displaced by an amount $\delta x = 0.25$ Å from equilibrium at $t = 0$. With this displacement and this potential, the initial below and above the barrier probability densities are $P_{\mathrm{R}}^{-}(t = 0) = 0.95$ and $P_{\mathrm{R}}^{+}(t = 0) = 0.05$, respectively. In Figure 5a, we show the evolution of the total reactant probability density, $P_{\mathrm{R,T}}(t)$. The dashed line is the total reactant probability density and the solid line best biexponential fit to $P_{\mathrm{R,T}}(t)$. The parameters from the fit lead to the values for the two rate constants of $\kappa^{-} = 2.2 \times 10^{-6}$ au for the tunneling rate and $\kappa^{+} = 2.3 \times 10^{-4}$ au for the above the barrier rate. The thin line in Figure 5b shows the evolution of the below the barrier component of the wavepacket, that is, $P_{\mathrm{R}}^{-}(t)$. The thicker solid line in Figure 4b shows the best fit of $P_{\mathrm{R}}^{-}(t)$ to a single, decaying, exponential. We extract out a rate constant for tunneling from the exponential fit in Figure 5b and find, in this case, $\kappa^{-} = 2.6 \times 10^{-6}$ au. We extract out a rate constant for above the barrier product production from the exponential fit to $P_{\mathrm{R}}^{+}(t)$ and find, in this case, $\kappa^{+} = 2.2 \times 10^{-4}$ au.

All the results described thus far were for symmetric double-well potentials, that is, potentials that result from diabatic states with the same frequencies, $\omega_{\mathrm{R}} = \omega_{\mathrm{P}}$. Now we see if the quantum dynamics is separable into tunneling and above the barrier components for asymmetric potentials. In Figure 6, we show results for a double well potential with $\omega_{\mathrm{R}} = 1500$ and 2000 cm^{-1} with nonadiabatic coupling strength of $g_0 = 150$ kJ mol^{-1}. In this case the barrier has a value of $E_{\mathrm{A}} = 45$ kJ mol^{-1}. We consider the dynamics of an initial reactant wavepacket that is displaced by an amount $\delta x = 0.25$ Å at $t = 0$. With this displacement and this potential, the initial below and above the barrier probability densities are $P_{\mathrm{R}}^{-}(t = 0) = 0.75$ and $P_{\mathrm{R}}^{+}(t = 0) = 0.25$, respectively. In Figure 6a, we show the evolution of the total reactant probability density, $P_{\mathrm{R,T}}(t)$. Again, the solid line is approximation to the total reactant probability density obtained when we separate it into it is $P_{\mathrm{R}}^{-}(t)$ and $P_{\mathrm{R}}^{+}(t)$components and the dashed line in Figure 6a is the exact total reactant probability density. Since the solid and dashed lines are in close agreement, the dynamics of the system is separable into below and above the barrier probability densities even for asymmetric double-well potentials. The solid line in Figure 6b shows the evolution of the below the barrier component of the wavepacket, that is, $P_{\mathrm{R}}^{-}(t)$. The dashed line in Figure 6b shows the best fit of $P_{\mathrm{R}}^{-}(t)$ to a single,

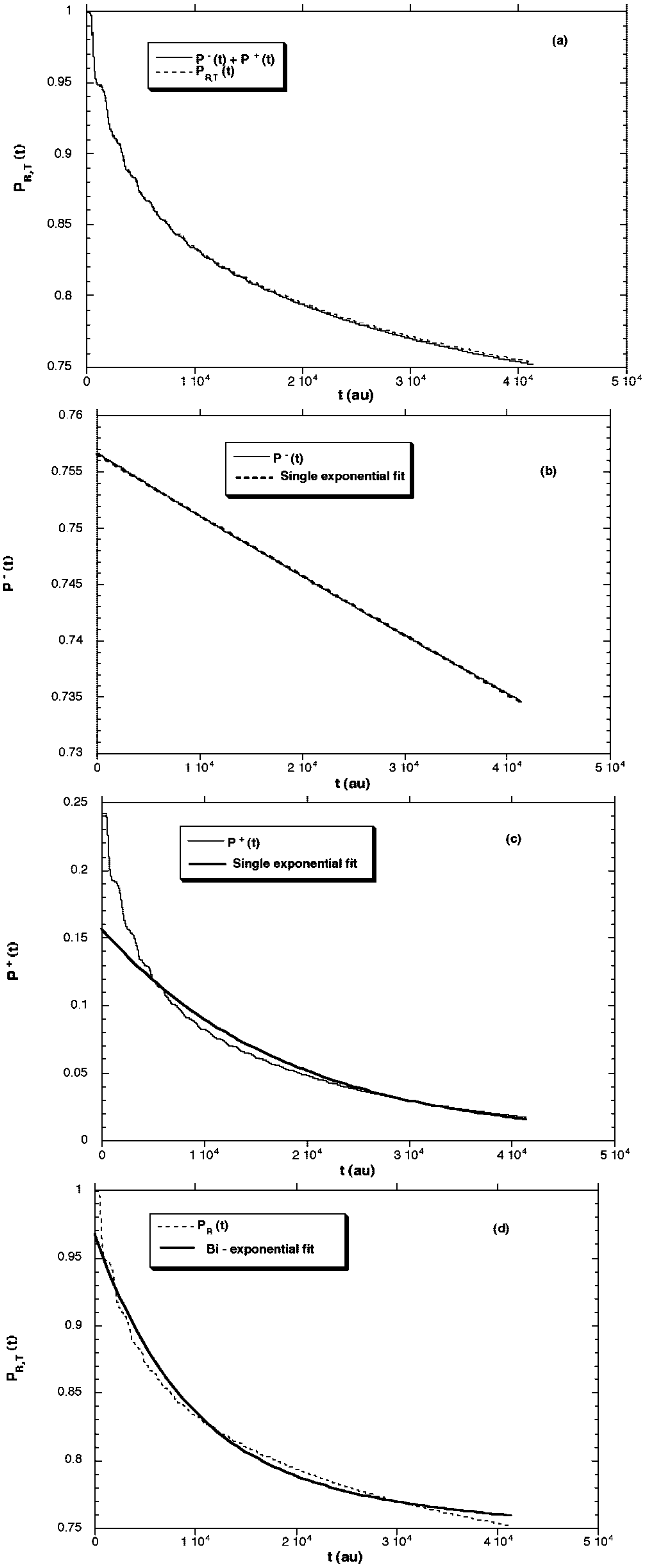


Figure 6. Evolution of the reactant probability density, $P_{\rm R}(t)$ in the asymmetric double-well potential as discussed in the text and an initial displacement of $\delta x = 0.25$ Å: (a) the exact total reactant probability density (dashed line) and the approximate probability density (solid line). (b) The evolution of the below the barrier probability density, $P_{\rm R}^{-}(t)$ (thin solid line) and the best single exponential fit (thick solid line). (c) Evolution of the above the barrier probability density, that is, $P_{\rm R}^{+}(t)$ (thin solid line) and the best exponential fit (thick solid line). (d) The biexponential fit to the reactant probability density, $P_{\rm R}(t)$, (solid line) and the exact total reactant probability density (dashed line).

decaying, exponential. Again, the tunneling dynamics is well represented by a single exponential. We extract out a rate constant for tunneling from the exponential fit in Figure 6b and find, in this case, $\kappa^- = 7 \times 10^{-7}$ au. This is an order of magnitude smaller than the tunneling rate constant for the symmetric potential with the same barrier, that is, $E_{\rm A} = 45$ kJ mol^{-1} discussed in Figure 2. The thin line in Figure 6c shows the time evolution of the above the barrier probability density, $P_{\rm R}^{+}(t)$. The thick solid line in Figure 6c shows the best single decaying exponential fit to $P_{\rm R}^{+}(t)$. Again, the single exponential fit of $P_{\rm R}^{+}(t)$ is not as good as that of $P_{\rm R}^{-}(t)$ the fit does follow the overall decay trend pretty well. We extract out a rate constant for above the barrier product production from the exponential fit in Figure 6c and find, in this case, $\kappa^+ = 6.5 \times 10^{-5}$ au. In Figure 6d, we show the decay of the total reactant probability density in time with the dashed line. The solid line is the best biexponential fit to $P_{\rm R,T}(t)$. The parameters from the fit lead to the values for the two rate constants of $\kappa^- = 6 \times 10^{-7}$ au for the tunneling rate and $\kappa^+ = 1 \times 10^{-4}$ au for the above the barrier rate. Again, these would be the rates that could be extracted from experimentally measured data. The rate constant for tunneling extracted in this way compares very well to the unobservable, but calculable, tunneling rate in Figure 6b of $\kappa^- = 7 \times 10^{-7}$ au.

Now we explore the isotope effect on the tunneling rate constant, κ^-. We study the same potential as in Figures 4. The potential parameters of $\omega_{\rm R} = \omega_{\rm P} = 2000$ cm^{-1} and a nonadiabatic coupling strength of $g_0 = 150$ kJ mol^{-1} leading to a barrier height of $E_{\rm A} = 125$ kJ mol^{-1}.

Now we change the mass of the system to that of a D atom, $m = 2\ m_{\rm H}$. When the system had the mass of the H atom there were 12 eigenstates below the barrier, now with the D atom there are 16 eigenstates below the barrier. This is consistent with the decrease in spacing between the energy levels with increasing mass. We study the quantum dynamics of the system when the initial wavepacket is displaced an amount $\delta x = 0.25$ Å at $t = 0$. In Figure 7a, we show the time evolution of the total reactant probability density, $P_{\rm R,T}(t)$ with the dashed line and the best biexponential fit with the solid line. The fit parameters lead to values of $\kappa^- = 5.4 \times 10^{-7}$ au and $\kappa^+ = 9.6 \times 10^{-5}$ au. Note that the tunneling rate constant was $\kappa^- = 2.2 \times 10^{-6}$ au for the same potential and initial conditions for the H atom. As expected the rate constant is significantly smaller, by nearly an order of magnitude, for tunneling of the heavier D atom. In Figure 7b, we show the time evolution of the below the barrier component of the wavepacket, that is, $P_{\rm R}^{-}(t)$. The thicker solid line in Figure 7b shows the best fit of $P_{\rm R}^{-}(t)$ to a single, decaying, exponential. Again, the tunneling dynamics is well represented by a single exponential. We extract out a rate constant for tunneling from the exponential fit in Figure 6b and find, in this case, $\kappa^- = 5.4 \times 10^{-7}$ au. This rate constant closely agrees to the experimentally observable value obtained from the biexponential fit. Although, not shown, we fit the above the barrier probability density to a single exponential and found that it agrees with that of the biexponential fit, that is, $\kappa^+ = 9.6 \times 10^{-5}$, within 5%.

We now say a little more about the assumption that the two dynamic events, that is, tunneling and over the barrier

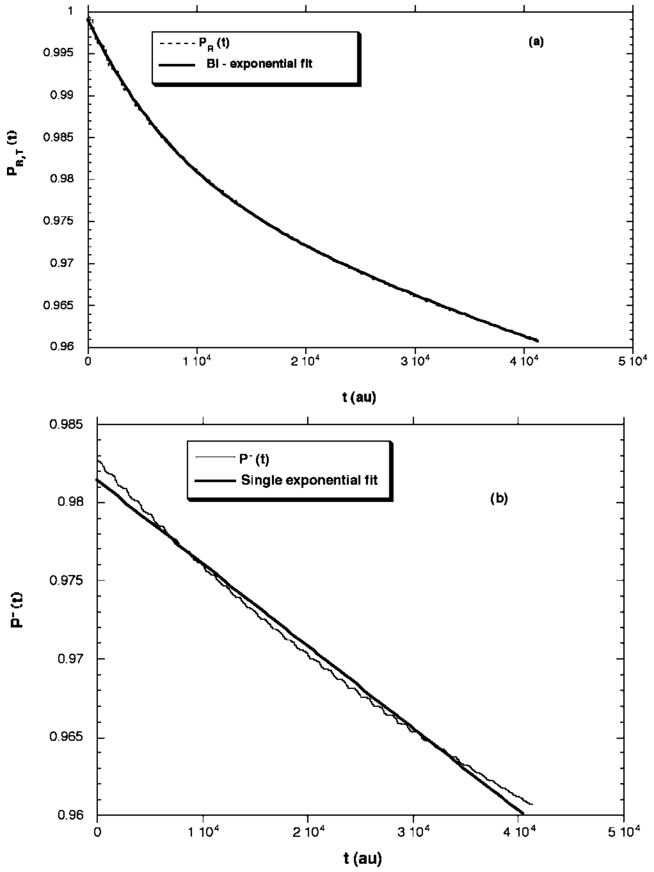


Figure 7. Evolution of the reactant probability density, $P_R(t)$ for deuterium in a double-well potential with a barrier height of 125 kJ mol^{-1} and an initial displacement of $\delta x = 0.25$ Å: (a) the exact total reactant probability density (dashed line) and the biexponential fit (solid line). (b) The evolution of the below the barrier probability density, $P_R^-(t)$ (thin solid line) and the best single exponential fit (thick solid line).

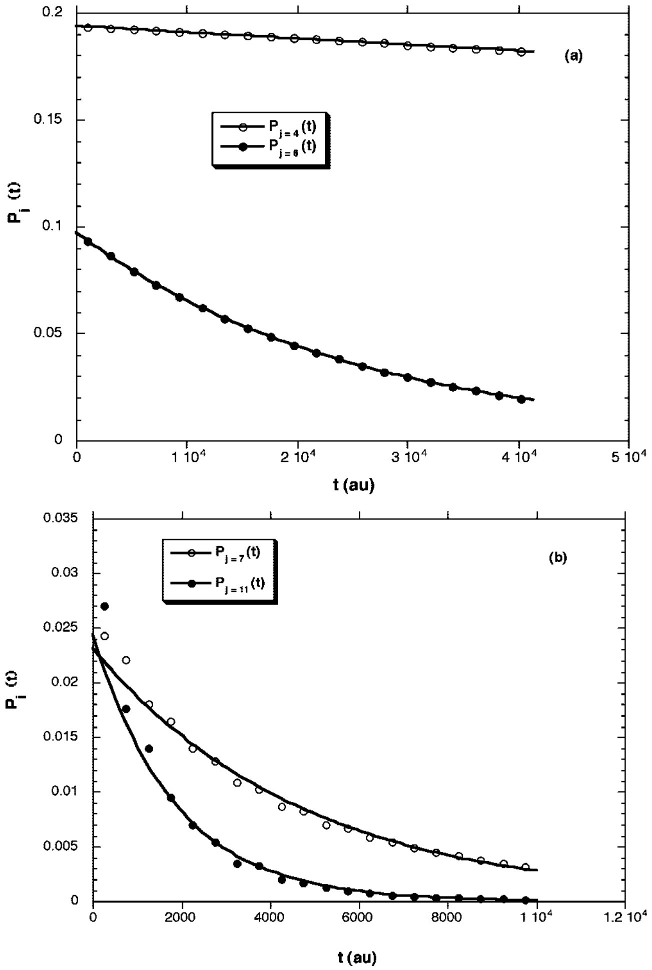


Figure 8. Evolution of the individual eigenstates as discussed in the text for a barrier of 45 kJ mol^{-1} and an initial displacement of $\delta x = 0.25$ Å: (a) the evolution of the $j = 4$ (open circles) and $j = 6$ (solid circles) below the barrier eigenstates. The solid lines are the best, single, exponential fits to each. (b) the evolution of the $j = 7$ (open circles) and $j = 11$ (solid circles) above the barrier eigenstates. The solid lines are the best, single, exponential fits to each.

production, can be treated as first order events. In Figure 8, we show the decay of some of the individual eigenstates for a barrier height of $E_A = 45$ kJ mol^{-1} and initial wavepacket displacement of $\delta x = 0.25$ Å. [The same system as that described by Figures 2.] As described previously, there are 6 eigenstates below the barrier and of these 6 only $j = 2, 4, 6$ contribute appreciably to the initial wavepacket superposition. In Figure 8a, we show the decay of the $j = 4$ and 6 eigenstates with the open and solid circles, respectively. The solid line shows a single, decaying, exponential fit to each. As shown by Figure 8a, each eigenstate decays exponentially and the aggregate decay of the below the barrier eigenstates, that is, the function $P_R^-(t)$, should be more accurately represented by the weighted sum of decaying exponentials as shown below.

$$P_R^-(t) = \sum_{j=1}^{m} |c_{j,0}|^2 \cdot e^{-\kappa_j t} \tag{35a}$$

In eq 35, the $\{c_{j,0}\}$are obtained from eq 20b, and the $\{\kappa_j\}$ are the decay constants for the eigenstates. In Figure 8a, the decay constants for the $j = 4$ and 6 eigenstates are $\kappa_4 = 1.6 \times 10^{-6}$ au and $\kappa_6 = 3.9 \times 10^{-5}$ au. Although not shown in Figure 8a, we have found the decay constant for the $j = 2$ eigenstate is $\kappa_2 = 1.8 \times 10^{-8}$ au. We define a weighted average of these below the barrier decay constants as

$$\langle \kappa^- \rangle = \sum_{j=1}^{m} |c_{j,0}|^2 \cdot \kappa_j \tag{36a}$$

This weighted average yields an average decay for the below the barrier eigenstates as $\langle \kappa^- \rangle = 5.5 \times 10^{-6}$ au. This compares favorably to the single exponential fit to $P_R^-(t)$, which had an exponent of $\kappa^- = 7.5 \times 10^{-6}$ au.

In Figure 8b, we show the decay of $j = 7$ and 11, both above the barrier eigenstates with the open and solid circles, respectively. The solid line shows a single, decaying, exponential fit to each. Again, both of the individual eigenstates decay exponentially. The $j = 7$ eigenstate is the first above the barrier eigenstate, and it decays with a decay constant of $\kappa_7 = 2.1 \times 10^{-4}$ au. This is a very revealing result. The energy separation between the last below the barrier eigenstate ($j = 6$) and the first above the barrier eigenstate ($j = 7$) is only 8 kJ mol$^{-1,}$ but the decay constant of the above the barrier, $j = 7$, eigenstate is nearly 2 orders of magnitude larger. For comparison, the $j = 7$ and 8 eigenstates are also separated by about 8 kJ mol^{-1} in energy, but the decay constant for the $j = 8$ eigenstate is $\kappa_8 = 2.3 \times 10^{-4}$ au, that is, nearly the same as the $j = 7$ eigenstate. Further, the $j = 11$ eigenstate, shown via the solid circles in Figure 8b, has a decay constant of $\kappa_{11} = 5.4 \times 10^{-4}$ au, only twice as large as the $j = 7$ eigenstate, even though it is nearly 40 kJ mol^{-1} higher in energy. In fact, all the above the barrier eigenstates between $j = 7$ and $j = 15$, have decay constants that range between $\kappa = 2.1 \times 10^{-4}$ au and 8×10^{-4} au. In other words, the above the barrier eigenstates decay on almost identical time scales and the aggregate decay, that is, $P_R^+(t)$, are well quantified by a single decay constant that can be obtained from the weighted average

$$\langle \kappa^+ \rangle = \sum_{j=m+1} |c_{j,0}|^2 \cdot \kappa_j \tag{36b}$$

The results of the analysis accompanying Figures 8 verify the conjecture that there are really two dynamic events that can lead to H atom transfer, tunneling and above the barrier production, and these events are well separated in time. The results also explain how a comparison of the time scales of these two events can be well quantified by treating each as a single, first-order event. Each above the barrier eigenstate is well represented by a decaying exponential and all decay with nearly the same rate constant. Thus, the aggregate decay of the above the barrier components given by the function $P_R^+(t)$ can be considered to decay with an overall decay constant given by the weighted average in eq 36b. The below the barrier eigenstates, the ones responsible for tunneling, also, individually, decay exponentially although there is a wide disparity between the decay rates among the below the barrier eigenstates. More importantly, though, all of the below the barrier eigenstates decay much slower than even the least energetic above the barrier eigenstate. This demonstrates that the main conjecture of the present work is correct: There are two dynamic events that can accompany H atom transfer; they operate on different time scales, and it is possible to quantify the rates of each of these two events.

The results also show that one must be very careful when assigning a time scale, a rate constant, or a portion of a rate constant resulting from product production through tunneling. The results show that even though tunneling occurs on a much different time scale than above the barrier production of product, there is a wide disparity between the rates of tunneling between the individual, below the barrier, eigenstates. We suggest a way to ascribe a rate constant for tunneling is to consider the tunneling as an overall first-order process that represents reactant decay with a rate constant, κ^-, given by the weighted average in eq 36a, although other definitions are possible.

IV. Conclusion

In this work we described a rationale for defining and determining a "tunneling rate" from experimental observables of an H atom transfer reaction. The key to this view is the energetic decomposition of an initial H atom wavepacket into above and below the barrier components. This wavepacket decomposition was defined in such a way as to satisfy two criteria. The first criterion is that these two wavepacket components must evolve independently of one another in time. The second criterion is that a time-evolving probability density can be assigned, separately, to each component. We have shown that both of these criteria can be satisfied if the two wavepacket components are formed by superpositions of the eigenstates of the potential energy under which they evolve.

We stress here that the major conclusion of this work is *not* that one can calculate a tunneling rate constant from a simulation using the wavepacket decomposition as described above. The major conclusion of this work is that such a simple, and perhaps obvious way to decompose an initial reactant wavepacket, leads to a rational way for an experimentalist to interpret observables in a kinetics experiment on an H atom transfer reaction. We have shown that a biexponential fit to the overall decay of reactant (the observable) can be used to extract out a rate constant for tunneling alone. The simulations merely serve to show that the less negative of the two exponents in the bi-exponential fit is the one that describes the tunneling rate, and the more negative describes over the barrier product production. We have shown that two, pre-exponential terms in the fit give information about the initial, nonequilibrium, state of the system. Again, the simulations merely serve to show that the pre-exponential term in front of the less negative exponential can be interpreted as the percent of the initial wavepacket that has energy less than the barrier and the other pre-exponential term gives the percent of the initial wavepacket above the barrier. As the pre-exponential term in front of the less negative exponential decreases, there is less of the initial H atom wavepacket below the barrier, and this can be interpreted as meaning that the system is prepared in a state farther from equilibrium at $t = 0$. Thus, an experimentalist can gain from a single experimental observable, the overall decay of reactant concentration, kinetic, *and* structural information about the H atom transfer reaction.

We have also found that, via simulations, that the energetic decomposition of the initial reactant wavepacket leads to a bi-exponential decay of reactant concentration for symmetric and asymmetric potentials and for differing isotopes of the H atom.

Another conclusion from this work is that a short-time exponential fit of reactant concentration decay has very little to do with the rate of tunneling. The initial decay of reactant concentration is dominated by the faster dynamic event, over the barrier product production.

Finally, we discuss how the coupling between the H atom reaction coordinate, x, and other degrees of freedom can effect the present results and conclusions. We will also sketch out a way toward treating such multidimensional systems.

We classify two types of coupling to the H atom reaction coordinate: (1) direct coupling and (2) bilinear coupling to vibrational modes. By direct coupling, we mean that the single-dimensional potential energy, $V^*(x)$, would be replaced by a multidimensional potential of the form, $V^*(x,\vec{y})$, where $\{\vec{y}\}$are some small set of other degrees of freedom that directly effect the evolution of the H atom wavepacket. For example, if each of the diabatic surfaces $V_P(x)$ and $V_R(x)$ were two-dimensional, that is,

$$V_R(x,y) = \frac{m\omega_R^2}{2}[(x-y) - (x_R^0 - y_R^0)]^2 \qquad (37)$$

the resulting double-well adiabatic potential would be a nontrivially coupled two-dimensional potential. One could envision that such direct coupling between the H atom coordinate, x, and the y degree of freedom would have a large effect on both tunneling and above the barrier product production. In fact, the effect might be large enough to make the time scales of the two dynamic events become too close to distinguish. This, by no means, should detract from the relevance of the present work. In fact, results of the present work would indicate that if there were *not* an observable separation of time scales in the H atom transfer event, then other effects such as the strong coupling mentioned above could be present.

The other type of coupling that could be present is bilinear coupling of the H atom transfer coordinate to a set of vibrational modes, $\{\vec{\eta}\}$. In this case the double-well potential, $V^*(x)$, would be replaced by a potential energy of the form,

$$V^*(x,\vec{\eta}) = V^*(x) + \frac{1}{2}\sum_{k=1}\omega_k^2\cdot(\eta_k - c_k\cdot x)^2 \qquad (38)$$

where ω_k and c_k are frequencies and coupling of the kth mode to the H atom coordinate, respectively. The major effect of such coupling would be to serve as a sink of energy for the H atom coordinate. If the coupling were strong enough then it might be possible for the above the barrier eigenstates to have decay constants approaching those of the below the barrier eigenstates. Thus, in the strong coupling case, the time scales of tunneling and above the barrier transfer could become too close to distinguish. Again, this should not detract from the relevance of the present work because the absence of two distinguishable time scales in the transfer event would indicate such strong coupling was present.

We now describe how one could generalize the present work to treat the two types of multidimensional systems described above. If the H atom reaction coordinate, x, was directly coupled to one or two degrees, of freedom, $\vec{y}$, then the generalization of the present work would be straightforward. For example, if the H atom reaction coordinate were coupled to a single degree of freedom, one would calculate the eigenenergies, $\{E_n\}$, and eigenfunctions, $\{\psi\}$, of the two-dimensional potential, $V^*(x,y)$. There would now be many more eigenstates above and below the barrier, but as in the present work, we would not have to propagate each one separately in time, just propagate two groupings: $\Psi^-(x,y,t)$ the group of below the barrier eigenstates and $\Psi^+(x,y,t)$ the group of above the barrier eigenstates. The propagation of each could be accomplished by solving the TDSE on a two-dimensional spatial grid instead of the single-dimensional grid used in the single-dimensional case.

If the H atom reaction coordinate, x, was bilinearly coupled to a set of vibrational modes, $\{\vec{\eta}\}$, the procedure described above could not be used. The reason is that, presently, there is an upper limit of 4−5 degrees of freedom for solving the TDSE on a spatial grid. The simplest and least accurate way to treat this type of multidimensional system would be to treat the oscillators classically. In this case, the above the below and above the barrier eigenstates,$\Psi^-(x,t)$ and $\Psi^+(x,t)$, would still be single dimensional, but they would be propagated via time-dependent potential, $V^*(x,\vec{\eta}_t;t)$, that depended parametrically on the instantaneous positions, $\{\eta_t\}$, of each vibrational mode. The trajectory of each vibrational mode would be determined via the Verlet algorithm, where each vibrational mode experienced a force, $F_j = -\sum_{k=1}\omega_k^2\cdot(\eta_k - c_k\cdot\langle x\rangle_t)$, that depended on the expectation value, $\langle x\rangle_t$, of the H atom position.

Acknowledgment. Financial support of this work was provided by the Henry Dreyfus Teacher-Scholar Awards Program.

References

(1) Klinman, J. P.; Liang, Z-X. *Curr. Opin. Struct. Biol.* **2004**, *14*, 648.

(2) Borman, S. *Chem. Eng. News.* **2004**, *82*, 35.

(3) Warshel, A.; Papazyan, A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13665.

(4) Gertlt, J. A.; Kreevoy, M. M.; Cleland, W. W.; Frey, P. A. *Chem. Biol.* **1997**, *4*, 259.

(5) Bruice, C. T. *Acc. Chem. Res.* **2002**, *35*, 1.

(6) Klinman, J. P. *Pure Appl. Chem.* **2003**, *75*, 601.

(7) Mincer, J. S.; Schwartz, S. D. *J. Phys. Chem. B* **2003**, *107*, 366.

(8) Benkovik, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1198.

(9) Hammes-Schiffer, S. *Biochemistry* **2002**, *41*, 13335.

(10) Antoniou, D.; Schwartz, S. D. *J. Chem. Phys. B* **2001**, *105*, 5553.

(11) Poulsen, T. D.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2003**, *107*, 9567.

(12) Alhambra, C.; Corchado, J.; Sanchez, M. L.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326.

(13) Warshel, A.; Chu, Z. T. *J. Chem. Phys.* **1990**, *93*, 4003.

(14) Hammes-Schiffer, S.; Tully, J. C. *J. Phys. Chem.* **1995**, *99*, 5793.

(15) Mincer, J. S.; Schwartz, S. D. *J. Phys. Chem. B* **2003**, *107*, 366.

(16) Hur, S.; Bruice, T. C. *J. Am. Chem. Soc.* **2003**, *125*, 10540.

(17) Hur, S.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12015.

(18) Reddy, S. Y.; Bruice, T. C. *J. Am. Chem. Soc.* **2003**, *126*, 2431.

(19) Kramers, H. A. *Physica* **1940**, *7*, 284.

(20) Hanggi, P.; Talkner, P.; Borkovec, M. *Rev. Mod. Phys.* **1990**, *62*, 251.

(21) Feit, M. D.; Fleck, J. A.; Steiger, A. *J. Comput. Phys.* **1982**, *47*, 412.

(22) Marquardt, D. W. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431.

CT800436U

J. Chem. Theory Comput. **2009,** *5,* 482–490

JCTC Journal of Chemical Theory and Computation

Investigating the Conformational Preferences of Transforming Growth Factor-β Isoforms using Targeted Molecular Dynamics Simulations

Daniel J. Warner,*[,†] Ian C. Paterson,[‡] and Richard B. Sessions[§]

Department of Medicinal Chemistry, AstraZeneca R&D Charnwood, Bakewell Road,
Loughborough, Leicestershire, LE11 5RH, U.K., Department of Oral and
Dental Science, Bristol Dental Hospital, University of Bristol, Lower Maudlin Street,
Bristol, BS1 2LY, U.K., and Department of Biochemistry, School of Medical Sciences,
University of Bristol, University Walk, Clifton, Bristol, BS8 1TD, U.K.

Received November 9, 2008

Abstract: We study the conformational preferences and mechanical properties of two isoforms of the cytokine transforming growth factor-β (-β1 and -β3) with atomistic detail and including the effects of explicit water. Targeted molecular dynamics simulations are used to perturb experimental "closed" conformations of both proteins into an "open" conformation, thus far only observed crystallographically for one of the two isoforms. The artificial restraints imposed by the protocol are later released, allowing the two covalently bound units of each homodimer to relax. Homology models of the two proteins are also constructed using the other as a template; models that are later subjected to the same process of perturbation into the open conformation and relaxation. On release, both simulations of transforming growth factor-β1 show a tendency to snap back toward the closed conformation, while those of transforming growth factor-β3 remain open for the remainder of the simulation, apparently consistent with measurements from a variety of experimental sources. Duplication of the simulations affords confidence that this observation reflects a genuine effect of the sequence, as opposed to an artifact of the conformations selected at the outset. The study provides a previously unseen level of detail, describing the structural and dynamic behavior of these proteins in solution, and brings us a step closer to understanding the complex relationship between sequence, structure, and signaling in this family of cytokines.

Introduction

TGF-β Signaling. Transforming growth factor-β (TGF-β) describes a family of multifunctional cytokines responsible for regulating a wide range of cellular processes including growth and differentiation.[1,2] Three mammalian isoforms of TGF-β have been identified (TGF-β1, -β2, and -β3), which exhibit a high degree of sequence identity (in the range of 70−76%). Despite this similarity, the differences that exist across the three isoforms are strictly conserved across species, reflecting each isoform's distinct biological properties, and hence their resistance to evolutionary pressures.[3,4] The type-I and type-II signaling receptors for TGF-β (TβR-I and TβR-II) each contain an extracellular ligand-binding, transmembrane, and intracellular Ser/Thr kinase domain.[5] Initially, TGF-β binds to the extracellular domain of TβR-II, forming a complex that is then able to recruit the second, type-I receptor.[6] Association of these two extracellular domains by TGF-β prompts transphorylation of the TβR-I kinase domain, which instigates a series of downstream signaling events.[5]

Structural Biology. X-ray crystal structures have been determined for both TGF-β2[7,8] and -β3[9] in isolation, whereas the structural information for TGF-β1 is derived from NMR spectroscopy.[10,11] These studies show all three

---

[*] To whom correspondence should be addressed. E-mail: dan.warner@astrazeneca.com.

[†] Department of Medicinal Chemistry, AstraZeneca R&D Charnwood.

[‡] Bristol Dental Hospital, University of Bristol.

[§] School of Medical Sciences, University of Bristol.

10.1021/ct800483q CCC: $40.75

isoforms to associate in the form of a homodimer, with the two monomers cross-linked by a single disulphide bond (Cys77−Cys77). The structures are typically described as resembling a pair of hands clasped together, with a collection of β-strands representing the fingers, while α-helices H1 and H3 from each monomer represent the thumbs and wrists, respectively.

Despite such high similarity between TGF-β1 and -β3 in terms of their sequences and experimentally determined conformations, their structures and dynamic behavior in solution show marked differences. ^{15}N NMR relaxation studies indicate that the sections of the sequence encompassing α-helices H1 and H3 display much greater mobility in TGF-β3 than they do in -β1.[12] Notably, α-helix H3 is made up of the least well conserved residues across the TGF-β isoforms, indicating that this is the region of the protein where the biological properties are most likely to be modulated. The structures of TGF-β in isolation appear to support this concept, in that despite having the same space group and arrangement within each unit cell, the residues comprising H1 have notably higher thermal B-factors in 1TGJ (TGF-β3)[9] than in the two crystallographic structures of TGF-β2.[7,8] The observed flexibility of H3 is a more surprising finding, as in the crystalline state this region represents one of the most rigid parts of the molecule. However, the restraint of this helix in TGF-β3 is thought to be an artifact of the crystal packing because a number of residues are involved in intermolecular interactions in the crystal lattice.[9]

Further evidence for the disordering of α-helices H1 and H3 of TGF-β3 in solution is provided from circular dichroism experiments.[13] This study shows the helical content of TGF-β3 in solution to be as low as 4%, far lower than would be expected from the crystal structures of either this isoform or TGF-β2. The authors conclude by stating that, in contrast to TGF-β2, -β3 has the ability to adopt two distinct conformations. Insight into the possible nature of this second conformation has, at least in part, become available with the publication of the crystal structure of TGF-β3 in complex with the extracellular domain of TβR-II.[14] In this structure, the two monomers undergo significant reorientation with respect to each other, with one subunit effectively rotating by 101° with respect to the other. As might be expected, this “open” structure contains far lower helical content than any of the previously determined “closed” structures of TGF-β in isolation because the α-helices that comprise the thumb and wrist are considerably more mobile. This flexibility brings the thermal B-factors more closely into line with the exaggerated mobilities observed in the NMR experiments, making the open conformation a more reliable template for models of TGF-β3 in solution.

A significant advance in understanding the relationship between structure and function was recently made with the solution, to a resolution of 3.0 Å, the structure of the complete TGF-β:TβR-I:TβR-II complex.[15] Crucially, this complex indicates that TGF-β must adopt a closed conformation to recruit the extracellular domains of its type-I and type-II receptors and thus instigate the downstream signaling cascade. Prior to this work, it had only been possible to construct putative models of this complex,[6,16] based on combination of the TGF-β3:TβR-II complex structure[14] with that of the closely related bone morphogenetic protein 2 (BMP-2) in complex with the extracellular domain of the BMP receptor 1A (BR1A).[17]

Methodology Selection. The striking differences between the TGF-β isoforms, particularly in terms of their structural and dynamic behavior in solution, make this group of proteins compelling subjects for studies by molecular dynamics (MD) simulation. We have used targeted molecular dynamics (TMD)[18,19] to promote the transition from the closed form of TGF-β3 to the open form to produce a model that is more representative of this protein’s behavior in solution.

Our reasons for adopting this approach are essentially 2-fold. First, while the actual number of residues that remain unresolved in the TGF-β3:TβR-II complex structure is fairly low, those sections of the sequence neighboring these residues appear frayed when compared to structures of the ligand alone and feature high crystallographic B-factors.[14] This raises a number of questions for anyone wishing to construct a homology model of the open conformation directly, in that it is unclear how far back to prune these termini to construct the missing fragment, and this potentially requires construction of a loop longer than would be recommended for most modeling packages. The second argument centers on the nature of the fragment to be inserted. The relative orientation of the monomers in the complex structure prohibits the introduction of a wrist helix as observed in the closed conformation, while construction of an extended loop would generate a starting point for a simulation with little chance of regaining this secondary structure, regardless of the shape of the conformational energy landscape.

As a point of comparison, we also performed the same experiment based on an experimental structure of TGF-β1, for which there is currently no evidence of such a second, partially denatured conformation existing. The conformational preferences and secondary structure composition of these two isoforms are presented, and their implications for recognition and signaling are discussed.

Methods

System Preparation. The minimized average NMR structure of TGF-β1[10] and X-ray crystal structure of TGF-β3[9] were downloaded from the RCSB Protein Data Bank[20] (entries 1KLC and 1TGJ). The 224 (2 × 112) residue dimers were parametrized within the leap module of AMBER 8[21] using the ff03 force field,[22] with disulphide bonds connecting the cystine residues at positions 7, 15, 44, and 48 with those at 16, 78, 109, and 111 in each monomer, respectively. The two remaining cystine residues at position 77 formed the only covalent bond connecting the two dimers. The histidine residues at 34, 40, and 68 in TGF-β1 and those at 34 and 58 in TGF-β3 were all designated as ε tautomers. Default ionization states were used for all other residues.

Homology models of the two proteins were also constructed on the basis of the structure of the alternate isoform, that is, a model of TGF-β3 was based on the 1KLC structure and that of TGF-β1 was based on the 1TGJ structure. This task was performed manually; atoms common to any mutated residues were preserved and renamed where necessary, while the remainder of the side-chains had their coordinates deleted. The leap module was allowed to complete any remaining atoms according to residue templates, and the systems were parametrized as above. All four structures were immersed in rectangular boxes containing over 8000 TIP3P[23] water molecules. Neutralization required the addition of 8 Cl^- ions to each of the TGF-β1 systems and 2 Na^+ ions to each of the TGF-β3 systems. The structures were allowed to relax via 50 steps of steepest descent, followed by conjugate gradient minimization until the energy gradient had been reduced to 0.5 kcal mol^{-1}.

Target structures representing the open conformation were based upon the X-ray crystal structure of the TGF-β3:TβR-II complex (PDB entry 1KTZ).[14] Because the structure of TGF-β3 in the complex is incomplete and to minimize any potential bias (see discussion), monomers from each of the systems above were superimposed onto the complex structure, creating a total of four target conformations (Figure 1). Those residues where the tertiary structure remained unchanged between the open and closed conformations were then identified (resides 18−44 and 81−108) and designated as the residues to which the TMD restraints would be applied (α-carbon atoms only).

Molecular Dynamics Protocol. All simulations were performed with a 2 fs time step and with the use of SHAKE to constrain all bonds to hydrogen at equilibrium values.[24] Electrostatics beyond a 12 Å cutoff were calculated using particle mesh Ewald.[25] In the initial phase, simulations were performed at constant temperature and volume (NVT ensemble), with the temperature regulated using Langevin dynamics and a collision frequency of 5 collisions ps^{-1}; 100 ps simulations were performed on the solvent and counterions at 100 K to stabilize interactions with the protein. The water was heated to 300 K over a further 100 ps, with a harmonic restraint of 100 kcal mol^{-1} $Å^{-2}$ to keep the protein atoms close to the original structures. These restraints were then gradually reduced over a further 600 ps, before allowing the simulations to run unrestrained for a period of 200 ps.

At this point, the protocol was switched to the NPT ensemble to allow equilibration of the density, and the temperature was coupled to an external bath using a weak coupling algorithm.[26] Temperature and pressure coupling was performed at 1 ps intervals, with isotropic position scaling ensuring the pressure remained constant at 1 bar. The simulations were allowed to run in this state for a period of 10 ns.

Following this period of unrestrained simulation of the closed conformations, restraints were applied to the structures to induce the open state. At each step, the current simulation structure was superimposed onto the target using the α-carbon atoms of residues 18−44 and 81−108 from both monomers, and the rmsd was calculated between the two sets of atomic coordinates. The total energy penalty for deviation from the target rmsd was set to 5 kcal mol^{-1} $Å^{-2}$ $atom^{-1}$ (550 kcal mol^{-1} $Å^{-2}$). Initially, the target rmsd was set to the rmsd at 11 ns (typically in the region of 7.5−8 Å) and then incrementally reduced over a period of 5 ns to a value of 2 Å. Finally, the structures were allowed to relax, through gentle reduction of the restraint over a period of 1 ns, before allowing the simulations to run unrestrained for a further 33 ns. Overall, deposition of coordinates at 100 ps intervals generated a total of 500 structures for analysis.

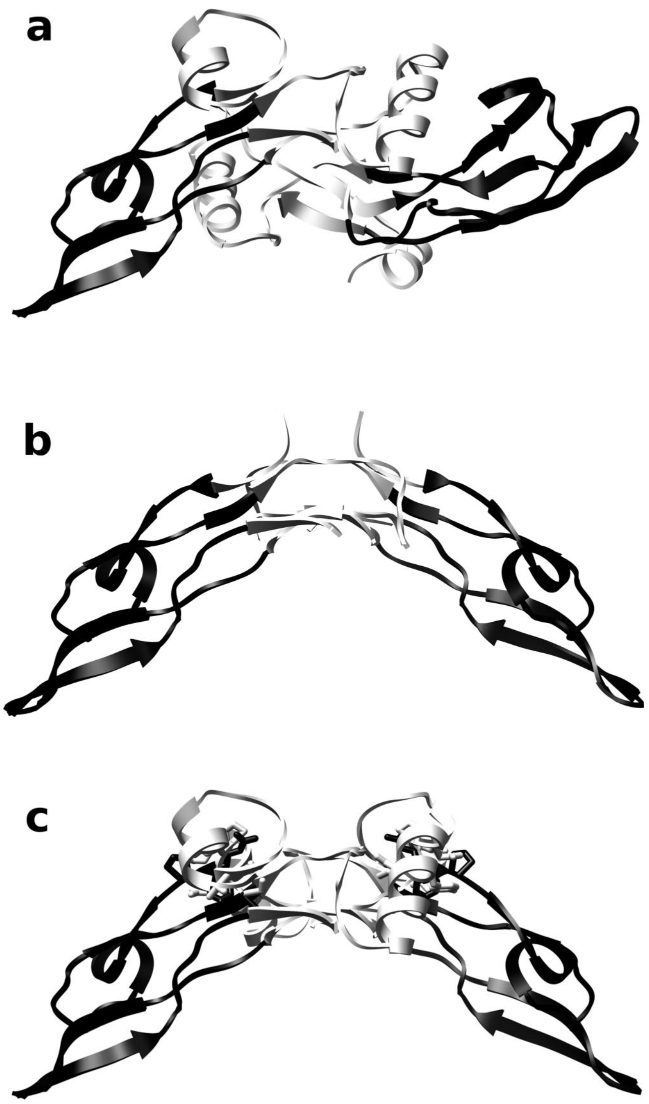


Figure 1. Construction of a model for the open conformation of TGF-β. (a) Crystal structure of TGF-β3 in isolation depicting the closed conformation.[9] (b) Crystal structure of TGF-β3 in complex with the extracellular domain of TβR-II (receptor not shown).[14] (c) Superimposition of monomers from a onto b to form a template for the open conformation. Residues exhibiting minimal displacement between b and c are rendered in black and were used as the α-carbon atoms to which the TMD restraints were applied (resides 18−44 and 81−108). Stick rendering highlights the side-chains involved in intermonomer steric overlap.

Trajectory Analysis and Figure Generation. The rmsd calculations, secondary structure, and native contact analyses were all performed using the AMBER module ptraj. Native contacts were defined as any pairwise atomic contacts (<5 Å) present in the initial structure (closed). In each system, contacts were identified within two individual subsets of residues comprising α-helix H3 of one monomer (residues

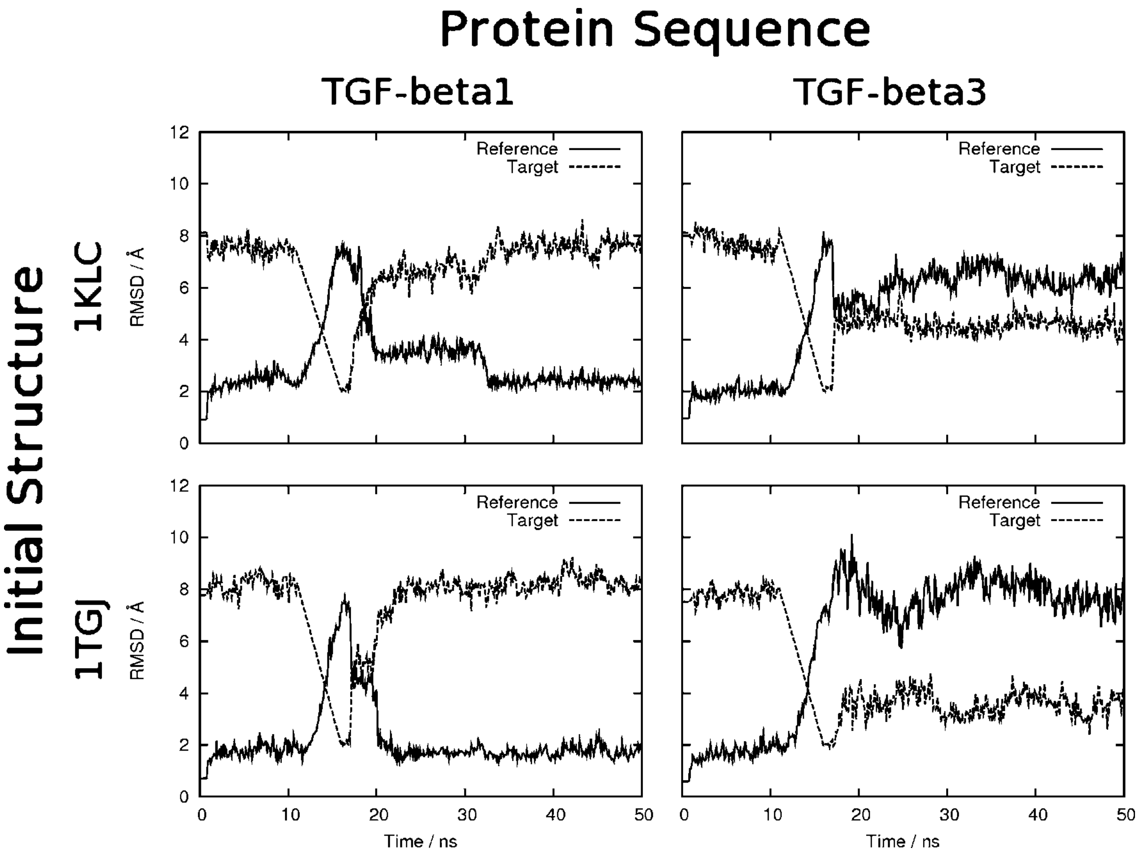


Figure 2. rmsd calculations for α-carbon atoms, tracking the conformational transition from the free, closed state to the restrained, open state and eventual release in the four simulations. Solid lines show comparisons of the structures to the closed, initial structure, calculated for all residues. The dotted lines show comparisons to the open, target structure, calculated only for those residues to which the restraint was applied (residues 18−44 and 81−108).

57−68 inclusive) and the full opposite monomer. The presence of such contacts was evaluated for each trajectory structure and averaged over the periods prior to and following the TMD restraints being applied (1−11 and 17−50 ns, respectively).

Computational glycine scanning was performed using the mm_pbsa perl scripts provided in the AMBER package. This method involves mutation of each residue to glycine through the removal all side-chain atoms except C_β, which is replaced by a hydrogen atom. Using this technique, it is possible to quickly estimate the effect of mutating each residue to glycine from a single trajectory. The GBSA method was used to evaluate the electrostatic and nonpolar components of the free energy of solvation,[27] thus the free energy difference between the wild-type protein, G_{wt}, and the mutant, G_{mut}, was represented by

$$\Delta G = G_{\mathrm{wt}} - G_{\mathrm{mut}}$$

where

$$G = E_{\mathrm{int}} + E_{\mathrm{vdW}} + E_{\mathrm{ele}} + G_{\mathrm{sol}} - TS$$

and

$$G_{\mathrm{sol}} = G_{\mathrm{GB}} + \gamma_{\mathrm{SASA}} + \beta$$

In this study, we assume that the difference in the overall entropy between the wild-type protein and the mutant is negligible, and so no attempt was made to calculate *TS*. In the calculation of G_{GB}, a dielectric constant of 1 was used for the interior of the protein, and 80 was used for the surrounding solvent; γ represents the surface tension applied to the solvent accessible surface area (SASA, calculated by the LCPO method[28]) and was set to 0.0072 kcal mol^{-1} Å^{-2}, with the offset, β set to 0 kcal mol^{-1}. Finally, the free energy differences were mean-centered on a per-residue basis to facilitate the identification of energetic hot-spots along the trajectory. Proline and disulphide bonded cystine residues were omitted from the analysis.

Molecular graphics images were produced using the UCSF Chimera[29] package from the Computer Graphics Laboratory, University of California, San Francisco. All plots were produced using Gnuplot 4.0.

Results

Backbone Structural Analyses. To assess the stability of the simulations, rmsd calculations were performed on the α-carbon atoms, comparing the original experimental structures with those generated during the course of the simulation (Figure 2). It can be seen that for a little over 10 ns from the outset, the simulated structures maintain an average deviation of around 2 Å from the initial conformation in all four systems. This provides evidence that the unrestrained simulations of the closed conformation were stable and, notably, that this was equally the case for the homology models as for the experimental structures. During this time, none of the four systems gave any indication of spontaneously adopting the open conformation.

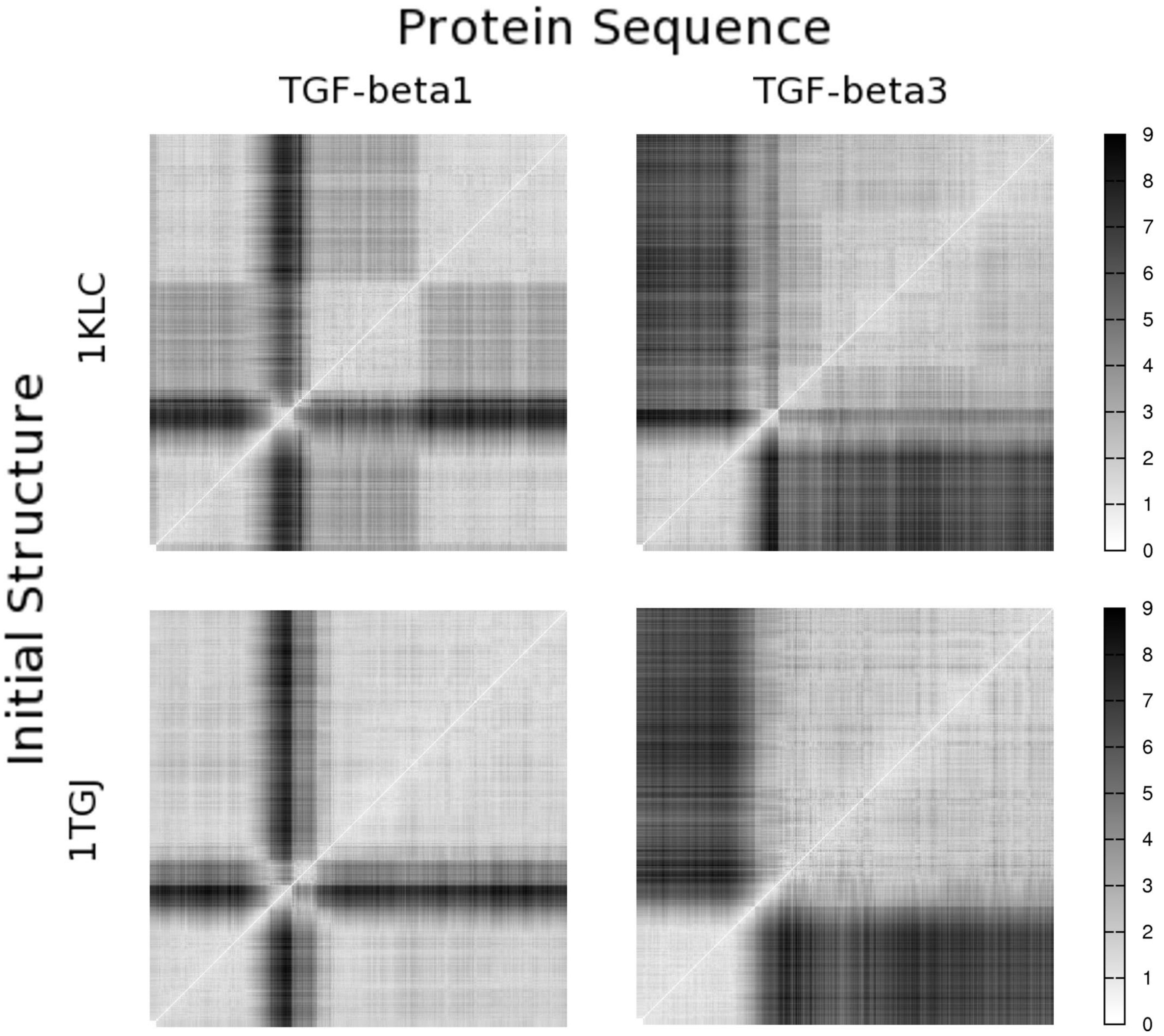


Figure 3. 2D rmsd plots comparing each snapshot with every other snapshot within a given simulation. Low rmsd values are represented as white squares, while high values are black. Simulation time begins at the bottom/left of the plots and ends at the top/right. Calculations were based on the α-carbon atoms of all residues.

From 11 ns, the rmsd between the simulated structures and the original (solid line) rises sharply as the TMD restraints are applied, and the relative orientation of the dimers is forced away from the closed state, toward the open conformation. Conversely, when comparisons are made with the target structure (dotted line), the rmsd falls as the target value is reduced, with the measured rmsd eventually reaching around 2 Å at a time of 16 ns.

It is from this point forward, as the structures are allowed to relax, that the behavior of the four simulations differs most significantly. In the two simulations of TGF-β1, the rmsd from the initial structure begins to fall as the restraints are released, indicating that the structures are moving back toward the original, closed conformation. By the end of the simulations the α-carbon atoms are, on average, around 2 Å from where they began, and exhibit a greater degree of similarity to the closed conformation than they do the target. In contrast to this, the TGF-β3 simulations showed no such tendency to revert back to the closed state, eventually finishing with rmsd values of 6.5 and 8.2 Å from their initial conformations. Both the simulation that started from the TGF-β3 crystal structure, and that which started from a homology model based on the TGF-β1 NMR structure, ended with structures with a greater similarity to the open target (with an rmsd of approximately 4 Å in both cases) than to the initial, closed conformation.

The 2D rmsd plots (Figure 3) provide a clear, visual means of identifying the proteins' transitions between different regions of conformational space. The clearly defined white squares in the bottom left corners of the plots represent the early part of the simulations, where all four systems occupied the clearly defined closed state. The large proportion of off-diagonal white points in this area shows that the structures

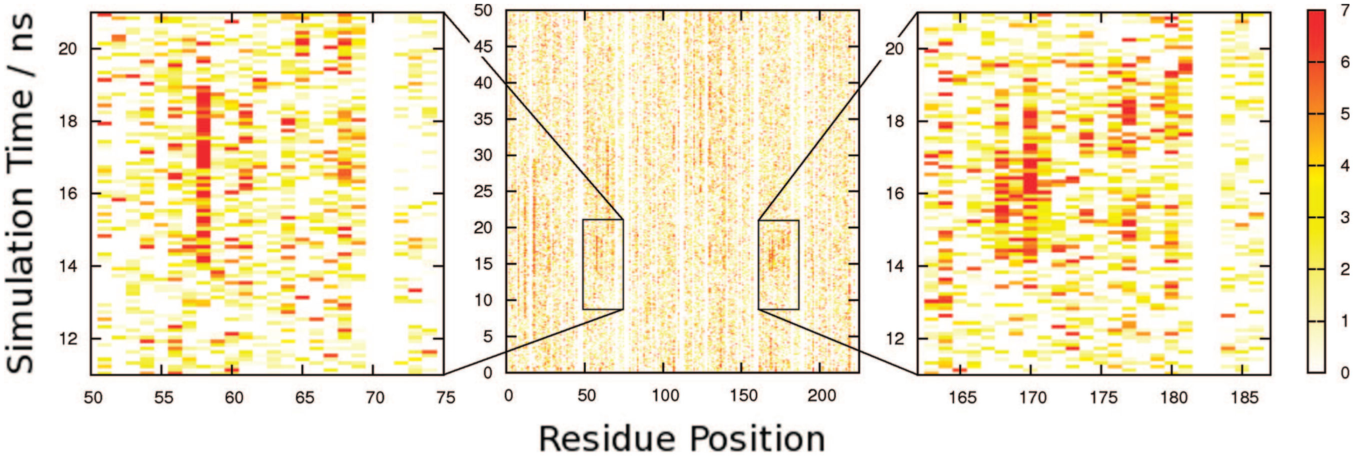


Figure 4. Computational glycine scanning applied to the simulation of TGF-β1 (taking the NMR structure,[10] 1KLC as the initial conformation). Points on the maps represent the free energy difference, ΔG between the wild-type and glycine mutant for each snapshot in the simulation, mean-centered for each residue. White points represent stable residues, where this value is less than 0 kcal mol^{-1}, and red points represent unstable residues, where this value exceeds 7 kcal mol^{-1}.

in the first 11 ns of any given simulation were comparatively very similar to each other.

As the restraints are applied and the conformations are encouraged to move toward the open state, the calculated deviations from the early structures begin to increase sharply, and consequently, the off-diagonal elements of the matrices become much darker in all four systems.

Once again, it is from this point on, when the TMD restraint is released, that the four simulations begin to differ. In the final phase of the simulations, it can be seen that the two simulations of TGF-β1 contain much lighter off-diagonal elements than do the simulations of TGF-β3. These lighter areas indicate the exploration of conformational space in the latter stages of the simulations that bears a high similarity to the initial, closed state. The apparent pattern of a dark cross against a light background is characteristic of the A → B → A state behavior. In contrast to this, the TGF-β3 simulations simply continue to explore space surrounding the twisted, open conformation, without returning to the experimental structures. It is this simple A → B state behavior that results in the apparent division of the 2D rmsd plots into four distinct corners.

Changes to the secondary structure content of the proteins during the course of the simulations were also calculated. From the outset, all four simulations had a much greater proportion of β-strands than α-helices, with a relative composition of around 40 and 15%, respectively. This composition is subject to a degree of fluctuation as the simulations progress, and while there is a general tendency for the strand content to deteriorate slightly in all simulations, there is no direct link between this variation and the application or release of the restraints. Ultimately, there is little change in the secondary structure of any of the proteins throughout the whole process, with all 4 systems retaining about 15% α-helix and 35% β-sheet at 50 ns.

Computational Glycine Scanning. The results from computational glycine scanning reveal the location of hot-spots in terms of the free energy difference between the wild-type and mutant protein (Figure 4). The red points indicate the location of structures where ΔG is high, that is to say, where the wild-type protein is considerably less stable than the glycine mutant. Particularly interesting in this study are those residues where the energy appears to spike at time points in the range 11−16 ns, because this is the time when TGF-β1 was forced to adopt the open conformation. The area of the plot where this effect appears most prevalent is in the region of residues 164−180 (Figure 4, right) and, to a lesser extent, 52−68 (Figure 4, left). These regions of the protein encompass the wrist helix of each monomer, which undergoes significant displacement during the course of the conformational transition. The most noticeable degree of instability affects residue 58 (residue 170 in monomer B), a buried tyrosine residue that undergoes mutation to histidine in TGF-β3.

Native Contact Analysis. A key feature of how the TGF-β3 structure is able to adapt to the open conformation is a change in the relative position of α-helix H3 in one of the two monomers. This translation avoids clashes with the opposite monomer, while maintaining native contacts at the dimer interface. To explain and quantify this effect, we tracked the presence of native atomic contacts within H3 and its opposite monomer over two periods, *pre* and *post* application of the TMD restraint. The results of this analysis are presented in Table 1.

As a result of reverting to the closed conformation, the two simulations of TGF-β1 show little difference in the prevalence of native atomic contacts before and after application of the restraints. In all four, such H3 plus monomer subsets, the loss of native contacts is in the region of just 1−2%. In contrast to this, the two TGF-β3 simulations both remained in the open state, and as would be expected, a significant proportion of native (closed state) contacts are lost. It appears from the data that in both simulations, the loss of native contacts is more marked in one monomer than in the other, that is, one α-helix shifts to accommodate the opposite monomer, thus retaining a number of closed state contacts at the dimer interface, while the alternate subset makes 8−10% less native contacts than it did in the closed conformation. The same effect is also apparent from the visual representation of the final structures presented in

Table 1. Number of Native Atomic Contacts As a Percentage of the Number Present in the Initial Structure[a]

initial structure	protein sequence	residue subset	pre-TMD	post-TMD	difference
1KLC	TGF-β1	H3 (monomer A) + monomer B	74.7%	73.1%	−1.6%
1KLC	TGF-β1	H3 (monomer B) + monomer A	74.2%	73.2%	−1.0%
1TGJ	TGF-β3	H3 (monomer A) + monomer B	81.6%	73.3%	**−8.3%**
1TGJ	TGF-β3	H3 (monomer B) + monomer A	80.5%	76.8%	−3.6%
1KLC	TGF-β3	H3 (monomer A) + monomer B	79.1%	76.2%	−2.9%
1KLC	TGF-β3	H3 (monomer B) + monomer A	79.8%	69.9%	**−9.9%**
1TGJ	TGF-β1	H3 (monomer A) + monomer B	81.7%	80.2%	−1.6%
1TGJ	TGF-β1	H3 (monomer B) + monomer A	82.2%	80.8%	−1.4%

[a] Values are quoted for the periods pre- and post-application and release of the TMD restraints (1−11 and 17−50 ns, respectively). The difference column quotes the loss of native contacts as a result of the conformational transition.

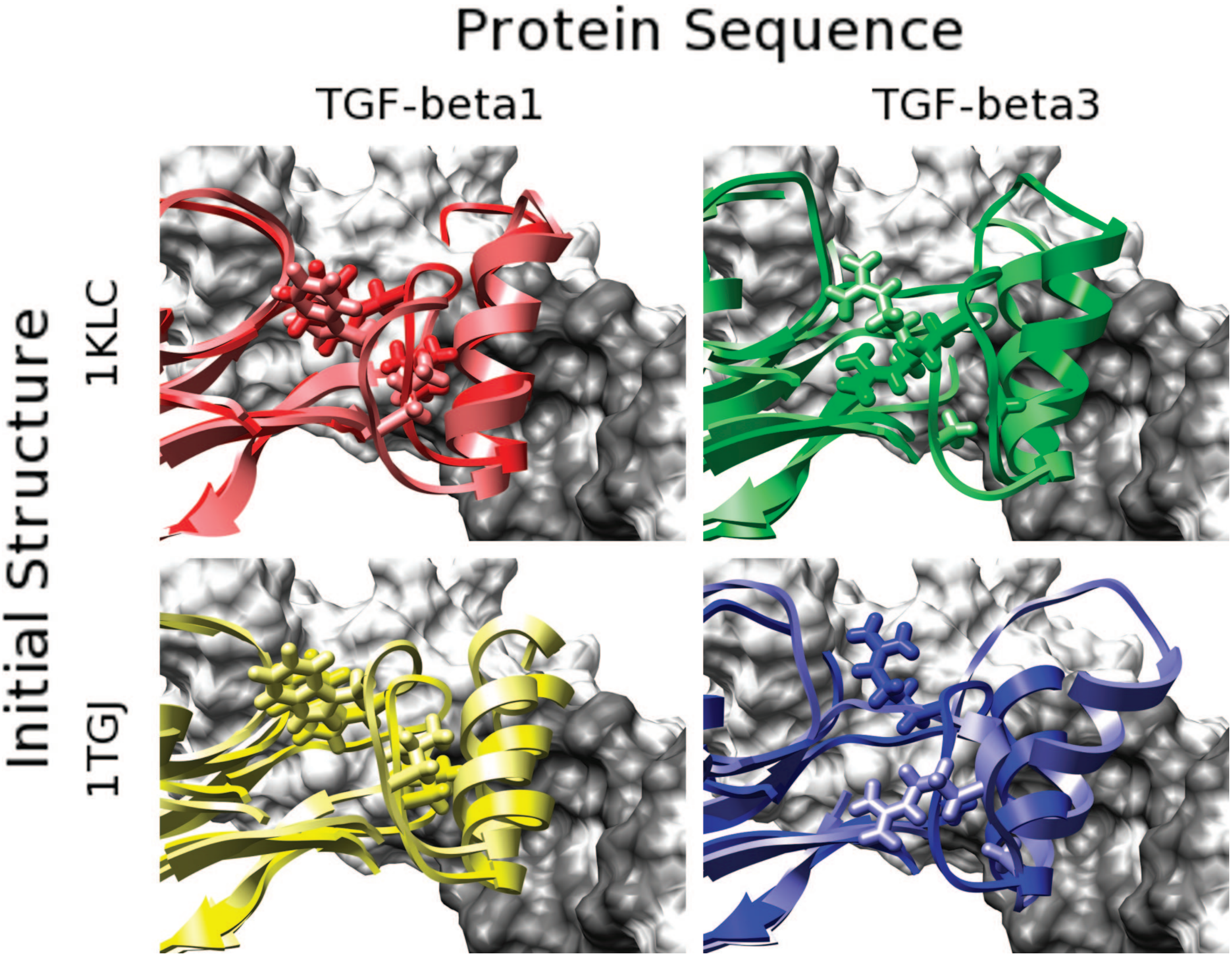


Figure 5. Cartoon representations of all four final structures following 50 ns of simulation. Monomer A in each system is shown in bold color, while monomer B is represented in a lighter shade. All 8 monomers are presented in the same frame of reference by overlaying each structure onto the final coordinates of TGF-β3 monomer B (1TGJ simulation). Monomer A from the same simulation is represented as a white surface, with those residues which form native contacts with α-helix H3 rendered in dark gray. Mutated residues in the loop preceding the helix are represented as sticks (residues 52 and 54).

Figure 5. In the TGF-β3 simulation that began from 1TGJ, it is monomer B (light blue cartoon) that shifts to regain contact with the concave groove in the opposite monomer (dark gray surface), while in the simulation that started from the 1KLC structure, it is monomer A (bold green cartoon) that undergoes this same transformation.

Along with the observed shifts of α-helix H3 in TGF-β3, we also see concerted changes in the conformation of the preceding loop, which encompasses residues Arg 52 and Ala 54. In both simulations, a change in the rotamer state of Arg 52 appears to be involved, featuring loss of the crystallographic interaction with Glu 12, and an approximate doubling in the population of the first solvation sphere (number of solvent residues <3.4 Å). The loss of the interaction with Glu 12 has the effect of releasing the tension in this loop, thus allowing the entire helix to change conformation. Notably in TGF-β1, where residues 52 and 54 are mutated to tryptophan and leucine, respectively, solvent exposure of Trp 52 remains at pretransition levels, and the loop undergoes no such conformational change.

Discussion

Simulation Stability. The initial phase of the 1D rmsd plots allows comparison of the stability of the simulations, not only between one isoform and the other, but also between the simulations that began from an experimental structure of a particular isoform, and the homology model that was built using the other as a template. One may have expected that simulation of the solution phase NMR structure of TGF-β1 would generate structures in closer agreement with the experimental structure than the simulation of TGF-β3 because of the need for TGF-β3 to overcome any crystal packing artifacts present in the structure and adapt to the presence of aqueous solvent. This proved not to be the case. An encouraging result is that the same level of deviation from the experimental structures is exhibited in the simulations of the homology models. This indicates that the process of building the models did not introduce any serious clashes in the system that could not be rectified through minimization or the carefully controlled warm-up process, presumably because of the high level of sequence similarity across the two isoforms.

Conformational Preferences of TGF-β Isoforms. The major differences in the two isoforms' structures and behavior only became apparent following the release of the TMD restraints. When the perturbed structures were presented with an opportunity to relax, both simulations of TGF-β1 immediately moved away from the induced open conformation and began to revert back toward the initial, closed state. This observation was in stark contrast to the two simulations of the other isoform studied, TGF-β3, which remained comparatively comfortably in the open state.

These observations indicate that the process of adopting the open conformation is energetically far less far favorable for TGF-β1 than it is for TGF-β3. It is particularly encouraging to see the conformations favored by the homology models mimicking those of the equivalent isoform, rather than the isoforms on which the models were based. This allows confidence that the observations are a direct result of the protein sequence and not simply an artifact stemming from the fact that the simulations were initiated from different conformations. The results are in excellent agreement with the current experimental evidence surrounding these two isoforms. After all, it was an X-ray crystal structure of TGF-β3 in complex with the type II TGF-β receptor that was used as the initial template for the open conformation of the homodimer. TGF-β1, on the other hand, has only ever been observed in the closed state.

It should be noted that selection of a TGF-β3 structure for the TMD template had the potential to introduce a bias into the experiment. That is to say, if the systems had been encouraged to adopt the precise backbone conformation of TGF-β3, simulations of other isoforms would be considerably destabilized regardless of the relative orientation of the monomers. For example, Thr 87 in TGF-β3 is substituted for proline in TGF-β1, a residue which is far more limited in terms of main-chain flexibility. This residue may have encountered serious difficulties in adopting the backbone conformation of the corresponding threonine in the template. However, such bias was avoided by superimposing each individual monomer from the closed structures onto the complex at the time of target construction, the result being that the backbone conformation of the target then became allied to the initial experimental structure. Consequentially, we can confidently attribute the preferences exhibited by the two isoforms of TGF-β to the nature of the intramolecular interactions within the protein.

Dependence of Secondary Structure on Monomer Orientation. The relative orientation of the monomers in the open conformation poses a direct conflict with the location of α-helix H3 as exhibited in the closed crystal structure (Figure 1). For this reason, the structural stability of this section is thought to be crucial in determining the ability of TGF-β isoforms to adopt the open conformation. In particular, it has been suggested that the four helix destabilizing residues Thr 57, Thr 60, Gly 63, and Thr 67 on α-helix H3 of TGF-β3 are responsible for disrupting the formation of this structure in solution.[12] Although no evidence for any loss of helical content was observed in these simulations, it is precisely this area of the protein where mutation of the side chains to hydrogen had the most stabilizing effect upon the induced open conformations of TGF-β1. The localized instability experienced by H3 in adopting the open conformation clearly needs to be overcome by some means. In this study we see how for TGF-β1, this is achieved by reverting to the closed state, while for β3, the helix still remains intact, yet is simply shifted sideways to accommodate the opposite monomer.

The reason TGF-β3 alone is able to tolerate the change in orientation of α-helix H3 lies (at least in part) in the nature of the residues that comprise the preceding loop. In the available experimental structures, position 52 appears semi-buried, an environment that clearly suits the highly lipophilic tryptophan residue in TGF-β1. However, during the process of adopting the open conformation and in an attempt to retain native contacts at the dimer interface, this loop becomes more exposed to solvent, and so the environment becomes far better suited to the more polar arginine residue in TGF-β3. Indeed, from a lipophilicity perspective, mutation at this position could not be more dramatic. It is, therefore, not surprising that there are significant differences in behavior, not only between one isoform and another, but also between observations made in solid crystals and those in a polar solvent.

It is likely that the simulations did not sample an adequate amount of time for full helical decomposition to be observed, and that ultimately this is what we would expect to happen. Nevertheless, the findings are still interesting in their own right. We have observed that in solution, TGF-β3 has the ability to adopt a high-energy intermediate conformation in which the two monomers are rotated but α-helix H3 remains intact. TGF-β1 on the other hand, lacks the ability to stabilize such a conformation, and immediately reverts back to the closed state. This implies that kinetics is likely to play a key role in the transition between open and closed forms of TGF-β isoforms.

Conclusion

We have successfully performed four TMD simulations on two TGF-β isoforms, -β1 and -β3. Both the experimental structures of each isoform and the two homology models remained stable throughout the course of a 11 ns MD simulation in explicit solvent. Furthermore, with the aid of a rmsd weighted restraint, we were able to promote the transition of both isoforms from the closed conformation to one we believe to be more representative of TGF-β3's structure in solution.

Following the simulations from this point, we observed distinct, sequence determined preferences for either the open or the closed form of TGF-β. Transition between the open and closed form is key to the biological function of TGF-β3 because it must adopt the closed conformation to promote assembly of the TGF-β:TβR-I:TβR-II signaling complex as has now been observed.[15] Evidence from NMR, CD, and now TMD simulations indicate that a second, open conformation is preferable in the case of TGF-β3. This is in contrast to the evidence surrounding TGF-β1, which is believed to exist entirely in the closed conformation. Here, we have presented an argument as to why this is the case, and highlighted key mutations at positions 52−58 inclusive that we believe are responsible for the change in behavior.

The protocol outlined in this paper has been shown to distinguish between the conformational preferences of TGF-β1 and -β3, taking just the experimental structure of one isoform as a starting point. Of significant interest now is likely to be the identification of a small subset of point mutations capable of reversing the behavior of a given isoform. This is clearly a potential avenue for further exploration, ideally with the support of experimental evidence to validate any predictions.

Acknowledgment. The authors would like to thank Breast Cancer Campaign for funding this research. D.J.W. would also like to thank Dr. Charles Laughton of the University of Nottingham for use of the Linux cluster on which the calculations were performed

References

(1) Massague, J. *Annu. Rev. Cell Biol.* **1990**, *6*, 597–641.

(2) de Caestecker, M. *Cytokine Growth Factor Rev.* **2004**, *15*, 1–11.

(3) Cheifetz, S.; Hernandez, H.; Laiho, M.; ten Dijke, P.; Iwata, K. K.; Massague, J. *J. Biol. Chem.* **1990**, *265*, 20533–20538.

(4) Massague, J.; Andres, J.; Attisano, L.; Cheifetz, S.; Lopez-Casillas, F.; Ohtsuki, M.; Wrana, J. L. *Mol. Reprod. Dev.* **1992**, *32*, 99–104.

(5) Wrana, J. L.; Attisano, L.; Wieser, R.; Ventura, F.; Massague, J. *Nature* **1994**, *370*, 341–7.

(6) Zúniga, J. E.; Groppe, J. C.; Cui, Y.; Hinck, C. S.; Contreras-Shannon, V.; Pakhomova, O. N.; Yang, J.; Tang, Y.; Mendoza, V.; López-Casillas, F. *J. Mol. Biol.* **2005**, *354*, 1052–1068.

(7) Daopin, S.; Piez, K. A.; Ogawa, Y.; Davies, D. R. *Science* **1992**, *257*, 369–373.

(8) Schlunegger, M. P.; Grütter, M. G. *Nature* **1992**, *358*, 430–434.

(9) Mittl, P. R.; Priestle, J. P.; Cox, D. A.; McMaster, G.; Cerletti, N.; Grütter, M. G. *Protein Sci.* **1996**, *5*, 1261.

(10) Hinck, A. P.; Archer, S. J.; Qian, S. W.; Roberts, A. B.; Sporn, M. B.; Weatherbee, J. A.; Tsang, M. L. S.; Lucas, R.; Zhang, B. L.; Wenker, J. *Biochemistry* **1996**, *35*, 8517–8534.

(11) Archer, S. J.; Bax, A.; Roberts, A. B.; Sporn, M. B.; Ogawa, Y.; Piez, K. A.; Weatherbee, J. A.; Tsang, M. L. S.; Lucas, R. *Biochemistry* **1993**, *32*, 1164–1171.

(12) Bocharov, E. V.; Korzhnev, D. M.; Blommers, M. J. J.; Arvinte, T.; Orekhov, V. Y.; Billeter, M.; Arseniev, A. S. *J. Biol. Chem.* **2002**, *277*, 46273–46279.

(13) Pellaud, J.; Schote, U.; Arvinte, T.; Seelig, J. *J. Biol. Chem.* **1999**, *274*, 7699–7704.

(14) Hart, P. J.; Deep, S.; Taylor, A. B.; Shu, Z.; Hinck, C. S.; Hinck, A. P. *Nat. Struct. Biol.* **2002**, *9*, 203–8.

(15) Groppe, J.; Hinck, C. S.; Samavarchi-Tehrani, P.; Zubieta, C.; Schuermann, J. P.; Taylor, A. B.; Schwarz, P. M.; Wrana, J. L.; Hinck, A. P. *Mol. Cell* **2008**, *29*, 157–168.

(16) Lin, S. J.; Lerch, T. F.; Cook, R. W.; Jardetzky, T. S.; Woodruff, T. K. *Reproduction* **2006**, *132*, 179–190.

(17) Keller, S.; Nickel, J.; Zhang, J. L.; Sebald, W.; Mueller, T. D. *Nat. Struct. Mol. Biol.* **2004**, *11*, 481–488.

(18) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291–308.

(19) Apostolakis, J.; Ferrara, P.; Caflisch, A. *J. Chem. Phys.* **1999**, *110*, 2099.

(20) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(21) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.

(22) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(24) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal Comput. Phys.* **1977**, *23*, 327.

(25) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(26) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(27) Tsui, V.; Case, D. A. *Biopolymers (Nucleic Acid Sci.)* **2001**, *56*, 257–291.

(28) Weiser, J.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1999**, *20*, 217–230.

(29) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

CT800483Q

J. Chem. Theory Comput. **2009,** *5,* 491–499

JCTC Journal of Chemical Theory and Computation

Gaussian Basis Set and Planewave Relativistic Spin−Orbit Methods in NWChem

Patrick Nichols,* Niranjan Govind,* Eric J. Bylaska, and W. A. de Jong

William R. Wiley Environmental Molecular Sciences Laboratory, Pacific Northwest
National Laboratory, 902 Battelle Boulevard, P.O. Box 999,
Mail Stop K8-91, Richland, Washington, 99352

Received July 21, 2008

Abstract: Relativistic spin−orbit density functional theory (DFT) methods have been implemented in the molecular Gaussian DFT and pseudopotential planewave DFT modules of the NWChem electronic-structure program. The Gaussian basis set implementation is based upon the zeroth-order regular approximation (ZORA) while the planewave implementation uses spin−orbit pseudopotentials that are directly generated from the atomic Dirac−Kohn−Sham wave functions or atomic ZORA-Kohn−Sham wave functions. Compared to solving the full Dirac equation these methods are computationally efficient but robust enough for a realistic description of relativistic effects such as spin−orbit splitting, molecular orbital hybridization, and core effects. Both methods have been applied to a variety of small molecules, including I_2, IF, HI, Br_2, Bi_2, AuH, and Au_2, using various exchange-correlation functionals. Our results are in good agreement with experiment and previously reported calculations.

I. Introduction

It is well established that scalar and spin−orbit relativistic effects have to be taken into account for accurate electronic structure calculations of actinides and other heavy elements. Relativistic effects are best described in electronic-structure calculations by solving the Dirac equation, whose solutions are made up of four-component spinor wave functions. However, four-component methods are not only fraught with problems such as variational collapse,[1] they are also an order of magnitude more expensive. The extra cost stems from the need to properly describe the small component wave function. In principle, the cost of the Dirac equation can be reduced by transforming it from a regular four-component eigenvalue equation into a nonregular two-component eigenvalue equation by decoupling the large and small components of the wave function. However, this nonregular two-component eigenvalue equation turns out to be difficult to solve, because it has a nontrivial normalization condition and the eigenvalue depends nonlinearly upon itself. Over the years, many approximations have been developed to circumvent these difficulties such as the Breit−Pauli Hamiltonian, direct perturbation theory, Dyall's modified Dirac method,[2] and the Douglas−Kroll−Hess (DKH) Hamiltonian.[3−5] Another approximation, the zeroth-order regular approximation (ZORA) method, has become one of the more popular of these approximations. The ZORA method was originally developed by Chang and Durand.[6] This method was rediscovered and further developed by Baerends, van Lenthe, and co-workers.[7] It is a two-component spinor approach for approximately solving the Dirac equation based upon regularizing the wave equation by ignoring the energy dependence of the effective mass of the electron. It has been shown that the solutions of the ZORA equation are one of the best two-component approximations to the fully relativistic Dirac solution for hydrogen-like systems. Since this is a two-component method, calculations can be performed using only the large component. This approximation has been shown to perform well for a wide variety of properties and is among the cheapest of the approximate spinor methods.[8,9]

The ZORA approximation and other regularized two-component methods can be used to solve for all the electrons (core + valence) in the system. However, this is not typically necessary, since the most significant relativistic effects are in the core region and can be encapsulated using core potentials (RECP)[10−12] or pseudopotentials.[13−15] These types

---

* Corresponding authors. E-mail: patrick.nichols@pnl.gov (P.N.), niri.govind@pnl.gov.

10.1021/ct8002892 CCC: $40.75

of approaches have been shown to give accurate structures, frequencies, and other properties that depend primarily on getting the valence electronic structure correct. For properties that directly involve core orbitals such as XPS and NMR, all-electron methods may be needed. Relativistic pseudopotentials with spin−orbit effects have been also been developed in the past. The HGH pseudopotentials are among the most well-known.[13] The Helmstreet, Fong, and Nelson pseudopotential is another example.[16] Theurich and Hill have also developed a pseudopotential for III−V semiconductors that is similar to the one implemented here for the norm-conserving case.[17,18] Ultrasoft pseudopotentials incorporating spin−orbit effects have also been studied, e.g. Corso and co-workers[14,15] and the work of Oda and Hosokawa.[19] Chelikowsky and co-workers have recently extended these approaches to real space grids.[20] Solid-state calculations using these approaches have all yielded favorable results for both geometries and spin−orbit splittings.

In this paper we present our implementation and applications of the relativistic two-component Hamiltonian within the Gaussian basis set and planewave frameworks. Both spin-free and spin−orbit versions of the Hamiltonian have been implemented. In the formalism and implementation section, we present a short overview of the theory and the details of our implementation in the framework of the NWChem program package.[21] To test our implementation, we have calculated the equilibrium geometries, harmonic frequencies, and dissociation energies for a set of small closed-shell molecules (I_2, Br_2, HI, IF, Au_2, Bi_2, AuH). We compare the results using various exchange-correlation functionals within the spin-free and spin−orbit approaches. The results are also compared with experimental data and published calculations.

II. Formalism and Implementation

A. ZORA Formalism. The two-component ZORA equation is given by

$$\left[\boldsymbol{\sigma}\cdot\mathbf{p}\frac{K(\mathbf{r})}{2m_e}\boldsymbol{\sigma}\cdot\mathbf{p}+V(\mathbf{r})\right]\psi_n(\mathbf{r})=\epsilon_n^{\mathrm{ZORA}}\psi_n(\mathbf{r}) \qquad (1)$$

where

$$K(\mathbf{r})=\left(1-\frac{V(\mathbf{r})}{2m_ec^2}\right)^{-1} \qquad (2)$$

After some algebra, eq 1 can be written as

$$\left[\mathbf{p}\frac{K(\mathbf{r})}{2m_e}\mathbf{p}+\frac{(K(\mathbf{r}))^2}{4m_e^2c^2}\boldsymbol{\sigma}\cdot(\nabla V(\mathbf{r})\times\mathbf{p})+V(\mathbf{r})\right]\psi_n(\mathbf{r})=\epsilon_n^{\mathrm{ZORA}}\psi_n(\mathbf{r}) \qquad (3)$$

As one can see above, the spin−orbit operator (term 2) in eq 3 is already present at this lowest-order of the expansion. Equation 1 can also be written as

$$\left[\frac{\mathbf{p}^2}{2m_e}+\mathbf{p}\left(\frac{K(\mathbf{r})-1}{2m_e}\right)\mathbf{p}+\frac{(K(\mathbf{r}))^2}{4m_e^2c^2}\boldsymbol{\sigma}\cdot(\nabla V(\mathbf{r})\times\mathbf{p})+V(\mathbf{r})\right]\psi_n(\mathbf{r})=\epsilon_n^{\mathrm{ZORA}}\psi_n(\mathbf{r}) \qquad (4)$$

where the nonrelativistic kinetic energy contribution has been isolated. The advantage of this decomposition is that the ZORA contribution can be treated as a correction which can be added to the kinetic-energy matrix elements and also recovers the correct limits in addition. The spin-free or scalar-relativistic equation can be obtained by eliminating the spin−orbit term, resulting in the following equation,

$$\left[\frac{\mathbf{p}^2}{2m_e}+\mathbf{p}\left(\frac{K(\mathbf{r})-1}{2m_e}\right)\mathbf{p}+V(\mathbf{r})\right]\psi_n(\mathbf{r})=\epsilon_n^{\mathrm{ZORA}}\psi_n(\mathbf{r}) \qquad (5)$$

It has been established that the one-electron energies of the ZORA equation, $\epsilon_i^{\mathrm{ZORA}}$, can be improved by scaling the ZORA energy.[8,9] This scaling essentially captures the effects of summing certain higher-order contributions to infinite order. The scaled one-electron energies $\epsilon_i^{\mathrm{scaled}}$ are given by

$$\epsilon_i^{\mathrm{scaled}}=\frac{\epsilon_i^{\mathrm{ZORA}}}{1+\langle\phi_i(\mathbf{r})|\boldsymbol{\sigma}\cdot\mathbf{p}\frac{(K(\mathbf{r}))^2}{4m_e^2c^2}\boldsymbol{\sigma}\cdot\mathbf{p}|\phi_i(\mathbf{r})\rangle} \qquad (6)$$

Within the density functional theory (DFT) framework, the total scaled ZORA energy E^s_{tot} can be written as

$$E_{\mathrm{tot}}^{\mathrm{scaled}}=\sum_{i=1}^{N}\left[\langle\phi_i(\mathbf{r})|\boldsymbol{\sigma}\cdot\mathbf{p}\frac{K(\mathbf{r})}{2m_e}\boldsymbol{\sigma}\cdot\mathbf{p}|\phi_i(\mathbf{r})\rangle-\epsilon_i^{\mathrm{scaled}}\langle\phi_i(\mathbf{r})|\boldsymbol{\sigma}\cdot\mathbf{p}\frac{(K(\mathbf{r}))^2}{4m_e^2c^2}\boldsymbol{\sigma}\cdot\mathbf{p}|\phi_i(\mathbf{r})\rangle\right]+\int \mathrm{d}r\,\rho(\mathbf{r})V_{\mathrm{ion}}(\mathbf{r})+\frac{1}{2}\iint \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'\,\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}+E_{\mathrm{xc}}[\rho(\mathbf{r})]+E_{nn} \qquad (7)$$

where the summation runs over the occupied orbitals. The above equation is identical to the nonrelativistic total energy except for the kinetic energy contribution which includes the ZORA correction.

B. Gaussian Basis Implementation. In our Gaussian basis set implementation, the molecular spinors are expanded in terms of real basis functions as follows,

$$\phi_i(\mathbf{r})=\sum_\mu\left\{c_{\mu i}^{\alpha}\begin{pmatrix}\chi_\mu(\mathbf{r})\\0\end{pmatrix}+c_{\mu i}^{\beta}\begin{pmatrix}0\\\chi_\mu(\mathbf{r})\end{pmatrix}\right\}=\begin{pmatrix}\phi_i^{\alpha}(\mathbf{r})\\\phi_i^{\beta}(\mathbf{r})\end{pmatrix} \qquad (8)$$

where χ_μ are Gaussian basis functions and $c_{\mu i}^{\alpha}$ and $c_{\mu i}^{\beta}$ are expansion coefficients which are, in general, complex. Since the ZORA corrections only affect the kinetic energy, we only concentrate on this term. The remaining terms are evaluated in the conventional manner within the molecular Gaussian DFT module in NWChem. Within the basis set approximation and using the decomposition from eq 4, the corrected kinetic energy elements of the Fock matrix can be written as

$$T_{\mu\nu}^{\mathrm{tot}}=\langle\chi_\mu(\mathbf{r})|\frac{\mathbf{p}^2}{2m_e}|\chi_\nu(\mathbf{r})\rangle+\langle\chi_\mu(\mathbf{r})\boldsymbol{\sigma}\cdot\mathbf{p}|\frac{K(\mathbf{r})-1}{2m_e}|\boldsymbol{\sigma}\cdot\mathbf{p}\chi_\nu(\mathbf{r})\rangle \qquad (9)$$

The spin-free or scalar-relativistic case can be obtained in a similar way using eq 5. In NWChem, the nonrelativistic part of the kinetic energy (first term in the above equation) is evaluated analytically where as the ZORA correction (second term) is calculated numerically on atom-centered grids.[25] Since the ZORA correction depends on the potential, it is not gauge invariant. There have been a number of attempts to address this issue.[22−24] In our implementation, we address this by using the atomic approximation of van Lenthe and co-workers.[26,27] Strictly speaking none of these methods are truly gauge invariant in the general sense, but they help minimize the problem. Within the atomic approximation, the ZORA corrections to the kinetic energy matrix elements are calculated using the superposition of densities of the atoms in the system. As a result, only intra-atomic contributions are involved and no gradient or second derivatives of these corrections need to be calculated. In addition, the corrections only have to be calculated once and stored. With this, we can re-express eq 9 as

$$T_{\mu\nu}^{\mathrm{tot}} = \langle \chi_\mu(\mathbf{r}) | \frac{\mathbf{p}^2}{2m_\mathrm{e}} | \chi_\nu(\mathbf{r}) \rangle + \langle \chi_\mu(\mathbf{r}) \boldsymbol{\sigma} \cdot \mathbf{p} | \frac{\tilde{K}(\mathbf{r}) - 1}{2m_\mathrm{e}} | \boldsymbol{\sigma} \cdot \mathbf{p} \chi_\nu(\mathbf{r}) \rangle \tag{10}$$

where $\tilde{K}(\mathbf{r}) = (1 - \tilde{V}(\mathbf{r})/2m_\mathrm{e}c^2)^{-1}$ and $\tilde{V} \approx V_{\mathrm{atom}}^{\mathrm{ne}} + V_{\mathrm{atom}}^{\mathrm{H}}$, the sum of the atomic nuclear-electron, Hartree potentials, respectively. The atomic densities are calculated for the neutral atoms at the Hartree−Fock (HF) level, and then, the ZORA potential corrections are calculated using this density. Note that we ignore the exchange-correlation contribution in the evaluation of the ZORA atomic corrections in the same way as van Lenthe and co-workers.[26,27,33] This approximation works well, as shown by our results. Our implementation differs from theirs in that they use a resolution of identity (RI) approach to calculate the ZORA corrections while we use atom-centered grids.

C. Planewave Implementation. The planewave implementation uses a two-component spinor expanded in planewaves as a basis to describe the eigenfunctions of the system.

$$\psi_i(\mathbf{r}) = \begin{bmatrix} \psi_i^\alpha(\mathbf{r}) \\ \psi_i^\beta(\mathbf{r}) \end{bmatrix} \tag{11}$$

where

$$\int \mathrm{d}^3 r \, (\psi_i^{\alpha *}(\mathbf{r}) \psi_i^\alpha(\mathbf{r}) + \psi_i^{\beta *}(\mathbf{r}) \psi_i^\beta(\mathbf{r})) = 1 \tag{12}$$

The Hamiltonian operator acts upon the two-component spinor planewaves in the usual fashion. For a single eigenfunction,

$$\begin{bmatrix} H_{\alpha\alpha} & H_{\alpha\beta} \\ H_{\beta\alpha} & H_{\beta\beta} \end{bmatrix} \begin{bmatrix} \psi_i^\alpha(\mathbf{r}) \\ \psi_i^\beta(\mathbf{r}) \end{bmatrix} = \begin{bmatrix} \epsilon \psi_i^\alpha(\mathbf{r}) \\ \epsilon \psi_i^\beta(\mathbf{r}) \end{bmatrix} \tag{13}$$

and the generalization to many eigenfunctions is of similar form. Relativistic effects are most dominant in the deep core region. Since we are mainly interested in the valence properties, one can encapsulate the core effects into a pseudopotential. In this work, we construct a two-component nonlocal pseudopotential. This potential will in general act on the spin up and spin down components of each eigenfunction in a nontrivial fashion. The first step in generating a relativistic pseudopotential is to solve for the two-component spinor all-electron wave functions and the self-consistent potential for the atom using either two-component radial Dirac or ZORA equation. The large two-component solutions are of the form

$$\psi_{njlm_j}(\mathbf{r}) = \frac{u_{njl}(r)}{r} \mathscr{Y}_l^{j,m_j}(\hat{r}) \tag{14}$$

where $u_{jl}(r)$ is the radial wave function, $\mathscr{Y}_l^{j,m_j}$ is a spin angular function, and the principal quantum numbers are n, j, l, and j_3 where $n \in [1, 2, 3, \ldots]$, $l \in [0, 1, 2, \ldots, n-1]$, $j = l \pm {}^1/_2$ and $m_j \in [-j, -j+1, \ldots, 0, \ldots, (j-1), j]$. The spin angular functions, which are eigenfunctions of J^2, J_z, and $\boldsymbol{\sigma}\cdot\mathbf{L}$, are defined to be for $j = l + {}^1/_2$

$$\mathscr{Y}_l^{j,m_j}(\hat{\mathbf{r}}) = \begin{bmatrix} \sqrt{\dfrac{j + m_j + \frac{1}{2}}{2l+1}} Y_l^{m_j - 1/2}(\hat{\mathbf{r}}) \\ \sqrt{\dfrac{j - m_j + \frac{1}{2}}{2l+1}} Y_l^{m_j + 1/2}(\hat{\mathbf{r}}) \end{bmatrix} \tag{15}$$

and for $j = l - {}^1/_2$

$$\mathscr{Y}_l^{j,m_j}(\hat{\mathbf{r}}) = \begin{bmatrix} -\sqrt{\dfrac{j - m_j + \frac{1}{2}}{2l+1}} Y_l^{m_j - 1/2}(\hat{\mathbf{r}}) \\ \sqrt{\dfrac{j + m_j + \frac{1}{2}}{2l+1}} Y_l^{m_j + 1/2}(\hat{\mathbf{r}}) \end{bmatrix} \tag{16}$$

where Y_l^m are the spherical harmonic functions.[28] We can convert the all-electron radial valence wave functions $u_{njl}(r)$ to radial pseudowave functions, $w_{jl}(r)$, and corresponding screened ionic pseudopotentials, $V_{jl}^{\mathrm{scr,ion}}$ by inverting the radial Schrödinger equation using the Hamann[29] or Troullier−Martins procedure.[30] The screened ionic pseudopotential components are then unscreened (ignoring the small component contribution to the valence density for Dirac solutions and the corresponding change in wave function normalization)

$$V_{jl}^{\mathrm{ion}}(r) = V_{jl}^{\mathrm{scr,ion}}(r) - V_{\mathrm{xc}}[\rho_{\mathrm{val}}(r)] - V_{\mathrm{H}}[\rho_{\mathrm{val}}(r)] \tag{17}$$

where $V_{\mathrm{xc}}[\rho_{\mathrm{val}}(r)]$ and $V_{\mathrm{H}}[\rho_{\mathrm{val}}(r)]$ are the exchange-correlation and Hartree functional potentials due to the pseudovalence density, $\rho_{\mathrm{val}}(r)$.

The two-component semilocal pseudopotential operators are of the following form,

$$\hat{V}^{\mathrm{ps}} = \sum_{j,l,m_j} |\mathscr{Y}_l^{j,m_j}\rangle \, V_{jl}^{\mathrm{ion}}(r) \langle \mathscr{Y}_l^{j,m_j}| \tag{18}$$

One can decompose V_{jl}^{ion} as follows

$$V_{jl}^{\mathrm{ion}}(r) = \bar{V}_l(r) + \langle \sigma \cdot \mathbf{L} \rangle_{jl} V_l^{\mathrm{so}}(r) \tag{19}$$

where $\bar{V}_l$ denotes the l-averaged scalar relativistic pseudopotential, V_l^{so} denotes the radial spin−orbit pseudopotential component, and

$$\langle \sigma \cdot \mathbf{L} \rangle_{jl} = \begin{cases} \dfrac{l}{2} & \text{for } j = l + \dfrac{1}{2} \\ -\dfrac{l+1}{2} & \text{for } j = l - \dfrac{1}{2} \end{cases} \tag{20}$$

are the expectation values of the spin−orbit operator. Given that

$$V_{j=l+1/2,l}^{\mathrm{ion}}(r) = \bar{V}_l(r) + \frac{l}{2} V_l^{\mathrm{so}}(r) \tag{21}$$

$$V^{\text{ion}}_{j=l-1/2,l}(r) = \bar{V}_l(r) - \frac{1}{2}(l+1)V^{\text{so}}_l(r) \tag{22}$$

one can readily solve for the l-averaged and spin−orbit potentials.

$$\bar{V}_l = \frac{(l+1)V^{\text{ion}}_{j=l+1/2,l} + lV^{\text{ion}}_{j=l-1/2,l}}{2l+1} \tag{23}$$

$$V^{\text{so}}_l(r) = \frac{2}{2l+1}(V^{\text{ion}}_{j=l+1/2,l} - V^{\text{ion}}_{j=l-1/2,l}) \tag{24}$$

The radial ionic pseudopotentials and radial pseudowave functions are shown in Figures 1 and 2, respectively, for the case of bismuth. For the s and d states, all three methods are in complete agreement. For the p states, the ZORA and Dirac results agree almost exactly while the scalar Dirac clearly shows the effects of the averaging procedure. The results for j-averaged pseudopotential, $\bar{V}_l$, are shown in Figure 3. Note that all three methods produce identical functions for this atom. In Figure 4, one sees that for the spin−orbit term, V^{so}_l, Dirac and ZORA once again provide identical results. Analogous to the form given by Hamann,[29] the full pseudopotential matrix element assumes the form,

$$\langle\phi_1|\hat{V}^{\text{ps}}|\phi_2\rangle = \sum_{j,l,m_j}\int \mathrm{d}^3r \int \mathrm{d}^3r'\ \phi_1^*(\mathbf{r})\mathscr{Y}_l^{j,m_j}(\hat{r})V^{\text{ion}}_{jl}(r)\delta(r-r')\mathscr{Y}_l^{j,m_j\ddagger}(\hat{r}')\phi_2(r') \tag{25}$$

where the ion is centered at the origin. The Kleinman−Bylander expansion of this expression[31] is given by

$$\langle\phi_1|\hat{V}^{\text{ps}}(\mathbf{r})|\phi_2\rangle = \langle\phi_1|V^{\text{local}}|\phi_2\rangle + \sum_{\kappa,J_3} C_\kappa\langle\phi_1|V^{\text{nonlocal}}_{jl}|\phi_2\rangle \tag{26}$$

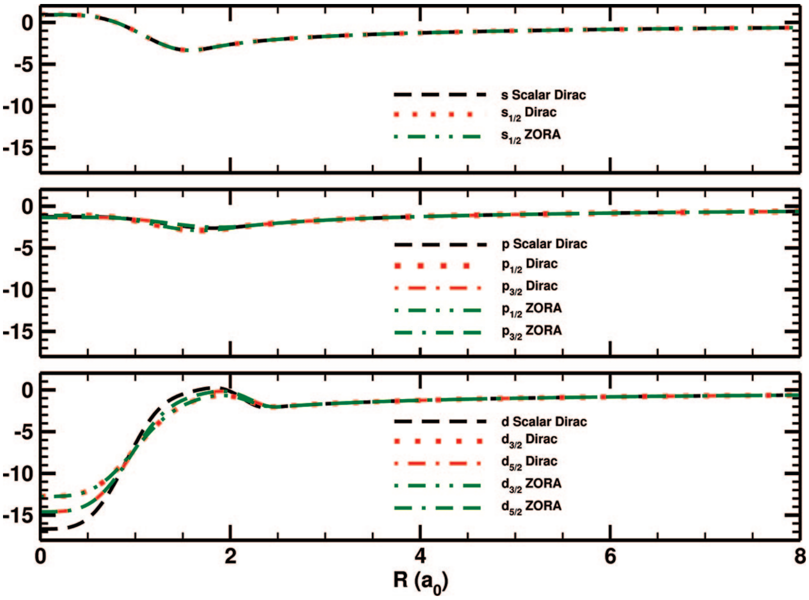


Figure 1. Ionic spin angular pseudopotentials (**L** = s, p, d) for Bismuth using Scalar Dirac, Dirac, and ZORA.

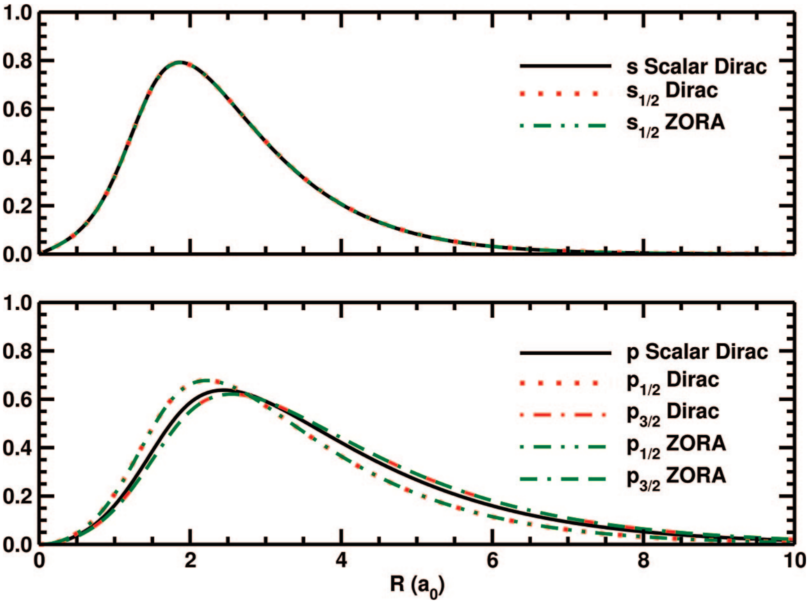


Figure 2. Radial component of the spin angular pseudowave functions (**L** = s, p) for bismuth using Scalar Dirac, Dirac, and ZORA.

where ϕ_1 and ϕ_2 are Pauli spinor planewave functions of the form

$$\phi_1(\mathbf{r}) = j_l(kr)\mathcal{Y}_l^{j,m_j}(\hat{\mathbf{r}}) \tag{27}$$

and the local potential matrix element is taken to be that of the nonrelativistic form, ($V^{\mathrm{local}}(r) = \bar{V}_{l=l_{\mathrm{local}}}(r)$),

$$\langle\phi_1|V^{\mathrm{local}}|\phi_2\rangle = \int \mathrm{d}^3r\,\phi_1(\mathbf{r})\bar{V}_l(r)Y_l^m(\hat{r})\phi_2(\mathbf{r}) \tag{28}$$

and

$$\langle\phi_1|V_{jl}^{\mathrm{nonlocal}}|\phi_2\rangle = \int \mathbf{d}^3r\,\phi_1^*(\mathbf{r})\mathcal{Y}_l^{j,m_j}(\hat{r})Q_{jl}^{\mathrm{ion}}(r)\int \mathrm{d}^3r'\,\mathcal{Y}_l^{j,m_j\ddagger}(\hat{r}')Q_{jl}^{\mathrm{ion}*}(r')\phi_2(\mathbf{r}') \tag{29}$$

The scalar radial term, $Q_{jl}^{\mathrm{ion}}(r)$, is defined as

$$Q_{jl}^{\mathrm{ion}}(r) = v_{jl}(r)\frac{w_{jl}(r)}{r} \tag{30}$$

the term C_{jl} is

$$C_{jl} = [4\pi\int_0^\infty r\,\mathrm{d}r\,Q_{jl}^{\mathrm{ion}}w_{jl}(r)]^{-1} \tag{31}$$

and the nonlocal pseudopotential term, $v_{jl}(r) = V_{jl}^{\mathrm{ion}}(r) - V^{\mathrm{local}}(r)$. It should be pointed out that, in this prescription, neither of the two nonlocal potentials corresponding to the orbital angular momentum eigenstate l_{local} vanish, which is the case for the nonrelativistic prescription of standard pseudopotentials. In fact, each has parts of the spin−orbit interaction. Also, one must note that the local potential must be chosen carefully to ensure that C_{jl} does not diverge. For example, if integral containing v_{jl} that appears in the denominator is small this coefficient will be very large. Any error in the projector will be unreasonably amplified. The approach taken here is to set C_{jl} to zero when this occurs. Another noteworthy fact is that the nonlocal potential is an outer product of two spinor wave functions effectively making it a 2 × 2 complex matrix.

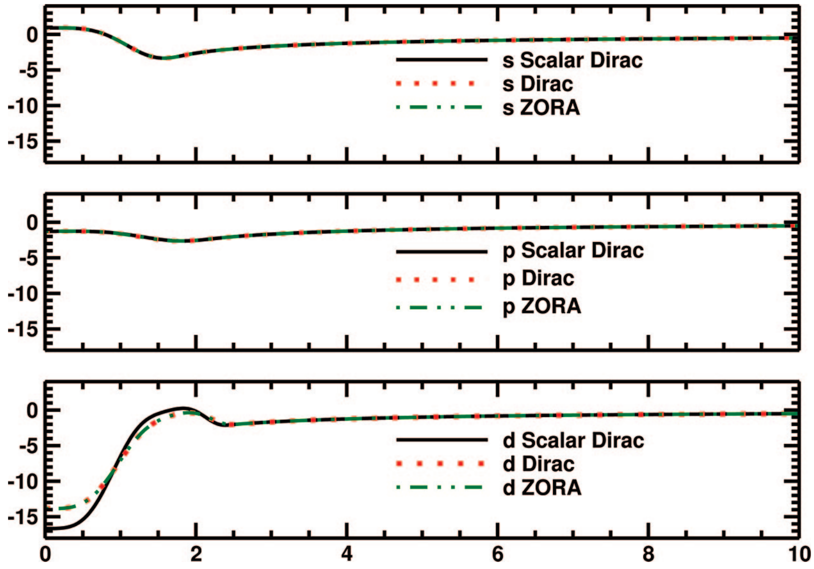


Figure 3. *l*-Averaged radial pseudopotentials (**L** = s, p, d) for bismuth using Scalar Dirac, Dirac, and ZORA.

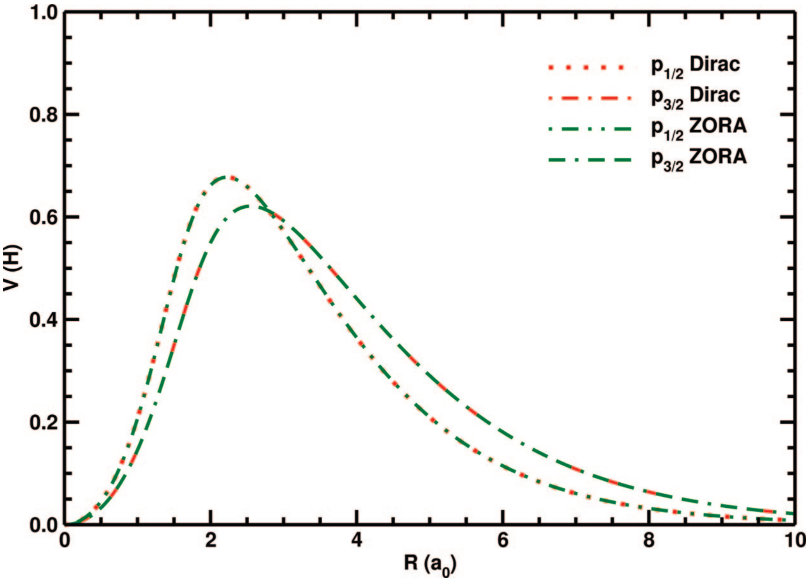


Figure 4. Radial spin−orbit component of pseudopotential (L = p) for bismuth using Scalar Dirac, Dirac, and ZORA.

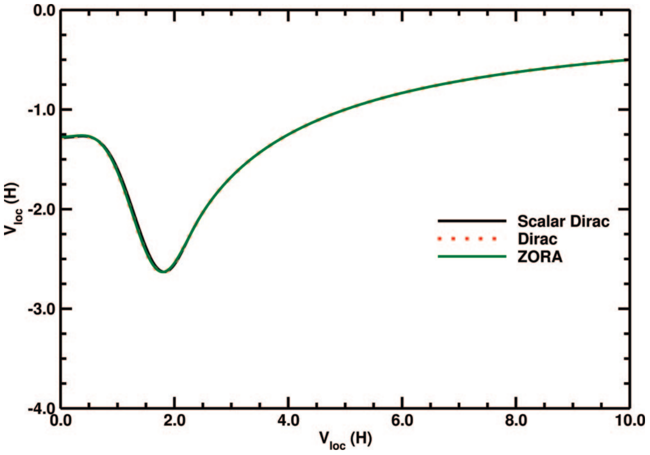


Figure 5. Radial local potential (**L** = p) for bismuth using Scalar Dirac, Dirac, and ZORA.

These quantities for the three methods are shown in Figure 6. As was the case, for the ionic pseudopotentials, the p and d states show very close to exact agreement between ZORA and Dirac, while the scalar Dirac method is clearly an averaged outcome for the two values of j for each l. The crucial feature of this new pseudopotential is that the projectors are two-component spinors. Thus, when the Kleinman−Bylander expansion is done, we are expanding in terms of the eigenvalues of the spin angular functions instead of scalar functions which are eigenfunctions of the angular momentum operator. Since our projectors are diagonal with respect to the spin−orbit term $\sigma\cdot\mathbf{L}$, the implementation of this term is direct and compiler can optimize the code more efficiently. Note that the spin−orbit interaction is now coupled in the nonlocal interaction in a non-perturbative fashion and that the spin-components of an arbitrary wave function are mixed by this interaction term. This approach allows the inclusion of spin−orbit effects in an accurate manner with only a small increase in computational costs. Since a spinor wave function with two components is needed for each eigenstate, the number of planewaves is effectively doubled (compared to a gamma point code, the cost is four times more components). Also the nonlocal potential now requires four times the amount of memory of a conventional code. While this increase is certainly not small, it is not prohibitive. The extra computational cost can be offset on parallel machines, since the nonlocal pseudopotential is easy to parallelize. The accuracy comes from the fact that the large majority of the spin−orbit effects comes from the core electrons. Assuming norm-conservation conditions are imposed, the electric field near each atom at or beyond the core cutoffs is nearly identical to one would find for an all-electron calculation since the core charge of the ion is preserved by the pseudopotential.

III. Applications

The Gaussian basis set calculations were performed with the development version of the NWDFT module of NWChem.[21] Molecules involving I, H, F, and Br were calculated using the 6-311G** basis set, and for those involving Au and Bi, large uncontracted basis sets[32] Au(26s20p19d16f) and Bi(26s23p19d14f) were used. Numerical integration was performed using the extremely fine integration grids available within NWChem.[25] All calculations were performed with spherical basis sets. In this paper, we have performed calculations with three different exchange-correlation functionals (LDA, PBE96, B3LYP). In each of the cases, the spin−orbit effects are discerned by comparing the spin-free case to that of the two-component spinor approach.

The plane wave calculations were done with the development version of the NWPW module of NWChem.[21] A cubic box of $20a_0$ was chosen with a cutoff energy of 101 Rydbergs for the wave function and 202 Rydbergs for the density for the PBE96 exchange-correlation functional calculations. Tests using this functional were also done using a smaller density grid (equal to that of the wave function) and showed insignificant differences for the reported properties of the molecules in the test suite confirming that the cutoff is adequate. However, changes in the vibrational frequencies of the gold dimer and gold hydride were noted. For comparison to the more accurate GGA(PBE96) results, calculations using the LDA or SVWN functional were completed using a cutoff of 101 Rydbergs for both the wave function and the density. The pseudopotential generation was done with the SVWN exchange-correlation functional. Since each test case is a closed shell molecular dimer, the initial wave function is taken to have multiplicity of one. The core radii, electronic configuration, maximum angular momentum, local pseudopotential used for the Kleinman-Bylander expansion, and the exact type of pseudopotential used for each molecule are given in Tables 4 and 5. Except for the H and F atoms, the core radii were generated based upon the default criteria of Hamann or Troullier−Martin for the corresponding pseudopotential.[29,30] This very stringent criteria was used in order to make the pseudopotentials as transferable as possible. It should be noted that these criteria will result in slightly different core radii for scalar relativistic and relativistic pseudopotentials. By default, the local pseudopotential used in the Kleinman−Bylander expansion was taken to be the pseudopotential for the maximum angular momentum. However, when ghost states were found a different angular momentum was chosen. The evidence for such a ghost state was taken to be the presence of a molecular bound state energy with a very low energy compared to the energy of the lone atom. For example, in Br the ghost state was 3 times lower in energy than the corresponding atomic orbital energy. In general, Hamann type pseudopotentials were used as a default and for all the relativistic two component pseudopotentials. Troullier−Martins type pseudopotentials were used only for cases where we had difficulty making the Hamann potentials work. However in most cases, there was no discernible difference in the quality between Hamann and Troullier−Martins potentials. In the radial Dirac or radial ZORA solver, the occupancies for the relativistic orbitals were determined by multiplying the total occupancy of the nonrelativistic orbital by the fraction of states for the relativistic orbital. For example, for a $5p^5$ state there are 6 states, the $5p_{1/2}$ state has 2 states while the $5p_{3/2}$ would have 4. Thus, the number of electrons in the $5p_{1/2}$ is $5 \times 2/6 = 5/3$, and the number of electrons in the $5p_{3/2}$ is $5 \times 4/6 = 10/3$. The yields a spherical potential for the relativistic case.

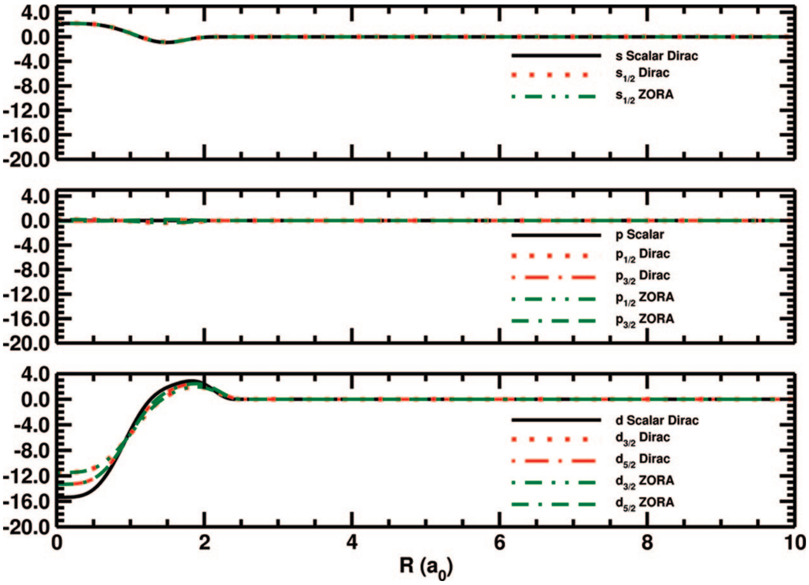


Figure 6. Radial components of the nonlocal spin-angular projectors (**L** = s, p, d) for bismuth using Scalar Dirac, Dirac, and ZORA.

Table 1. Bond Lengths (Å)

	I_2	IF	HI	Br_2	Bi_2	AuH	Au_2
experiment[34]	2.67	1.91	1.61	2.28	2.66	1.52	2.42
PW-PBE96-HGH-SO	2.67	1.97	1.61	2.28	2.57	1.54	2.53
PW-LDA-SR	2.67	1.87	1.63	2.29	2.62	1.60	2.47
PW-PBE96-SR	2.68	1.99	1.60	2.28	2.63	1.52	2.50
PW-PBE96-SO	2.61	1.90	1.62	2.27	2.57	1.59	2.61
GB-PBE96-ZORA-SR	2.71	1.94	1.63	2.32	2.65	1.55	2.54
GB-PBE96-ZORA-SO	2.73	1.95	1.63	2.33	2.68	1.55	2.52
GB-LDA-ZORA-SR	2.68	1.92	1.62	2.29	2.61	1.54	2.46
GB-LDA-ZORA-SO	2.70	1.92	1.63	2.29	2.64	1.54	2.45
GB-B3LYP-ZORA-SR	2.73	1.93	1.62	2.33	2.64	1.55	2.54
GB-B3LYP-ZORA-SO	2.75	1.93	1.62	2.33	2.69	1.55	2.53
GAMESS-UK-PBE96-GGA-SR[33]	2.70	1.91	1.61		2.61	1.53	2.46
GAMESS-UK-PBE96-GGA-SO[33]	2.72	1.92	1.62		2.64	1.53	2.45

Table 2. Vibrational Frequencies (cm^{-1})

	I_2	IF	HI	Br_2	Bi_2	AuH	Au_2
experiment[34]	215	610	2309	325	173	2305	191
PW-PBE96-HGH-SO	216	634	2263	327	199	2212	173
PW-LDA-SR	224	616	2249	331	187	3531	549
PW-PBE96-SR	217	610	2270	322	206	2250	185
PW-PBE96-SO	225	620	2315	323	201	2006	167
GB-PBE96-ZORA-SR	209	607	2247	308	192	2204	173
GB-PBE96-ZORA-SO	196	603	2228	305	177	2206	175
GB-LDA-ZORA-SR	221	639	2261	325	202	2267	193
GB-LDA-ZORA-SO	207	635	2239	321	168	2295	195
GB-B3LYP-ZORA-SR	209	626	2287	310	198	2186	169
GB-B3LYP-ZORA-SO	198	623	2268	307	166	2241	173

Table 3. Dissociation Energies (eV)

	I_2	IF	HI	Br_2	Bi_2	AuH	Au_2
experiment[34]	1.56	2.92	3.20	2.00	2.03	3.36	2.31
PW-LDA-SR	2.69	3.97	3.90	3.06	3.38	3.99	3.44
PW-PBE96-SR	2.13	2.95	3.40	2.59	4.20	3.28	2.40
PW-PBE96-SO	1.30	2.80	3.05	2.03	4.05	2.72	1.87
GB-PBE96-ZORA-SR	2.38	3.50	3.03	2.43	2.62	3.53	2.55
GB-PBE96-ZORA-SO	1.87	3.29	2.87	2.22	2.08	3.52	2.52
GB-LDA-ZORA-SR	2.92	4.30	3.76	3.10	3.51	3.64	3.06
GB-LDA-ZORA-SO	2.26	4.01	3.42	2.82	2.85	3.62	3.02
GB-B3LYP-ZORA-SR	2.52	3.05	3.19	2.11	2.45	3.48	3.48
GB-B3LYP-ZORA-SO	1.96	2.81	2.96	1.86	1.83	3.48	3.48

The Dirac solver was used for all calculations except for those noted below. To the best of the authors knowledge, this is first test of a norm conserving spin−orbit pseudopotential for most of these molecular dimers (except for a very recent paper[20] which calculated Au_2 and AuH). Hence for most of these molecules, the results can only be compared with the HGH spin−orbit pseudopotential [13]and Gaussian basis set ZORA DFT results. For Bi_2, I_2, and Au_2, tests were conducted at a lower cutoff (32 × 32 × 32 Fourier grid) to determine the optimal values for the local potential, maximum angular momentum, and pseudopotential type.

The results for the bond lengths are presented in Table 1 while vibrational frequencies are presented in Table 2. The dissociation energies are given in Table 3. For most of these molecules, the spin−orbit effect is small. This is expected since these molecules are closed shell. One will note that in some cases, the spin−orbit does not seem to improve things, though the correction is not significant and this disagreement might easily result from other factors. In general, one might expect that the spin−orbit interaction will drive the bond length to be slightly longer. The rationale for this type of effect is that the spin−orbit interaction splits the $2(2l + 1)$ degenerate atomic orbitals for a given l into two separate groups having total

Table 4. Planewave Pseudopotentials

atom	local	L_{max}	config	type
Br	d	d	$4s^24p^5$	TM
Bi	d	d	$6s^26p^26d^1$	Ham
I	d	d	$5s^25p^5$	Ham
Au	s	d	$5s^25p^65d^{10}$	TM
F	d	d	$2s^22p^5$	Ham
H	p	p	$1s^1$	Ham
Br	p	d	$4s^24p^5$	Ham
Bi	p	d	$6s^26p^3$	TM
I	p	d	$5s^25p^5$	Ham
Au	s	d	$5s^25p^65d^{10}$	Ham

Table 5. Planewave Radial Cutoffs for Pseudopotentials (a_0)

	Br	Bi	I	H	Au	F
s	1.405	1.114	1.0130	0.800	0.844	0.700
p	1.657	1.466	1.2062	0.800	0.951	0.700
d	1.657	1.958	1.4733		1.215	0.700
$s_{1/2}$	0.843	1.866	1.013		0.506	
$p_{1/2}$	0.979	2.211	1.171		0.535	
$p_{3/2}$	0.999	2.542	1.224		0.585	
$d_{3/2}$	1.706	2.542	1.466		0.714	
$d_{5/2}$	1.715	2.542	1.481		0.740	

angular momenta of $j = l + {}^1/_2$ and $j = l - {}^1/_2$. The $j = l - {}^1/_2$ component is $2l$ degenerate and is slightly pushed in toward the nucleus with respect to the nonrelativistic or scalar relativistic reference, while the $j = l + {}^1/_2$ component is $2l + 2$ degenerate and pushed away from the nucleus. This effect can be seen in Figure 2 for the Bi pseudowave functions. The net result is that one may expect the bonding region to move away from the nuclei and create a slightly longer bond when spin−orbit effects are included. This type of effect is seen for the I_2 and Bi_2 dimers in the Gaussian ZORA basis set calculations; however, this effect is not readily observed for the other dimers. This type of effect was not observed in the planewave basis set calculations. The most likely reason for this is that the differences associated with the subtle differences in parameters for generating the scalar relativistic and spin−orbit relativistic pseudopotentials are large enough to wash out this effect. Similar results have been observed in a previous plane-wave study by Oda and coworkers[19] using ultrasoft fully relativistic pseudopotentials for solid-state calculations. For Pb and Pt an increase in the equilibrium lattice constant is observed while for Au a decrease is noted. There are many competing effects that do not directly derive from the spin−orbit potential which might have influences on the bond lengths such as the image interaction or box effects. For most cases, there is good agreement between experiment and theory. The Gaussian ZORA appears to yield much smaller spin−orbit effects than the relativistic Dirac planewave pseudopotential. Comparing these results to that of the all-electron DFT ZORA implementation in GAMESS-UK,[33] one sees very close agreement. The differences are less than five percent in all cases. Also one sees close agreement with the results for the planewave method with HGH pseudopotentials (note these results were generated by NWChem and, while very close to those published originally in the HGH paper,[13] are not identical).

For dissociation energies, the same trend is observed as with the bond lengths. As one would expect, LDA tends to give better geometries while GGA yields better dissociation energies. The inclusion of spin−orbit seems to improve the dissociation energy somewhat. Once again the Gaussian ZORA agree well with the results from Faas et al.[33] As for the harmonic frequencies, most of these follow a trend similar to what is observed in other quantities. In agreement with physical intuition, a shorter bond length leads to a higher harmonic frequency and vice versa. One notable exception was found in the planewave calculations for the dimers IF, AuH, and Au_2 where initial tests yielded very large frequencies even when the dissociation energies and bond lengths were nearly equal the experimental values. The problem was found to be that F and Au both have very small compact pseudowave functions using the default parameters of Hamann or Troullier−Martin. It was found that the frequencies for Au_2 and AuH could be made to approach the experimental values when larger cutoff radii were used to define the Au pseudopotential. Unfortunately, increasing the cutoff radii resulted in large errors in the dissociation energy. The lack of transferability of the Au pseudopotential is likely the result of the core states of Au being highly polarizable, and a very large cutoff is needed to resolve the compact core functions. To address this effect, the 5s, 5p, and 5d orbitals were included in the valence space. Even though these pseudopotentials produced better results than without the 5s and 5p state. Table 1 still shows a need for improvement. For Gaussian basis sets, a very diffuse set is needed to properly represent the metallic bonding aspects of this molecule. Nevertheless, the results are reasonable. Past calculations such a that of Naveh et al. found bond lengths of 2.55 for this dimer.[20] Scalar ZORA calculation by van Lenthe et al.[8] found 2.458 and 2.517 Å for LDA and GGA, respectively. The Gaussian ZORA DFT results of 2.45 and 2.52 Å are very close to the previously reported results. The planewave result of 2.50 Å is nearly the same as the scalar ZORA; however, the inclusion of spin−orbit effects actually makes the agreement with the experimental bond length, 2.42 Å, slightly worse. Note that an experimental value of 2.47 Å has also been reported. For AuH, the calculated values of 1.539 for scalar ZORA, 1.55 for the real-space method, 1.55 for the present Gaussian ZORA DFT, and 1.52 for the present planewave method all compare somewhat well with the experimental value of 1.52 Å. Once again for the gold pseudopotential, the planewave results are slightly elongated with the addition of the spin−orbit coupling for bond lengths and the addition of spin−orbit in the planewave case actually leads to slightly less accurate predictions. This addition has no visible effect in the Gaussian ZORA DFT approach.

IV. Conclusions

Relativistic two-component methods have been implemented in the NWChem program within the molecular Gaussian DFT and pseudopotential planewave DFT modules. These methods have been shown to produce reliable results for the small closed-shell molecules we have considered in this paper. These two methods are also complementary. The all-electron Gaussian ZORA DFT method allows one to calculate relativistic properties of molecules with albeit expensive costs compared with RECP DFT while the planewave method allows a more economical approach. Also the planewave method will allow calculations of surfaces, solids, and solutions in a completely ab initio manner, while the Gaussian basis set will allow the calculation of properties associated with core electrons. These two implementations will be available within the quantum chemistry program NWChem. The planewave method builds on previous implementations[13−19] of similar pseudopotentials and opens the possibility of performing spin−orbit calculations on thousands of processors. Likewise, the Gaussian basis set all-electron ZORA implementation provides similar capabilities. Future work includes the application of these methods to open shell heavy atom compounds, in particular, actinide and lanthanide compounds. While these methods are certainly cheap compared with more traditional approaches, they provide chemists with the ability to include a realistic description of relativistic effects such as spin−orbit splitting, molecular orbital hybridization, and core effects.

Acknowledgment. This research was supported by the BES Heavy Element Chemistry Program of the U.S. Department of Energy, Office of Science (No. DE-AC06-76RLO 1830). E.J.B. would like to acknowledge the DOE BES Geosciences Program for helping support the development of the AIMD and analysis programs. The Pacific Northwest National Laboratory is operated by the Battelle Memorial Institute. Some of the calculations were performed on the MPP2 computing system at the Molecular Science Computing Facility in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at PNNL. EMSL operations are supported by the DOE's Office of Biological and Environmental Research. We also wish to thank the Department of Energy for a grant of computer time at the National Energy Research Scientific Computing Center (Berkeley, CA).

References

(1) Barysz, M. *J. Chem. Phys.* **2000**, *113*, 4003.

(2) Dyall, K. G. *J. Chem. Phys.* **1994**, *100*, 2118.

(3) Douglas, M.; Kroll, N. M. *Ann. Phys. (N.Y.)* **1974**, *82*, 89.

(4) Wolf, A.; Reiher, M.; Hess, B. A. *J. Chem. Phys.* **2002**, *117*, 9215.

(5) Hess, B. A. *Phys. Rev. A* **1985**, *32*, 756. **1986**, *33*, 3742. .

(6) Chang, Ch.; Pelissier, M.; Durand, Ph. *Phys. Scr.* **1986**, *34*, 394.

(7) van Lenthe, E.; Baerends, E. J.; Snijders, J. G.; Zhang, J. *J. Chem. Phys.* **1993**, *99*, 4597.

(8) Van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1994**, *101*, 9783.

(9) van Lenthe, E.; Snijders, J. G.; Baerends, E. J. *J. Chem. Phys.* **1996**, *105*, 6505.

(10) Pacios, I. F.; Christiansen, P. A. *J. Chem. Phys.* **1985**, *82*, 2664.

(11) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.

(12) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270.

(13) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B* **1998**, *58*, 3641.

(14) Dal Corso, A.; Conte, A. M. *Phys. Rev. B* **2005**, *71*, 115106.

(15) Dal Corso, A. *Phys. Rev. B* **2007**, *76*, 054308.

(16) Hemstreet, L. A.; Fong, C. Y.; Nelson, J. S. *Phys. Rev. B* **1993**, *47*, 4238.

(17) Theurich, G.; Spaldin, N. *Phys. Rev. B* **2001**, *64*, 073106.

(18) Theurich, G.; Spaldin, N. *Phys. Rev. B* **2002**, *66*, 115208.

(19) Oda, T.; Hosokawa, A. *Phys. Rev. B* **2005**, *72*, 224428.

(20) Naveh, D.; Kronik, L.; Tiago, M. L.; Chelikowsky, J. R. *Phys. Rev. B* **2007**, *76*, 153407.

(21) Bylaska, E. J.; de Jong, W. A.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Apra, E.; Windus, T. L.; Hammond, J.; Nichols, P.; Hirata, S.; Hackler, M. T.; Zhao, Y.; Fan, P.-D.; Harrison, R. J.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Wu, Q.; Van Voorhis, T.; Auer, A. A.; Nooijen, M.; Brown, E.; Cisneros, G.; Fann, G. I.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J. A.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Pollack, L.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. *NWChem, A Computational Chemistry Package for Parallel Computers*; version 5.1, Pacific Northwest National Laboratory: Richland, WA, 2007; 99352−0999.

(22) van Wüllen, C. *J. Chem. Phys.* **1998**, *109*, 392.

(23) van Lenthe, E.; Ehlers, A.; Baerends, E.-J. *J. Chem. Phys.* **1999**, *110*, 8943.

(24) Filatov, M.; Cremer, D. *J. Chem. Phys.* **2005**, *122*, 44104.

(25) (a) Murray, C. W.; Handy, N. C.; Laming, G. L. *Mol. Phys.* **1993**, *78*, 997. (b) Mura, M. E.; Knowles, P. J. *J. Chem. Phys.* **1996**, *104*, 9848. (c) Treutler, O.; Alrhichs, R. *J. Chem. Phys.* **1995**, *102*, 346. (d) Lebedev, V. I.; Laikov, D. N. *Doklady Mathematics* **1999**, *366*, 741. (e) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053. (f) Stratmann, R. E.; Scuseria, G.; Frisch, M. J. *Chem. Phys. Lett.* **1996**, *257*, 213. (g) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. *Chem. Phys. Lett.* **1993**, *209*, 506–512.

(26) Van Lenthe, J. H.; Faas, S.; Snijders, J. G. *Chem. Phys. Lett.* **2000**, *328*, 107.

(27) Van Lenthe, J. H.; van Lingren, J. N. J. *Int. J. Quantum Chem.* **2006**, *106*, 2525.

(28) Sakurai, J. J. *Advanced Quantum Mechanics*, 10th ed.; Addison Wesley Publishing Co.: Reading, MA, 1967.

(29) Hamann, D. R. *Phys. Rev. B* **1989**, *40*, 2980.

(30) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993.

(31) Kleinman, L.; Bylander, D. M. *Phys. Rev. Lett.* **1982**, *48*, 1425.

(32) Tatewaki, H.; Koga, T. *Chem. Phys. Lett.* **2000**, *328*, 473.

(33) Faas, S.; Van Lenthe, J. H.; Hennum, A. C.; Snijders, J. G. *J. Chem. Phys.* **2000**, *113*, 4052.

(34) Huber, K. P.; Herzberg, G. (data prepared by Gallagher, J. W.; Johnson, R. D., III.) Constants of Diatomic Molecules. In NIST Chemistry WebBook, NIST Standard Reference Database Number 69; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, June 2005; http://webbook.nist.gov (accessed June 1, 2008).

CT8002892

JCTC Journal of Chemical Theory and Computation

Auxiliary Basis Sets for Density-Fitted MP2 Calculations: Correlation-Consistent Basis Sets for the 4d Elements

J. Grant Hill[†] and James A. Platts*

School of Chemistry, Cardiff University, Park Place, Cardiff CF10 3AT, U.K.

Received October 1, 2008

Abstract: Auxiliary basis sets for use in density fitting second-order Møller−Plesset perturbation theory and other correlated ab initio methods have been developed for the 4d transition metal elements Y−Tc and Rh−Pd (sets for Ru, Ag, and Cd are already available), to be used in conjunction with the correlation consistent basis sets with pseudopotentials cc-pV*n*Z-PP and aug-cc-pV*n*Z-PP. Correlation energy calculations for a test set of small- to medium-sized transition metal complexes encompassing a variety of oxidation states show that the error in using these auxiliary basis sets is around 3−4 orders of magnitude smaller than the error due to orbital basis set size. The effect of truncating the auxiliary basis sets to remove higher angular momentum functions is also considered.

1. Introduction

With the advent of correlation consistent (cc) basis sets for the first and second row transition metals,[1−3] the application of correlated ab initio methods such as second-order Møller−Plesset perturbation (MP2) and coupled-cluster (CC) theories to transition metal complexes has become more attractive as a systematically improvable increase in basis set toward the basis set limit is now possible.[4−9] Even with the use of pseudopotentials to replace the chemically inert core electrons, the large size of many typical transition metal complexes means that conventional post-Hartree−Fock (HF) calculations rapidly become prohibitively expensive. The density fitting (DF) approximation[10] of electron repulsion integrals as applied to some post-HF methods expands orbital product densities in an optimized auxiliary density fitting basis set, and in the case of MP2 typically reduces calculation times by around an order of magnitude with a negligible loss of accuracy.

Also known as the resolution-of-the-identity (RI) approximation in some circles, density fitting has been implemented for a wide range of electronic structure theories, with correlated ab initio examples including MP2,[11,12] coupled-cluster,[13] and approximate coupled-cluster singles-and-doubles model CC2.[14] Lately, progress has also been made in combining density fitting with local electron correlation techniques in order to produce further computational savings.[15−17] Unlike density-fitted HF or density functional theory, the same optimized auxiliary fitting bases can be successfully employed across the range of these post-HF methods, with the exception that density fitted local CC methods can benefit from selecting a larger auxiliary basis from the same series. An alternative to the use of optimized auxiliary basis sets is to automatically generate the auxiliary basis[18] in the same manner as that which can be employed for Cholesky decomposition of electron integrals.[19] Although these sets do not hold any bias toward a particular quantum chemical method they are generally larger, and therefore incur a higher computational cost, than sets specifically optimized for post-HF calculations.

Dunning originally optimized cc basis sets for first-row atoms.[20] Further work also provided cc sets for p block elements,[21−24] along with options for the description of anions and noncovalent interactions,[25,26] among many others. In their simplest form (valence only correlated, without being augmented with additional diffuse functions) this family of basis sets is usually denoted as cc-pV*n*Z, where the cardinal number, *n*, refers to the ζ level of the basis set (D, T, Q, 5, and so on). This increase in cardinal number corresponds to a systematic approach to the complete basis set (CBS) limit, and when combined with the hierarchy of post-HF ab initio methods this leads to an obvious progression toward more

---

* To whom correspondence should be addressed. E-mail: platts@cf.ac.uk.

[†] Present address: Department of Chemistry, Washington State University, Pullman, Washington 99164-4630.

10.1021/ct8005584 CCC: $40.75

accurate calculations. DF auxiliary basis sets for correlated ab initio methods (herein referred to as MP2 fitting basis sets) have been optimized specifically to be used with cc orbital basis sets for the majority of lighter elements,[27−29] enabling systematic improvement of calculations at a significantly reduced computational cost.

In the present work MP2 fitting basis sets are optimized for the 4d transition metal elements Y−Tc and Rh−Pd for use with the recently proposed cc-pV*n*Z-PP and aug-cc-pV*n*Z-PP,[2] where PP signifies the use of a pseudopotential, orbital basis sets. Similar MP2 fitting sets for the remaining 4d elements Ru[30] and Ag−Cd[31] are available elsewhere.

2. Basis Set Construction and Optimization

The cc-pV*n*Z-PP and aug-cc-pV*n*Z-PP (where n = D, T, Q, and 5) orbital basis sets and pseudopotentials (replacing 28 of the core electrons) used for constructing and optimizing the MP2 fitting basis sets were obtained from the William R. Wiley Environmental Molecular Sciences Laboratory Basis Set Exchange.[32] All optimizations were performed with the analytic gradients for the optimization of auxiliary basis sets[28] implemented in the RICC2 module[14,33] of the TURBOMOLE package.[34,35] Two guidelines previously employed for other cc MP2 fitting basis sets were adopted in the construction of the current sets, namely, that the number of functions in the auxiliary set should be less than four times the number of functions in the orbital basis and that the number of auxiliary functions in each specific basis set should be the same for all atoms considered.[27] Because of the variety of oxidation states any given transition metal element may be found in, the accuracy of MP2 fitting basis sets was verified by comparing the mean error, standard deviation, maximum error, and mean unsigned error introduced by the DF approximation at the MP2 level with the same statistics for the error in orbital basis set incompleteness for conventional MP2 over a test set of representative transition metal complexes. The later error was assessed as the difference in energy between a given orbital basis set and an estimate of the CBS limit obtained via a basis set extrapolation of quadruple- and quintuple-ζ energies utilizing the formula $E_n = E_{\mathrm{CBS}} + An^{-3}$ of Helgaker and co-workers,[36,37] where n is the cardinal number of the basis set. For the density fitting error to be regarded as negligible compared to the orbital basis set incompleteness all of the statistical measures should be at least 2 orders of magnitude smaller than the orbital basis error outlined above.[27,29]

Table 1. Ratio of the Number of GTOs in the MP2 Fitting Basis Sets Compared to the Number of GTOs in the Generally Contracted Orbital Basis

	orbital basis	MP2 fitting basis	ratio of functions
cc-pVDZ-PP	[4s4p3d1f]	(8s8p6d6f4g2h)	2.8
cc-pVTZ-PP	[5s5p4d2f1g]	(10s10p9d7f6g3h2i)	2.8
cc-pVQZ-PP	[6s6p5d3f2g1h]	(11s11p10d8f7g5h3i2k)	2.5
cc-pV5Z-PP	[7s7p6d4f3g2h1i]	(12s12p11d10f9g6h5i3k2l)	2.3
aug-cc-pVDZ-PP	[5s5p4d2f]	(9s9p7d7f5g3h)	2.5
aug-cc-pVTZ-PP	[6s6p5d3f2g]	(11s11p10d8f7g4h3i)	2.5
aug-cc-pVQZ-PP	[7s7p6d4f3g2h]	(12s12p11d9f8g6h4i3k)	2.2
aug-cc-pV5Z-PP	[8s8p7d5f4g3h2i]	(13s13p12d11f10g7h6i4k3l)	2.1

To ensure that the MP2 fitting basis sets produced satisfactory results for transition metal complexes with a range of different oxidation states, a procedure similar to that outlined previously[28,29] was followed. The elements Y, Zr, and Tc, which are generally found in a single oxidation state, had exponents divided into two categories, and initially all exponents were optimized for a cation of the element corresponding to the oxidation state, before exponents that contribute to the outermost atomic orbitals (AOs) and with high angular momentum were reoptimized for the neutral atom with all other exponents fixed. This layered optimization is modified for the atoms Nb, Mo, Rh, and Pd, where the initial optimization is for the cation producing the highest oxidation state generally found in chemical compounds and an additional intermediate set of exponents are optimized for the cation typically corresponding to the most commonly found oxidation state of the element. For a more complete explanation of the procedure, readers are referred to ref 29. For the aug-cc-pV*n*Z-PP sets, the corresponding cc-pV*n*Z-PP auxiliary basis is held fixed while an additional diffuse exponent for each angular momentum quantum number is optimized for the anion of the element. For the testing of the MP2 fitting basis sets all non-4d elements are treated with the cc-pV*n*Z or aug-cc-pV*n*Z AO orbital basis sets consistent with the cardinal number used for the transition metal element, with the corresponding auxiliary basis sets of Weigend et al.[27] employed in the density fitting.

3. Results and Discussion

The total number of Gaussian-type orbitals (GTOs) included in the MP2 fitting basis sets, along with the number of GTOs in the corresponding generally contracted orbital basis, are presented in Table 1. It should be noted that the number of functions is the same as for the MP2 fitting basis sets available for Ru, Ag, and Cd. It can be seen from the ratio of the orbital and auxiliary functions that at no point does the number of auxiliary functions approach the upper guideline limit of four times the number of orbital functions, and that as the number of functions in the orbital basis increases the relative number of functions required in the auxiliary basis is reduced. Just as with previous cc-type MP2 fitting basis sets,[27−29] to ensure sufficient accuracy in the DF it was necessary to include auxiliary functions with an angular momentum quantum number of $l_{\mathrm{occ}} + l_{\mathrm{bas}}$, where l_{occ} and l_{bas} are the highest occupied angular momentum for the atom and largest angular momentum included in the orbital basis set, respectively.

While the auxiliary basis sets presented in this paper have been optimized for atoms and ions, verification of their accuracy for use in molecular applications requires a test set of representative transition metal complexes containing the relevant 4d elements. This test set has been extracted from the more comprehensive (in terms of elements covered) set presented previously[38] and consists of YF, YF_3, YO, ZrF, ZrF_3, ZrO, ZrO_2, NbF_3, NbO, NbO_2, NbO_2F, $Mo(CO)_6$, MoF_3, MoH, MoO_2, MoO_3, Tc_2O_7, TcO, TcO_3F, RhF, RhF_4, RhF_6, RhO, $Pd(CO)_4$, PdF, and PdO_2. The errors in MP2 correlation energy due to orbital basis set incompleteness

Table 2. Relative Percentage and Absolute (kcal mol^{-1}) Errors in the MP2 Correlation Energy due to Orbital Basis Set Size, Assessed by Comparing the cc Basis Set with a Complete Basis Set Estimate, for a Test Set of Transition Metal Complexes[a]

	$\bar{\Delta}^{AO}$	Δ^{AO}_{std}	Δ^{AO}_{max}	$^{ABS}\bar{\Delta}^{AO}$
cc-pVDZ-PP	36.64	5.99	46.87	308.69
cc-pVTZ-PP	16.84	3.67	24.93	138.06
cc-pVQZ-PP	8.24	1.86	12.67	67.54
cc-pV5Z-PP	4.22	0.95	6.49	34.58
aug-cc-pVDZ-PP	33.07	5.64	42.70	267.83
aug-cc-pVTZ-PP	15.35	3.55	23.21	121.07
aug-cc-pVQZ-PP	7.46	1.86	11.77	58.46
aug-cc-pV5Z-PP	3.82	0.95	6.03	29.93

[a] See text for further details.

Table 3. Absolute Errors (kcal mol^{-1}) in the MP2 Correlation Energy due to the Density Fitting Approximation for a Test Set of Transition Metal Complexes[a]

	$^{ABS}\bar{\Delta}^{DF}$	$^{ABS}\Delta^{DF}_{std}$	$^{ABS}\Delta^{DF}_{max}$
cc-pVDZ-PP	0.0662	0.0925	0.4013
cc-pVTZ-PP	0.0188	0.0184	0.0716
cc-pVQZ-PP	0.0028	0.0029	0.0117
cc-pV5Z-PP	0.0022	0.0027	0.0107
aug-cc-pVDZ-PP	0.0746	0.0902	0.3866
aug-cc-pVTZ-PP	0.0119	0.0137	0.0550
aug-cc-pVQZ-PP	0.0030	0.0030	0.0136
aug-cc-pV5Z-PP	0.0028	0.0034	0.0171

[a] See text for further details.

Table 4. Relative Percentage Errors Introduced by the Density Fitting Approximation for a Test Set of Transition Metal Complexes[a]

	$\bar{\Delta}^{DF}$	Δ^{DF}_{std}	$\overline{\lvert\Delta^{DF}\rvert}$	Δ^{DF}_{max}
cc-pVDZ-PP	−0.0086	0.0067	0.0090	−0.0273
cc-pVTZ-PP	−0.0022	0.0015	0.0023	−0.0067
cc-pVQZ-PP	−0.0002	0.0007	0.0004	−0.0032
cc-pV5Z-PP	−0.0001	0.0004	0.0003	−0.0018
aug-cc-pVDZ-PP	−0.0094	0.0056	0.0096	−0.0255
aug-cc-pVTZ-PP	−0.0007	0.0014	0.0014	−0.0030
aug-cc-pVQZ-PP	0.0003	0.0003	0.0003	0.0009
aug-cc-pV5Z-PP	0.0003	0.0002	0.0003	0.0009

[a] See text for further details.

for the double-, triple-, quadruple-, and quintuple-ζ basis sets are shown in Table 2, where the relative percentage error for each complex has been calculated as $\Delta_i^{AO} = (E^{CBS} - E^{VnZ})/E^{CBS} \times 100\%$, and statistical data over the full test set is presented as the mean ($\bar{\Delta}^{AO}$), standard deviation (Δ^{AO}_{std}), and the largest error within the test set (Δ^{AO}_{max}). Also shown is the absolute error between the MP2 correlation energy for a given basis set and the estimate of the CBS limit ($^{ABS}\bar{\Delta}^{AO}$). It should be noted that conventional MP2 calculations on $Pd(CO)_4$ with the aug-cc-pV5Z-PP basis were beyond the computational resources available, because of the literature geometry possessing no symmetry, and hence the statistics shown in Table 2 for the augmented basis sets do not include results for this system. An extensive test of the errors due to the use of the orbital basis sets is not a goal of the current investigation and tests of atomic and molecular applications are available elsewhere;[8] it is sufficient that the errors displayed in Table 2 are consistent with the magnitude of errors observed for molecules containing second row[28] and 3d[29] atoms.

The absolute error in the correlation energy due to the density fitting approximation has been assessed for all eight basis sets over the test set of transition metal complexes. These errors are presented, as a mean ($^{ABS}\bar{\Delta}^{DF}$), standard deviation ($^{ABS}\Delta^{DF}_{std}$), and maximum error ($^{ABS}\Delta^{DF}_{max}$) over the test set, in Table 3. It can be seen that these absolute errors in the correlation energy are small, especially compared to the absolute error in correlation energy due to the orbital basis (cc-pVDZ-PP has an estimated mean error of 308.6870 kcal mol^{-1} in the orbital basis yet the mean error from density fitting is 0.0662 kcal mol^{-1}). On increasing the orbital basis from cc-pVDZ-PP to cc-pVTZ-PP the mean change in energy drops by 170.6235 kcal mol^{-1}, but as this is more than 3 orders of magnitude greater than the absolute error in correlation energy due to density fitting it seems clear that the error, in absolute terms, introduced by density fitting is small and should have a negligible effect on correlation energy based basis set extrapolations. Readers should be aware that these errors, and all others regarding the MP2 fitting basis sets reported in the current investigation, include contributions from the non-4d atoms (and their respective orbital and MP2 fitting basis sets) that are part of the complexes in the test set and are thus not entirely due to the new sets presented. However, as real world applications are likely to include such atoms in addition to 4d transition metals, they are more indicative of the magnitude of errors that may occur in further applications.

Further analysis of the errors introduced by density fitting is carried out by inspecting the percentage error relative to the orbital basis set, which can be expressed in terms of the correlation energy as $\Delta_i^{DF} = (E^{corr} - E^{DF-corr})/E^{corr} \times 100\%$, where E^{corr} and $E^{DF-corr}$ are the conventional and density fitting MP2 correlation energies, respectively. These errors can be found in Table 4, summarized as the mean of the error over the test set ($\bar{\Delta}^{DF}$), the standard deviation (Δ^{DF}_{std}), the mean unsigned error ($\overline{\lvert\Delta^{DF}\rvert}$), and the largest error within the test set (Δ^{DF}_{max}). As there is no guarantee that the density fitting approximation will consistently under- or overestimate the conventional correlation energy, the mean error of Table 2 should be compared with the mean unsigned error of Table 4 to properly assess the quality of the fitting.

Table 4 indicates that the relative error introduced by the density fitting approximation with all eight MP2 fitting basis sets is small enough to be insignificant in terms of the overall correlation energy. In all cases the statistical measures of the errors are between 3 and 4 orders of magnitude smaller than the error due to orbital basis set incompleteness highlighted in Table 2. Because of the rounding of values in Table 4, it is not immediately obvious that aug-cc-pV5Z-PP produces a more accurate fitting than aug-cc-pVQZ-PP, but it is evident at a higher precision, for example, that the aug-cc-pVQZ-PP Δ^{DF}_{max} is 0.000347 while that for aug-cc-pV5Z-PP is 0.000286. This improvement can be seen graphically in Figure 1, which plots the normalized Gaussian distributions of the relative errors due to density fitting for all eight auxiliary basis sets.

Both Figure 1 and Table 4 show that as the cardinal number of the basis set is increased DF-MP2 produces

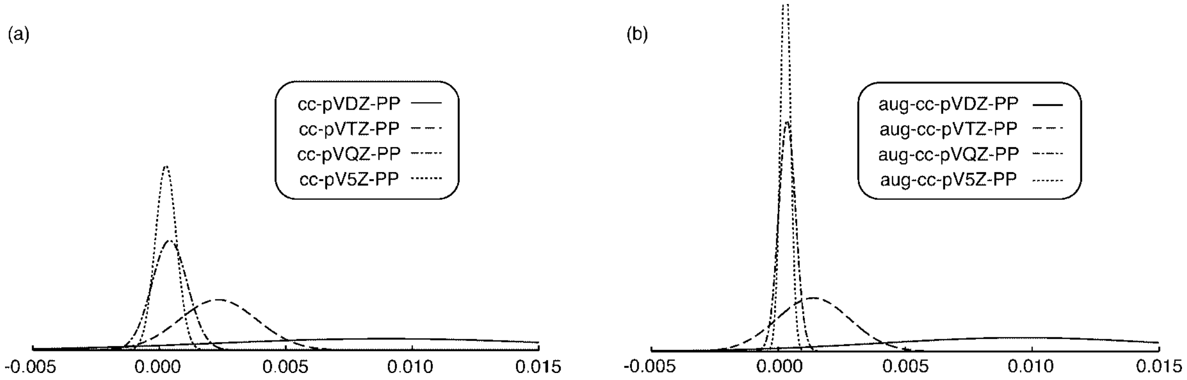


Figure 1. Normalized Gaussian distributions of the relative percentage errors in MP2 valence correlation energy introduced via the density fitting approximation for a test set of small and medium sized transition metal complexes. (a) Errors for the cc-pV *n* Z-PP basis sets. (b) Errors for the aug-cc-pV *n* Z-PP basis sets. The normal distributions are defined by $y(x) = (1)/(\sigma\sqrt{2\pi})\cdot e^{-(1)/(2)((x-\bar{x})/(\sigma))^2}$, where σ represents the standard deviation and $\bar{x}$ the mean.

Table 5. Relative Percentage and Absolute (kcal mol^{-1}) Errors Introduced by the Density Fitting Approximation Using Truncated Auxiliary Basis Sets for a Test Set of Transition Metal Complexes; l_{max} Indicates the Angular Momentum Functions at which the Basis Has Been Truncated[a]

	l_{max}	$\overline{\lvert\Delta^{DF}\rvert}$	Δ^{DF}_{std}	Δ^{DF}_{max}	$^{ABS}\bar{\Delta}^{DF}$
cc-pVQZ-PP	7	0.0004	0.0007	−0.0032	0.0028
	6	0.0244	0.0255	−0.0952	0.1722
cc-pV5Z-PP	8	0.0003	0.0004	−0.0018	0.0022
	7	0.0095	0.0099	−0.0371	0.0693
	6	0.0656	0.0677	−0.2444	0.4839
aug-cc-pVQZ-PP	7	0.0003	0.0003	−0.0009	0.0030
	6	0.0256	0.0295	−0.1070	0.1771
aug-cc-pV5Z-PP	8	0.0003	0.0002	0.0009	0.0028
	7	0.0103	0.0127	−0.0451	0.0680
	6	0.0653	0.0746	−0.2607	0.4482

[a] See text for further details.

correlation energies that are closer to the conventional energy and have a lower standard deviation. Comparison of parts a and b of Figure 1, which are plotted at the same scale, indicates that there is also a small improvement in the overall accuracy of the fitting for augmented sets relative to standard cc and that this is noticeably more pronounced with the quadruple- and quintuple-ζ sets.

While *k* and *l* angular momentum functions were required to optimize the quadruple- and quintuple-ζ MP2 fitting basis sets, most quantum chemistry packages currently only support *i* functions as a maximum. To gauge the errors associated with such maxima imposed on the MP2 fitting basis sets presented, the errors were calculated for the larger basis sets truncated to a maximum *l* of 6 or 7 (*i* and *k* functions, respectively) and are reported in Table 5. The same statistical measures for the nontruncated sets are also repeated in the table for ease of reference, along with the mean of the absolute error (kcal mol^{-1}) relative to the conventional MP2 correlation energy.

The statistical data in Table 5 show that truncating the MP2 fitting basis sets has a large effect on the magnitude of the errors introduced by the density fitting approximation. The truncation of the quadruple-ζ sets increases the mean and standard deviation by almost 2 orders of magnitude, truncating the quintuple-ζ sets such that $l_{max} = 7$ increases the mean errors by between 1 and 2 orders of magnitude while truncating to $l_{max} = 6$ increases it by more than 2 orders of magnitude. These increases in the error due to the density fitting approximation are displayed pictorially in the form of normalized Gaussian distributions for the aug-cc-pVQZ-PP and aug-cc-pV5Z MP2 fitting basis sets and truncations

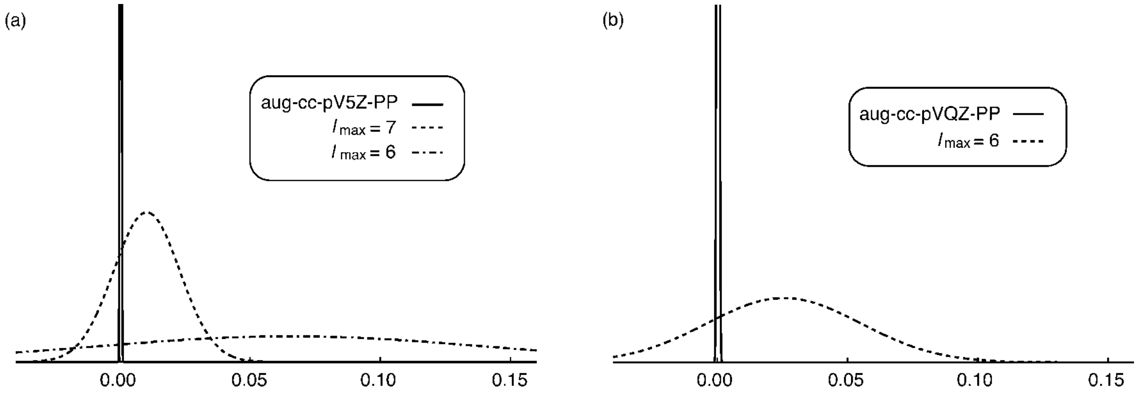


Figure 2. Relative percentage errors due to the density fitting approximation in the MP2 valence correlation energy of a test set of small and medium sized transition metals when truncated MP2 fitting basis sets are employed. l_{max} indicates the angular momentum functions at which the basis has been truncated. (a) aug-cc-pV5Z MP2 fitting basis and truncations. (b) aug-cc-pVQZ-PP MP2 fitting basis and truncations. The normal distributions are defined by $y(x) = (1)/(\sigma\sqrt{2x})e^{-(1)/(2)((x-\bar{x})/(\sigma))^2}$, where σ represents the standard deviation and $\bar{x}$ the mean.

in Figure 2. Equivalent plots for the non-augmented sets show much the same trends and thus are not presented here. In general, truncating the augmented MP2 fitting basis sets produces a larger increase in the mean unsigned error, standard deviation, and maximum error than truncating the nonaugmented sets, perhaps suggesting that high angular momentum functions have a larger effect on the fitting of diffuse functions in the orbital set. It is also worth noting that all of the truncated auxiliary basis sets overestimate the correlation energy for every complex in the test set.

Both Table 5 and Figure 2 display a large reduction in the quality of the fitting, when the MP2 fitting basis sets are truncated, such that the relative percentage errors are larger than those shown for cc-pVDZ-PP in Table 4. A comparison with Table 2 indicates that when both quadruple- and quintuple-ζ MP2 fitting sets are truncated such that $l_{\max} = 6$, the mean unsigned error and standard deviation are around 2 orders of magnitude smaller than the error in the orbital basis. It should be noted that the maximum relative percentage error within the test set for the truncated MP2 fitting basis sets is only around 1 order of magnitude smaller than that of the respective orbital basis. Inspecting the absolute errors from Table 5 reveals that truncating the quintuple-ζ basis sets to $l_{\max} = 6$ produces a mean error of almost 0.5 kcal mol^{-1}, which compares to a mean error from Table 2 of 29.9 kcal mol^{-1} in the orbital basis when it is compared to the CBS estimate. Despite this MP2 fitting basis meeting the initial criteria suggested earlier for a sufficiently accurate fit, such large basis sets are generally only employed when high accuracy is desired and it seems that, for the majority of applications, it should only be used with a sizable degree of caution. The extent of both the relative and absolute errors for the remaining truncated basis sets suggests that they could be applied more readily, but again some caution is advisable. Because of the vastly superior accuracy of the nontruncated MP2 fitting basis sets it is recommended that truncated basis sets should not be utilized when higher angular momentum functions are supported, despite their smaller size.

4. Conclusions

MP2 fitting auxiliary basis sets for density fitted correlated post-HF ab initio methods have been optimized for use with the cc orbital basis sets for the 4d transition metal elements Y−Tc and Rh−Pd. These cc-pV*n*Z-PP and aug-cc-pV*n*Z-PP MP2-fit (where n = D, T, Q, and 5) sets have been shown to provide accurate density fitting with the MP2 method for small to medium sized transition metal complexes with a range of oxidation states. The relative error introduced by the use of these sets compared to conventional MP2 is negligible, generally between 3 and 4 orders of magnitude smaller than the error due to orbital basis set incompleteness and, as shown by the absolute errors, is also an insignificant percentage of the overall correlation energy. Thus, the auxiliary basis sets presented can be used with confidence in accurately reproducing conventional correlation energies.

While truncating these MP2 fitting basis sets may be desirable from the point of view of utilizing them in a larger range of quantum chemical packages, removing functions from the quintuple-ζ sets such that $l_{\max} = 6$ introduces significant errors. Truncating quadruple-ζ to $l_{\max} = 6$ and quintuple-ζ to $l_{\max} = 7$ also introduces errors considerably larger than before truncation, and thus the application and desired degree of accuracy should be taken into account before employing them.

All eight of the MP2 fitting auxiliary basis sets presented in this work will be made available via the Environmental Molecular Sciences Laboratory Basis Set Exchange[32] Web site, http://gnode2.pnl.gov/bse/portal, and can also be found in the Supporting Information.

Acknowledgment. The EPSRC National Service for Computational Chemistry Software[39] is acknowledged for the CPU time and access to the TURBOMOLE package it provided.

Supporting Information Available: Tables depicting MP2 fitting auxiliary basis sets for Y, Zr, Nb, Mo, Tc, Rh,and Pd. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Balabanov, N. B.; Peterson, K. A. *J. Chem. Phys.* **2005**, *123*, 064107.

(2) Peterson, K. A.; Figgen, D.; Dolg, M.; Stoll, H. *J. Chem. Phys.* **2007**, *126*, 124101.

(3) Peterson, K. A.; Puzzarini, C. *Theor. Chem. Acc.* **2005**, *114*, 283.

(4) Ohnishi, Y.; Nakao, Y.; Sato, H.; Sakaki, S. *J. Phys. Chem. A* **2007**, *111*, 7915.

(5) Solomonik, V. G.; Stanton, J. F.; Boggs, J. E. *J. Chem. Phys.* **2008**, *128*, 244104.

(6) Lee, E. P. F.; Dyke, J. M.; Mok, D. K. W.; Chau, F. *J. Phys. Chem. A* **2008**, *112*, 4511.

(7) Li, S.; Dixon, D. A. *J. Phys. Chem. A* **2008**, *112*, 6646.

(8) Figgen, D.; Peterson, K. A.; Stoll, H. *J. Chem. Phys.* **2008**, *128*, 034110.

(9) Williams, T. G.; Wilson, A. K. *J. Chem. Phys.* **2008**, *128*, 054108.

(10) Whitten, J. L. *J. Chem. Phys.* **1973**, *58*, 4496.

(11) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514.

(12) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143.

(13) Rendell, A. P.; Lee, T. J. *J. Chem. Phys.* **1994**, *101*, 400.

(14) Hättig, C.; Weigend, F. *J. Chem. Phys.* **2000**, *113*, 5154.

(15) Werner, H.-J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, *118*, 8149.

(16) Schütz, M.; Manby, F. R. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3349.

(17) Kats, D.; Korona, T.; Schütz, M. *J. Chem. Phys.* **2007**, *127*, 064107.

(18) Aquilante, F.; Lindh, R.; Pedersen, T. B. *J. Chem. Phys.* **2007**, *127*, 114107.

(19) Beebe, N. H. F.; Linderberg, J. *Int. J. Quantum Chem.* **1977**, *12*, 683.

(20) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(21) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98*, 1358.

(22) Wilson, A. K.; Peterson, K. A.; Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1999**, *110*, 7667.

(23) Dunning, T. H.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, *114*, 9244.

(24) Peterson, K. A. *J. Chem. Phys.* **2003**, *119*, 11099.

(25) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(26) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1994**, *100*, 2975.

(27) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175.

(28) Hättig, C. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59.

(29) Hill, J. G.; Platts, J. A. *J. Chem. Phys.* **2008**, *128*, 044104.

(30) Hill, J. G.; Platts, J. A. *J. Chem. Phys.* **2008**, *129*, 134101.

(31) Hättig, C.; Hellweg, A. Unpublished results. Available within TURBOMOLE V5.10.

(32) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. *J. Chem. Inf. Model.* **2007**, *47*, 1045.

(33) Hättig, C. *J. Chem. Phys.* **2003**, *118*, 7751.

(34) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.

(35) *TURBOMOLE V5.10*; see http://www.cosmologic.de/QuantumChemistry/main_turbomole.html.

(36) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639.

(37) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.

(38) Weigend, F. *J. Comput. Chem.* **2008**, *29*, 167.

(39) EPSRC National Service for Computational Chemistry Software; see http://www.nsccs.ac.uk.

CT8005584

JCTC Journal of Chemical Theory and Computation

A Scheme for the Evaluation of Electron Delocalization and Conjugation Efficiency in Linearly π-Conjugated Systems

Maurizio Bruschi,[†] Peter A. Limacher,[‡] Jürg Hutter,[§] and Hans Peter Lüthi[‡,*]

Department of Environmental Science, Università degli Studi di Milano-Bicocca, Piazza della Scienza 1, I-20126 Milano, Italy; Laboratory for Physical Chemistry, ETH Zürich, Wolfgang-Pauli-Strasse 10, CH-8093 Zürich, Switzerland; and Physical Chemisty Institute, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received October 15, 2008

Abstract: In this study, we present a scheme for the evaluation of electron delocalization and conjugation efficiency in lineraly π-conjugated systems. The scheme, based on the natural bond orbital theory, allows monitoring the evolution of electron delocalization along an extended conjugation path as well as its response to chemical modification. The scheme presented is evaluated and illustrated by means of a computational investigation of π-conjugation in *all-trans* polyacetylene [PA; H(−CH═CH)$_n$−H], polydiacetylene [PDA, H(−C≡C−CH═CH)$_n$−H], and polytriacetylene [PTA, H(−C≡C−CH═CH−C≡C)$_n$−H] with up to 180 carbon atoms, all related by the number of ethynyl units incorporated in the chain. We are able to show that for short oligomers the incorporation of ethynyl spacers into the PA chain increases the π-delocalization energy, but, on the other hand, reduces the efficiency with which π-electron delocalization is promoted along the backbone. This explains the generally shorter effective conjugation lengths observed for the properties of the polyeneynes (PDA and PTA) relative to the polyenes (PA). It will also be shown that the reduced conjugation efficiency, within the NBO-based model presented in this work, can be related to the orbital interaction pattern along the π-conjugated chain. We will show that the orbital interaction energy pattern is characteristic for the type and the length of the backbone and may therefore serve as a descriptor for linearly π-conjugated chains.

1. Introduction

For the rational design of functionalized π-conjugated polymers, which are extensively being investigated as materials for application in nanoscience, molecular electronics, photonics, and other areas,[1,2] a more thorough understanding of electron delocalization and conjugation efficiency is crucial. In this class of materials, polyacetylene [PA; H(−CH═CH)$_n$−H] is the prototypical compound. The insertion of ethynyl units into a PA scaffold leads to rodlike polymers, such as the polydiacetylenes [PDAs, H(−C≡C−CH═CH)$_n$−H],[3,4] and the polytriacetylenes [PTAs, H(−C≡C−CH═CH−C≡C)$_n$−H][5−7] (see Scheme 1). It has been observed that these polyeneynes (PDAs, PTAs) show better thermal stability and solubility than the original polyenes (PAs), and, in addition, the capability to form functional derivatives with fully planar, sterically unhindered frameworks.[6]

Obviously, the incorporation of ethynyl groups in the PA scaffold will affect the molecular as well as the electronic structure of these compounds. Properties such as the bond length alternation (BLA),[8] the maximum absorption wavelength ($E_{\max}$),[9] and the second-order hyperpolarizabilities (γ)[10] were reported to differ substantially between polyenes and polyeneynes. It was also shown that γ and other

---

* Corresponding author.

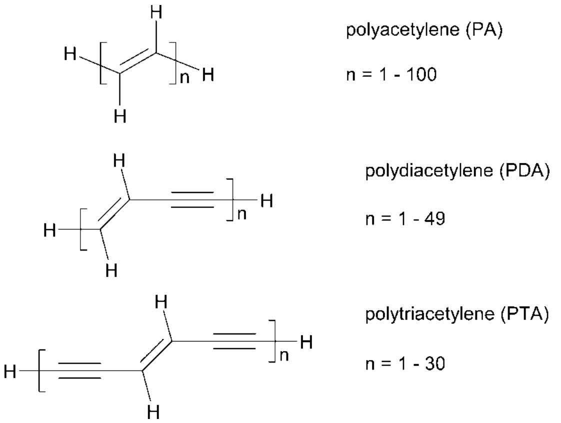
[†] Università degli Studi di Milano-Bicocca.

[‡] ETH Zürich.

[§] University of Zürich.

10.1021/ct8004358 CCC: $40.75

Scheme 1. The Series of Oligomers Considered in This Study



properties saturate at shorter effective conjugation length (ECL) relative to PA, indicating that conjugation efficiency, i.e., the ability to promote electron delocalization along the backbone, is less pronounced.

These observations prompted us to explore the evolution of π-conjugation in PA, PDA, and PTA. For that matter, we studied the response of selected physical observables to the length of the backbone, and related these observations to the evolution of the delocalization energy along the π-conjugated chain using a new scheme based on the natural bond orbital (NBO) analysis.[11−13] The NBO analysis was shown to be a useful tool for the investigation of π-conjugation in earlier work.[8,14−16]

In this work we present and validate a new NBO-based scheme consisting of three different protocols for the evaluation of electron delocalization and conjugation efficiency in linearly π-conjugated chains. We will see that each π-conjugated backbone shows an orbital interaction energy profile which is characteristic for its type (PA, PDA, or PTA) as well as for its length. The orbital interaction profile may thus serve as a valuable molecular descriptor for this class of compounds. Furthermore, the scheme is general and transportable to other π-conjugated systems, including donor−acceptor functionalized ones.

2. Computational Details

In this work we study the oligomers of PA, PDA, and PTA with up to 180 carbon atoms. In the case of PDA, the hydrogen atom of the $\mathrm{C{=}C{-}H}$ terminus is replaced by an ethynyl unit for better comparison with the other oligomers (note that with this definition the monomers of PDA and PTA are identical). Systems with 18, 30, 42, etc. carbon atoms are observed for all of the three types of backbones.

The geometry optimizations and the NBO analysis were performed within the DFT framework,[17] using the hybrid Becke three-parameters B3LYP functional[18−20] with a 6-31G(d,p) basis set of Gaussian orbitals.[21] The optical absorption properties were computed using Zerner‘s INDO/S method[22] on the B3LYP/6-31G(d,p) optimized geometries. In the ZINDO calculations, all occupied and virtual orbitals were retained in the active space. The excitation energies listed refer to the first allowed transition, which, in all the cases considered, is a $\pi-\pi^*$ transition.

Polarizabilities (α) and second-order hyperpolarizabilities (γ) were computed on the B3LYP geometries using the coupled-perturbed (CP) Hartree−Fock approach[23] based on a semiempirical MNDO Hamiltonian.[24] The longitudinal polarizabilities and hyperpolarizabilities correspond to the tensor elements in the direction of the charge transfer axis (α_{zz} and γ_{zzzz} in our reference system) of the molecule. In order to compare computed and measured γ values, the theoretical values were multiplied by factors derived from the phenomenological approach.[25]

The DFT and ZINDO calculations were carried out using the Gaussian03 package,[26] while the calculations of polarizabilities and second-order hyperpolarizabilities were performed with the program MOPAC2000.[27] The NBO analysis was performed using the program NBO 5.0.[28]

The tendency of B3LYP to overestimate electron delocalization[29−32] as well as its failure to reproduce the polarizabilities and hyperpolarizabilities[33,34] of linearly conjugated systems are well-known. Various forms of long-range-corrected exchange functionals have been shown to overcome these problems.[35−37] At the same time, it is not obvious that there is *one single* DFT functional that makes excellent predictions for *all* observables considered here. In a recent study on PA chains, Tozer and co-workers demonstrate that the Coulomb-attenuated CAM-B3LYP DFT method[38] provides a much better estimate of the BLA,[35] but, on the other hand, also confirm the findings of Ma et al.[39] that the standard B3LYP HOMO−LUMO energies are much better suited to model the excitation energies for the long chains. This observation is of relevance here, as the NBO orbital interaction energies directly depend on the orbital energy differencies.

Even though it is a small basis set, the 6-31G(d,p) basis has frequently been used in similar investigations, since the addition of more diffuse functions rather quickly leads to linear dependency issues.[35,36] It was recently shown[36] that the trend observed for the long-range-corrected B3LYP and the MP2 longitudinal polarizability and hyperpolarizability of PDA as a function of chain length is well preserved when switching from a 6-31G(d) to a 6-31+G(d) basis set.

Since this study is about reproducing trends in similar series of compounds rather than to present highest quality data on these observables for the oligomers of PA, PDA, and PTA, we resorted to semiempirical methods that appear to reproduce the available experimental data rather well (see also considerations on method performance in the respective sections of this article or ref 8). For short oligomers of PA, PDA, and PTA, for which experimental absorption spectra are available, the ZINDO calculations are in reasonable agreement with the experimental lowest allowed excitation energies. The polarizabilities and hyperpolarizabilities computed by means of an MNDO Hamiltonian reproduce the trend observed in the experiment rather well.[6,8]

Since PDA and PTA contain alternating single, double, as well as triple bonds, the traditional definition of the BLA needs to be extended. In earlier work on PTA oligomers,[8] we had introduced a generalized parameter δR defined as the difference between the averages of the saturated and the unsaturated bond lengths:

$$\delta R = [(C_D - C_T) + (C_T - C_T)]/2 - [(C{\equiv}C) + (C{\equiv}C)]/2 \tag{1}$$

where $(C_D - C_T)$ and $(C_T - C_T)$ represent the lengths of the single bonds in between a double and a triple bond or two triple bonds, respectively. For PDA, which only shows one kind of single bond, eq 1 is reduced to

$$\delta R = (C_D - C_T) - [(C{=}C) + (C{\equiv}C)]/2 \tag{2}$$

whereas for PA the standard definition applies

$$\delta R = (C-C) - (C{=}C) \tag{3}$$

The values of δR are always those in the *center* of the oligomers. On account of the different definitions given above, only the evolution of the δR for a given chain type will be observed; comparisons of δR for different types of backbones are to be made with care.

The properties of the oligomers considered in this work, at least to smaller chain length, have already been investigated extensively by other researchers[9,33,40−54] using different methodologies and levels of theory.

3. Details of the Analysis of Electron Delocalization and Conjugation Efficiency

The basic approach used for the analysis of electron delocalization in π-conjugated systems applied here is outlined in detail in earlier work.[15,55] In these articles we show that the π-delocalization energy can be evaluated either by means of the deletion of the weakly occupied π^* NBOs contained in the path, or by the summation of the interaction energies between the adjacent unsaturated orbitals along the pathway. The scheme presented also applies for σ-conjugation. In this work, however, we will exclusively focus on π-electron delocalization, since in the *through* conjugated systems considered here, in-plane σ-hyperconjugation is not expected to play a major role.

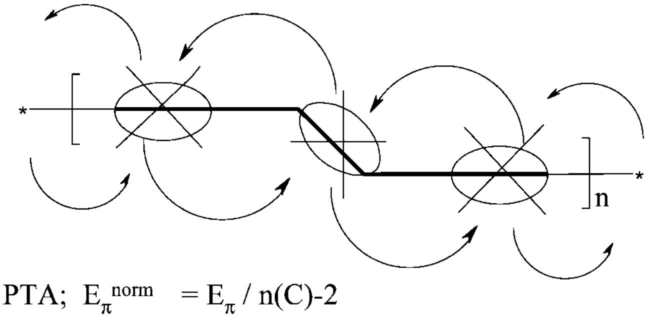
Whereas for the (quantitative) evaluation of π-electron delocalization various approaches have been proposed,[56−62] the conjugation efficiency, to the best of our knowledge, lacks a clear definition. It is generally understood as the ability of a backbone to promote electron delocalization. The ECL and the value of the property under consideration at infinite chain length offer a quantitative measure for the conjugation efficiency of a given backbone. A system with a backbone that promotes electron delocalization well, will, in general, show a large ECL as well as a significant evolution of the property considered.

The first estimate of conjugation efficiency therefore is in the response of the properties to chain length extension, i.e., the slope of the property as a function of chain length. For the systems studied here, those backbones whose properties show the strongest response to chain length extension will also show the largest ECL and the strongest evolution of the respective property.

In this research, we are using three different protocols to analyze π-conjugation presented in the next section.

π-Delocalization Energy Per Carbon Atom. The total π-delocalization energy (E_π) is evaluated by calculating the energies resulting from the deletion of *all* vertical π^* orbitals along the chain. To allow a direct comparison of the E_π values for the different oligomers, E_π is normalized to the number of $\pi \rightarrow \pi^*$ donor−acceptor orbital interactions in the chain, which corresponds to $n(C) - 2$, where $n(C)$ is the number of carbon atoms in the oligomer ($E_\pi^{\text{norm}} = E_\pi/n(C) - 2$; see Scheme 2).

Scheme 2. Schematic Representation of the PTA Polymer with the $\pi \rightarrow \pi^*$ Interactions Accounted for the Calculation of E_π^{norm}[a]



[a] In the scheme n(C) stands for the number of carbon atoms in the oligomer. For PDA and PA, the interactions are accounted for in the same manner.

The orbital interaction energy profiles make use of the donor−acceptor view of the chemical bond in NBO theory, which renders itself nicely to explore electron delocalization along a polymer chain. The (second order) orbital interaction energies (SOIE) between π and π^* donor/acceptor orbital pairs are given by the expression:

$$E_{\text{SOIE}_{ij}} = q_i \frac{F(i,j)^2}{\varepsilon_i - \varepsilon_j} \tag{4}$$

where i and j are the donor and the acceptor NBOs under consideration, respectively, ε_i and ε_j are the corresponding orbital energies, q_i is the donor orbital occupancy, and $F(i,j)$ is the element of the Fock matrix in the NBO basis connecting the interacting orbitals.

The interaction energies between neighboring unsaturated bonds along the backbone for 30-atom chains of the three oligomers considered here are shown in Figure 1. The evolution of the SOIE interactions along the chain defines a profile which is characteristic for the chain type and also for the chain length.

In PA, where the chain consists of identical unsaturated bonds, the orbital interactions will be very similar, varying slowly when moving from the terminal to the central part of the oligomer. If the chain is long enough, the orbital interaction energies in the core of the oligomer will reach an asymptotic value, which is the same for all oligomers of that type. On the other hand, in PDA or PTA the orbital interactions energies between double and triple bonds, and between triple and triple bonds, will differ considerably, leading to a more “rugged” profile of the chain. Again, all these specific interactions energies will vary slowly when moving toward the center of the oligomer, to finally reach their respective asymptotic values. Therefore, we will obtain interaction energy profiles that are characteristic for the type of chain as well as for its length. Note that in Figure 1 these

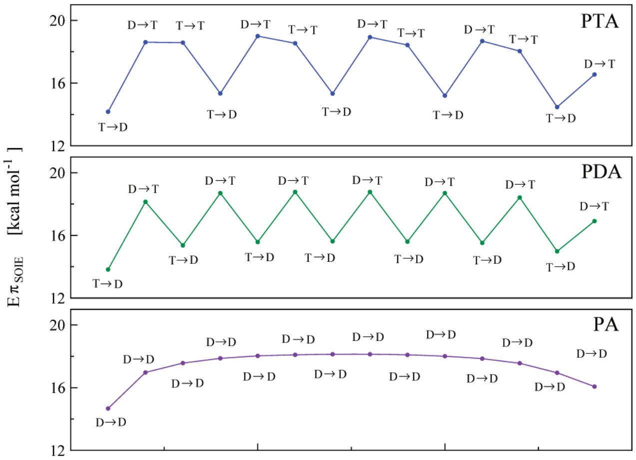


Figure 1. $\pi \rightarrow \pi^*$ SOIEs profile (in kcal mol^{-1}) along a fixed-length 30 carbon atoms backbone for PA (purple dots), PDA (green dots), and PTA (blue dots). The labels D → D, T → D, T → T, and D → T stand for the $\pi_{C=C} \rightarrow \pi^*_{C=C}$, $\pi_{C\equiv C} \rightarrow \pi^*_{C=C}$, $\pi_{C\equiv C} \rightarrow \pi^*_{C\equiv C}$, and $\pi_{C=C} \rightarrow \pi^*_{C\equiv C}$ interactions, respectively. The graph reports the $\pi \rightarrow \pi^*$ SOIEs determined by scanning the backbone from left to right. The profile corresponding to the scan in opposite direction is the mirror image to the one reported.

profiles are scanned from "left to right" along the chain, without considering the (symmetric) interactions in opposite direction.

Response of the Backbone to the Extension of the *π*-Conjugated Chain. To determine the response of the backbone to chain extension, the oligomers are partitioned into a core backbone and into two terminal π-conjugated units, as shown in Scheme 3. This allows the definition of a parameter $\Delta E\pi_{SOIE}$

$$\Delta E\pi_{SOIE} = \sum E\pi_{SOIE}^{core}(n+2) - \sum E\pi_{SOIE}(n) \qquad (5)$$

where $\sum E\pi_{SOIE}^{core}(n + 2)$ is the sum of the $E\pi_{SOIE}$ energies between all vertical π orbitals of the core backbone in the oligomer $n + 2$, and $\sum E\pi_{SOIE}(n)$ is the sum of the $E\pi_{SOIE}$ energies between all vertical π orbitals of the oligomer of size n. The $E\pi_{SOIE}$ interactions to be included in the summations are shown in Scheme 3. For PTA and PDA an extra monomer unit covers six, respectively four more carbon atoms. For PA, the additional monomer unit was selected to cover six carbon atoms (as in PTA). The reference system is the geometry optimized oligomer of size n. This means that $\Delta E\pi_{SOIE}$ also takes into account the contribution of the geometry relaxation in response to the extension of the π-conjugation pathway. Clearly, the values of $\Delta E\pi_{SOIE}$ depend on the length of the terminal units assumed in the definition. However, we found that for different lengths of the terminal segments, the comparison of the values of $\Delta E\pi_{SOIE}$ and their evolution with respect to the number of carbon atoms is very similar, and therefore leads to the same observations (not shown). Also, the values of the SOIEs are functions of the $\varepsilon_i - \varepsilon_j$ energy gap (see eq 1), which, in turn, is dependent on the theoretical method adopted. However, from previous studies[15,55] we know that the B3LYP orbital interaction energies computed for the purpose of the comparison of similar compounds (polyenes and polyeneynes) appear to be adequate.

4. Evolution of the Molecular Properties as a Function of the Chain Length

Bond-Length Alternation. The evolution of δR, as defined in eqs 1−3, is reported as a function of the reciprocal number of carbon atoms (1/C) in Figure 2. δR of PA decreases to reach a constant value (zero gradient for δR vs 1/C) at about 33 double bonds. For PDA and PTA the response of δR to chain length extension is nearly identical; δR converges to an effective conjugation length (ECL) of 20−22 double bonds for both compounds,[8] a value substantially smaller that the one of PA. The extrapolated infinite chain BLA for PDA and PTA is significantly larger than that of PA (0.106 and 0.078, respectively, versus 0.056 Å), but, due to the differences in the definition of the respective δR values, the comparison has to be made with care.

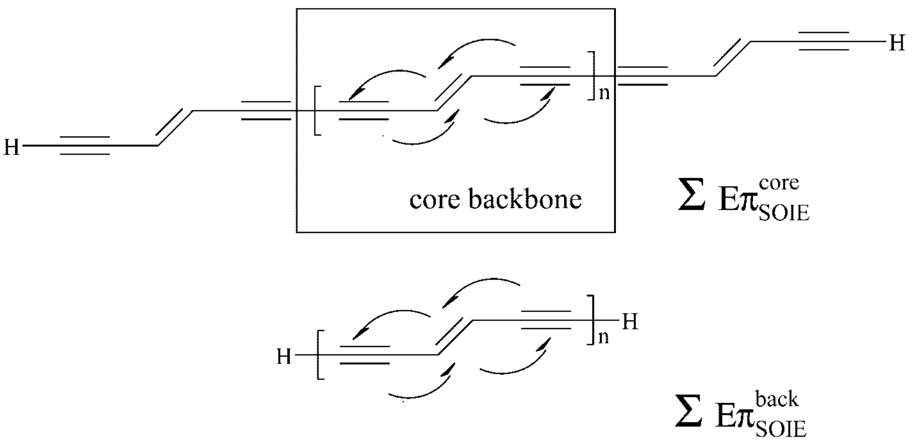
The BLA value extrapolated for a polymer of infinite chain length is in agreement with the finding of Tozer et al. using the same functional and basis set.[35] Using the CAM-B3LYP[38] and BHHLYP[18] functionals, these authors find an infinite chain value for the BLA of 0.087 and 0.089 Å, which is well within the experimental range of 0.08 ± 0.03 Å.[63] These results confirm the tendency of the generic B3LYP method to overestimate delocalization and presumably also the ECL. On the other hand, the present study does not confirm the presence of a minimum of the BLA before reaching the ECL (infinite chain value) as suggested by Tozer and co-workers.[35]

The evolution of the *first allowed excitation energy* (E_{max}) and the *HOMO−LUMO energy gap* as a function of the reciprocal number of carbon atoms of the oligomers investigated are reported in Figure 3. For short oligomers we observe excitation energies for PA that are close to those of PDA and PTA. As the oligomer chain increases, the excitation energy of PA decreases most rapidly, and, for the largest oligomers considered in this study, it is significantly lower than that calculated for the other two types of backbones. The excitation energy of PDA responds slightly faster than that of PTA to chain extension, but converges to the same value at a very large ECL (>75 unsaturated bonds). The extrapolated ZINDO/S value for the excitation energy of PA is equal to about 1.8 eV, in agreement with the experimental optical gap of 1.5−1.8 eV.[64,65]

The longitudinal polarizability α_L of a linearly π-conjugated compound is proportional to the length of the system for large oligomers.[33,44,66−68] This means that for these compounds the polarizability per unit length approaches an asymptotic value. Indeed, for all of the oligomers investigated in this work, the values of α_L divided by the number of carbon atoms of the backbone (α_L/C) appear to slowly converge to an asymptotic limit (Figure 4a). For short oligomers the values of α_L/C are very similar for all compounds. Upon increasing chain length, the value of α_L/C of PA evolves most rapidly, reaching an ECL of 38−40 double bonds, a value significantly higher than that extrapolated for PDA and PTA (28−30 unsaturated bonds). This is in good agreement with experiment[9] and other theoretical predictions.[45,52]

For short π-conjugated oligomers the longitudinal second-order hyperpolarizability γ_L follows the power law $\gamma_L = \gamma_m n^a$

Scheme 3. Schematic Representation of the Partitioning of a PTA Oligomer into a Core Backbone and Two Extra Terminal π-Conjugated Units[a]



[a] $\Delta E\pi_{\mathrm{SOIE}}$ is calculated as the difference between the $E\pi_{\mathrm{SOIE}}$ of the core backbone ($\Sigma E\pi^{\mathrm{core}}_{\mathrm{SOIE}}$) and $E\pi_{\mathrm{SOIE}}$ of an oligomer without the two terminal units ($\Sigma E\pi^{\mathrm{back}}_{\mathrm{SOIE}}$). For PDA and PA, $\Delta E\pi_{\mathrm{SOIE}}$ is evaluated in the same manner.

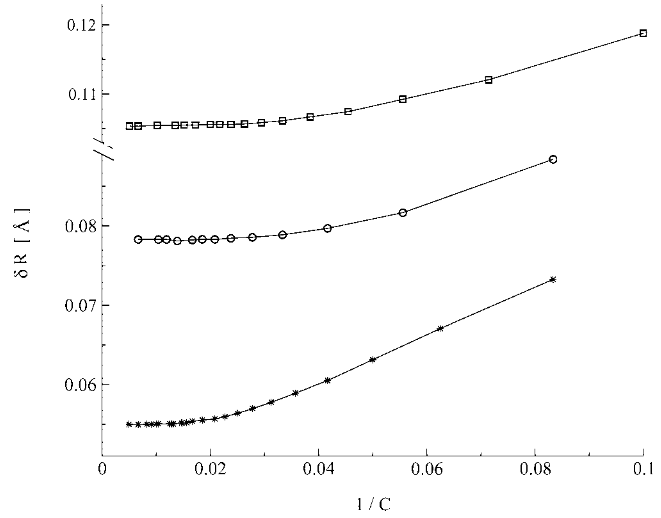


Figure 2. Evolution of the calculated δR parameter (in Å) as function of the reciprocal number of carbon atoms ($1/C$) for the PA (*), PDA (□),and PTA (○) oligomers at the B3LYP/6-31G(d,p) level of theory. The alternation parameter is determined for the central unit using eqs 1−3.

(where n is the number of oligomer units, and γ_{m} and a are constants).[69−71] For large oligomers, the γ_{L} values deviate from this power law, thus indicating inward saturation. In previous work[8] we showed that the γ values calculated for the PTA oligomers up to 54 carbon atoms are in satisfactory agreement with those determined from third harmonic generation (THG) and degenerate four-wave mixing (DFWM) measurements.[72] This observation is confirmed also for the longer chains considered here. Figure 4b reports the evolution of the values of γ_{L} per carbon atom of the backbone (γ_{L}/C) with respect to oligomer size. The trend is similar to that discussed above for α_{L}/C. The γ_{L}/C value of PA evolves much more rapidly than that of any other type of backbone. The γ_{L}/C values of the PDA and PTA oligomers again are very similar and converge to the same ECL. The extrapolation yields an ECL slightly larger than that evaluated for the BLA, but more similar to the one discussed for the polarizability.

In summary, we see that for all of the properties investigated in this work (Table 1), the incorporation of an ethynyl group into a PA chain has a significant effect on the evolution of the backbone properties. This is reflected by the slope (gradient) of the various properties with respect to chain length extension as well as by the difference in ECL between PA and PDA/PTA. For PTA, we confirm the experimental observation of generally shorter ECLs relative to PA.[6,10] On the other hand, the insertion of a second triple bond when going from PDA to PTA, shows no additional impact on the values of the properties, at least as far as their evolution as a function of the chain length is concerned. For the excitation energies and the polarizabilities, we even observe convergence to the same value for the two backbones.

5. Analysis of π-Electron Delocalization in PA, PDA, and PTA Oligomers

For all compounds investigated, the π-delocalization energy per donor−acceptor interaction (E_{π}^{norm}; see Scheme 2 for definition of term) monotonically increases with increasing oligomer size (Figure 5). For short oligomers, PTA shows the largest E_{π}^{norm}, followed by PDA and PA. However, E_{π}^{norm} of PA evolves most rapidly with increasing oligomer size. Therefore, for longer chains PA surpasses PDA to nearly reach the infinite chain value of E_{π}^{norm} for PTA. This result supports the observation made on the basis of the computed molecular properties that the incorporation of triple bonds in the PA chain increases the π-delocalization energy, but that in larger chains of PDA and PTA π-electron delocalization is promoted less efficiently.

From the response of the backbone to the elongation of the chain length in terms of orbital interactions ($\Delta E\pi_{\mathrm{SOIE}}$; eq 5), we see that for all oligomers investigated, $\Delta E\pi_{\mathrm{SOIE}}$ converges to an asymptotic value (Figure 6). The addition of further units only has a constant effect (delocalization energy gained by connecting fragments to a saturated core) and does not lead to further enhancement of π-electron

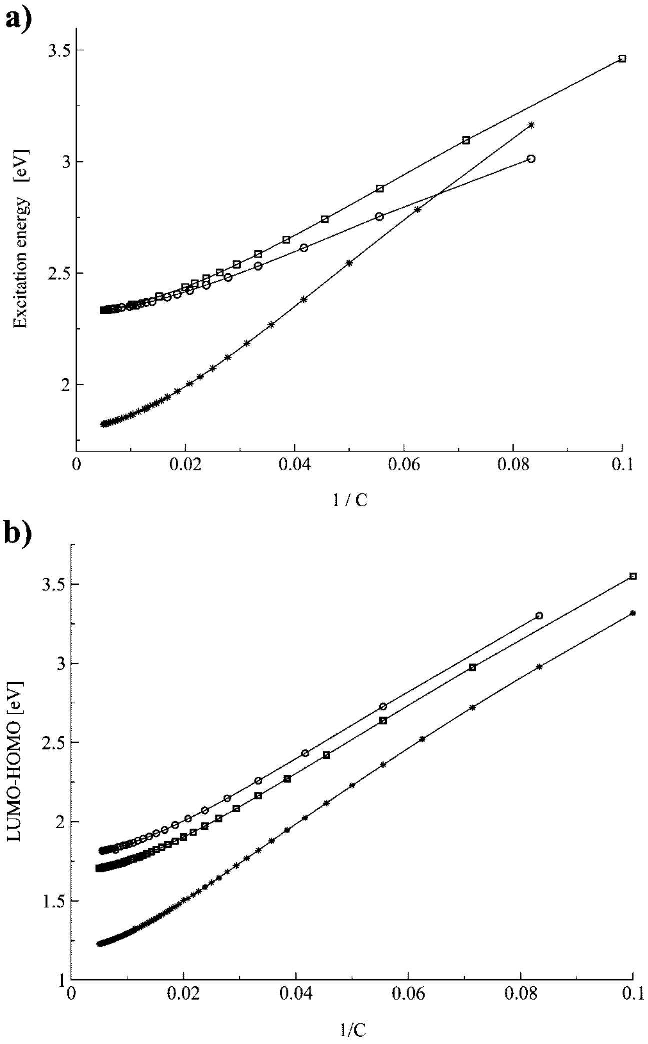


Figure 3. (a) Evolution of the calculated first allowed excitation energy (E_{max}) (in eV) as function of the reciprocal number of carbon atoms (1/*C*) for the PA (*), PDA (□), and PTA (○) oligomers computed at ZINDO level. (b) Evolution of the B3LYP LUMO-HOMO energy gap (in eV) as function of the reciprocal number of carbon atoms (1/*C*) for the PA (*), PDA (□), and PTA (○) oligomers.

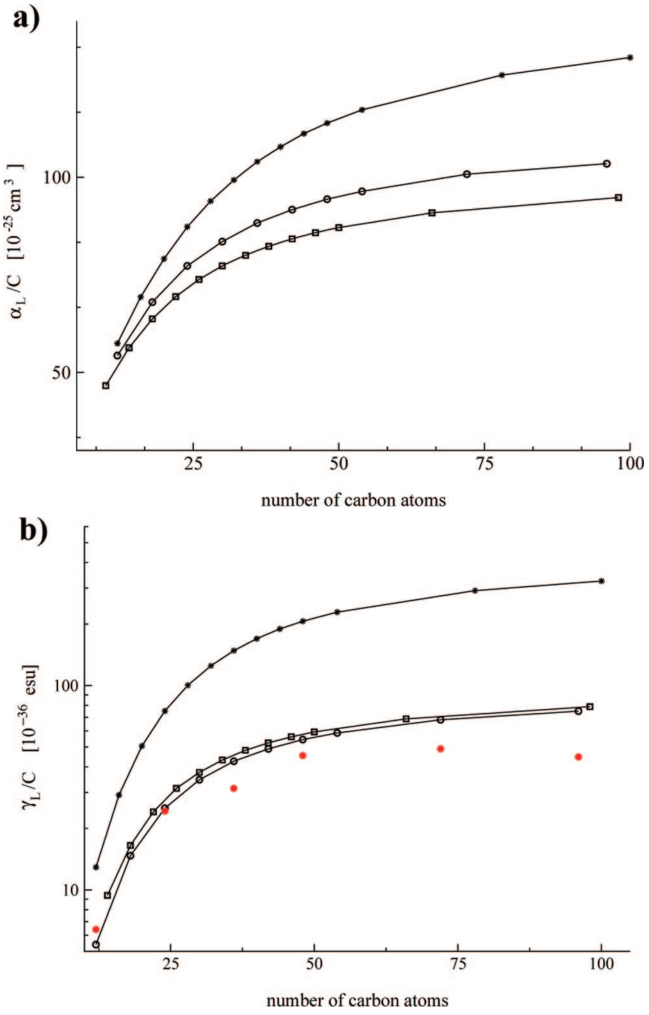


Figure 4. (a) Evolution of the calculated longitudinal polarizability per carbon atom of the backbone (α_L/C) (in 10^{-25} cm^3) as a function of the number of carbon atoms for PA (*), PDA (□), and PTA (○) oligomers computed at the CPHF-MNDO level of theory. (b) Evolution of the calculated longitudinal second-order hyperpolarizability per carbon atom of the backbone (γ_L/C) (in 10^{-34} esu) as a function of the number of carbon atoms for the PA (*), PDA (□), and PTA (○) oligomers computed at the CPHF-MNDO level of theory. The experimental values available for PTA are listed as single filled dots.

delocalization in the backbone. Again, we see that $\Delta E\pi_{SOIE}$ of PA increases most rapidly to take a value significantly larger (about 8 kcal mol^{-1}) than that of the polyeneynes. The evolution of $\Delta E\pi_{SOIE}$ for PDA and PTA is similar, with PTA converging to a value of 2 kcal mol^{-1} lower than that of PDA. Given the much larger gradient of $\Delta E\pi_{SOIE}$ with respect to chain length, the PA backbone again appears to show the more enhanced π-conjugation efficiency, which is consistent with the observed evolution of the molecular properties.

Finally, the analysis of π-electron delocalization along the backbone in terms of $\pi \rightarrow \pi^*$ orbital interactions gives us more insight in how π-electron delocalization propagates along the different types of chains. The interaction energy profiles for the polyenes and the polyeneynes are vastly different. Whereas we have a "smooth" interaction energy profile for PA, we observe an oscillating ("rugged") profile for PDA and PTA. In the polyeneynes the $\pi_{C\equiv C} \rightarrow \pi^*_{C\equiv C}$ (T → T) and the $\pi_{C=C} \rightarrow \pi^*_{C\equiv C}$ (D → T) interaction energies are considerably larger than the $\pi_{C'C} \rightarrow \pi^*_{C=C}$ (T → D) ones. This interaction is also weaker than the D → D one in PA. In the orbital interaction model used here, the comparatively weak T → D interactions in the polyeneynes will reduce delocalization relative to the polyene and therefore lower their conjugation efficiency.

For short oligomers, the sum of orbital interactions in the polyeneynes is still larger than in PA, leading to a larger value for the overall delocalization energy. For the extended oligomers, the reduced promotion of electron delocalization due to the relatively weak T → D interactions will reverse this situation in favor of PA (see also Figure 5). The observation that the D → D interaction energies in PA are lower than the T → T interaction energies in PTA is also in agreement with the conclusions reported in a recent study of von Schleyer et al.,[73] in which the π-delocalization energy

Table 1. Conjugation Efficiency in Terms of the Effective Conjugation Length (ECL), the Extrapolated Value for a Polymer of Infinite Chain Length, and the Gradient of the Properties with Respect to Chain Length Extension (Slope) for the Three Types of Backbones Investigated in This Work[a]

	PA			PDA			PTA		
	ECL[b]	$n = \infty$[c]	slope[d]	ECL[b]	$n = \infty$[c]	slope[d]	ECL[b]	$n = \infty$[c]	slope[d]
δR	~32	0.056	0.301 (32)	~21	0.106	0.199 (26)	~21	0.780	0.214 (24)
excitation energy	~49	1.81	18.6 (54)	~37	2.33	12.9 (50)	~34	2.33	9.5 (54)
LUMO−HOMO energy gap	~64	1.22	22.2	~50	1.70	19.0	~52	1.80	18.8
α/C	~57	135	−	~49	96	−	~51	107	−
γ/C	~82	348	−	~77	86	−	~77	82	−
$\Delta E\pi_{\mathrm{SOIE}}$	~32	20.8	−	~21	12.7	−	~21	10.7	−
E_{π}^{norm}	−	15.8	−20.0	−	15.4	−11.9	−	15.9	−14.2

[a] δR in Å, excitation energies in eV, LUMO-HOMO gaps in eV, α/C in 10^{-25} esu, γ/C in 10^{-36} esu, $\Delta E\pi_{\mathrm{SOIE}}$, and E_{π}^{norm} in kcal mol^{-1}. [b] The ECL indicates the number of unsaturated bonds for which the observable considered exceeds 99.9% of the infinite chain value. [c] Extrapolated value for a polymer of infinite chain length based on 4th-order polynomials (correlation coefficient greater than 0.999). [d] Slope of the linear segment of the curves (in 1/C representation) with the maximum number of carbon atoms considered for the linear regression.

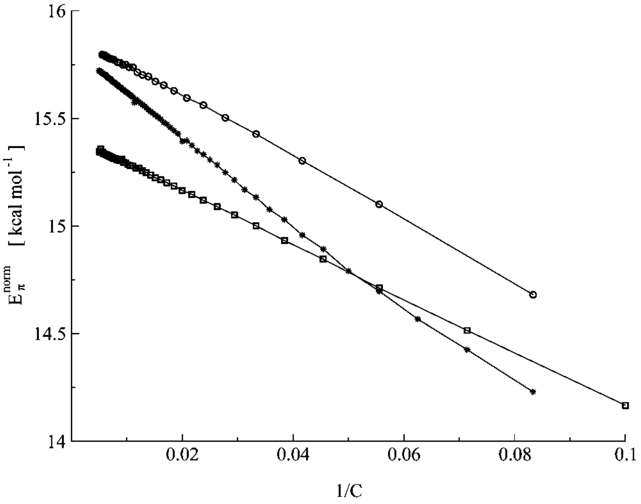


Figure 5. E_{π}^{norm} values as a function of reciprocal number of carbon atoms for the PA (*), PDA (□), and PTA (○) oligomers, computed at the B3LYP/6-31G(d,p) level of theory (see also Scheme 2).

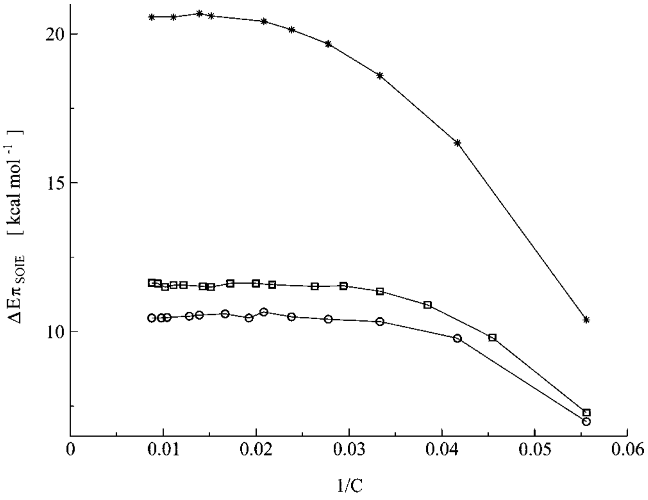


Figure 6. Evolution of $\Delta E\pi_{\mathrm{SOIE}}$ (eq 5) as a function of the reciprocal number of carbon atoms for the PA (*), PDA (□), and PTA (○) oligomers computed at the B3LYP/6-31G(d,p) level of theory. The definition of backbones and terminal groups is given in Scheme 3.

of 1,3-butadiyne and 1,3-butadiene were compared by means of the computed heats of hydrogenation.

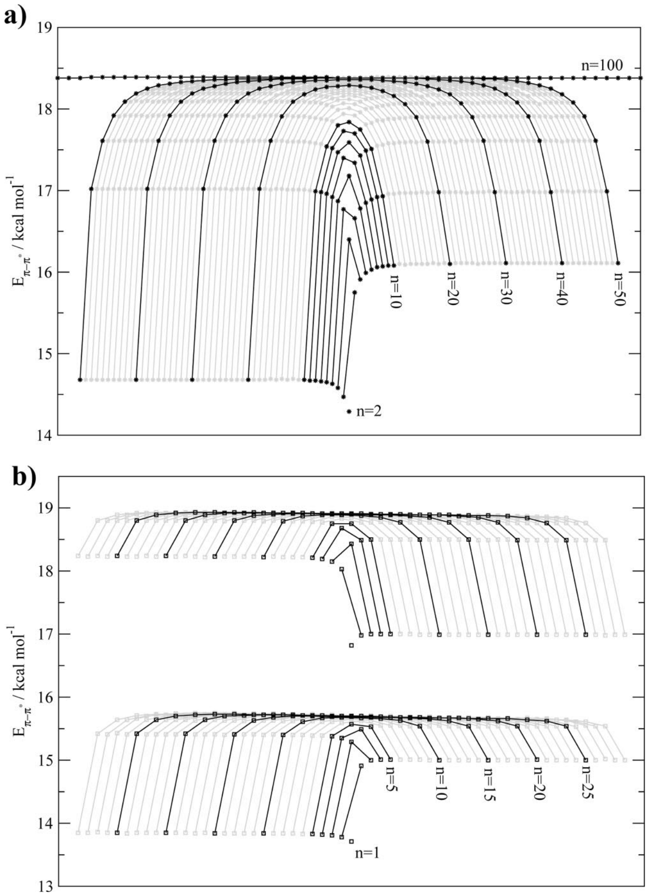


Figure 7. $\pi \rightarrow \pi^*$ interaction energies (in kcal mol^{-1}) for all oligomers of PA (up to n = 50) and PDA (up to n = 29). As in Figure 1, the backbone is scanned from left to right, only showing the symmetry-unique interaction energies. The graph a starts with butadiene (n = 2; listed in the center), which shows just one unique orbital−orbital interaction, and ends with the interaction energies of the pentadecamer (n = 50) of PA. In addition, the interaction energies in the core of the hectamer (n = 100) are shown. In panel b, the double−triple bond (upper part) and triple−double bond interactions (lower part) are shown from the monomer up to the n = 29 system.

From Figure 7a,b, which shows *all* orbital interaction energies for *all* oligomers of PA and PDA, we see that for both compounds each specific interaction converges to a specific asymptotic value. For PA, the interaction energy between the outermost double bond (donor) and its next neighbor (acceptor), for example, converges to a value of 14.6 kcal mol^{-1} for chains larger than 20 atoms, whereas the interaction between orbitals in the center (core) of the chain converges to a value 3.8 kcal mol^{-1} higher for chains longer than 30 atoms. Obviously, the asymptotic value of the interaction energies increases as one moves toward the core of the backbone. (see also Figure 1 for the 30 carbon atom chain); at the same time we observe that the interaction energies converge more slowly.

For PDA and PTA (not shown), we observe a very similar convergence pattern for the orbital interactions. Relative to PA, the asymptotic values for the interactions in the core of the PDA backbone are somewhat higher for the D → T bond interactions (19.0 kcal mol^{-1}), but distinctly lower (15.8 kcal mol^{-1}) for the T → D bond interactions. Also, the margin between the interaction energies among the inner (core) and the outer bonds is much smaller in PDA (and also in PTA). This observation is yet another expression of the reduced response of PDA and PTA to chain length extension.

Finally, the stability of the converged values of the interaction energies is an illustration of the reliability as well as the numerical stability of the models used here.

In summary, from the three schemes of analysis we see that the incorporation of an ethynyl group into a PA chain increases the total delocalization energy, but, at the same time, introduces relatively weak T → D interactions in the backbone, which are responsible for the reduced conjugation efficiency in PDA. The insertion of a second triple bond (PDA → PTA) does not add any more “damage”, and therefore the properties of PDA and PTA show similar conjugation efficiency.

6. Conclusions

The schemes of analysis introduced here show that the insertion of ethynyl groups into a PA chain increases the total π-delocalization energy for short PDA and PTA oligomers, but, very importantly, reduces the efficiency with which π-electron delocalization is promoted. The less efficient π-electron delocalization is also responsible for the generally shorter ECL observed for the properties of the polyeneynes relative to the polyenes. Accordingly, all physical observables computed for PA evolve more rapidly as a function of the oligomer size, and converge to an ECL larger than those computed for PDA and PTA. The insertion of a second ethynyl group into the PDA chain does not further modulate the efficiency with which π-electron delocalization is promoted. The loss of conjugation efficiency observed for the polyeneynes is the result of the relatively weak orbital interactions involving the triple bond π orbitals as donors and the double bond π* orbitals as acceptors (T → D interaction; see Figures 1 and 6). These weak orbital interactions represent the *bottleneck* in the delocalization of charge along the chain. The orbital interaction energy profile appears to be a very promising scheme of analysis, also because it is characteristic for the compound investigated. It may thus render the basis for a molecular descriptor for this class of compounds.

Acknowledgment. This research is supported by a grant of the Swiss Science Foundation (200020-116678). The authors acknowledge a generous allocation of computer time by the Competence Center for Computational Chemistry (C4), as well as stimulating discussions with Prof. François Diederich (ETH Zürich) and his research group. This research has also benefited from a visit of one of the authors (P.A.L.) at the laboratory of Prof. Kurt V. Mikkelsen (University of Copenhagen) funded through a Short Term Scientific Mission (STSM) of COST Action D37.

References

(1) *Handbook of Conducting Polymers*, 3th ed.; Skotheim, T. A., Reynolds, J., Eds.; CRC Press: Boca Raton, FL, 2007.

(2) Blythe, T.; Bloor, D. *Electrical Properties of Polymers*; Cambridge University Press: Cambridge, UK, 2005.

(3) *Polydiacetylene*; Cantow, H. J., Ed.; Springer-Verlag: Berlin, 1984.

(4) Bässler, H. *Polydiacetylenes*; Blood, D., Chance, R. R., Eds.; Nijhoff: Dordrecht, 1985, p 135.

(5) Schreiber, M.; Anthony, J.; Diederich, F.; Spahr, M. E.; Nesper, R.; Hubrich, M.; Bommeli, F.; Degiorgi, L.; Wachter, P.; Kaatz, P.; Bosshard, C.; Günter, P.; Colussi, M.; Suter, U. W.; Boudon, C.; Gisselbrecht, J.-P.; Gross, M. *Adv. Mater.* **1994**, *6*, 786.

(6) Martin, R. E.; Gubler, U.; Cornil, J.; Balakina, M. Y.; Boudon, C.; Bosshard, C.; Gisselbrecht, J.-P.; Diederich, F.; Günter, P.; Gross, M.; Brédas, J.-L. *Chem. Eur. J.* **2000**, *6*, 3622.

(7) Nielsen, M. B.; Diederich, F. *Chem. Rev.* **2005**, *105*, 1837.

(8) Bruschi, M.; Giuffreda, M. G.; Lüthi, H. P. *ChemPhysChem* **2005**, *6*, 511.

(9) Fujii, A.; Hydarat, R.; Sonoda, T.; Fujisawa, T.; Ozaki, M.; Vardeny, Z. V.; Teraguchi, M.; Masuda, T.; Yoshino, K. *Synt. Met.* **2001**, *116*, 95.

(10) Samuel, I. D. W.; Ledoux, I.; Dhenaut, C.; Zyss, J.; Fox, H. H.; Schrock, R. R.; Silbey, R. J. *Science* **1994**, *265*, 1070.

(11) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066.

(12) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.

(13) Weinhold, F.; Landis, C. R. *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*; Cambridge University Press: Cambridge, UK, 2005.

(14) Limacher, P. A.; Lüthi, H. P. *J. Phys. Chem. A* **2008**, *112*, 2913.

(15) Bruschi, M.; Giuffreda, M. G.; Lüthi, H. P. *Chem. Eur. J.* **2002**, *8*, 4216.

(16) Bruschi, M.; Giuffreda, M. G.; Lüthi, H. P. *Chimia* **2005**, *59*, 539.

(17) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.

(18) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(19) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(20) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(21) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(22) Zerner, M. C. *Semiempirical Molecular Orbital Methods; Reviews of Computational Chemistry*; VCH: New York, 1991; *Vol. 2*; p 313.

(23) Kurtz, H. A.; Steward, J. J. P.; Dieter, K. M. J. *Comput. Chem.* **1990**, *11*, 82.

(24) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

(25) Willetts, A.; Rice, E., J.; Burland, E. J.; Shelton, M., D. *J. Chem. Phys.* **1992**, *97*, 7590.

(26) 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.

(27) Stewart, J. J. *MOPAC 2000*; Fujitsu Ltd: Tokyo, 1999.

(28) Glendening, E. D. Badenhoop, J., K. Reed, A. E. Carpenter, J. E. Bohmann, J. A. Morales, M. Weinhold, F. *NBO 5.0*; Theoretical Chemistry Institute, University of Wisconsin: Madison, WI, 2001.

(29) Geskin, M., V.; Dkhissi, A.; Brédas, J. L. *Int. J. Quantum Chem.* **2003**, *91*, 351.

(30) Geskin, V. M.; Grozema, F. C.; Siebbeles, S. D. A.; Beljonne, D.; Brédas, J. L.; Cornil, J. *J. Phys Chem. B* **2005**, *109*, 20237.

(31) Lundberg, M.; Siegbahn, P. E. M. *J. Chem. Phys.* **2005**, *122*, 224103.

(32) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. *Phys. Chem. Lett.* **2008**, *100*, 146401.

(33) Champagne, B.; Perpète, E. A.; van Gisbergen, S. J. A.; Baerends, J. E.; Snijders, J. G.; Soubra-Ghaoui, C.; Robins, K.; Kirtman, B. *J. Chem. Phys.* **1998**, *109*, 10489.

(34) Champagne, B.; Perpète, A. E.; Jacquemin, D.; van Gisbergen, S. J. A.; Baerends, J. E.; Soubra-Ghaoui, C. *J. Phys Chem. A* **2000**, *104*, 4755.

(35) Peach, M. J. G.; Tellgren, E. I.; Salek, P.; Helgaker, T.; Tozer, D. J. *J. Phys. Chem. A* **2007**, *111*, 11930.

(36) Kirtman, B.; Bonness, S.; Ramirez-Solis, A.; Champagne, B.; Matsumoto, H.; Sekino, H. *J. Chem. Phys.* **2008**, *128*, 114108.

(37) Jacquemin, D.; Perpète, E. A.; Scalmani, G.; Frisch, M. J.; Kobayashi, R.; Adamo, C. *J. Chem. Phys.* **2007**, *126*, 144105.

(38) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51.

(39) Ma, J.; Li, S.; Jiang, Y. *Macromolecules* **2002**, *35*, 1109.

(40) Barford, W. *Electronic and Optical Properties of Conjugated Polymers; International Series of Monographs on Physics*; Clarendon Press: Gloucestershire, UK, 2005.

(41) *Conjugated Polymers: The Novel Science and Technology of Highly Conducting and Nonlinear Optically Active Materials*; Brédas, J.-L., Silbey, R., Eds.; Kluwer Academic Publishers: Dordrecht, 1991.

(42) Suhai, S. *Phys. Rev. B* **1995**, *51*, 16553.

(43) Choi, C. H.; Kertesz, M.; Karpfen, A. *J. Chem. Phys.* **1997**, *107*, 6712.

(44) Champagne, B.; Perpète, E. A. *Int. J. Quantum Chem.* **1999**, *75*, 441.

(45) Lu, D. Q.; Marten, B.; Ringnalda, M.; Friesner, R. A.; Goddard, W. A. *Chem. Phys. Lett.* **1996**, *257*, 224.

(46) Su, W. P.; Schrieffer, J. R.; Heeger, A. J. *Phys. Rev. Lett.* **1979**, *42*, 1698.

(47) Su, W. P.; Schrieffer, J. R.; Heeger, A. J. *Phys. Rev. B* **1980**, *22*, 2099.

(48) Mintmire, J. W.; White, C. T. *Phys. Rev. B* **1987**, *35*, 4180.

(49) Perpète, E. A.; Champagne, B. *J. Mol. Struct. (Theochem)* **1999**, *487*, 39.

(50) Shirakawa, H.; Louis, E. J.; MacDiarmid, A. G.; Chiang, C. K.; Heeger, A. J. *J. Chem. Soc., Chem. Commun.* **1977**, *16*, 578.

(51) Shirakawa, H. *Angew. Chem., Int. Ed.* **2001**, *40*, 2575.

(52) Ma, J.; Li, S.; Jiang, Y. *Macromolecules* **2002**, *35*, 1109.

(53) Tretiak, S.; Mukamel, S. *Chem. Rev.* **2002**, *102*, 3171.

(54) Bredas, J.-L.; Beljonne, D.; Coropceanu, V.; Cornil, J. *Chem. Rev.* **2004**, *104*, 4971.

(55) Giuffreda, M. G.; Bruschi, M.; Lüthi, H. P. *Chem. Eur. J.* **2004**, *10*, 5671.

(56) Mo, Y.; Peyerimhoff, S. D. *J. Chem. Phys.* **1998**, *109*, 1687.

(57) Gobbi, A.; Frenking, G. *J. Am. Chem. Soc.* **1994**, *116*, 9275.

(58) Mo, Y.; Jiao, H.; Lin, Z.; Schleyer, P. v. R. *Chem. Phys. Lett.* **1998**, *289*, 383.

(59) Fernández, I.; Frenking, G. *Chem. Eur. J.* **2006**, *12*, 3617.

(60) Schleyer, P. v. R.; Maerker, C.; Dransfeld, A.; Jiao, H.; van Eikema Hommes, N. J. R. *J. Am. Chem. Soc.* **1996**, *118*, 6317.

(61) Fernández, I.; Frenking, G. *J. Org. Chem.* **2006**, *71*, 2251.

(62) Heine, T.; Schleyer, P. v. R.; Corminboeuf, C.; Seifert, G.; Reviakine, R.; Weber, J. *J. Phys Chem. A* **2003**, *107*, 6450.

(63) Yannoni, C. S.; Clarke, T. C. *Phys. Rev. Lett.* **1983**, *51*, 1191.

(64) Hori, Y.; Noda, K.; Kobayashi, S.; Taniguchi, H. *Tetrahedron Lett.* **1969**, 3563.

(65) Fincher, C. R.; Chen, C. E.; Heeger, A. J.; MacDiarmid, A. G.; Hastings, J. B. *Phys. Rev. Lett.* **1982**, *48*, 100.

(66) Perpète, E. A.; Champagne, B. *J. Chem. Phys.* **1997**, *107*, 2463.

(67) Champagne, B.; Mosley, D.; Vračko, M.; Andrè, J.-M. *Phys. Rev. A* **1995**, *52*, 178.

(68) Kudin, N., K.; Car, R.; Resta, R. *J. Chem. Phys.* **2005**, *122*, 134907.

(69) Kanis, D. R.; Ratner, M. A.; Marks, T. J. *Chem. Rev.* **1994**, *94*, 195.

(70) Morley, J. O. *Int. J. Quantum Chem.* **1993**, *46*, 19.

(71) Brédas, J.-L.; Adant, C.; Tackx, P.; Persoons, A.; Pierce, B. M. *Chem. Rev.* **1994**, *94*, 243.

(72) Gubler, U.; Bosshard, C.; Günter, P.; Balakina, M. Y.; Cornil, J.; Brédas, J.-L.; Martin, R. E.; Diederich, F. *Opt. Lett.* **1999**, *24*, 1599.

(73) Jarowski, P. D.; Wodrich, M. D.; Wannere, C. S.; Schleyer, P. v. R.; Houk, K. N. *J. Am. Chem. Soc.* **2004**, *126*, 15036.

CT8004358

J. Chem. Theory Comput. **2009,** *5,* 515–529 **515**

JCTC Journal of Chemical Theory and Computation

Comprehensive Energy Analysis for Various Types of π-Interaction

N. Jiten Singh, Seung Kyu Min, Dong Young Kim, and Kwang S. Kim*

Center for Superfunctional Materials, Department of Chemistry, Pohang University of Science and Technology, Pohang 790-784, Korea

Received November 4, 2008

Abstract: We have investigated various types of π-interactions, where one of the interacting π-systems is represented by an aromatic benzene molecule. The system includes Rg-π, CH-π, π-π(D), π-π(T), H-π(T), π^+-π(D), π^+-π(T), H$^+$-π(T), π^{+2}-π(D), M$^+$-π, and M^{+2}-π complexes, where Rg denotes a rare gas or noble atom, M denotes a metal, and D/T indicates displaced-stacked/T-shaped structure. The microsolvation effect is also considered. We note that the interaction between a cationic π system and a neutral π system (π_{cation}-π interaction) is so far ambiguously considered as either π-π or cation-π interaction. In terms of total binding energy, the π_{cation}-π interaction is weaker than the cation−π interaction, but much stronger than the π-π interaction. When the hydrophilic (N−H)$^+$ or (C−H)$^+$ group in a singly charged π^+ system (as in protonated histidine, arginine, pyridine, or dimethyl imidazolium) interacts with a π-system, the complex favors a T-shaped form [π^+-π(T) complex]. However, in the presence of polar solvating molecules or counteranions, these species interact with the (N−H)$^+$/(C−H)$^+$ group, while the π^+ system interacts with the neutral aromatic ring. Then, the displaced-stacked form [π^+-π(D) complex] is favored or otherwise nearly isoenergetic to the π^+-π(T) form. The π^+-π systems are stabilized mainly by both dispersion and electrostatic energies. Ternary diagrams using either attractive energy components or both attractive and repulsive energy components show that the π^+-π(D) complexes have more contribution from dispersion energy but less contribution from induction energy than the π^+-π(T) complexes, while both complexes have similar percentage contributions from electrostatic and exchange energy components. In particular, the π^+-π(D) complexes are found to be distinctly different from the π-π complexes and the non-π organic or metallic cation-π complexes.

1. Introduction

Nonbonding intermolecular interactions involving aromatic rings[1−5] are pivotal to the stabilization of proteins, enzyme−drug complexes, DNA−protein complexes, organic supramolecules, and nanomaterials.[1−11] In the last two decades, extensive studies have been done to understand the nature of π-complexes including the Rg-π interaction,[12,13] H-π/CH-π interaction,[6,14,17−22] π-π interaction,[7,23−43] cation-π interaction,[44−52] and anion-π interaction.[53−57] Extensive investigations have been made on the energetic and geometrical significance of π-interactions in stabilizing π systems.[58−85] Theoretical interpretations based on high level ab initio calculations have been instrumental in understanding the nature of π-interactions. The strength of the π-interactions is determined by the combined effect of attractive forces (electrostatic, dispersive, and inductive) and repulsive forces (exchange repulsion). Each of these components shows distinguished differences in physical origin, magnitude, and directionality. The dominant attractive energy component in the π-π interaction is the dispersion energy,[23,37,43,68] while that in the cation-π interaction is the electrostatic and induction energies.[44−52]

Interactions between the cationic π-systems (π_{cation}) and neutral π systems are present in many biological and

---

* Correponding Author e-mail: kim@postech.ac.kr.

10.1021/ct800471b CCC: $40.75

macromolecular systems and organic and metallo-organic systems. For example, the interactions play a pivotal role in various ways like (i) forming protein−drug complexes,[86−92] (ii) stabilizing proteins, particularly protein surfaces,[93−95] (iii) stabilizing protein−protein complexes,[96,97] (iv) assisting catalytic activities at the catalytic site of proteins,[96,97] (v) stabilizing protein−DNA complexes,[98,99] (vi) molecule-recognizing artificial receptors for neutral or π-cationic guests,[100,101] (vii) interlocking macromolecules in molecular machines,[102,103] (viii) self-assembling supramolecular structures and packing crystal structures,[104,105] and (ix) controlling conformations in regio- and stereoselective organic synthesis.[106,107]

In experiments, the concept of host−guest interaction between a neutral aromatic ring (benzene and naphthalene derivatives or tetrathiafulvalene) and the tetracationic cyclophane[100,101] has been explored as a fundamental component in constructing diverse catenanes and rotaxanes.[102,103] Upon interacting with a π-system, the quinolinium cation is bound to it more strongly than quinoline.[108] Nitrogen-containing aromatic cations are widely used for room temperature ionic liquids[109,110] as well as anion recognition.[111,112] The π-aromatic ring of triclosan, a general purpose biocide, interacts strongly with a positively charged nicotinamide ring of nicotinamide adenine dinucleotide(NAD^+) located in the active site of enoyl-acyl carrier protein reductase (ENR).[86] In protein structures, the interactions of neutral aromatic amino acids (Phe, Tyr, Trp) with a positively charged π-cloud of arginine are more commonly observed than those with cationic lysine, and their stacked geometries are preferred over the T-shaped structures,[114] whereas in the gas phase the T-shaped structures are more stable. In many of the proteins and protein−ligand complexes, both T-shaped and displaced-stacked structures are equally observed for the interactions of a positively charged histidine residue with the Phe, Tyr, Trp, or Ade residues.[115] It is not yet clearly understood why stacked structures are much more common. The interaction between benzene and pyridinium/imidazolium was suggested as the cation-π interaction[116] where the large dispersion energy term was addressed. However, ambiguity remains whether the interaction between a cationic π-system and a neutral π system would be treated as either conventional cation-π interaction or π-π interaction. Here, we report that it needs to be described as a special type to be denoted by the π_{cation}-π interaction. In particular, the interaction in π^+-π(D) complexes (to be denoted simply as the π^+-π(D) interaction) cannot be described simply as the cation-π interaction or the π-π interaction.[117,118] This finding is based on the comprehensive analysis of interaction energy components for the Rg-π, CH-π, π-π(D), π-π(T), H-π(T), π^+-π(D), π^+-π(T), H^+-π(T), π^{+2}-π(D), M^+-π, and M^{+2}-π complexes.

2. Computational Methods

The geometries of complexes were optimized at the basis-set-superposition−error (BSSE) corrected MP2 level of theory using the aug-cc-pVDZ basis set (to be abbreviated as aVDZ). The lowest energy structures were found from the investigation of the energies of various structures with respect to distances and angles (Supporting Information). These were further calculated using the MP2/aug-cc-pVTZ (to be abbreviated as aVTZ). The complete basis set (CBS) limit values for the MP2 binding energies were evaluated based on the extrapolation method exploiting that the basis set error in the electron correlation energy is proportional to N^{-3} for the aug-cc-pV*N*Z (or aV*N*Z) basis set.[119] Though most extrapolation methods would not be free from inherent over/under−estimation problems, the extrapolation method based on the theoretical understanding is found to be quite reliable.[120] The single point energies using the coupled cluster theory with single, double, and perturbative triple excitations (CCSD(T)) with the aVDZ basis set were also obtained on the BSSE-corrected MP2/aVDZ geometries. Given that the difference in binding energy between MP2/aV*N*Z and CCSD(T)/aV*N*Z does not change significantly with increasing basis set size, the CCSD(T)/CBS binding energies were evaluated from the MP2/CBS ones using the difference between CCSD(T)/aVDZ and MP2/aVDZ binding energies.[120] Ab initio calculations were done using the Gaussian suite of programs[121] (for MP2 calculations) and Molpro package[122] (for CCSD(T) calculations). Symmetry adapted perturbation theory (SAPT) calculations[123,124] at the MP2/6−31+G* or MP2/aVDZ† (where † means the diffuse basis function of H was removed from the full aug-cc-pVDZ basis set) level were carried out to evaluate the energy components.

The decomposition for the total interaction energy (E_{tot}) is made as follows:[124,125]

$$\begin{aligned} E_{\text{tot}} &= E_{\text{es}} + E_{\text{exch}} + E_{\text{ind}} + E_{\text{disp}} + \delta_{\text{int,resp}}{}^{\text{HF}} \\ &= E_{\text{es}} + E_{\text{exch}}{}^{*} + E_{\text{ind}*} + E_{\text{disp}*} + \delta_{\text{int,resp}}{}^{\text{HF}} \end{aligned} \tag{1}$$

where

$$E_{\text{es}} = E_{\text{es}}^{(10)} + E_{\text{es,resp}}^{(12)} \tag{2}$$

$$E_{\text{ind}} = E_{\text{ind}}^{(20)} \tag{3}$$

$$E_{\text{disp}} = E_{\text{disp}}^{(20)} \tag{4}$$

$$E_{\text{exch}} = E_{\text{exch}}^{(10)} + E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)} + E_{\text{exch,ind,resp}}^{(20)} + E_{\text{exch,disp}}^{(20)} \tag{5}$$

Here, the first number 1 or 2 in parentheses in superscript indicates the first or second order perturbation term; the second number 0/1/2 in parentheses in superscript indicates the zeroth/first/second order correction; Notation "resp" in subscript indicates that a given component has been computed including the coupled Hartree−Fock response for the perturbed system. As in our earlier work[16,43,57] and others,[125] the modified decomposition for the interaction energy (E_{tot}) (in which the superscript "*" indicates the effective energy component) is made as follows:

$$E_{\text{ind}*} = E_{\text{ind}}^{(20)} + E_{\text{exch,ind,resp}}^{(20)} + \delta_{\text{int,resp}}{}^{\text{HF}} \tag{6}$$

$$E_{\text{disp}*} = E_{\text{disp}}^{(20)} + E_{\text{exch,disp}}^{(20)} \tag{7}$$

$$E_{\text{exch}*} = E_{\text{exch}}^{(10)} + E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)} \tag{8}$$

Although E_{ind}, E_{disp} and E_{exch} are well defined, we need to use $E_{\text{ind}*}$, $E_{\text{disp}*}$, and $E_{\text{exch}*}$ to properly classify different types of π-interactions, i.e., for the purpose of better clustering each characteristic type. To this end, the $\delta_{\text{int,resp}}{}^{\text{HF}}$ is added to $E_{\mathrm{ind}*}$, while it is not directly related to the induction but still more correlated to the induction than other terms.

According to our previous investigations for π-π interactions, we note that the base set dependency of E_{es}, $E_{\mathrm{ind}*}$, and $E_{\mathrm{exch}*}$ is small, while that of $E_{\mathrm{disp}*}$ is substantial.[16] Thus, the realistic $E_{\mathrm{disp}**}$ for π-π interactions is obtained by adding the difference between the CCSD(T)/CBS and SAPT(MP2/6−31+G*) interaction energies to $E_{\mathrm{disp}*}$.[43] However, we find that $E_{\mathrm{ind}*}$ needs to be corrected in the case of strongly cationic types of π-interactions. Hence, we use the following general formula which is applicable to all of the π-complexes to calculate the respective realistic energy components with respect to the E_{tot}(CCSD(T)/CBS).

$$E_{\mathrm{disp}**} = E_{\mathrm{disp}*}(\mathrm{SAPT(MP2)/basis}) + \kappa_{\mathrm{disp}}(E_{\mathrm{tot}}(\mathrm{CCSD(T)/CBS}) - E_{\mathrm{tot}}(\mathrm{SAPT(MP2)/basis})) \quad (9)$$

$$E_{\mathrm{ind}**} = E_{\mathrm{ind}*}(\mathrm{SAPT(MP2)/basis}) + \kappa_{\mathrm{ind}}(E_{\mathrm{tot}}(\mathrm{CCSD(T)/CBS}) - E_{\mathrm{tot}}(\mathrm{SAPT(MP2)/basis})) \quad (10)$$

On the basis of energy values from SAPT(MP2)/small basis and SAPT(MP2)/large basis, we calculate the ratios of the energy component changes to the total energy changes (denoted by κ_{disp} and κ_{ind} for the dispersion and induction terms). There are insignificant changes in the electrostatic and exchange energy components in all π-complexes (Table S1 of the Supporting Information). In general, in the neutral π complexes and the charged π^+-π complexes, almost solely the dispersion energy is underestimated, while in highly cationic π complexes (i.e., multiply charged cations such as Mg^{2+} and Ca^{2+} and highly ionizable cations such as Li^+ and Na^+ at a very short distance from the benzene molecule) the induction energies are underestimated. Hence, $\kappa_{\mathrm{disp}}/\kappa_{\mathrm{ind}}$ ($\kappa_{\mathrm{disp}} + \kappa_{\mathrm{ind}} = 1$) ranges from 0 to 1 depending on the nature of π-complexes. In Rg-π, CH-π, π-π, π-π, and π^+-π complexes, κ_{disp} is 1, while κ_{ind} is 0. For the Rg-π, CH-π, π-π and π^+-π, π^{+2}-π and H-π complexes, only the dispersion energy terms show significant changes; thus, $\kappa_{\mathrm{disp}} = 1$ and $\kappa_{\mathrm{ind}} = 0$. For the NH_4−benzene complex (H^+-π complex), κ_{ind} is 0.5 and κ_{disp} is 0.5. On the other hand, for the Me_4N^+ complex, which has three methyl groups interacting with the π-system, κ_{disp} is 1 and κ_{ind} is 0, as in the CH-π system. The κ values in the M^+-π and M^{+2}-π complexes depend on the size of the cation. For the Cs^+-π complex, $\kappa_{\mathrm{disp}} = 1$ and $\kappa_{\mathrm{ind}} = 0$; for the K^+-π complex, $\kappa_{\mathrm{disp}} = 0.6$ and $\kappa_{\mathrm{ind}} = 0.4$; for the Na^+-π complex, $\kappa_{\mathrm{disp}} = 0.4$ and $\kappa_{\mathrm{ind}} = 0.6$; for the Li^+-π complex, $\kappa_{\mathrm{disp}} = 0.1$ and $\kappa_{\mathrm{ind}} = 0.9$. For the Ca^{2+}-benzene complex, $\kappa_{\mathrm{disp}} = 0.2$ and $\kappa_{\mathrm{ind}} = 0.8$; for the Mg^{2+}-benzene complex, $\kappa_{\mathrm{disp}} = 0$ and $\kappa_{\mathrm{ind}} = 1$. The details of the proper κ values are in the Supporting Information. All of these estimations are used to obtain more realistic values toward the CCSD(T)/CBS limit. However, it should be noted that the corrections are not substantial, and so the raw data without such corrections do not significantly change our results.

We also investigated the ternary diagram analysis for energy component percentages (fractions in %) with respect to E_{attr} ($= E_{\mathrm{es}} + E_{\mathrm{ind}**} + E_{\mathrm{disp}**}$), which are defined as $f_{\mathrm{es}} = E_{\mathrm{es}}/E_{\mathrm{attr}}$, $f_{\mathrm{ind}} = E_{\mathrm{ind}**}/E_{\mathrm{attr}}$, $f_{\mathrm{disp}} = E_{\mathrm{disp}**}/E_{\mathrm{attr}}$, and $f_{\mathrm{exch}} = -E_{\mathrm{exch}*}/E_{\mathrm{attr}}$. In addition, another set of energy component percentages (fractions in %) is also investigated in consideration of all the attractive and repulsive energy components; $f'_{\mathrm{es,ind}} = (E_{\mathrm{es}} + E_{\mathrm{ind}**})/(E_{\mathrm{attr}} - E_{\mathrm{exch}**})$, $f'_{\mathrm{disp}} = E_{\mathrm{disp}**}/(E_{\mathrm{attr}} - E_{\mathrm{exch}*})$, and $f'_{\mathrm{exch}} = -E_{\mathrm{exch}*}/(E_{\mathrm{attr}} - E_{\mathrm{exch}*})$. For the proper classification of the π-interactions, the clustering technique is very important, and so a few well-chosen characteristic parameters (which could be represented by a linear or nonlinear combination of well-known physical quantities) should be sought out. Indeed, we note that energy component percentages are also useful parameters.

3. Results and Discussion

A. Model Systems of Various Types of π-Interactions. We have chosen various model systems for the Rg-π, CH-π, π-π(D), π-π(T), H-π(T), π^+-π(D), π^+-π(T), H^+-π(T), π^{+2}-π(D), M^+-π, and M^{+2}-π complexes (Figure 1), where diverse molecular species interact with a benzene molecule as simplified examples of these very complex systems. Since the electron-rich benzene is not interacting with anions, the electron-deficient π systems involving in the anion-π interactions are not discussed here. To facilitate our discussion, the π-interaction in the A-π(D/T) complexes will be denoted simply as the A-π(D/T) interaction, where A is a partner atom/molecule/ion. To model the Rg-π interaction, Rg is chosen with He and Kr. CH_4, C_2H_6, and C_2H_2 are chosen for the CH-π interaction. For the π-π(D/T) interaction, we consider benzene, toluene (Tol), and C_6H_5CN. For the H-π(T) interaction, we consider the representative cases of NH_3, H_2O, HCl, and CH_3OH. Notation H^+-π(T) is used for the complex between a non-π organic cation and a neutral π-molecule. For this system, NMe_4^+, NH_4^+, and NH_4^+−w are considered, where "w" denotes a water molecule. These non-π organic cations, where the positive charge is more concentrated on the hydrogen atom of the organic molecule cation, are differentiated from the cases of neutral H-π(T) interactions as seen with the complexes of a neutral π-system with water, ammonia, alcohols, etc. For the M^+-π interaction, M is chosen with Cs, K, Na, and Li, while for the M^{+2}-π interaction, M is chosen with Ca and Mg. These well-studied cation-π interactions need to be compared with the π_{cation}-π interactions which have been commonly observed but hardly investigated. To this end, we have calculated various model complexes involving with these interactions at the same level of theory.

To understand the nature of the π_{cation}-π interaction, we carry out a detailed and systematic theoretical investigation of diverse intermolecular model systems in which a benzene ring forms the π_{cation}-π interaction with various positively charged π model systems (Figure 2). The first set of models consists of singly charged moieties. Here, the interaction of methyl imidazolium/*N*-methyl-guanidinium with benzene is considered as the representative example for the interaction between the positively charged His/Arg and the neutral aromatic amino moiety (Phe, Tyr, or Trp). The second set of models consists of doubly charged π-moieties. The third set of models consists of the π_{cation}-π complexes interacting

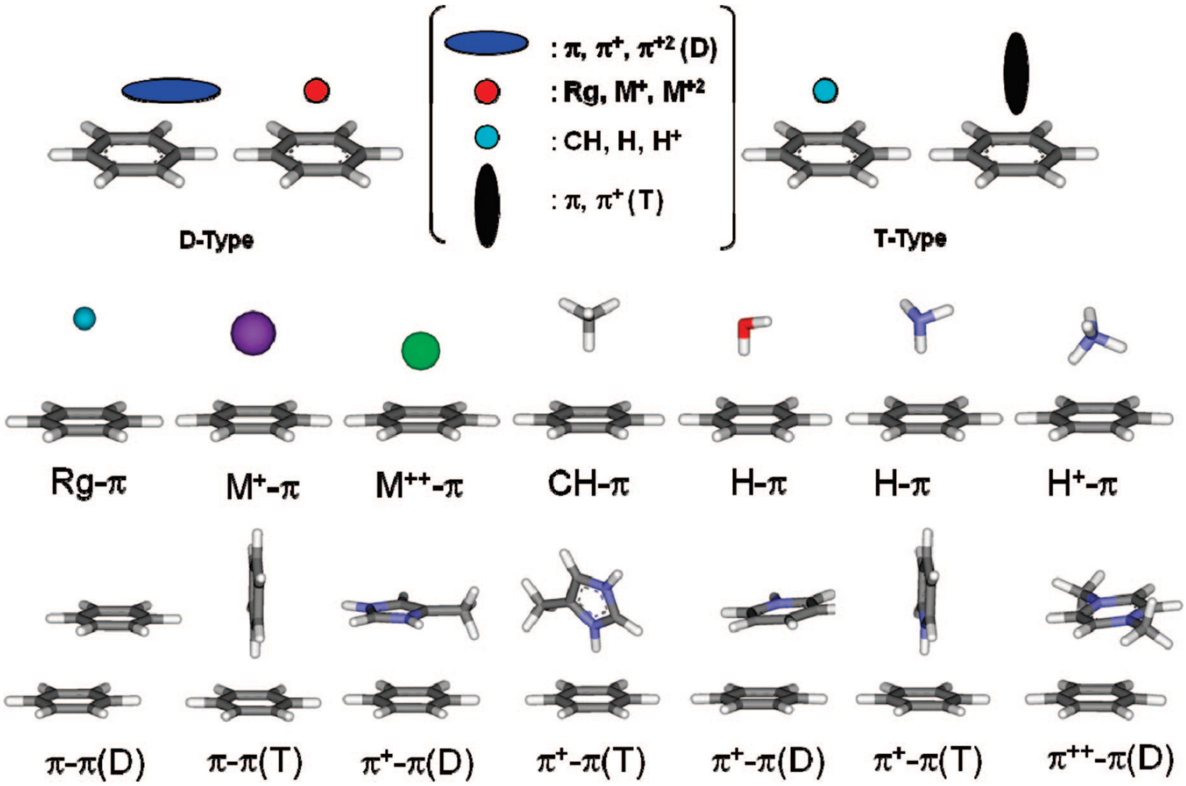


Figure 1. Schematic of various types of π-interaction (See Table 1 for the details of notation).

with water molecules (w) or counteranions (Cl^-/Br^-) for the investigation of the effect of solvent molecules and counterions.

B. Structures and Solvent/Counterion Effects in Terms of Total Interaction Energies. We discuss the MP2/aVDZ BSSE-corrected geometries and the most reliable CCSD(T)/CBS binding energies (to be denoted as $-E_{tot}$) of various π-complexes, because the MP2 energies are consistent with the CCSD(T) energies (Table 1).

Figure 3 shows the MP2/aVDZ BSSE-corrected geometries of the π_{cation}-π complexes in the following order. The methyl-guanidinium ($ArgH^+$ moiety; a, a_t), methyl imidazolium ($HisH^+$ moiety; b, b_t,), dimethyl-imidazolium ($ImMe_2^+$; c, c_t), and pyridinium (PyH^+; d, d_t) complexes have both displaced-stacked (D) and T-shaped (T) forms, where "(T)" is explicitly added to the complex name [e.g., $ArgH^+$(T)], while "(D)" is often dropped for brevity's sake unless there is confusion. For the $ArgH^+$/ $HisH^+$/PyH^+ complexes, the $(N-H)^+$ group directly points toward the benzene ring plane due to the presence of the positively charged H-atom attached to the N atom, so the most energetically favorable structure is T-shaped. For $ImMe_2^+$, the $(C-H)^+$ directly points to the benzene ring plane. As the H atom in the N−H group is replaced by the methyl group, the 1-methyl pyridinium complex ($PyMe^+$; e), 1,1′-dimethyl- [4,4′] bipyridinium ($BPMe_2^{+2}$; f), and 1,4-dimethyl pyrazinium ($PyMe_2^{+2}$; g) have no option to form T-shaped geometries, so they have the displaced-stacked structures. Furthermore, in the presence of polar solvents (water) or counterions, even the $ImMe_2^+$/PyH^+ complexes change to the displaced-stacked forms. Structures h,h′/i,i′ and h_t/i_t show two different displaced-stacked forms and one T-shaped form of the $ArgH^+$ moiety interacting with one/two water molecule(s), respectively. In structure h/i, the $ArgH^+$ moiety is significantly displaced from the benzene ring, and the water molecule forms the H-π interaction. In structure h′/i′, the $ArgH^+$ moiety is stacked with the benzene ring, and the water molecule(s) do not involve in the H-π interaction. In the case of $HisH^+$ complex, in the presence of two water molecules, the T-shaped isomer is still more stable than, but nearly isoenergetic to, the displaced-stacked one. Accordingly, both T-shaped and displaced-stacked structures would be compatible depending on the coordination environments. Structures j/k and j_t/k_t in Figure 3 show displaced-stacked and T-shaped isomers of the $HisH^+$ moiety interacting with one/two water molecule(s), respectively. Structures l and m show displaced-stacked forms of the pyridinium complexes coordinated by a water molecule and a bromide ion, respectively. Structure n displays the imidazolium complex coordinated by a chloride ion. For these π-complexes, we find the following trend:

The typical binding energies of the π^+-π(D) structures (~8−11 kcal/mol) and the π^+-π(T) structures (~9−14 kcal/mol) are much larger than the typical H-bonding energy (~5 kcal/mol for the water dimer),[126] CH-π (~1.5−3 kcal/mol) and H-π binding energy (~2−4 kcal/mol), but smaller than the typical cation-π (H^+-π and M^+-π) binding energy (~9−23 kcal/mol). In the cases of π^+-π complexes, the most stable structure in the presence of one or two water molecules is often topologically different from that in the gas phase. However, the conventional cation-π complexes (non-π organic H^+-π or metallic M^+-π complexes) retain their structural forms even in the presence of one or two water molecules coordinated to the cationic species.

Table 1. Geometrical Parameters (BSSE-corrected MP2/aVDZ) and Interaction Energies (MP2/CBS and CCSD(T)/CBS) of the π_{cation} Model Systems Interacting with a Benzene Molecule, Which Are Compared with the Rg-π, CH-π, H-π, π-π, and Cation-π Interactions Involving with a Benzene Molecule[a]

				MP2			CCSD(T)	
type	complex	θ	r_v (r_d)	aVDZ	aVTZ	CBS	aVDZ	CBS
Rg-π	He		3.46 (0.00)	−0.17	−0.21	−0.23	−0.16	−0.23
	Kr		3.64 (0.00)	−1.43	−1.84	−2.02	−0.87	−1.47
CH-π(T)	CH_4		6.80 (0.00)	−1.52	−1.68	−1.74	−1.21	−1.43
	C_2H_6	−0.9	3.66 (0.26)	−2.42	−2.69	−2.81	−1.85	−2.23
	C_2H_2	90	3.54 (0.00)	−3.04	−3.32	−3.45	−2.44	−2.85
H−π(T)	NH_3		3.60 (0.30)	−2.24	−2.48	−2.57	−1.93	−2.26
	H_2O		3.45 (0.42)	−2.99	−3.27	−3.39	−2.74	−3.14
	HCl		3.67(0.00)	−4.21	−4.61	−4.77	−3.37	−3.93
	CH_3OH		3.37(0.37)	−4.06	−4.47	−4.64	−2.74	−3.14
π-π(T)	C_6H_6 (T)	90	4.93 (0.74)	−3.32	−3.59	−3.72	−2.46	−2.83
	Tol(T)	90	4.84 (0.95)	−3.84	−4.13	−4.27	−2.88	−3.31
	C_6H_5CN(T)	90	4.91 (0.09)	−5.30	−5.73	−5.96	−4.05	−4.72
π-π(D)	C_6H_6	0	3.41 (1.55)	−4.29	−4.71	−4.93	−2.06	−2.61
	Tol	1.8	3.38 (1.57)	−5.42	−5.82	−6.03	−3.40	−4.00
	C_6H_5CN	4.2	3.42 (1.37)	−6.49	−7.04	−7.34	−3.50	−4.35
π^+-π(D)	$ArgH^+$	0.0	3.46 (1.32)	−10.30	−9.68	−9.42	−9.32	−8.44
	$HisH^+$	0.0	3.36 (1.11)	−10.11	−10.72	−10.98	−8.05	−8.92
	$ImMe_2^+$	10.2	3.36 (1.33)	−11.99	−12.71	−13.01	−9.98	−11.01
	PyH^+	6.5	3.36 (1.01)	−10.93	−11.48	−11.71	−8.12	−8.91
	$PyMe^+$	7.6	3.36 (1.20)	−11.39	−12.00	−12.25	−8.49	−9.35
	$ArgH^+$-w	−1.2	3.22 (2.80)	−9.46	−10.00	−10.23	−8.34	−9.11
	$ArgH^+$-w'	12.5	3.45 (1.23)	−9.25	−9.66	−9.84	−8.14	−8.73
	$ArgH^+$-w_2	−2.6	3.17 (2.84)	−8.79	−9.36	−9.60	−7.72	−8.53
	$ArgH^+$-w_2'	10.5	3.44 (1.26)	−8.23	−8.66	−8.85	−7.15	−7.77
	$HisH^+$-w	4.0	3.33 (0.16)	−9.65	−10.18	−10.40	−7.61	−8.36
	$HisH^+$-w_2	1.2	3.31 (0.16)	−8.96	−9.50	−9.73	−6.90	−7.66
	$ImMe_2^+$-Cl^-	8.8	3.29 (1.44)	−10.64	−11.39	−11.71	−8.02	−9.09
	PyH^+-w	6.9	3.37 (1.47)	−10.24	−10.83	−11.07	−7.93	−8.76
	PyH-Br^-	8.6	3.37 (1.31)	−9.03	−9.64	−9.90	−7.97	−8.85
π^+-π(T)	$ArgH^+$(T)	90	3.96 (0.46)	−14.35	−15.00	−15.27	−13.04	−13.96
	$HisH^+$(T)	73.6	4.14 (0.25)	−14.40	−15.11	−15.40	−12.67	−13.68
	$ImMe_2^+$ (T)	86.7	4.52 (0.10)	−11.90	−12.55	−12.83	−11.10	−12.03
	PyH^+(T)	90	4.38 (0.00)	−15.82	−16.60	−16.93	−14.03	−15.14
	$ArgH^+$-w(T)	89.8	3.98 (0.55)	−12.91	−13.54	−13.80	−11.69	−12.58
	$ArgH^+$-w_2(T)	89.8	4.01 (0.56)	−8.72	−9.35	−9.62	−7.61	−8.50
	$HisH^+$-w(T)	74.7	4.18 (0.25)	−13.15	−13.81	−14.09	−11.56	−12.49
	$HisH^+$-w_2(T)	76.9	4.21 (0.13)	−9.63	−10.42	−10.75	−8.16	−9.28
H^+−π(T)	Me_4N^+		4.28[b] (0.00)	−9.86	−10.28	−10.46	−8.89	−9.49
	NH_4^+		2.94 (0.06)	−18.80	−20.05	−20.58	−19.64	−21.41
	NH_4^+-w		2.94 (0.07)	−15.14	−15.88	−16.19	−14.25	−15.29
π^{++}-π(D)	$PyMe_2^{+2}$	8.8	3.16 (1.13)	−25.31	−26.27	−26.68	−20.85	−22.23
	$BPMe_2^{+2}$	6.7	3.32 (1.13)	−17.48	−19.33	−20.11		−16[c]
M^+-π	Cs^+		3.24 (0.00)	−15.52	−14.81	−14.51	−14.09	−13.08
	>K^+		2.90 (0.00)	−17.12	−18.74	−19.42	−16.47	−18.77
	Na^+		2.48 (0.00)	−22.33	−22.80	−23.00	−22.28	−22.95
	Li^+		1.92 (0.00)	−35.48	−37.37	−38.16	−35.44	−38.13
	K^+-w		2.94 (0.00)	−15.10	−16.32	−16.83	−14.50	−16.22
	Na^+-w		2.52 (0.00)	−22.28	−22.64	−22.79	−19.19	−19.70
	Li^+-w		1.98 (0.00)	−28.77	−30.16	−30.74	−28.63	−30.30
M^{++}-π	Ca^{2+}		2.39 (0.00)	−77.85	−81.43	−82.93	−76.32	−81.40
	Mg^{2+}		1.98 (0.00)	−111.94	−115.24	−116.63	−112.16	−116.85
	Ca^{2+}-w		2.44 (0.00)	−66.28	−68.82	−69.90	−64.93	−68.55
	Mg^{2+}-w		2.02 (0.00)	−94.30	−96.57	−97.53	−94.48	−97.71

[a] θ: angle (in degrees) tilted from the parallel stacking, r_v: vertical stacking distance for ring systems or distance from the heavy atom of non-π systems (or the middle carbon atom in the case of $ArgH^+$) to the benzene ring plane, r_d: off-center displacement. Distances in Å; energies in kcal/mol. CCSD(T)/CBS E_{tot}'s are obtained by applying the correction term (the difference of the BSSE-corrected interaction energies between MP2/aVDZ and CCSD(T)/aVDZ) to the BSSE-corrected MP2/CBS interaction energies (based on the extrapolation scheme utilizing the basis set error in the electron correlation energy proportional to N^{-3} for the aug-cc-pV*N*Z basis set (reference [34])). [b] Distance between the nitrogen atom and the benzene centroid. The distance from the carbon atom to the benzene ring plane is 3.38 Å. [c] Roughly estimated from the MP2/CBS estimated interaction energy (−20.11 kcal/mol) [For $PyMe_2^{+2}$, the CCSD(T)/CBS energies is ~4.5 kcal/mol less than the MP2/CBS values].

The T-shaped forms of $ArgH^+$(T), $HisH^+$(T), and PyH^+(T) ($-E_{\text{tot}}$: 14−15 kcal/mol) are ~5−6 kcal/mol more stable than the corresponding displaced-stacked forms. $ImMe_2^+$(T) is ~1 kcal/mol more stable than the displaced-stacked one. However, other π^+-π complex structures favor the displaced-stacked forms. The binding energy is ~8−11 kcal/mol for the π^+-π interaction of the singly charged systems and ~22 kcal/mol for the π^{+2}-π interaction of the doubly charged

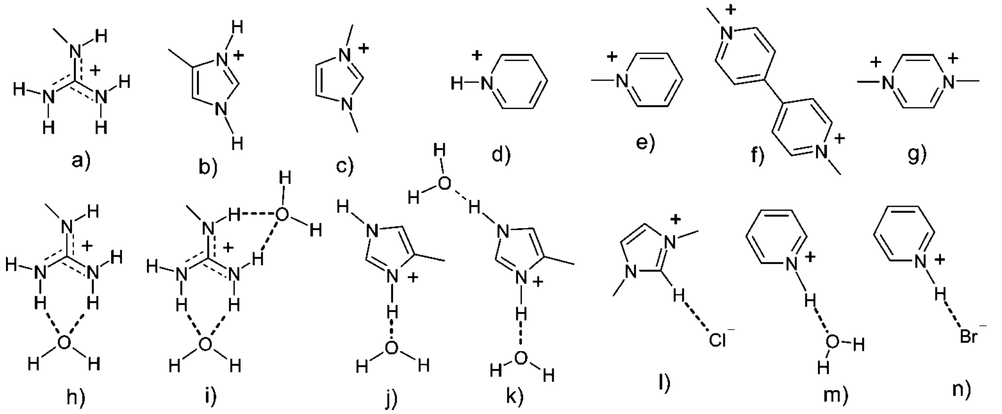


Figure 2. Selected positively charged π-moieties and the associated solvent molecules/counteranions. (a) *N*-methyl-guanidinium (representing the side chain of the protonated arginine), (b) methyl imidazolium (representing the side chain of the protonated histidine), (c) dimethyl imidazolium, (d) pyridinium, (e) 1-methyl pyridinium, (f) 1,1′-dimethyl-[4,4′] bipyridinium, (g) 1,4-dimethyl pyrazinium, (h,i) *N*-methyl-guanidinium with one or two water molecules, (j,k) methyl imidazolium with one or two water molecules, (l) dimethyl imidazolium with a chloride anion, and (m,n) pyridinium with one water molecule or a bromide anion.

systems in a single ring. Doubly charged $BPMe_2^{+2}$ with a positive charge in each of the two rings has the binding energy of ~16 kcal/mol, which is ~7 kcal/mol more than the binding energy of $PyMe^+$ and ~6 kcal/mol less than that of $PyMe_2^{+2}$ (doubly charged on a single ring).

For the π^+-π(D) structures, the vertical distance between two stacked rings (r_v) is 3.3−3.5 Å for the singly charged systems (except for $ArgH^+$−w/w_2(D) which has slightly shorter distances of 3.2 Å due to the large off-center displacement r_d), ~3.3 Å for the doubly charged system on two rings, and ~3.2 Å for the doubly charged system on a single ring, which are shorter than the distance (3.41 Å) observed for the neutral benzene-benzene dimer in the displaced-stacked form. It would be noted that the $ArgH^+$ moiety is a weakly π-conjugated system as compared with aromatic ring systems (such as $HisH^+$, PyH^+, and $ImMe_2^+$) discussed here, so the binding energies of the $ArgH^+$(−w/w_2) system are slightly smaller than aromatic ring systems.

The solvent and counterion effects play an important role in determining the structures of the π^+-π complexes. For the methyl-guanidium, in the absence of water, the T-shaped $ArgH^+$ moiety is 4.5 kcal/mol more stable than the displaced-stacked one. For the complex with one water molecule, the displaced-stacked $ArgH^+$−w(h) is about 0.5 kcal/mol more stable than the displaced-stacked $ArgH^+$−w(h′), but still 3.5 kcal/mol less stable than the T-shaped $ArgH^+$−w(T)(h_t). For the complex with two water molecules, the displaced-stacked $ArgH^+$−w2(i) is about 0.8 kcal/mol more stable than the displaced-stacked $ArgH^+$−w2(i′), and is isoenergetic to or slightly more stable than the T-shaped $ArgH^+$−w2(T)(i_t). For the protonated methyl-imidazolium, the T-shaped isomer in the presence of one/two water molecules ($HisH^+$−w(T)(j_t)/ $HisH^+$−w2(T)(k_t)) is ~4.1/1.7 kcal/mol more stable than the displaced-stacked isomer [$HisH^+$−w(j)/ $HisH^+$−w2(k)]. Either in the presence of both a counteranion (such as an acetate group) and a water molecule or in the presence of more than two coordinating species (water or counteranions or carbonyl oxygen of amide bond, etc.), the T-shaped isomers of $ArgH^+$/$HisH^+$ would eventually be isoenergetic to or slightly less stable than the displaced-stacked isomers. This result is in good agreement with the previous analysis of proteins and protein−ligand complexes for the interactions of neutral aromatic amino acids with positively charged π-cloud of protonated arginine/histidine, where the displaced-stacked structure is preferred to the T-shaped structure or otherwise both structures are equally observable.[111−114]

Furthermore, when the T-shaped structures of the PyH^+/$ImMe_2^+$−benzene complexes are optimized in the presence of one water molecule or a counterion, they change to the displaced-stacked structures. Thus, in the presence of polar solvent molecules and counteranions in the vicinity of positively charged π-systems, the T-shaped forms are no longer stable and so the displaced-stacked forms are favored because the polar solvent molecules and counterions interact with the charged moiety of the charged π-ring.

The Cambridge structural data analysis (CSD version 5.29, November 2007)[127,128] for the interaction between pyridinium (including substituted and fused ring systems) and benzene (including substituted and fused ring systems) within the centroid-to-centroid intramolecular distance of 5 Å shows that ~90% of the complexes are found to have displaced-stacked structural forms among the total 701 hits. The hydrogen atom in $(N-H)^+$ is in general coordinated to solvent molecules, counteranions, or other partner molecules. We have not found a single example where the hydrogen atom of $(N-H)^+$ directly points to the ring plane of the benzene ring. About 10% of the complexes are found to have angled T-shape structural forms where the hydrogen atom(s) of the C atom(s) of the pyridinium is pointing toward the benzene ring plane. However, the hydrogen atom in $(N-H)^+$ forms a strong hydrogen bond with a solvent molecule, a counteranion, or other partner molecules. Thus, our results are in good agreement with what are observed in the crystal structures.

C. Energy Components. Table 2 lists E_{tot} and SAPT energy components of various π-model systems, and these average values for various types of π-interactions are listed in Table 3. Figure 4a shows the 3-dimensional (3D) plot of E_{tot} as the function of E_{exch*} and E_{attr} and Figure 4b shows the 3D plot of E_{tot} as the function of E_{es*} and E_{ind**}. Figure

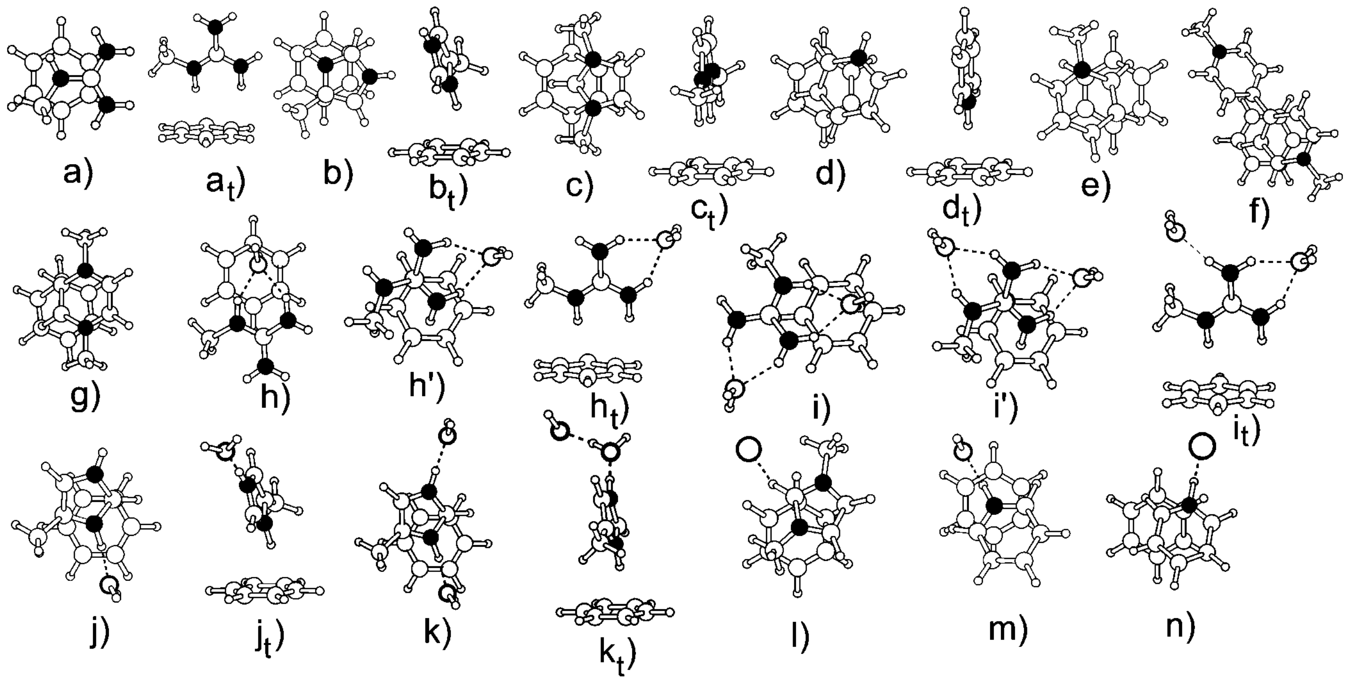


Figure 3. Calculated geometries (BSSE-corrected MP2/aug-cc-pVDZ) of model complex systems whose positively charged aromatic ring interacts with a neutral aromatic benzene ring: displaced stacked (D) and T-shaped (T) forms of *N*-methyl-guanidinium complexes (ArgH$^+$(D), ArgH$^+$(T)) (a$_t$,a); displaced stacked and T-shaped forms of the methyl imidazolium complexes (HisH$^+$(D), HisH$^+$(T)) (b$_t$,b); dimethyl imidazolium complex in displaced-stacked and T-shape forms (ImMe$_2{}^+$(D), ImMe$_2{}^+$(T)) (c$_t$,c); displaced stacked and T-shaped forms of the pyridinium complexes (PyH$^+$(D), PyH$^+$(T)) (d$_t$,d); 1-methyl pyridinium complex (PyMe$^+$) (e); the 1,1′-dimethyl-[4,4′] bipyridinium complex (BPMe$_2{}^{+2}$) (f); 1,4-dimethyl pyrazinium complex (PyMe$_2{}^{+2}$) (g); *N*-methyl-guanidinium complexes coordinated by one (w) and two water (w2) molecules in two displaced-stacked and T-shaped forms (h, h′, h$_t$ and i, i′, i$_t$,); methyl imidazolium complexes coordinated by one (w) and two water (w2) molecules in displaced-stacked and T-shaped forms (j, j$_t$ and k, k$_t$); dimethyl imidazolium complex coordinated by chloride (ImMe$_2{}^+$−Cl$^-$) (l); and pyridinium complexes coordinated by water and bromide (PyH$^+$-w, PyH$^+$-Br$^-$) (m, n). In BPMe$_2{}^{+2}$, one ring is twisted by ~37 ° with respect to the other ring interacting with the neutral benzene ring. The other structure that both rings are planar is ~2 kcal/mol less stable. Structures h′ and i′ are slightly less stable than structures h and i, respectively.

4c plots the correlation of $E_{\rm tot}$ with (i) $E_{\rm es*}$, (ii) $E_{\rm ind**}$, (iii) $E_{\rm es,ind}$ ($E_{\rm es,ind} = E_{\rm es*} + E_{\rm ind**}$), and (iv) $E_{\rm disp,exch}$ ($E_{\rm disp,exch} = E_{\rm disp**} + E_{\rm exch*}$). The $E_{\rm es*}$ (square of correlation factor, $R^2 = 0.921$), $E_{\rm ind**}$ ($R^2 = 0.993$), $E_{\rm es,ind}$ ($R^2 = 0.997$), and $E_{\rm disp,exch}$ ($R^2 = 0.927$) show correlations with $E_{\rm tot}$, while $E_{\rm disp**}$ and $E_{\rm exch*}$ do not show good correlations with $E_{\rm tot}$. Hence, $E_{\rm tot}$ shows a good correlation with $E_{\rm ind**}$ and in particular, $E_{\rm es,ind}$. In the $E_{\rm tot}$ vs $E_{\rm es*}$ regression, the quadratic polynomial regression of $E_{\rm es*}$ with respect to $E_{\rm tot}$ drastically improves the correlation ($R^2 = 0.980$ for $E_{\rm es*}$). This indicates the importance of the square terms of $E_{\rm es*}$, where $E_{\rm es*}$ is related to site charges, which implicitly reflects the induction term proportional to the square of the electric field correlated to the site charges. Hence, the $E_{\rm tot}$ vs $E_{\rm es,ind}$ linear regression results in an excellent correlation ($R^2 = 0.997$). This indicates that the difference in $E_{\rm tot}$ between different types of π-interaction can be reasonably explained by $E_{\rm es,ind}$. However, it does not mean that $E_{\rm es,ind}$ can explain small differences between same types of π-interaction such as π-π-(D), since $E_{\rm es,ind}$ alone cannot properly distinguish small energy difference in $E_{\rm tot}$ due to substitution in aromatic ring.[43] $E_{\rm disp**}$ and $E_{\rm exch*}$ independently show very poor correlations with $E_{\rm tot}$ (Figure 4c-iv), while $E_{\rm disp,exch}$ shows an anticorrelation with $E_{\rm tot}$. Overall, the $E_{\rm es*}$ component shares a large portion of $E_{\rm tot}$, but its variation depending on the types of π-interaction is not substantial. However, the $E_{\rm ind**}$ component changes drastically depending on the types of π-interaction. In this regard, $E_{\rm ind**}$ is the crucial energy component to differentiate different types of π-interaction, while other components also play significant roles in differentiating them. The detailed analysis of Tables 2 and 3 as well as Figure 4 is given in the Supporting Information. Here, we discuss only the main essence of the findings in the next paragraphs.

In going from the Rg-π complexes to the M^{+2}-π complexes, the increasing order in the values of $-E_{\rm es*}$ or $-E_{\rm ind**}$ is not correlated with the increasing/decreasing order in the values of $-E_{\rm disp**}$ or $E_{\rm exch*}$. A general observation from the above analysis is that $-E_{\rm es*}$ and $-E_{\rm ind**}$ are the smallest in the Rg-π complexes, followed by the CH-π, π-π, and H-π complexes, while they are large in the π^{+2}-π(D) and M^{+2}-π complexes. $-E_{\rm es*}$ is larger than $-E_{\rm ind*}$ in all the M^{+2}-π complexes with the exception of the Li$^+$ and Li$^+$−w complexes. For the M$^+$-π interaction, there is a large variation in $-E_{\rm ind**}$ (5−30 kcal/mol) because the Cs$^+$/Li$^+$ ion is at a very long/short distance from the benzene molecule. In all of the charged π-complexes, the electrostatic energy is important, and so is the induction energy except for the π^+-π(D) complexes.

The π^+-π(D), π-π(D), and π^{+2}-π(D) complexes have large dispersion energy ($E_{\rm disp**}$, ~9−13 kcal/mol), while the M$^+$-π complexes have small dispersion energy (~0.6−6 kcal/mol).

Table 2. Interaction Energies (E_{tot}: CCSD(T)/CBS), Energy Components (E_{es^*}, $E_{\mathrm{ind}^{**}}$, $E_{\mathrm{disp}^{**}}$, E_{exch^*}; $E_{\mathrm{attr}} = E_{\mathrm{es}^*} + E_{\mathrm{ind}^{**}} + E_{\mathrm{disp}^{**}}$) at the SAPT(MP2)/6-31+G* level or a VDZ$^+$ level, and Fractional Values of Energy Components Reported in Percentages [$f_{\mathrm{es}} = E_{\mathrm{es}^*}/E_{\mathrm{attr}}$, $f_{\mathrm{ind}} = E_{\mathrm{ind}^{**}}/E_{\mathrm{attr}}$, $f_{\mathrm{disp}^{**}} = E_{\mathrm{disp}^{**}}/E_{\mathrm{attr}}$, $f_{\mathrm{exch}} = -E_{\mathrm{exch}^*}/E_{\mathrm{attr}}$, $f'_{\mathrm{es,ind}} = (E_{\mathrm{es}^*} + E_{\mathrm{ind}^{**}})/(E_{\mathrm{attr}} - E_{\mathrm{exch}^*})$, $f'_{\mathrm{disp}^{**}} = E_{\mathrm{disp}^*}/(E_{\mathrm{attr}} - E_{\mathrm{exch}^*})$, $f'_{\mathrm{exch}^*} = -E_{\mathrm{exch}^*}/(E_{\mathrm{attr}} - E_{\mathrm{exch}^*})$] of the π-Model Systems Interacting with a Benzene Molecule[a]

type	complex	E_{tot}	E_{es^*}	$E_{\mathrm{ind}^{**}}$	$E_{\mathrm{disp}^{**}}$	E_{exch^*}	f_{es}	f_{ind}	f_{disp}	f_{exch}	$f'_{\mathrm{es,ind}}$	f'_{disp}	f'_{exch}
Rg-π	He	−0.23	−0.04	−0.02	−0.32	0.15	9.9	5.9	84.1	40.5	11.3	59.9	28.8
	Kr	−1.47	−0.91	−0.17	−2.67	2.28	24.2	4.4	71.3	60.8	17.8	44.3	37.8
CH-π(T)	CH_4	−1.43	−0.91	−0.25	−2.46	2.19	25.1	6.9	68.0	60.5	19.9	42.4	37.7
	C_2H_6	−2.23	−1.53	−0.39	−4.11	3.80	25.3	6.5	68.2	63.0	19.5	41.8	38.7
	C_2H_2	−2.85	−2.23	−0.83	−3.10	3.32	36.2	13.5	50.3	53.8	32.3	32.7	35.0
H−π(T)	NH_3	−2.26	−1.78	−0.43	−2.57	2.52	37.3	8.9	53.8	52.6	30.3	35.3	34.5
	H_2O	−3.34	−2.73	−0.79	−2.81	2.99	43.2	12.5	44.4	47.3	37.8	30.1	32.1
	HCl	−3.93	−3.41	−1.33	−3.73	4.53	40.3	15.7	44.0	53.5	36.5	28.7	34.9
	CH_3OH	−4.11	−3.40	−1.06	−4.65	5.00	37.3	11.7	51.0	54.9	31.6	32.9	35.4
π -π(T)	C_6H_6 (T)	−2.83	−2.11	−0.53	−4.17	3.98	31.0	7.8	61.2	58.5	24.5	38.6	36.9
	Tol(T)	−3.31	−2.35	−0.59	−5.08	4.71	29.3	7.4	63.3	58.7	23.1	39.9	37.0
	C_6H_5CN(T)	−4.72	−4.24	−1.28	−6.10	6.91	36.5	11.0	52.5	59.5	29.8	32.9	37.3
π-π(D)	C_6H_6	−2.61	−2.58	−0.84	−7.57	8.38	23.5	7.6	68.9	76.3	17.6	39.1	43.3
	Tol	−4.00	−3.26	−0.98	−9.66	9.90	23.5	7.0	69.5	71.2	17.8	40.6	41.6
	C_6H_5CN	−4.35	−5.02	−1.08	−9.61	11.36	32.0	6.9	61.2	72.3	22.5	35.5	42.0
π^+-π(D)	$ArgH^+$	−8.44	−5.99	−2.38	− 7.49	7.42	37.8	15.0	47.2	46.8	36.0	32.2	31.9
	$HisH^+$	−8.92	−7.36	−2.89	−8.86	10.86	38.5	15.1	46.4	56.8	34.2	29.6	36.2
	$ImMe_2^+$	−11.01	−8.61	−3.19	−10.57	11.36	38.5	14.2	47.3	50.8	35.0	31.3	33.7
	PyH^+	−8.91	−8.68	−3.76	−9.09	12.61	40.3	17.5	42.2	58.6	36.4	26.6	36.9
	$PyMe^+$	−9.35	−8.64	−3.57	−10.02	12.87	38.9	16.1	45.1	57.9	34.8	28.6	36.7
	$ArgH^+$−w	−9.11	−7.89	−3.90	−7.10	9.78	41.8	20.6	37.6	51.8	41.1	24.8	34.1
	$ArgH^+$−w′	−8.73	−7.94	−3.40	−7.31	9.91	42.6	18.2	39.2	53.1	39.7	25.6	34.7
	$ArgH^+$−w_2	−8.53	−7.45	−3.45	−7.26	9.64	41.0	19.0	40.0	53.1	39.2	26.1	34.7
	$ArgH^+$−w_2'	−7.77	−7.18	−2.76	−7.36	9.53	41.5	16.0	42.5	55.1	37.0	27.4	35.5
	$HisH^+$−w	−8.36	−7.38	−2.80	−9.04	10.87	38.4	14.6	47.1	56.5	33.8	30.1	36.1
	$HisH^+$−w_2	−7.66	−6.88	−2.43	−9.15	10.79	37.3	13.1	49.6	58.5	31.8	31.3	36.9
	$ImMe_2^+$-Cl^-	−9.09	−11.08	−3.43	− 11.53	16.94	42.6	13.2	44.3	65.1	33.8	26.8	39.4
	PyH^+−w	−8.76	−7.86	−3.02	−9.28	11.41	39.0	15.0	46.0	56.6	34.5	29.4	36.1
	PyH-Br^-	−8.85	−9.17	−2.47	−11.94	14.73	38.9	10.5	50.6	62.5	30.4	31.2	38.5
π^+-π(T)	$ArgH^+$(T)	−13.96	−10.88	−7.40	− 7.61	11.92	42.0	28.6	29.4	46.0	48.4	20.1	31.5
	$HisH^+$(T)	−13.68	−10.22	−7.22	−7.86	11.62	40.4	28.5	31.1	45.9	47.2	21.3	31.5
	$ImMe_2^+$ (T)	−12.03	−8.34	−5.06	−7.79	9.16	39.3	23.9	36.8	43.2	44.1	25.7	30.2
	PyH^+(T)	−15.14	−10.98	−7.74	−8.06	11.64	41.0	28.9	30.1	43.5	48.7	21.0	30.3
	$ArgH^+$−w(T)	−12.58	−9.86	−6.16	−7.47	10.90	42.0	26.2	31.8	46.4	46.6	21.7	31.7
	$ArgH^+$-w_2(T)	−8.50	−9.18	−5.39	−4.22	10.30	48.8	28.7	22.5	54.8	50.1	14.5	35.4
	$HisH^+$−w(T)	−12.49	−9.39	−6.17	−7.59	10.66	40.6	26.6	32.8	46.0	46.0	22.4	31.5
	$HisH^+$−w_2(T)	−9.28	−8.94	−5.61	−4.76	10.02	46.3	29.0	24.6	51.9	49.6	16.2	34.2
H^+-π(T)	Me_4N^+	−9.49	−7.33	− 4.04	−6.37	8.25	41.3	22.8	35.9	46.5	43.7	24.5	31.7
	NH_4^+	−21.20	−13.27	−14.08	−6.85	13.00	38.8	41.2	20.0	38.0	57.9	14.5	27.5
	NH_4^+−w	−15.29	−12.00	−10.57	−5.97	13.24	42.0	37.0	20.9	46.4	54.0	14.3	31.7
π^{++}-π(D)	$PyMe_2^{+2}$	−22.23	−17.56	−14.16	−13.15	22.64	39.1	31.6	29.3	50.5	47.0	19.5	33.5
M^+-π	Cs^+	−13.08	−12.97	−5.62	− 5.63	11.13	53.5	23.2	23.3	46.0	52.6	15.9	31.5
	K^+	−18.77	−13.59	−9.98	−3.57	8.36	50.1	36.8	13.1	30.8	66.4	10.1	23.6
	Na^+	−22.95	−15.64	−13.84	−0.59	6.81	52.0	46.0	2.0	22.7	79.9	1.6	18.5
	Li^+	−38.13	−17.84	−29.97	−0.59	10.28	36.9	61.9	1.2	21.2	81.5	1.0	17.5
	K^+−w	−16.22	−12.28	−8.18	−3.27	7.51	51.8	34.5	13.8	31.6	65.5	10.5	24.0
	Na^+−w	−19.70	−13.93	−11.19	−0.76	6.00	53.8	43.2	2.9	23.2	78.8	2.4	18.8
	Li^+−w1	−30.60	−15.29	−22.23	−1.26	8.18	39.4	57.3	3.3	21.1	79.9	2.7	17.4
M^{++}-π	Ca^{2+}	−81.40	−36.17	−69.62	− 6.28	30.67	32.3	62.1	5.6	27.4	74.1	4.4	21.5
	Mg^{2+}	−116.85	−40.11	−102.82	−0.39	26.47	28.0	71.7	0.3	18.5	84.2	0.2	15.6
	Ca^{2+}−w	−68.55	−33.26	−57.35	−5.24	27.29	34.7	59.8	5.5	28.5	73.6	4.3	22.2
	Mg^{2+}−w	−97.71	−36.32	−84.20	−1.22	24.03	29.8	69.2	1.0	19.7	82.7	0.8	16.5

[a] SAPT(MP2)/6−31+G* is used for the energy component calculations of π-π(D), π-π(T), π^+-π(D), π^+-π(T), π^{+2}-π(D), M^+-π, and M^{+2}-π complexes, while SAPT(MP2)/aVDZ′ is used for Rg-π, CH-π, H-π(T), and H^+-π(T) complexes, as the total energies were highly underestimated for these complexes at the SAPT(MP2)/6−31+G* level. The E_{es^*}, $E_{\mathrm{ind}^{**}}$, $E_{\mathrm{disp}^{**}}$, and E_{exch^*} values at the SAPT(MP2)/aVDZ† level are, respectively: −10.22, −8.18, −8.57, 11.83 for PyH^+(T); −8.44, −4.00, −9.44, 13.00 for PyH^+(D); −8.34, −3.74, −10.31, and 13.07 for $PyMe^+$. These values are consistent with the SAPT(MP2)/6−31+G* results, as in the table. While the base set dependency of E_{es^*} and E_{exch^*} is not significant, E_{disp^*} and E_{ind^*} depend on the basis set. Thus, the realistic $E_{\mathrm{disp}^{**}}$ and $E_{\mathrm{ind}^{**}}$ are obtained as described in computational details and Supporting Information.

The Rg-π, CH-π(T), and H-π(T) complexes have small exchange energy, while the π^{+2}-π(D) and M^{+2}-π complexes have very large exchange energy. Given that the dispersion energy depends on the size of a partner molecule or the total number of electrons or the sum of atomic numbers, the values of $-E_{\mathrm{disp}^{**}}$ of the H-π(T) interactions composed of only small species in this study are much smaller than those of other types of π-interaction. If similar sizes are considered, then the CH-π(T) and H-π(T) interactions would be more or less equivalent in $-E_{\mathrm{disp}^{**}}$ to the π-π(T) interaction.

In the Rg-π complex (Rg = He, Kr), $-E_{\mathrm{disp}^{**}}$ (0.32−2.7 kcal/mol) is partly canceled by E_{exch^*}, gaining only a small

Table 3. Average Interaction Energies (E_{tot}: CCSD(T)/CBS), Energy Components (SAPT(MP2)/6-31+G*), and Energy Component Percentages for the Various Types of π-Interactions, Which Are Ordered in Terms of Increasing Induction Energy Component Percentage f_{ind}

	E_{tot}	E_{es*}	E_{ind**}	E_{disp**}	E_{exch*}	f_{es}	f_{ind}	f_{disp}	$f'_{es,ind}$	f'_{disp}	f'_{exch}
Rg-π	−0.8	−0.5	−0.1	−1.5	1.2	17.1	5.2	77.7	14.6	52.1	33.3
π-π (D)	−3.7	−3.6	−1.0	−8.9	9.9	26.3	7.2	66.5	19.3	38.4	42.3
π -π(T)	−3.6	−2.9	−0.8	−5.1	5.2	32.3	8.7	59.0	25.8	37.2	37.1
CH-π(T)	−2.2	−1.6	−0.5	−3.2	3.1	28.9	9.0	62.1	23.9	39.0	37.1
H-π (T)	−3.4	−2.8	−0.9	−3.4	3.8	39.5	12.2	48.3	34.0	31.7	34.2
π^+-π (D)	−8.8	−8.0	−3.1	−9.0	11.3	39.8	15.6	44.6	35.5	28.6	35.8
π^+-π(T)	−12.2	−9.7	−6.3	−6.9	10.8	42.6	27.6	29.9	47.6	20.4	32.0
π^{++}-π(D)	−22.2	−17.6	−14.2	−13.1	22.6	39.1	31.6	29.3	47.0	19.5	33.5
H^+-π(T)	−15.3	−10.9	−9.6	−6.4	11.5	40.7	33.7	25.6	51.9	17.8	30.3
M^+-π	−22.8	−14.5	−14.4	−2.2	8.3	48.2	43.3	8.5	72.1	6.3	21.6
M^{++}-π	−91.1	−36.5	−78.5	−3.3	27.1	31.2	65.7	3.1	78.6	2.4	18.9

amount of binding energy. However, as the electron density in the space between benzene and He/Kr is redistributed outside the benzene−Kr complex (while it is slightly more distributed outside benzene), there is small $-E_{es*}$ due to small charge transfer from the benzene ring to the Rg (natural bond orbital charge of −0.003 au for Kr at the CCSDT(T)/aVDZ level) as well as small polarization-driven $-E_{ind**}$, resulting in a substantial binding energy contribution in the small total binding energy. In the CH-π complexes, $-E_{disp**}$ is larger than $-E_{es*}$, and $-E_{ind**}$ is relatively small.

The π-π(D) complexes have larger $-E_{disp**}$ (∼8−10 kcal/mol) than the π-π(T) complexes (∼4−6 kcal/mol), which are larger than corresponding $-E_{es}$ (∼2−5 kcal/mol) and $-E_{ind*}$ (∼0.5−1 kcal/mol). However, the H-π(T) complexes tend to have smaller $-E_{disp**}$ (2.5−4.6 kcal/mol) than the π-π(T) complexes in this study. However, given that the partner molecules in the H-π(T) complexes are smaller than those in π-π(T) complexes, there would be no significant differences in E_{disp**} between H-π(T) and π-π(T) complexes.

In the π^+-π(D) complexes, $-E_{disp**}$ (∼7−12 kcal/mol) is only slightly larger than $-E_{es*}$, while $-E_{ind**}$ is small and E_{exch*} is relatively large. These are contrasted to the π^+**-**π(T) complexes, where $-E_{es*}$ (∼8−11 kcal/mol) is larger than $-E_{disp**}$. In this regard, the H^+-π(T) complexes are similar to those of π^+-π(T) complexes. The $-E_{es*}$ of small organic cations (NH_4^+ and NH_4^+−w) are nearly similar to $-E_{ind**}$, while for bigger organic cations (Me_4N^+) $-E_{es*}$ is larger than $-E_{ind**}$. In the π^+-π(T)/ H^+-π(T) complexes, $-E_{exch*}$ and $-E_{ind**}$ are much larger than those in the π-π(D), π-π(T) and π^+-π(D) complexes. Alternatively, in the π^{++}**-**π(D) complex of $PyMe_2^{+2}$, the $-E_{es*}$ (18 kcal/mol) is larger than $-E_{ind**}$ and $-E_{disp**}$, but smaller than E_{exch*}.

In the M^+**-**π complexes, $-E_{es*}$ (∼12−16 kcal/mol) is larger than $-E_{ind**}$ excluding the Li^+ and Li^+−w complexes where $-E_{es*}$ is smaller than $-E_{ind**}$. These are contrasted to the M^{++}-π complexes where $-E_{ind**}$ (∼57−102 kcal/mol) is larger than $-E_{es*}$. In both M^+-π and M^{+2}-π complexes, $-E_{disp**}$ is much smaller than E_{exch*}. For the M^{+2}-π complexes, the magnitude of the sum of the attractive energy components ($-E_{attr}$) is much larger than E_{exch*}, hence the binding energy is the largest among the π-complexes studied here.

One may note at this point that the π-π interaction has large dispersion energy, the cation-π interaction has large electrostatic and induction energies, and the π^+-π interaction has large dispersion and electrostatic energies. Since the total energies vary significantly depending on the types of π-complexes, the degree of contribution of these energy components to the total binding energy is another way to analyze different types of π-complexes. Thus, to more clearly understand the similarities and differences between these important π interactions, we look into the percentage contributions from each of the attractive and repulsive energy components using the ternary diagram analysis.

D. Ternary Diagram Analysis. Since each π-interaction is represented by five components (E_{tot}, E_{es}, E_{ind*}, E_{disp**}, and E_{exch*}), a clustering scheme in multidimensional spaces is utilized to differentiate different types of π-interaction. Due to the complexity involved with the multidimensional clustering techniques, it would be useful to exploit three component ternary diagrams which are popular for the interpretation of the phase transition in physical chemistry.[129] Figure 5 shows some of selected ternary diagrams drawn using the combination of variously selected energy components.

In Figure 5a, the ternary diagram using the attractive energy components (E_{es*}, E_{ind**}, E_{disp**}) shows the percentage of $E_{es*}/E_{ind**}/E_{disp**}$, i.e., $f_{es}/f_{ind}/f_{disp}$ (in %) in the total attractive interaction energy for different types of π-interaction. The average percentages of $f_{es}/f_{ind}/f_{disp}$ are ∼17/**5**/78% for Rg-π, ∼26/**7**/67% for π-π(D), ∼32/**9**/59% for π-π(T), ∼29/**9**/62% for CH-π, ∼40/**12**/48% for H-π, ∼40/**16**/44% for π^+-π(D), ∼43/**27**/30% for π^+-π(T), ∼39/**31**/30% for π^{+2}-π(D), ∼41/**34**/25% for H^+-π(T), ∼48/**43**/9% for M^+-π, and ∼31/**66**/3% for M^{+2}-π (Table 3), which are listed in the increasing order of the average induction energy percentage. The main points that could be drawn from figure 5a (Table 2) are highlighted below (the detailed analysis is in Supporting Information).

The variation of f_{es} between different types of π-interaction is rather small as compared to the variations of f_{ind}. Whereas, the f_{disp} and f_{exch} are almost in the reverse order of f_{ind}. The dispersion energy is the highly dominant component in the Rg-π (∼78%), CH-π (∼62%), and π-π(D) (∼67%) complexes and the dominant component (∼59%) in the π-π(T) complexes, whereas all their induction components are very small (5−9%). The f_{exch} increases with increasing contacting area between benzene and the partner molecule and with increasing charge density in the contacting area. More

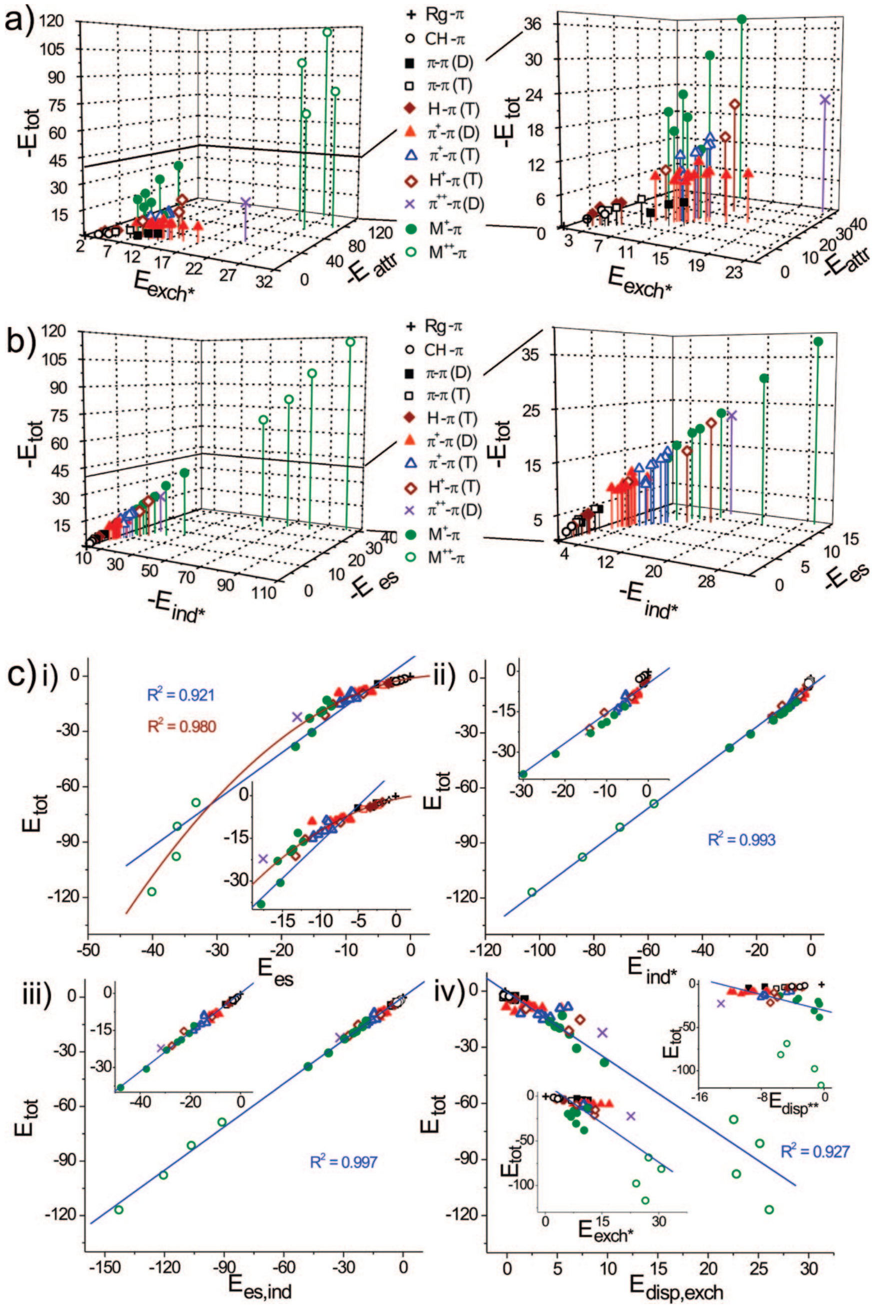


Figure 4. 3D-plots of (a) $-E_{tot}$ as the function of E_{exch^*} and $-E_{attr}(E_{attr} = E_{es^*} + E_{ind^{**}} + E_{disp^{**}})$ and (b) $-E_{tot^*}$ as the function of $-E_{ind^{**}}$ and $-E_{es^*}$ for the Rg-π, CH-π(T), π-π(D), π-π(T), H-π(T), π^+-π(D), π^+-π(T), H$^+$-π(T), π^{+2}-π(D), M$^+$-π, and M^{+2}-π interactions which are described in Tables 1 and 2. (c) Linear or multiple linear polynomial regression curves between (i) E_{tot} and E_{es^*}, (ii) E_{tot} and $E_{ind^{**}}$, (iii) E_{tot} and $E_{es,ind}$ (= $E_{es^*} + E_{ind^{**}}$), and (iv) E_{tot} and $E_{disp,exch}$ (= $E_{disp^{**}} + E_{exch^*}$). Insets in (i), (ii), and (iii) are the expanded figures in the region of small values of $-E_{tot}$. Insets in (iv) are the linear regression curves for E_{tot} vs $E_{disp^{**}}$ and E_{tot} vs E_{exch^*}. The regression equations with very high degree of correlation are as follows: $E_{tot} = -1.216 + 0.584 \times E_{es^*} - 0.052 \times E_{es^*}{}^2$ (R^2: 0.9798); $E_{tot} = -4.441 + 1.110 \times E_{ind^{**}}$ (R^2: 0.9921); $E_{tot} = 0.102 + 0.793 \times E_{es,ind}$ (R^2: 0.9967). Thus, if any component of E_{tot}, E_{es}, $E_{ind^{**}}$, or $E_{es,ind}$ is known, other components can be approximately estimated.

positively charged species have smaller f_{exch}, while the displaced stacking complexes such as π-π(D), π^+-π(D), and π^{+2}-π(D) have larger f_{exch} than the corresponding T-shaped complexes.

In the π^+-π(D) complexes, the dispersion energy is still a major energy component (∼44%) and the contribution from the electrostatic energy is comparable (∼40%), while the induction energy is still small (∼16%). Meanwhile, in the π^+-π(T) complexes, the dispersion energy is reduced to ∼30%, but the induction increases (∼28%), while the electrostatic energy is the major component (∼43%). These complexes are quite similar to the H$^+$-π(T) complexes, where the dispersion energy is ∼26%, the induction energy is ∼34%, and the electrostatic energy is ∼41%. The π^{++}-π(D) complexes have ∼29% contribution from the dispersion energy which is smaller than the

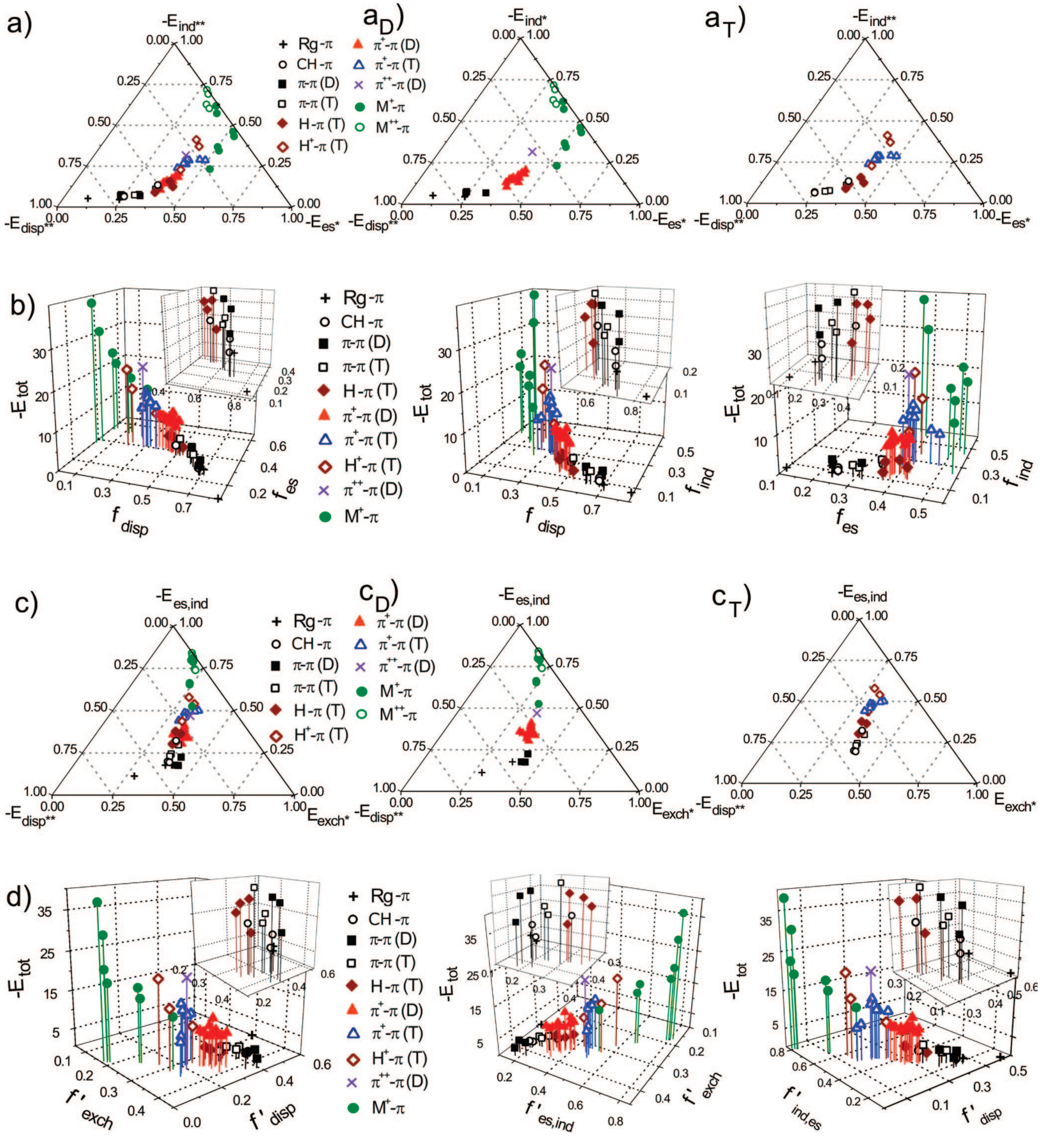


Figure 5. (a) Ternary diagrams of the attractive energy component terms (E_{es^*}, $E_{ind^{**}}$ and $E_{disp^{**}}$) for the Rg-π, CH-π(T), π-π(D), π-π(T), H-π(T), π^+-π(D), π^+-π(T), H$^+$-π(T), π^{+2}-π(D), M$^+$-π, and M^{+2}-π interactions for all the systems studied here (a), for only the displaced stacked systems (a$_D$), and for only the T-shaped systems (a$_T$). (b) 3D-plot of E_{tot} vs any two fractional values f_{disp}, f_{es} and f_{ind} of E_{es^*}, $E_{ind^{**}}$, and $E_{disp^{**}}$, respectively. (c) Ternary Diagrams plotted using f'_{exch}, $f'_{es,ind}$, and f'_{disp} for all the systems studied here (c), for only the displaced stacked systems (c$_D$), and for only the T-shaped systems (c$_T$). (d) 3D-plot of E_{tot} vs any two of f'_{disp}, $f'_{es,ind}$, and f'_{exch} [inset for Rg-π, CH-π, H-π(T), and π-π(T)].

contributions from the induction energy (~32%) and the electrostatic energy (~39%).

On the other extreme, the M$^+$-π complexes have small contribution from dispersion energy (~9% except for Cs$^+$ of 23%), large induction energy (~41% except for the moderate value of 23% for Cs$^+$), and very large electrostatic energy (~50%), and the M^{++}-π complexes have very small dispersion energy (~3%), very large induction energy (~66%), and large electrostatic energy (~31%).

When we consider individual π-complexes on the ternary diagram (Figure 5a), there are a few overlapping points between various forms of π-complexes. However, when the displaced-stacked π-complexes along with the stacked metal cation-π complexes (Figure 5a$_D$) are separated from the T-shaped π-complexes (Figure 5a$_T$), the π-π(D), π^+-π(D), π^{+2}-π(D), M$^+$-π, and M^{+2}-π complexes are well separated from each other except for very few cases.

The 3D-plots of E_{tot} (Figure 5b) with respect to fractional contribution of any two attractive energy components of the ternary diagram (Figure 5a) resolves the overlapping points between π-π(T) and π^+-π(D) complexes by the differences in height of E_{tot}. The π^+-π(T) complexes have much smaller heights in $-E_{tot}$ than the π^{+2}-π(D) complexes. Though the π^+-π(D) and H-π(T) complexes are partially overlapped on the ternary diagram in Figure 5a, the π^+-π(D) complexes have much higher height in $-E_{tot}$ than the H-π(T) complexes in Figure 5b. One of the closely located points of the Rg-π complexes (Kr complex) at the close vicinity of the π-π(D) complexes (C_6H_6 and Tol complexes) is distinguished from the differences in their height of $-E_{tot}$.

Finally, we have drawn a second set of ternary (Figure 5c) diagrams with the inclusion of the exchange repulsion energy terms in addition to the attractive energy components, where the fractional terms are given by $f'_{es,ind} = (E_{es*} + E_{ind**})/(E_{attr} - E_{exch*})$, $f'_{disp} = E_{disp**}/(E_{attr} - E_{exch*})$, and $f'_{exch} = -E_{exch*}/(E_{attr} - E_{exch*})$. Although the M^{+2}-$\pi$ complexes have much larger binding energy than the M$^+$-π complexes, the ternary diagrams in Figure 5c show that both complexes (with the exception of the Cs$^+$−benzene complex) are clustered on the top of diagrams with the maximum contribution of $-E_{es,ind}$ and the least contributions from $-E_{disp**}$ and $-E_{exch*}$. The Cs$^+$−benzene complex is located close to the π^+-π(T), H$^+$-π(T), and π^{+2}-π(D) complex. The π^+-π(T), H$^+$-π, and π^{+2}-π(D) complexes are located in the middle of the ternary diagram along the $-E_{es,ind}$ axis. In this context, the π^+-π(T) complexes have similar characteristics to non-π organic cation-π complexes, but are distinctly different from the metallic cation-π complexes. This also suggests the importance of the exchange repulsion terms toward the understanding of various π-systems. The π^+-π(D) complexes are located in between the π^+-π(T) complexes on the top of the diagram and the π-π complexes on the bottom (Figure 5c). However, the π^+-π(D) complexes overlap with the H-π(T) complexes, and the π^{+2}-π(D) complexes overlap with the π^+-π(T) and H$^+$-π(T) complexes. In the ternary diagram of displaced-stacked π-complexes along with the stacked metal cation-π complexes (Figure 5c$_D$), the Rg-π, π-π(D), π^+-π(D), π^{+2}-π, and (M$^+$-π, M^{+2}-π) complexes are separated from each other with the exception of the Cs$^+$−benzene complex which lies very close to the π^{+2}-π complex. The overlaps between M^{+2}-π and the M$^+$-π complexes, between Cs$^+$-benzene and the π^{+2}-π complex, and between Kr and C_6H_6/Tol complexes are distinguished by the differences in height of $-E_{tot}$ (Figure 5d). The C_2H_2 and NH_3 complexes are still located very close to each other with nearly similar heights. Hence, the NH-π complex of NH_3 is very similar in nature to the CH-π complex of C_2H_2.

On the basis of the above analysis, though the π^+-π(T) complexes could be treated in the same category of non-π organic cation-π complexes, it is however distinctly different from the metallic cation-π systems. The π^+-π(D) complexes have different characteristics from the π^+-π(T) complexes and are well separated from both the neutral π-π(D) complexes with small E_{ind**} or $E_{es,ind}$ on the bottom and the metallic cation M$^+$-π complexes with large E_{ind**} or $E_{es,ind}$ on the top in the ternary diagrams (Figure 5, parts a and c). Hence, the π^+-π(D) complexes need to be considered as a special type of π-interaction which lies in between the well-known π-π and cation-π complexes.

4. Concluding Remarks

On the basis of extensive structural and energetic analysis of various model systems as the representatives of Rg-π, CH-π(T), π-π(D), π-π(T), H-π(T), π^+-π(D), π^+-π(T), H$^+$-π(T), π^{+2}-π(D), M$^+$-π, and M^{+2}-π interactions, the nature of these interactions is systematically investigated and compared for the similarities and differences between different types of π-interactions in terms of the energy components. Among all of the energy components, induction energies were found to have the highest correlation with the total binding energies of the π-complexes, which is followed by electrostatic energies. Although dispersion and repulsive exchange energies independently show very poor correlations with total binding energies, their sum shows an improved correlation with total binding energies due to the cancelation effect. The magnitude of the energy components are important for quantitative understanding of the strength of binding patterns in π-complexes, while the ternary diagrams employed with selected energy components which show distinctive features for different types of π-interactions help in clustering different π-interactions on different spots in the diagrams. The ternary diagrams along with the 3D-plots of E_{tot} with respect to fractional contribution of any two contributing component terms help resolve most of the overlapping points in ternary diagrams for a qualitative and quantitative classification of different π-complexes except for a few closely overlapped points between CH-π and H-π complexes which are somewhat similar in characteristics to each other.

From the present comprehensive analysis of various types of π-interaction, we have demonstrated that the π^+-π interaction is a special type of π-interaction because the π^+-π complexes show different features from the conventional π-π complexes and cation-π complexes (for both organic non-π cations and metallic cations). Even the π^+-π(D) complexes show different characteristics from the π^+-π(T) complexes. The π^+-π(D) complexes, which are more commonly observed in the crystal structures than the π^+-π(T) complexes, have large dispersion and electrostatic energies and moderate induction energy, while the π^+-π(T) complexes have large electrostatic energy and substantial dispersion and induction energies. Namely, the π^+-π(D) complexes have larger dispersion energy but smaller induction energy than the π^+-π(T) complexes. The π^+-π interaction has both characteristics of conventional π-π and cation-π interactions but cannot be represented by their combined interactions. It is much weaker than the commonly known cation-π interaction, but much stronger than the π-π interaction.

For the π^+-π complexes with the positively charged π-system having the (N−H)$^+$ or (C−H)$^+$ group such as in methyl imidazolium, *N*-methyl-guanidinium, pyridinium, and dimethyl imidazolium, the T-shaped structure [π^+-π(T) complexes] are more stable than the displaced-stacked structure [π^+-π(D) complexes] in the absence of solvent molecules and counteranions, whereas the former tends to be less stable than (or isoenergetic to) the latter in the presence of the solvent molecules and counteranions. This explains why the stacked structures are much more common in the π^+-π interaction. This π^+-π interaction helps in assembling π-systems since the direct interactions with water or polar solvent molecules are suppressed. We believe that this comprehensive analysis of various types of π-interaction will be useful for future research works involving π-interactions in molecular assembly, molecular recognition and sensing, and optimal design of drugs and functional materials.

Synopsis

A comprehensive energy analysis of various π-interactions and comparison of their properties in terms of the attractive and repulsive energy components with the aid of ternary diagrams demonstrate that the π_{cation}-π interaction is a special type of π-interaction which can be represented by neither the conventional π-π interaction nor the conventional cation-π interaction.

Acknowledgment. This work was supported by GR-L(KICOS), KOSEF (EPB Center: R11-2008-052-01000), and BK21(KRF). Most calculations were carried out using supercomputers at KISTI (KSC-2008-K08-0002).

Supporting Information Available: The calculated isomers on various π_{cation}-π model systems, details of energetic analysis, and "Complete refs 66 121, and 122. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hobza, P.; Selzle, H. L.; Schlag, E. W. *Chem. Rev.* **1994**, *94*, 1767–1784.

(2) Tarakeshwar, P.; Choi, H. S.; Kim, K. S. *Chem. Rev.* **2000**, *100*, 4145–4185.

(3) Brutschy, B. *Chem. Rev.* **2000**, *100*, 3891–3920.

(4) Heβelmann, A.; Jansen, G.; Schütz, M. *J. Am. Chem. Soc.* **2006**, *128*, 11730.

(5) Fiethen, A.; Jansen, G.; Heβelmann, A.; Schütz, M. *J. Am. Chem. Soc.* **2008**, *130*, 1802.

(6) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23–28.

(7) Gazit, E. *FASEB J.* **2002**, *16*, 77–83.

(8) Hong, B. H.; Lee, J. Y.; Lee, C.-W.; Kim, J. C.; Bae, S. C.; Kim, K. S. *J. Am. Chem. Soc.* **2001**, *123*, 10748–10749.

(9) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210–1250.

(10) Singh, N. J.; Lee, H. M.; Hwang, I.-C.; Kim, K. S. *Supramol. Chem.* **2007**, *19*, 321–332.

(11) Singh, N. J.; Lee, H. M.; Suh, S. B.; Kim, K. S. *Pure Appl. Chem.* **2007**, *79*, 1057–1075.

(12) Hobza, P.; Bludský, O.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1992**, *97*, 335–340.

(13) Tarakeshwar, P.; Kim, K. S.; Kraka, E.; Cremer, D. *J. Chem. Phys.* **2001**, *115*, 6018–6029.

(14) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1990**, *93*, 5893–5897.

(15) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. *J. Am. Chem. Soc.* **2001**, *123*, 3323–3331.

(16) Lee, E. C.; Hong, B. H.; Lee, J. Y.; Kim, J. C.; Kim, D.; Kim, Y.; Tarakeshwar, P.; Kim, K. S. *J. Am. Chem. Soc.* **2005**, *127*, 4530–4537.

(17) Vaupel, S.; Brutschy, B.; Tarakeshwar, P.; Kim, K. S. *J. Am. Chem. Soc.* **2006**, *128*, 5416–5426.

(18) Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27–32.

(19) Shibasaki, K.; Fujii, A.; Mikami, N.; Tsuzuki, S. *J. Phys. Chem. A* **2007**, *111*, 753–758.

(20) Ringer, A. L.; Figgs, M. S.; Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10822–10828.

(21) Tsuzuki, S.; Honda, K.; Fujii, A.; Uchimaru, T.; Mikami, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2860–2865.

(22) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2000**, *122*, 11450–11458.

(23) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525–5534.

(24) Hunter, C. A. *Angew. Chem., Int. Ed.* **1993**, *32*, 1584–1586.

(25) Hunter, C. A. *Chem. Soc. Rev.* **1994**, *23*, 101–109.

(26) Cockroft, S. L.; Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *J. Am. Chem. Soc.* **2005**, *127*, 8594–8595.

(27) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1990**, *93*, 5893–5897.

(28) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1993**, *97*, 3937–3938.

(29) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 3500–3506.

(30) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790–18794.

(31) Špirko, V.; Engkvist, O.; Soldán, P.; Selzle, H. L.; Schlag, E. W.; Hobza, P. *J. Chem. Phys.* **1999**, *111*, 572–582.

(32) Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802–11808.

(33) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.

(34) Tsuzuki, S.; Uchimaru, T.; Sugawara, K.-I.; Mikami, M. *J. Chem. Phys.* **2002**, *117*, 11216–11221.

(35) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104–112.

(36) Tsuzuki, S.; Uchimaru, T.; Mikami, M. *J. Phys. Chem. A* **2006**, *110*, 2027–2033.

(37) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887–10893.

(38) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2003**, *107*, 8377–8379.

(39) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200–10207.

(40) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690–7697.

(41) Sinnokrot, M. O.; Sherrill, C. D. *J Phys Chem A* **2006**, *110*, 10656–10668.

(42) Ringer, A. L.; Sinnokrot, M. O.; Lively, R. P.; Sherrill, C. D. *Chem. Eur. J.* **2006**, *12*, 3821–3828.

(43) Lee, E. C.; Kim, D.; Jurecka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. *J. Phys. Chem. A* **2007**, *111*, 3446–3457.

(44) Dougherty, D. A.; Stauffer, D. *Science* **1990**, *250*, 1558–1560.

(45) Kumpf, R. A.; Dougherty, D. A. *Science* **1993**, *261*, 1708–1710.

(46) Dougherty, D. A. *Science* **1996**, *271*, 163–168.

(47) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303–1324.

(48) Zacharias, N.; Dougherty, D. A. *Trends. Pharmacol. Soc.* **2002**, *23*, 281–287.

(49) Kim, D.; Hu, S.; Tarakeshwar, P.; Kim, K. S.; Lisy, J. M. *J. Phys. Chem. A* **2003**, *107*, 1228–1238.

(50) Lee, J. Y.; Lee, S. J.; Choi, H. S.; Cho, S. J.; Kim, K. S.; Ha, T. K. *Chem. Phys. Lett.* **1995**, *232*, 67–71.

(51) Kim, K. S.; Lee, J. Y.; Lee, S. J.; Ha, T.-K.; Kim, D. H. *J. Am. Chem. Soc.* **1994**, *116*, 7399–7400.

(52) Tsuzuki, S.; Yoshida, M.; Uchimaru, T.; Mikami, M. *J. Phys. Chem. A* **2001**, *105*, 769–773.

(53) Quinonero, D.; Garau, C.; Frontera, A.; Ballester, P.; Costa, A.; Deya, P. M. *J. Phys. Chem. A.* **2005**, *109*, 4632–4637.

(54) Quinonero, D.; Garau, C.; Rotger, C.; Frontera, A.; Ballester, P.; Costa, A.; Deya, P. M. *Angew. Chem., Int. Ed.* **2002**, *41*, 3389–3392.

(55) Kim, D.; Tarakeshwar, P.; Kim, K. S. *J. Phys. Chem. A* **2004**, *108*, 1250–1258.

(56) Kim, D.; Lee, E. C.; Kim, K. S.; Tarakeshwar, P. *J. Phys. Chem. A* **2007**, *111*, 7980–7986.

(57) Kim, D. Y.; Singh, N. J.; Kim, K. S. *J. Chem. Theor. Comput.* **2008**, *4*, 1401–1407.

(58) Janda, K. C.; Hemminger, J. C.; Winn, J. S.; Novick, S. E.; Harris, S. J.; Klemperer, W. *J. Chem. Phys.* **1975**, *63*, 1419–1421.

(59) Vernon, M. F.; Lisy, J. M.; Kwok, H. S.; Krajnovich, D. J.; Tramer, A.; Shen, Y. R.; Lee, Y. T. *J. Phys. Chem.* **1981**, *85*, 3327–3333.

(60) Ruhl, E.; Bisling, P. G. F.; Brutschy, B.; Baumgartel, H. *Chem. Phys. Lett.* **1986**, *126*, 232–237.

(61) Felker, P. M.; Maxton, P. M.; Schaeffer, M. W. *Chem. Rev.* **1994**, *94*, 1787–2160.

(62) Cozzi, F.; Ponzini, F.; Annunziata, R.; Cinquini, M.; Siegel, J. S. *Angew. Chem., Int. Ed.* **1995**, *34*, 1019–1020.

(63) Ren, R.; Jin., Y.; Kim, K. S.; Kim, D. H. *J. Biomol. Struct. Dyn.* **1997**, *15*, 401–405.

(64) Kim, E.; Paliwal, S.; Wilcox, C. S. *J. Am. Chem. Soc.* **1998**, *120*, 11192–11193.

(65) Hong, B. H.; Lee, J. Y.; Cho, S. J.; Yun, S.; Kim, K. S. *J. Org. Chem.* **1999**, *64*, 5661–5665.

(66) Kim, K. S.; et al. *J. Am. Chem. Soc.* **2002**, *124*, 14268.

(67) Nakamura, K.; Houk, K. N. *Org. Lett.* **1999**, *1*, 2049–2051.

(68) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768–4774.

(69) Tarakeshwar, P.; Lee, S. J.; Lee, J. Y.; Kim, K. S. *J. Chem. Phys.* **1998**, *108*, 7217–7223.

(70) Wang, Y.; Hu, X. *J. Am. Chem. Soc.* **2002**, *124*, 8445–8451.

(71) Manojkumar, T. K.; Choi, H. S.; Hong, B. H.; Tarakeshwar, P.; Kim, K. S. *J. Chem. Phys.* **2004**, *121*, 841–846.

(72) Grimme, S. *Chem. Eur. J.* **2004**, *10*, 3423–3429.

(73) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

(74) Grimme, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 3430–3434.

(75) Manojkumar, T. K.; Kim, D.; Kim, K. S. *J. Chem. Phys.* **2005**, *122*, 014305.

(76) Adamovic, I.; Li, H.; Lamm, M. H.; Gordon, M. S. *J. Phys. Chem. A* **2006**, *110*, 519–525.

(77) Piacenza, M.; Grimme, S. *J. Am. Chem. Soc.* **2005**, *127*, 14841–14848.

(78) Mishra, B. K.; Sathyamurthy, N. *J. Phys. Chem. A* **2005**, *109*, 6–8.

(79) Diastro, R. A.; von Helden, G.; Steele, R. P.; Head-Gordon, M. *Chem. Phys. Lett.* **2007**, *437*, 277–283.

(80) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 4209–4212.

(81) Heβelmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 014103.

(82) Piacenza, M.; Grimme, S. *Chem. Phys. Chem.* **2005**, *6*, 1554–1558.

(83) Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10345–10354.

(84) Sato, T.; Tsuneda, T.; Hirao, K. *J. Chem. Phys.* **2005**, *123*, 104307.

(85) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* 2008, DOI: 10.1021/ar700111a.

(86) Levy, C. W.; Roujeinikova, A.; Sedelnikova, S.; Baker, P. J.; Stuitje, A. R.; Slabas, A. R.; Rice, D. W.; Rafferty, J. B. *Nature* **1999**, *398*, 383–384.

(87) Marcotrigiano, J.; Gingras, A. C.; Sonenberg, N.; Burley, S. K. *Cell* **1997**, *89*, 951–961.

(88) Schumacher, M. A.; Miller, M. C.; Grkovic, S.; Brown, M. H.; Skurray, R. A.; Brennan, R. G. *Science* **2001**, *294*, 2158–2163.

(89) Biot, C.; Buisine, E.; Rooman, M. *J. Am. Chem. Soc.* **2003**, *125*, 13988–13994.

(90) Kapoor, L. S M.; Surolia, N.; Surolia, A.; Suguna, K. *J. Mol. Biol.* **2004**, *343*, 147–155.

(91) Murray, D. S.; Schumacher, M. A.; Brennan, R. G. *J. Biol. Chem.* **2004**, *279*, 14365–14371.

(92) Ghosh, P.; Park, C.; Peterson, M. S.; Bitterman, P. B.; Polunovsky, V. A.; Wagner, C. R. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2177–2180.

(93) Crowley, P. B.; Golovin, A. *Proteins* **2005**, *59*, 231–239.

(94) Argiriadi, M. A.; Morisseau, C.; Hammock, B. D.; Christianson, D. W. *Proc. Natl. Acad. Sci. U.S. A.* **1999**, *96*, 10637–10642.

(95) Thomas, A.; Meurisse, R.; Charloteaux, B.; Brasseur, R. *Proteins* **2002**, *48*, 628–634.

(96) Li, H. L.; Galue, A.; Meadows, L.; Ragsdale, D. S. *Mol. Pharmacol.* **1999**, *55*, 134–141.

(97) Okada, A.; Miura, T.; Takeuchi, H. *Biochemistry* **2001**, *40*, 6053–6060.

(98) Wintjens, R.; Lievin, J.; Rooman, M.; Buisine, E. *J. Mol. Biol.* **2000**, *302*, 395–410.

(99) Biot, C.; Wintjens, R.; Rooman, M. *J. Am. Chem. Soc.* **2004**, *126*, 6220–6221.

(100) Stoddart, J. F. *Pure Appl. Chem.* **1998**, *60*, 467–472.

(101) Odell, B.; Reddington, M. V.; Slawin, A. M. Z.; Spencer, N.; Stoddart, J. F.; Williams, D. J. *Angew. Chem., Int. Ed.* **1998**, *27*, 1547–1550.

(102) Raymo, F. M.; Stoddart, J. F. *Chem. Rev.* **1999**, *99*, 1643–1663.

(103) Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. F. *Angew. Chem. Int. Ed.* **2002**, *41*, 898–952.

(104) Kaifer, A. E. *Acc. Chem. Res.* **1999**, *32*, 62–71. Balzani, V.; Credi, A.; Raymo, F. M.; Stoddart, J. F. *Angew. Chem., Int. Ed.* **2000**, *39*, 3349–3391.

(105) Nelson, A.; Belitsky, J. M.; Vidal, S.; Joiner, C. S.; Baum, L. G.; Stoddart, J. F. *J. Am. Chem. Soc.* **2004**, *126*, 11914–11922.

(106) Yamada, S.; Morita, C. *J. Am. Chem. Soc.* **2002**, *124*, 8184–8185.

(107) Birman, V. B.; Uffman, E. W.; Jiang, H.; Li, X.; Kilbane, C. J. *J. Am. Chem. Soc.* **2004**, *126*, 12226–12227.

(108) Petti, M. A.; Shepodd, T. J.; Barrans, R. E., Jr.; Dougherty, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 6825–6840.

(109) Dupont, J.; de Souza, R. F.; Suarez, P. A. Z. *Chem. Rev.* **2002**, *102*, 3667–3692.

(110) Hu, Z.; Margulis, C. J. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 831–836.

(111) Abouderbala, L. O.; Belcher, W. J.; Boutelle, M. G.; Cragg, P. J.; Steed, J. W.; Turner, D. R.; Wallace, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5001–5006.

(112) Chellapan, K.; Singh, N. J,; Hwang, I. C.; Lee, J. W.; Kim, K. S. *Angew. Chem., Int. Ed.* **2005**, *44*, 2899–2903. (a) *Angew. Chem.* 2005, *117*, 2959−2963.

(113) Yoon, J.; Kim, S. K.; Singh, N. J.; Kim, K. S. *Chem. Soc. Rev.* **2006**, *35*, 355–360.

(114) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459–9464.

(115) Cauët, E.; Rooman, M.; Wintjens, R.; Liévin, J.; Biot, C. *J. Chem. Theory. Comput.* **2005**, *1*, 472–483.

(116) Tsuzuki, S.; Mikami, M.; Yamada, S. *J. Am. Chem. Soc.* **2007**, *129*, 8656–8662.

(117) Singh, N. J. Ph. D *dissertation*, Pohang University of Science and Technology, 2005, 123−156.

(118) Kim, K. S. Creative Research Initiative Proposal (KOSEF), 2005.

(119) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639–9646.

(120) Min, S. K.; Lee, E. C.; Lee, H. M.; Kim, D. Y.; Kim, D.; Kim, K. S. *J. Comput. Chem.* **2008**, *29*, 1208–1221.

(121) *Gaussian 03, Revision C.02*, Frisch, M. J. et al., Gaussian, Inc., Wallingford CT, 2004.

(122) MOLPRO, version 2006.1. a package of *ab initio* programs, Werner, H.-J. et al.

(123) SAPT 2006: “An *Ab Initio* Program for Many-Body Symmetry-Adapted Perturbation Theory Calculations of Intermolecular Interaction Energies”, Bukowski R.; Cencek, W.; Jankowski, P.; Jeziorski, B.; Jeziorska, M.; Kucharski, S. A.; Lotrich, V. F.; Misquitta, A. J.; Moszynski, R.; Patkowski, K.; Podeszwa, R.; Rybak, S.; Szalewicz, K.; Williams, H. L.; Wheatley, R. J.; Wormer, P. E. S.; •Zuchowski, P. S.

(124) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887–1930.

(125) He*β*elmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 014103.

(126) Lee, H. M.; Suh, S. B.; Lee, J. Y.; Tarakeshwar, P.; Kim, K. S. *J. Chem. Phys.* **2000**, *112*, 9759–9772.

(127) Allen, F. H. *Acta Crystallogr. B* **2002**, *58*, 380–388.

(128) Allen, F. H.; Motherwell, W. D. S. *Acta Crystallogr., B* **2002**, *58*, 407–422.

(129) Shen, H. W.; Eisenberg, A. *J. Phy. Chem. B* **1999**, *103*, 9473–9487.

CT800471B

JCTC Journal of Chemical Theory and Computation

Molecular Mechanics Force Field for Octahedral Organometallic Compounds with Inclusion of the Trans Influence

Ivan Tubert-Brohman, Maurus Schmid, and Markus Meuwly*

University of Basel, Klingelbergstrasse 80, 4052 Basel, Switzerland

Received September 19, 2008

Abstract: Efficient calculation of the properties of metal-containing complexes relevant to catalysis is of major interest for better characterizing and optimizing the catalysts. For this, a new force field, called VALBOND-TRANS here, is proposed. It is based on the existing VALBOND force field of Landis and co-workers, extended by adding terms that account for electronic effects such as the trans influence of ligands on bond lengths and relative energies. Parameters and results for model octahedral complexes of Ru, Os, Rh, and Ir are determined and discussed. The model is then applied to the study of reactive intermediates involved in asymmetric hydrogenation catalyzed by iridium complexes with chiral phosphinooxazolines (PHOX) ligands. The new force field explores and capitalizes on the separation of electronic and steric effects on the stability of different diastereomers and reproduces DFT results which are consistent with experimental observations.

1. Introduction

Theoretical methods can be very helpful for understanding and predicting structure and reactivity in organometallic chemistry. Density functional theory (DFT) methods, in particular, have been very successful for the computational study of transition metal complexes.[1] However, DFT calculations are CPU-intensive, which makes the treatment of large complexes difficult. Such calculations are often limited to geometry optimizations of gas-phase structures or to model systems with smaller and sometimes unrealistic ligands. Dynamics simulations and calculations over large libraries of compounds are often impractical. Therefore, having a faster, even if approximate method is desirable. Molecular mechanics (MM) force fields such as OPLS,[2] AMBER,[3] and CHARMM[4] have become standard methods in biomolecular chemistry and are routinely used for molecular dynamics simulations of systems with up to 1 million atoms.[5] However, the development of general force fields for transition metal complexes has been relatively limited because of the unique difficulties presented by these types of compounds:[6−15] metals can have a variety of coordination numbers, π-binding ligands can bind in various ways, and electronic effects such as the Jahn−Teller distortion and the trans influence need to be considered. Although considerable work has been done toward including the Jahn−Teller distortion in molecular mechanics,[16,17] the molecular mechanics of the trans influence remains largely unexplored. Here, we present a model for the trans influence and implement it in a force field for octahedral complexes of iridium and three other platinum-group metals (Ru, Os, and Rh). We then test our method on real complexes relevant to asymmetric hydrogenation. The current work is a conventional force field in the sense that it does not allow the formation and breaking of bonds, which are essential for the modeling of chemical reactions. This work, however, can be readily combined with the adiabatic reactive molecular dynamics (ARMD) approach recently developed by our group.[18] In ARMD, a conventional force field description of the reactants and products is used for each dynamics time step, and the system is propagated along the lower energy state, with specific rules for treating crossover between the reactant and product states.

Asymmetric hydrogenation is one of the most important catalytic methods for the preparation of optically active compounds, both in the laboratory and in industry.[19] Iridium catalysts with chiral N,P ligands are a class of highly efficient catalysts that have expanded the scope of asymmetric

---

* Corresponding author e-mail: m.meuwly@unibas.ch.

10.1021/ct800392n CCC: $40.75

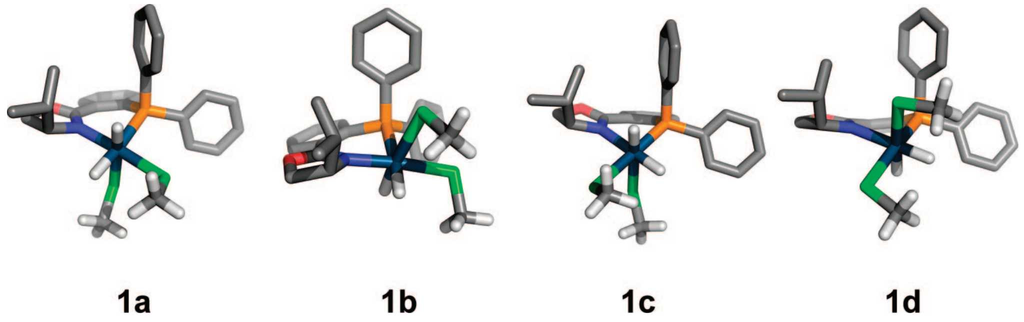


Figure 1. DFT-calculated structures of $[Ir(PHOX)(H)_2(CH_3Cl)_2]^+$ complexes.[24]

hydrogenation to include unfunctionalized alkenes.[19−23] In a recent study, we used DFT calculations in combination with NMR data to show that the hydrogenation reaction may proceed through dihydride intermediates **1a-1d** (Figure 1).[24] Here we revisit these compounds to see whether it is possible to capture the necessary electronic effects within a force field representation that can also be extended to other related compounds.

VALBOND, developed by Landis and co-workers, is a force field that computes the angle bending energies based on valence bond theory.[25] Its main premise is that the bending energy is a function of the overlap of the hybrid orbitals on the central atom, using hybridizations that are computed using certain rules. Unlike the simple harmonic approximation used by many force fields for the bending potential, the VALBOND bending function gives reasonable results at very large distortions. VALBOND also supports hypervalent compounds by means of a 3-center-4-electron bonding model.[26] As a result, it can describe geometries found in organometallic compounds, such as octahedral, trigonal biplanar, and square planar. VALBOND has been shown to work for transition metal compounds of a largely covalent nature, such as hydrides and alkyls.[27] For these compounds, VALBOND uses a 12-electron rule, where only sd hybrids are considered, and excess electrons are involved in hypervalent bonds. In order to have a complete force field, VALBOND has to be combined with another force field such as CHARMM[4] or UFF,[28] to account for the bond stretching, torsional, and nonbonded energy terms. VALBOND does not account for the trans influence, but the valence bond ideas behind the force field led us to investigate a plausible way of incorporating it.

Another approach to the molecular mechanics of coordination compounds is the inclusion of force field terms corresponding to the ligand field stabilization energy, as done in the ligand field molecular mechanics (LFMM) model of Deeth and co-workers[17] and the SIBFA-LF method of Gresh and co-workers.[29] The LFMM method has been applied to Werner-type complexes of Cu^{2+}, Ni^{2+}, Co^{3+}, and Mn^{2+}; it is able to account for a number of coordination geometries and for the Jahn−Teller effect and can also incorporate the contribution of the s orbitals that is important for complexes such as WH_6 and $[CuCl_4]^{2-}$. LFMM has also been successfully applied to the study of type I copper proteins.[30] However, we chose VALBOND for our work because it has been tested more extensively with compounds resembling the ones involved in catalytic hydrogenation, such as alkyl and hydride complexes of metals from the second and third transition series.[27]

The trans influence has been defined as "the tendency of a ligand to selectively weaken the bond trans to itself".[31] It is an electronic ground-state phenomenon, unlike the trans effect, which is defined in terms of the rate of a substitution reaction.[32] Although much of the work on trans influences and trans effects has focused on square planar complexes, these effects are also important for octahedral complexes.[32] Experimental studies of the trans influence generally concentrate on the lengthening of the trans bond as measured from crystallographic methods. The trans influence can also affect the relative stability of different diastereomers, but this has not been studied as extensively experimentally due to the difficulties of gathering thermochemical data for metal compounds.[33] For this reason, theoretical studies and molecular mechanics methods have often concentrated on the geometric influence. An early example is an MM3 study which found that the central nitrogen in terpyridine forms an unusually short Ru−N bond which causes the Ru−N bond trans to it to stretch; the authors simulated this effect within the force field by adding a pseudobond to constrain the N−N distance;[34] this is a case where the trans influence is caused not by the electronic nature of the ligand but by the strain caused by its tridentate nature. As a note on terminology, we use the terms *structural trans influence* to refer to the bond lengthening effect and *thermodynamic trans influence* to refer to the effect on the relative stability of different diastereomers.

We believe that current theoretical methods and computer power are ripe for a systematic study of the thermodynamic trans influence. The main goal of the present work is to explore the thermodynamic trans influence on the relative energies of different diastereomers of octahedral organometallic compounds and to implement a force field that can predict which isomers are most stable. We chose the octahedral geometry not only because of its ubiquity but also because of its rich stereochemistry: a complex with six different ligands has fifteen possible diastereomers, and predicting which one is most stable involves an intricate mix of steric and electronic effects. As a secondary goal, we also considered the geometric trans influence and incorporated it into our force field, but at this stage we were not concerned with spectroscopic properties or with using the most accurate possible functions and parameters for bond stretching and for bonding angles.

The current work is structured as follows. First, we review the theory underlying VALBOND and the trans influence and describe modifications and additions we envisage. Next, we describe the parametrization of the new method using a training set of simple model compounds. Then we present results for two test cases involving real complexes of relevance to asymmetric hydrogenation. Finally, general conclusions and suggestions for future developments are discussed.

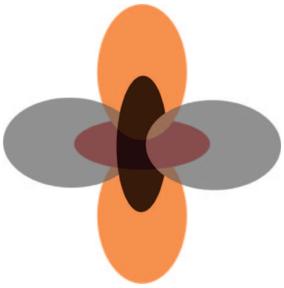


Figure 2. Cartoon representation of the overlap between two sd^2 orbitals at 90°. For normal bonds, the overlap is minimal at 90°, the equilibrium geometry. For hypervalent bonds, the equilibrium geometry corresponds to the maximum overlap at 180°.

2. Theory

Qualitatively, the trans influence may be attributed to the effect of different ligands on the resonance structures representing a 3-center-4-electron bond:[35,36]

$$\text{X: M-Y} \leftrightarrow \text{X-M :Y} \tag{1}$$

In this picture, the ligand that forms the stronger bond to the metal will weaken the bond to the weaker ligand. This leads to *antisymbiosis*, where "stronger" ligands prefer to bind trans to "weaker" ones, or softer ligands prefer to bind trans to harder ones.[37]

Based on the idea of the trans influence acting through a 3-center-4-electron bond, a natural place for incorporating this effect into VALBOND is in the part of the algorithm that treats hypervalent molecules. Note that, under this model, the trans influence will only exist for hypervalent complexes; however, most transition metal complexes, including all octahedral and square complexes of groups 8−10 are hypervalent according to the VALBOND formalism, due to its use of the 12-electron rule instead of the 18-electron rule.

The VALBOND equations that are directly relevant to our modifications are eqs 2−4, below, which describe hypervalent bonding. For a detailed explanation and for the nonhypervalent VALBOND equations, the reader is referred to the original literature.[25−27]

$$E(\alpha) = BOF \times k_\alpha(1 - \Delta(\alpha + \pi)^2) \tag{2}$$

$$\Delta = \frac{1}{1+m+n}\left(1 + m\cos\alpha + \frac{n}{2}(3\cos^2\alpha - 1)\right) \tag{3}$$

$$c_i = \frac{\prod_{i=1}^{hype} \Delta_i^2}{\sum_{j=1}^{res}\prod_{i=1}^{hype} \Delta_i^2} \tag{4}$$

Here $E(\alpha)$ is the bending energy of a hypervalent bond, BOF is the bond order factor (always 0.25 for hypervalent bonds), α is the angle, k_α is a force constant, and Δ is the overlap between two orbitals with sp^{m}d^n hybridization (Figure 2). VALBOND iterates over all possible resonance structures (*res*) that distribute the hypervalent bonds among the different ligands in different ways, giving each configuration i a weight c_i proportional to the product of Δ^2 over all hypervalent bonds (*hype*). The key equation is eq 2, which defines the bending energy of a 3-center-4-electron bond as proportional to the overlap between one orbital and another orbital shifted by 180°. This term favors linear geometries in hypervalent bonds, because $\Delta(\alpha+\pi) = 1$ when $\alpha = 180°$, which leads to a minimum of $E(\alpha)$ at 180°. Because the trans influence also acts along hypervalent bonds and is expected to be maximal at 180°, it seems reasonable to include the $\Delta(\alpha+\pi)^2$ term from eq 2 in a function for representing the thermodynamic trans influence. A simple and efficient function that was successful in practice is

$$E_{trans} = \sum_{i=1}^{hype} p_{AB}\Delta(\alpha + \pi)^2 \tag{5}$$

where p_{AB} is an adjustable parameter which depends on the types of ligands and metal involved. These parameters are determined by least-squares fitting to DFT energies of model compounds. Eq 5, like eq 2, is used for each resonance configuration and weighed accordingly.

For the structural trans influence, preliminary tests showed that the bond lengths could best be fitted by a function of the type

$$r_{A(B)} = r^0_A(1 + s_A \cdot i_B/100) \tag{6}$$

where $r_{A(B)}$ is the equilibrium bond length of a bond from the metal to a ligand of type A which is trans to a ligand of type B, r^0_A is the reference or unperturbed bond length to a ligand of type A, i_B is the "trans influence intensity" due to ligands of type B, and s_A is the "lengthening sensitivity" of bonds to ligands of type A. Like the trans influence on the energy, the trans influence on the bond length is implemented in the force field as part of the hypervalent bonding calculation and includes the overlap function as

$$r_{A(B)} = r^0_A(1 + s_A \cdot \Delta(\alpha + \pi)^2 \cdot i_B/100) \tag{7}$$

In the molecular mechanics calculations, $r_{A(B)}$ is used for the equilibrium bond length, r_0, in the standard harmonic bond energy function, $E_{bond}(r) = k_B(r - r_0)^2$.

3. Parametrization

Thermodynamic Trans Influence. The model was parametrized against a training set comprising 129 series of diastereomers of octahedral complexes of Ru, Os, Rh, and Ir with oxidation states II and III, for a total of 532 structures (Table 1). Each series consists of all diastereomers of a given complex, ranging from two isomers for complexes of type MX_4Y_2 (cis/trans isomerism) and MX_3Y_3 (fac/mer isomerism), to up to fifteen diastereomers for complexes with six different ligands. The choice of

Table 1. Composition of the Training Set

	number of series	number of isomers
Ru(II)	20	60
Ru(III)	24	134
Os(II)	16	49
Os(III)	13	69
Rh(III)	24	86
Ir(III)	32	134
total	129	532

compounds in the training set was motivated by complexes from the Cambridge Structural Database[38] and from the review by Coe and Glenwright,[32] with all ligands replaced with small model ligands: Cl^-, Br^-, H^-, H_2O, NH_3, PH_3, CH_3^-, CH_3OCH_3, CH_3Cl, $CH_2{=}CH^-$, and $C_6H_5^-$. This truncation, in addition to speeding up the calculations, minimizes steric interactions between the ligands, which allows us to better focus on electronic effects. All the reference energies and structural data were obtained from B3LYP[39] calculations, which were carried out using GAUSSIAN 03,[40,41] with the LANL2DZ effective core potential[42] for the metals and the 6−31G(d,p) basis set for all other atoms. The grid=ultrafine option was used for consistency with previous work.[24]

The DFT energy of each compound relative to the maximum energy isomer in its series was fit to the following function

$$\Delta E = \sum_{trans} p_{AB} \tag{8}$$

by means of a least-squares procedure. This function is equivalent to eq 5 for perfectly octahedral geometries. The relative energies used in the parametrization can be interpreted as energies of isomerization from the least stable isomer to a given isomer of a compound. For example, Figure 3 shows the series of compounds with formula $[Ir(PH_3)_3(H)(CH_3)(Cl)]$, along with their relative energies computed using eq 8 and the reference DFT values. For Ru and Os, which are found in two different oxidation states (II and III) in the training set, separate parameters were fitted for each oxidation state. Note, however, that the same parameters ("atom types") were used for all carbon ligands (CH_3^-, $CH_2{=}CH^-$, and $C_6H_5^-$) and for all oxygen ligands (H_2O and CH_3OCH_3); the validity of this assumption will be scrutinized further below. Because only relative energies are of interest, a common zero of energy was chosen by setting all the parameters involving the chloride anion to zero. The resulting parameters together with their average errors are shown in Table 2. The 109 adjustable parameters were fitted to 403 relative energies, with an overall mean absolute deviation (MAD) of 1.31 kcal/mol and an rmsd of 1.78 kcal/mol. The correlation between the values calculated using eq 8 and the DFT data is shown in Figure 4. The linear regression is close to $y = x$ with $R^2 = 0.93$. Most of the isomerization energies are within a range of 25 kcal/mol; the four points between −40 and −30 kcal/mol all correspond to compounds with the CH_3Cl ligand, which binds very weakly and is particularly sensitive to the thermodynamic trans influence. When considering the values of the parameters, the antisymbiotic effect is especially noticeable for hydride and methyl, which have a strong trans influence. Having H or C trans to H or C is unfavorable by up to 20 kcal/mol, relative to having them trans to chloride; conversely, having C or H trans to weaker trans-influencing ligands such as N and O is generally favorable, by up to −10 kcal/mol.

Although two different parameters were required for chlorine-binding ligands (one for the chloride ion and one for chloromethane), parameters for neutral oxygen ligands were found to be reasonably transferable—the same parameter can be used for water and for dimethyl ether as a ligand. Similarly, different carbon ligands are reproduced satisfactorily despite changes in hybridization and the addition of electron-donating or electron-withdrawing groups. As an illustration, Table 3 shows the isomerization energies for several series of nine compounds with formula $[Ir(H)(PH_3)(NH_3)_2(OH_2)R]^+$, where R is a formally anionic carbon ligand. The same trend is followed for all the R ligands considered, with an overall mean absolute deviation of 0.9 kcal/mol.

A possible drawback of eq 8 is that, for N ligand types, up to $N(N-1)/2$ parameters are required per metal; that is, the number of parameters scales roughly as N^2. It would be desirable to have atom-based parameters which scaled as N. We investigated functions such as $p_{AB} = p_A - p_B$, based on the idea of antisymbiosis,[37] but the results, while initially promising, were not sufficiently general for our purposes. Other simple functions such as $p_{AB} = p_A \cdot p_B$ were also tested with similar results.

Structural Trans Influence. The parameters in eq 6 were fit against all 2622 metal−ligand bond lengths in the training set using a least-squares method. The resulting parameters and average errors are summarized in Table 4. The overall mean absolute deviation is 0.019 Å (rmsd = 0.027 Å), which is less than one-tenth of the observed range of bond lengths for each bond type (for example, the ligands exerting the strongest trans influence, hydride and alkyl, can lengthen bonds trans to them by about 0.2 Å). The correlation between the predicted and the reference bond lengths is shown in Figure 5a. For comparison, a fit where the trans influence term is omitted, by setting i_B or s_A to zero, was also calculated; only the values of r^0 (not shown) were optimized for this comparison. This way we can compare our new approach with what would be the best possible result that can be obtained with a "constant r^0" approach. With the trans influence neglected, the mean absolute deviation increases to 0.044 Å (rmsd = 0.061 Å). The correlation between the bond lengths calculated without the trans influence and the reference values is shown in Figure 5b.

Examination of the structural trans influence parameters reveals a slight reciprocal correlation between intensity and sensitivity (correlation coefficient $R = 0.53$); this agrees with the analysis of the trans influence in terms of resonance structures discussed above, because ligands with a stronger trans influence will form bonds with a more covalent character, which can reasonably be expected to be less flexible than the more polar bonds formed with the weaker trans-influencing ligands. Another striking feature of the parameter values are the large differences between the para-

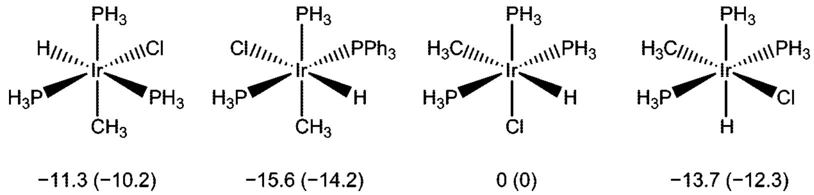


Figure 3. Isomers with formula $[Ir(PH_3)_3HCH_3Cl]$, with relative computed energies using eq 8 in kcal/mol (DFT values in parentheses).

Table 2. Parameters, Mean Absolute Deviation (MAD), and Root-Mean-Square Deviation (RMSD)[a]

parameter	Ru(II)	Ru(III)	Os(II)	Os(III)	Rh(III)	Ir(III)
p_{BrBr}	0.30	−0.07	0.41	−1.34	0.54	3.35
p_{CBr}		0.19		−1.06	3.14	6.41
p_{CC}	5.95		4.32		20.61	18.61
p_{CH}		10.28		15.30	19.65	18.30
p_{CO}		−8.34		−7.50	−9.55	−6.16
p_{HH}	16.17		18.12		15.95	20.42
p_{NBr}	−0.54	−2.00		−1.12	0.67	2.28
p_{NC}	−4.64	−3.73	−5.79	−1.08	−3.68	−2.81
p_{NH}	−1.38	−3.67	0.15	3.92	−5.82	−4.27
p_{NN}	−6.46	−10.28	−7.77	−6.93	−8.88	−9.58
p_{NO}	−6.63	−8.18	−9.60	−5.20	−3.35	−5.50
p_{OBr}		−0.33		−1.47	−1.35	−0.14
p_{OH}	−9.26	−8.81	−11.26	−2.54	−9.87	−7.35
p_{OO}	−7.50	−5.49	−14.22	−1.26	2.38	1.28
p_{PBr}	1.01	−0.37	2.16	−0.15	0.47	1.92
p_{PC}	−0.31	2.68	1.14	7.88	0.63	2.73
p_{PH}	4.74		8.32		1.51	4.58
p_{PN}	−3.43	−11.54	−2.99	−7.18	−7.29	−7.44
p_{PO}	−6.66	−13.20	−8.89	−10.43	−6.97	−7.35
p_{PP}	4.45	−9.04	9.30	−8.20	−2.53	0.26
p_{PCl^*}[b]						−3.45
p_{NCl^*}						−3.66
p_{HCl^*}						−6.18
$p_{Cl^*Cl^*}$						6.64
N	40	110	33	56	62	102
MAD	0.97	1.95	1.05	1.08	1.05	1.23
rmsd	1.26	2.47	1.48	1.45	1.36	1.52

[a] All values are in kcal/mol. [b] Cl* refers to the formally neutral chlorine atom in CH_3Cl.

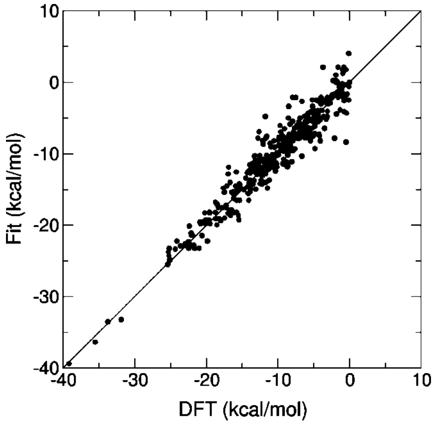


Figure 4. Isomerization energies computed using eq 8 versus the reference DFT values. The solid line corresponds to the function $y = x$.

Table 3. Relative Energies of Compounds with Formula $[Ir(H)(PH_3)(NH_3)_2(OH_2)R]^+$ (kcal/mol)

	fit	DFT energy for ligand R				
isomer	(eq 8)	CH_3^-	$C_2H_4^-$	$C_6H_5^-$	p-$C_6H_4OMe^-$	p-$C_6H_4NO_2^-$
1	−16.5	−15.7	−15.4	−14.1	−13.9	−14.3
2	−9.1	−9.3	−8.8	−8.6	−8.5	−8.6
3	−19.6	−20.9	−19.6	−20.5	−20.7	−20.1
4	−12.4	−13.4	−13.4	−10.3	−10.6	−12.9
5	−23.0	−23.4	−22.7	−22.2	−23.4	−23.4
6	−23.2	−22.5	−21.6	−21.0	−20.7	−21.0
7	0	0	0	0	0	0
8	−4.0	−3.9	−4.4	−3.8	−3.9	−3.9
9	−19.8	−20.1	−20.8	−19.5	−19.7	−19.2

Table 4. Parameters and Average Errors for the Bond Lengths Calculated Using Eq *6*

parameter	Ru(II)	Ru(III)	Os(II)	Os(III)	Rh(III)	Ir(III)
r^0_{Br} (Å)	2.543	2.396	2.566	2.414	2.448	2.363
r^0_C	2.084	2.062	2.020	2.064	2.048	2.011
r^0_{Cl}	2.425	2.283	2.430	2.307	2.362	2.298
$r^0_{Cl^*}$						2.252
r^0_H	1.565	1.553	1.567	1.586	1.510	1.484
r^0_N	2.114	2.087	2.081	2.087	2.060	1.954
r^0_O	2.159	2.081	2.140	2.068	2.102	1.977
r^0_P	2.272	2.237	2.219	2.242	2.242	2.151
s_{Br}	0.83	0.85	0.78	1.00	1.16	1.22
s_C	1.31	0.32	2.33	0.45	0.76	0.74
s_{Cl}	1	1	1	1	1	1
s_{Cl^*}						1.69
s_H	1.55	0.74	2.00	0.75	1.11	1.11
s_N	1.70	0.92	2.14	0.97	1.41	1.50
s_O	1.80	1.31	2.00	1.52	1.62	1.73
s_P	1.04	1.31	1.60	0.97	1.33	1.14
i_{Br}	2.06	5.28	2.09	4.45	2.77	6.25
i_C	5.76	13.02	4.52	10.66	9.07	11.48
i_{Cl}	2.01	4.87	2.14	4.30	2.32	5.81
i_{Cl^*}						4.13
i_H	5.99	14.80	4.68	11.68	8.23	11.41
i_N	1.84	4.04	2.05	3.65	1.78	5.42
i_O	0.95	2.12	1.63	1.89	0.99	4.41
i_P	3.25	6.25	3.35	6.10	3.54	7.42
N	306	426	294	420	516	660
MAD (Å)	0.020	0.023	0.020	0.021	0.019	0.015
rmsd (Å)	0.027	0.032	0.026	0.029	0.027	0.022

meters for oxidation states II and III of Ru and Os. The bond lengths for these metals in oxidation state III have larger errors; a detailed inspection of some of the worst cases revealed a marked Jahn−Teller distortion, particularly for Ru(III). This is expected for low-spin d^5 complexes. The Jahn−Teller effect may also explain why the errors in the isomerization energies were largest for Ru(III) complexes.

An analysis of the deviations as a function of total charge revealed a small correlation. Adding a term $q \cdot p_M$ to eq 6,where p_M is a metal-dependent parameter and q is the total charge, reduces the overall MAD by 0.001 Å and the rmsd by 0.002 Å. Because the improvement due to considering the charge is modest, we decided to keep the original form of eq 6 for simplicity.

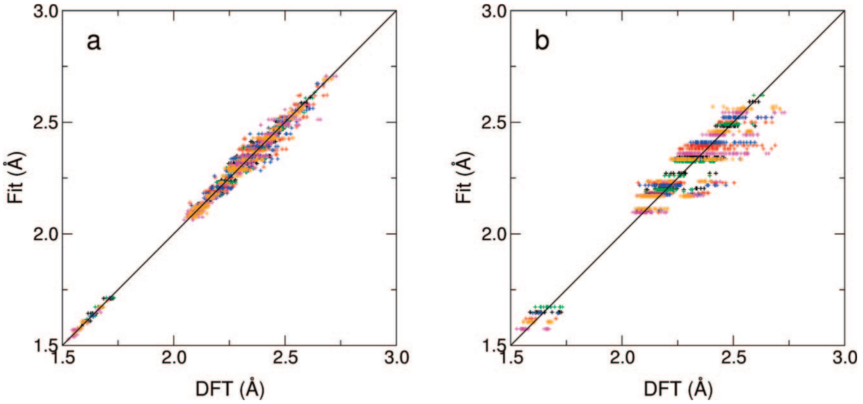


Figure 5. a) Bond lengths computed using eq 6 versus the reference DFT values. b) Fit to eq 6 neglecting the trans influence (i.e., $s_A = 0$ or $i_B = 0$). The solid lines correspond to $y = x$. Color code: black, Ru(III); red, Ru(IV); green, Os(II); blue, Os(III); magenta, Rh(III); orange, Ir(III).

Table 5. Relative Energies (in kcal/mol) for Complexes **1a**-**1d** from DFT, VALBOND-TRANS, and VALBOND Calculations

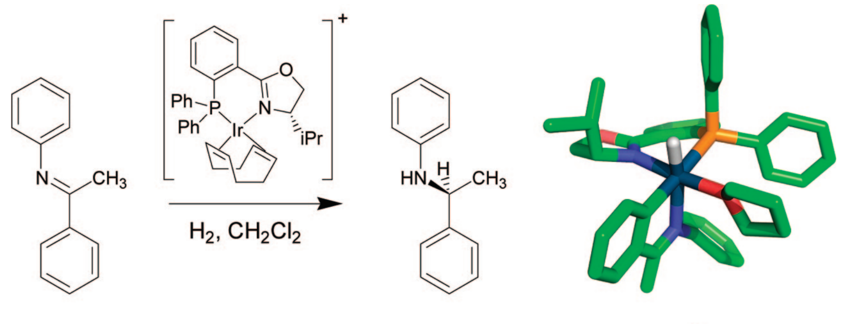
		VALBOND-TRANS		VALBOND	
compound	DFT	no charge	NPA	no charge	NPA
1a	8.6	7.5	7.1	−1.3	−2.0
1b	12.1	10.6	11.0	2.1	2.1
1c	0	0.0	0.0	0.0	0.0
1d	1.2	1.0	0.4	1.4	1.1

4. Application to Systems of Practical Relevance

To test the practicality of the method (referred to as VALBOND-TRANS), two series of complexes, the iridium dihydrides **1** and the monohydrides **2**, were investigated in more detail using DFT and VALBOND-TRANS. To this end, VALBOND and the extensions described above were implemented into CHARMM version 34a1.[4]The CHARMM22 parameter set was used.[43] When CHARMM force field parameters were missing, a) bond lengths were estimated as the sum of the covalent radii, and force constants were set to a default value; b) dihedrals were generally neglected, except for some dihedrals for triphenylphosphine and PHOX, which were obtained from energy scans on model compounds; c) partial charges were either set to zero or natural population analysis (NPA) charges from a DFT calculation were used;[44] and d) Lennard-Jones parameters were obtained from analogous CHARMM atom types. For simplicity, all angles, including those in the organic ligands, were treated using VALBOND, so no additional CHARMM angle bending parameters were necessary. All the nonstandard parameters that were used are listed in the Supporting Information.

As a first test, we revisited the iridium dihydrides involved in asymmetric hydrogenation, **1a**-**1d**, that we had studied previously.[24] The calculated DFT and VALBOND-TRANS energies are shown in Table 5. To assess the improvement due to the modifications presented in this work, the relative energies calculated using VALBOND without the trans terms are also shown. The qualitative stability sequence **1c** > **1d** > **1a** > **1b** is reproduced perfectly using VALBOND-TRANS, and the magnitudes of the relative energies are also reproduced well, with the largest deviation being 1.4 kcal/mol, occurring for the highest-energy isomer. Without the trans term, the stability sequence is not reproduced, and the largest error increases to 10 kcal/mol. An important observation is that the difference between isomers **c** and **d** and between **a** and **b** can be assumed to be largely steric, because they both have the same pairs of trans ligands (e.g, for **c** and **d**, P is trans to Cl, N trans to H, and Cl trans to H). The steric energy difference is accounted for by the nonbonded (electrostatic and van der Waals) interactions in the CHARMM force field, which are represented using Coulomb's law and a Lennard-Jones potential. To test the influence of the partial charges on the results, the calculation was performed using two different charge models. In one, all the atoms had a partial charge of zero; in the other, the partial charges were set to the NPA charges from the B3LYP calculation on compound **1c**. The dielectric constant was set to $\varepsilon = 4$, to account to some extent for the polarizability of the complex itself and for the large magnitude of the partial charges typical of the NPA method. Perhaps fortuitously, in this case the calculation with no charges agrees with the DFT results about as well as the calculation with the NPA charges.

As a second test, we considered complex **2**, which has been observed during the hydrogenation of imines using Ir-PHOX catalysts, as shown in Figure 6.[45] This complex has twenty possible diastereomers, which makes it an interesting test case. Again, the VALBOND-TRANS calculations were carried out using two different charge models, one with no charges and one with NPA charges, also with $\varepsilon = 4$. To prevent the choice of a single reference isomer from biasing the evaluation of the average deviations, we considered the isomerization between each possible pair of isomers, for a total of 190 isomerization energies spanning a range from −20 to +20 kcal/mol. The mean absolute deviations between VALBOND-TRANS and DFT are 4.8 kcal/mol with no charges and 4.7 kcal/mol with NPA charges. The partial charges are largely configurationally independent; using the charges derived from any of the isomers produced similar results (not shown), which indicates that it is reasonable to use the same partial charges for every isomer. The results using NPA charges are shown in Figure 7a,c; the trend in isomerization energies is qualitatively correct but not within chemical accuracy and there are significant outliers. How-



2

Figure 6. Catalytic imine hydrogenation, and structure of complex **2**, isolated when using THF as a solvent.

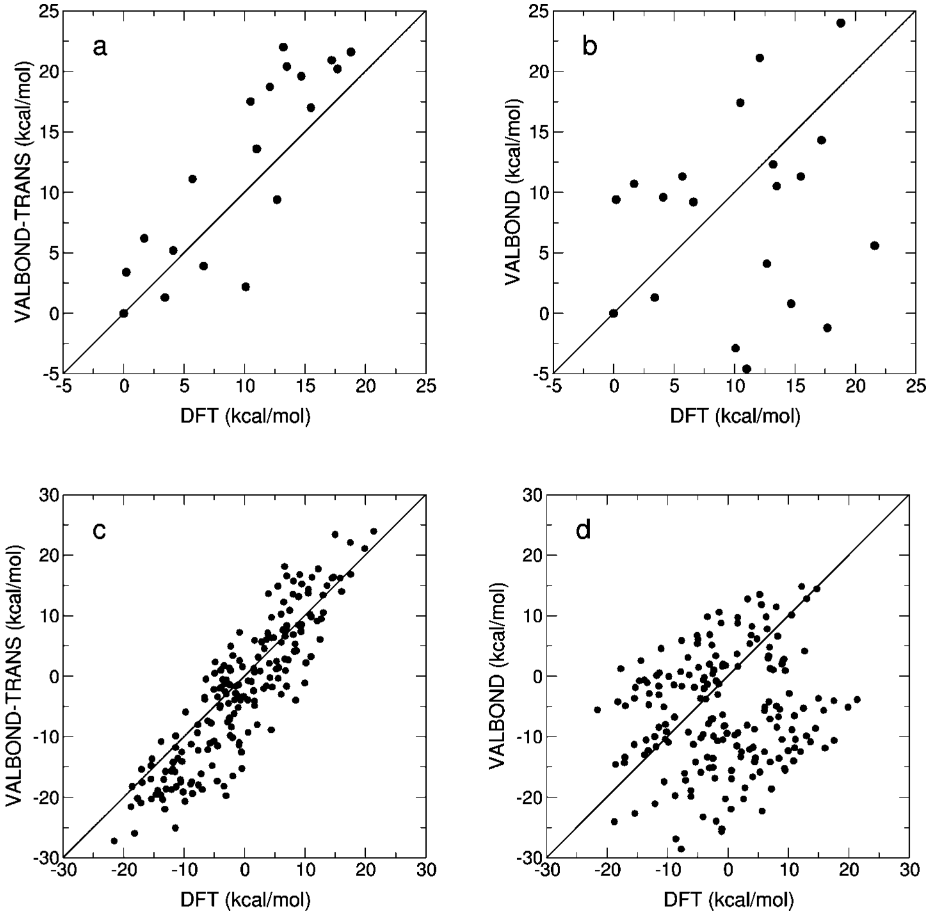


Figure 7. a) Relative energies of the 20 isomers of **2** using VALBOND-TRANS with NPA charges vs DFT. b) Same as a) but without including the new treatment of the trans influence. c) All isomerization energies between the 20 isomers of **2**, using VALBOND-TRANS vs DFT. d) Same as d) but without the trans influence. The solid lines correspond to the function $y = x$. The R^2 values for a-d are 0.78, 0.03, 0.79, and 0.02, respectively.

ever, the lowest-energy isomers are identified correctly using VALBOND-TRANS. When VALBOND without the trans terms is used instead, the low-energy isomers are not identified correctly and the correlation between the VALBOND and DFT energies is extremely poor, as shown in Figure 7b,d (R^2 = 0.02 and MAD = 10.7 kcal/mol). A calculation with $\varepsilon = 1$ resulted in a mean absolute deviation of 10.8 kcal/mol, and an analysis of the energy terms indicated that the electrostatic repulsions were overestimated, overwhelming all the other energy terms. This special sensitivity to the electrostatic terms in the case of compound **2** may stem from its very crowded ligand environment.

The ability of the force field to reproduce the overall shape of the complexes was assessed qualitatively by superimposing the DFT- and VALBOND-TRANS-optimized structure that minimizes the rmsd (see Figure 8). A deviation that is noticeable in both cases is a slight difference in the twisting of the oxazoline ring as well as slightly different torsions for the phenyl rings. While such a measure is useful for visual comparison, the magnitude of the rmsd is, however, dominated by deviations of the more flexible parts away from the metal center. Therefore, a more direct comparison of the bond lengths and angles was made. Plots showing the correlation between the metal−ligand bond lengths using

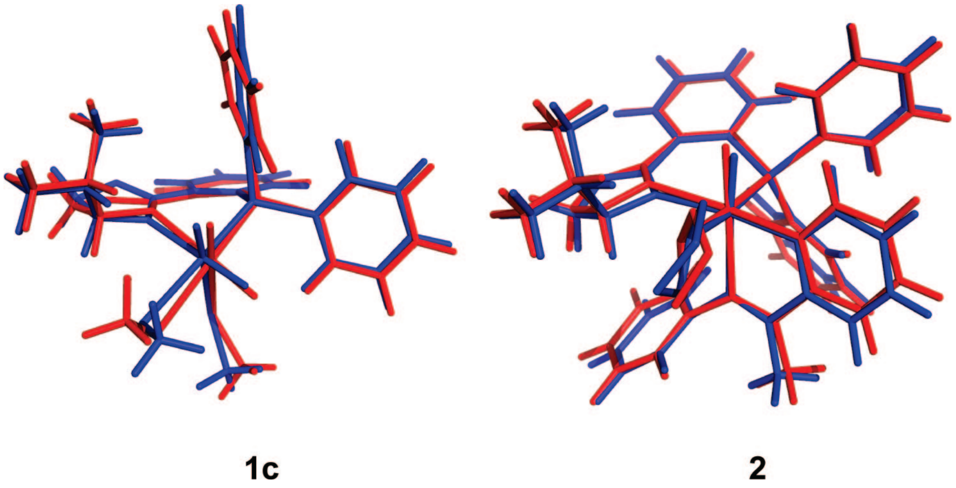


Figure 8. Comparison between the DFT-optimized structures (blue) and the VALBOND-TRANS-optimized structure (red) for compounds **1c** and the lowest energy isomer of **2**. The RMSDs are 0.53 Å and 0.34 Å, respectively; for discussion see text.

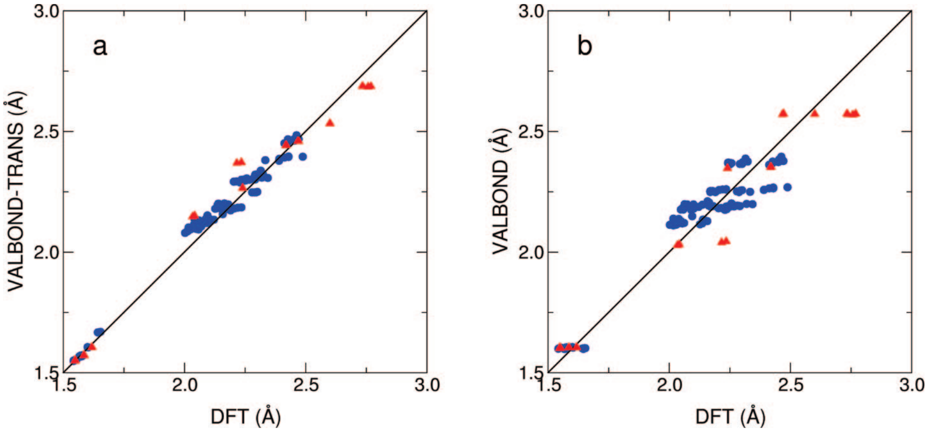


Figure 9. a) Bond lengths calculated using VALBOND-TRANS vs DFT for all isomers of compounds **1** (red triangles) and **2** (blue circles). b) Same as a) but without including the new treatement of the trans influence. The solid lines correspond to the function $y = x$.

VALBOND-TRANS and VALBOND vs DFT are shown in Figure 9. For both compounds **1** and **2**, the MAD for the bond lengths was 0.045 Å using VALBOND-TRANS, compared with 0.083 Å using VALBOND without the trans influence terms. For the angles, the MAD for compounds **1** and **2** were 6.4° and 4.0°, respectively, regardless of whether the trans influence terms were included or not. The largest deviations in the angles occur because normal sd^2 bending term in VALBOND is soft, while the hypervalent bending term in VALBOND is hard. As a result, large distortions such as the N−Ir−C angle, which is forced away from the ideal 90° due to the constraint of forming part of a five-member ring, have a large effect on the position of the trans atoms. For example, an N−Ir−H angle in the lowest energy isomer of **2** is 171.5° in the DFT geometry but 179.3° in the VALBOND geometry. Adjustment of the VALBOND parameters can reduce this problem, but for simplicity we decided to keep the default parameters in this work.

5. Conclusions

We have proposed and tested an extension of VALBOND combined with the CHARMM22 force field to account for the trans influence in octahedral complexes, based on valence-bond concepts such as 3-center-4-electron bonds. The proposed function fits the 403 isomerization energies in the training set with a mean absolute deviation of 1.3 kcal/mol, a result which is close to chemical accuracy using a minimal number of atom types. We tested whether the resulting parameters are transferable to two series of related complexes with large chiral ligands. In the case of the dihydride complex **1**, VALBOND-TRANS reproduces the DFT results very well, while for compound **2** the results were more qualitative but still identified the most stable isomers correctly. This demonstrates that the force field allows for the separation of electronic and steric effects to investigate the stability of different diastereomers. Since the steric effect can be accounted for using traditional force field nonbonded terms, and it is the steric interactions with the chiral ligands which drive enantioselectivity, the proposed force field may become a powerful tool for the study of enantioselective reactions involving transition metal catalysts. There are several areas where further work will be beneficial: 1) the development of a fast and systematic way of determining the partial atomic charges, instead of using NPA charges from a DFT calculation; 2) the inclusion of Jahn−Teller distortions; 3) an extension of the method for pi and sigma complexes, with ligands such as ethylene and H_2, respectively; 4) adiabatic reactive molecular dynamics, to potentially locate transition states; 5) extension to other metals and geometries; and 6) a more extensive reparametrization of the entire force field, including all relevant CHARMM and VALBOND parameters. The elements used in this work were chosen because of their importance to catalysis, the great stereochemical diversity allowed by the octahedral geometry, and their relatively simple electronic ground states. However, we expect that the method can be generalized to square-planar and trigonal bipyramidal geometries.

Acknowledgment. The authors thank the Swiss National Science Foundation for financial support. DFT calculations were carried out with a generous allocation of computing time on machines of the Swiss Center for Scientific Computing (CSCS), Manno, Switzerland. I.T.-B. was supported by a Marie Curie Fellowship (MIF1-CT2006-039088) from the European Commission.

Supporting Information Available: Tables of all the compounds in the training set and their DFT-computed energies and the nonstandard CHARMM parameters used in this work. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Strassner, T.; Taige, M. A. *J. Chem. Theory Comput.* **2005**, *1*, 848–855.

(2) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(3) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.

(4) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(5) Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. *Structure* **2006**, *14*, 437–449.

(6) Comba, P.; Remenyi, R. *Coord. Chem. Rev.* **2003**, *238*−*239*, 9–20.

(7) Boeyens, J. C. A.; Comba, P. *Coord. Chem. Rev.* **2001**, *212*, 3–10.

(8) Comba, P.; Zimmer, M. *J. Chem. Educ.* **1996**, *73*, 108–110.

(9) Bernhardt, P. V.; Comba, P. *Inorg. Chem.* **1992**, *31*, 2638–2644.

(10) Comba, P. *Coord. Chem. Rev.* **1993**, *123*, 1–48.

(11) Landis, C. R.; Root, D. M.; Cleveland, T. *Rev. Comput. Chem.* **1995**, *6*, 73–148.

(12) Zimmer, M. *Chem. Rev.* **1995**, *95*, 2629–2649.

(13) Norrby, P. O.; Brandt, P. *Coord. Chem. Rev.* **2001**, *212*, 79–109.

(14) White, D. P.; Douglass, W. *Molecular Mechanics Modeling of Organometallic Catalysts. In Computational Organometallic Chemistry*; Cundari, T. R., Ed.; Marcel Dekker:2001; pp 237−274.

(15) Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O. *J. Chem. Theory Comput.* **2008**, *4*, 1313–1323.

(16) Comba, P.; Zimmer, M. *Inorg. Chem.* **1994**, *33*, 5368–5369.

(17) Deeth, R. J. *Coord. Chem. Rev.* **2001**, *212*, 11–34.

(18) Danielsson, J.; Meuwly, M. *J. Chem. Theory Comput.* **2008**, *4*, 1083–1093.

(19) Roseblade, S. J.; Pfaltz, A. *Acc. Chem. Res.* **2007**, *40*, 1402–1411.

(20) Lightfoot, A.; Schnider, P.; Pfaltz, A. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 2897–2899.

(21) Pfaltz, A.; Blankenstein, J.; Hilgraf, R.; Hormann, E.; McIntyre, S.; Menges, F.; Schonleber, M.; Smidt, S. P.; Wustenberg, B.; Zimmermann, N. *Adv. Synth. Catal.* **2003**, *345*, 33–44.

(22) Kaellstroem, K.; Munslow, I.; Andersson, P. G. *Chem. Eur. J.* **2006**, *12*, 3194–3200.

(23) Cui, X.; Burgess, K. *Chem. Rev.* **2005**, *105*, 3272–3296.

(24) Mazet, C.; Smidt, S. P.; Meuwly, M.; Pfaltz, A. *J. Am. Chem. Soc.* **2004**, *126*, 14176–14181.

(25) Root, D. M.; Landis, C. R.; Cleveland, T. *J. Am. Chem. Soc.* **1993**, *115*, 4201–4209.

(26) Cleveland, T.; Landis, C. R. *J. Am. Chem. Soc.* **1996**, *118*, 6020–6030.

(27) Landis, C. R.; Cleveland, T.; Firman, T. K. *J. Am. Chem. Soc.* **1998**, *120*, 2641–2649.

(28) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(29) Piquemal, J.-P.; Williams-Hubbard, B.; Fey, N.; Deeth, R. J.; Gresh, N.; Giessner-Prettre, C. *J. Comput. Chem.* **2003**, *24*, 1963–1970.

(30) Deeth, R. J. *Inorg. Chem.* **2007**, *46*, 4492–4503.

(31) Pidcock, A.; Richards, R. E.; Venanzi, L. M. *J. Chem. Soc. A* **1966**, *1707*, 1710.

(32) Coe, B. J.; Glenwright, S. J. *Coord. Chem. Rev.* **2000**, *203*, 5–80.

(33) Steele, W. V. *Annu. Rep. Prog. Chem., Sect. A Inorg. Chem.* **1975**, *71*, 103–118.

(34) Brandt, P.; Norrby, T.; Aakermark, B.; Norrby, P.-O. *Inorg. Chem.* **1998**, *37*, 4120–4127.

(35) Weinhold, F.; Landis, C. R. *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*; Cambridge University Press: 2005; pp 473−474.

(36) Harvey, J. N.; Heslop, K. M.; Orpen, A. G.; Pringle, P. G. *Chem. Commun. (Cambridge, U.K.)* **2003**, *278*, 279.

(37) Pearson, R. G. *Inorg. Chem.* **1973**, *12*, 712–713.

(38) Allen, F. H. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380–388.

(39) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(40) 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.01*.

(41) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO Version 3.1*.

(42) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270–283.

(43) MacKerell, A. D., Jr.; 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., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(44) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735–746.

(45) Barrios-Landeros, F.; Pfaltz, A. Ir-catalyzed enantioselective hydrogenation of imines: Catalytic intermediates and mechanistic study. Presented at the 23rd International Conference on Organometallic Chemistry, Rennes, France, 2008.

CT800392N

JCTC Journal of Chemical Theory and Computation

Fully Integrated Approach to Compute Vibrationally Resolved Optical Spectra: From Small Molecules to Macrosystems

Vincenzo Barone,*[,†,‡] Julien Bloino,[†] Malgorzata Biczysko,[†] and Fabrizio Santoro[‡]

Dipartimento di Chimica "Paolo Corradini" and CR-INSTM Village, Universita di Napoli Federico II, Complesso Univ. Monte S. Angelo, via Cintia, 80126 Napoli, Italy, and Istituto per i Processi Chimico-Fisici, Area della Ricerca-CNR, via G. Moruzzi, 56124 Pisa, Italy

Received November 5, 2008

Abstract: A general and effective time-independent approach to compute vibrationally resolved electronic spectra from first principles has been integrated into the Gaussian computational chemistry package. This computational tool offers a simple and easy-to-use way to compute theoretical spectra starting from geometry optimization and frequency calculations for each electronic state. It is shown that in such a way it is straightforward to combine calculation of Franck−Condon integrals with any electronic computational model. The given examples illustrate the calculation of absorption and emission spectra, all in the UV−vis region, of various systems from small molecules to large ones, in gas as well as in condensed phases. The computational models applied range from fully quantum mechanical descriptions to discrete/continuum quantum mechanical/molecular mechanical/polarizable continuum models.

1. Introduction

Nowadays the characterization of complex biological systems or nanomaterials of direct technological and social interest relies more and more on computational approaches, e.g., for the evaluation and rationalization of structural, energetic, electronic, and dynamic features.[1−6] On the experimental side accurate information can be gained, in principle, by a number of spectroscopic techniques, magnetic as well as optical. Nevertheless, up to very recently direct comparisons between experimental and computed spectroscopic data have been rather scarce. Integrated approaches, capable of accurately simulating spectra but at the same time easily accessible to nonspecialists, are highly desirable. Such tools would allow for the exploitation of the recent and ongoing developments that are taking place in the field of computational spectroscopy[7−15] resulting in easy and, ideally, automatic vis-a-vis comparisons between experimental and theoretical results. In the present work, we introduce an approach to model quantitatively vibronic spectra, in line with such a demand. The integration among the different computational steps is particularly straightforward, since all calculations are performed within the same computational package: this enables the fully automatic computation and visualization of vibrationally resolved UV−visible spectra, a feature that is routinely available for other spectroscopic ranges (e.g., IR/Raman).

In the framework of the Franck−Condon (FC) principle,[16−18] time-independent ab initio approaches to simulate vibronic spectra are based on the computation of overlap integrals (known as FC integrals) between the vibrational wave functions of the electronic states involved in the transition. The computation of FC integrals requires a detailed knowledge of the multidimensional potential energy surfaces (PES) of both electronic states or, within the harmonic approximation, at least computation of equilibrium geometry structures and vibrational properties. Till recently, computations of vibronic spectra have been limited to small systems or approximated approaches, mainly a consequence of the difficulties in obtaining accurate descriptions of excited electronic states of polyatomic molecules. Recent develop-

---

* To whom correspondence should be addressed. E-mail: baronev@unina.it.

[†] Universita di Napoli Federico II.

[‡] Istituto per i Processi Chimico Fisici.

10.1021/ct8004744 CCC: $40.75

ments in electronic structure theory for excited states within the time-dependent density functional theory (TD-DFT)[19,20] and resolution-of-the-identity approximation of coupled cluster theory (RI-CC2)[21] have paved the route toward the simulation of spectra for significantly larger systems. When treating such large systems, the inclusion of vibrational contributions becomes very challenging, since the number of vibrational states to be taken into account increases steeply with the dimension of the molecule and the spectral width. Nonetheless, most of the possible vibronic transitions do not contribute significantly to the spectrum. Therefore the availability of effective selection criteria to individuate a priori the most relevant vibronic transitions within the dense bath of possible final states, can make feasible the calculation of the spectrum line shape also for these systems. Several schemes have been proposed,[22−27] ranging from the simplest approach, based solely on the energy window of the spectrum,[22,23] up to rigorous prescreening techniques based on analytically derived sum rules.[27] To maximize the efficiency of calculations, it is necessary to adopt a fast and a priori selection scheme of general applicability for a variety of different systems that is able to correctly choose all the non-negligible transitions. To satisfy these conditions we derived a general and robust tool starting from a method recently introduced by some of us[25,26] in the frame of harmonic approximation, which has been proven to provide very accurate spectra of medium-to-large systems with a limited computational cost. This new computational tool has been integrated into the Gaussian package.[53] A particular care has been taken to avoid the introduction of any prefixed constraint neither for the number of allowed quanta in a single mode nor for the number of simultaneously excited vibrations and for the spectrum energy range. Moreover, an automated procedure controls the spectrum convergence to avoid unnecessary calculations, and it alerts the user when incorrect settings might cause missing of important transitions. Thanks to these characteristics the integrated approach here presented can be applied to simulate in a fully automatic manner any vibronic spectrum for systems where nonadiabatic couplings are negligible and harmonic approximation is reliable.

Our integrated approach starts with the computation of the equilibrium geometries, frequencies, and normal modes for both electronic states involved in the electronic transition. A dedicated procedure performs reorientation of molecular coordinates to minimize the effect of rotations between the displaced equilibrium structures. After superposition of molecular geometries, the requested vibrationally resolved absorption or emission spectrum is generated. While the presented computational tool is developed within the harmonic approximation, it can be extended to take into account anharmonic effects. As a first step in this direction a correction scheme to derive excited state's anharmonic frequencies from ground-state data has been implemented.

In general, the accuracy of a simulated spectrum depends on the quality of the description of both the initial and the final electronic states of the transition. This is obviously related to the proper choice of a well-suited computational model: a reliable description of equilibrium structures, harmonic frequencies, normal modes, and electronic transition energy is necessary. In this respect, the advantage of the integrated approach we present here is the possibility of combining various computational schemes to create user-defined and/or problem-tailored approaches as for example the refinement of the electronic transition energy at a higher level of theory than the one used for the computation of structures and frequencies. Moreover, if a stand-alone software is used to compute the spectrum, some errors and/or inaccuracies in computations may be introduced by data processing and limited printing precision in the outputs of the electronic structure codes. The integrated tool here presented solves these problems by directly working with machine precision data available inside the Gaussian computational package. On the other side, the possibility of using available visualization or data analyzing tools is strictly related to the ease of use and accessibility of the approach to the nonspecialists. The aforementioned advantages of an integrated approach are illustrated with a few examples showing simulations of absorption and emission spectra, all in the UV−vis region, in the gas and condensed phases.

The paper is organized as follows. Section 2 describes the general theoretical frame of computation of FC and Herzberg−Teller spectra along with the details of the current implementation. Computational models that have been applied to the determination of structures, vibrations, and energies to provide the needed information for spectra calculation are gathered in section 3. Section 4 reports the application of this integrated tool to the simulation of: the vibronically resolved $S_1 \leftarrow S_0$ absorption spectrum of anisole (section 4.1), the photodetachment spectrum of SF_6^- (section 4.2), the emission phosphorescence the spectrum of $T_1 \rightarrow S_0$ transition of chlorophyll c2 (section 4.3), the UV absorption spectrum of $n \rightarrow \pi^*$ transition of acrolein in gas phase and in aqueous solution (section 4.4), and the photoelectron spectra of the isolated adenine molecule and adenine adsorbed on the Si(100) surface (section 4.5).

2. Spectra Calculation

2.1. Theory. In this section we briefly summarize the general mathematical frame for the spectra computations. The absorption spectrum, defined as the rate of energy absorption by a single molecule per unit radiant energy density is given by the expression[8]

$$\sigma_{\mathrm{abs}}(\omega) = \frac{4\pi^2\omega}{3}\sum_{\mathrm{f}} |\langle\Psi_{\mathrm{i}}|\mu|\Psi_{\mathrm{f}}\rangle|^2\delta(E_{\mathrm{f}} - E_{\mathrm{i}} - \hbar\omega) \qquad (1)$$

On the same foot the emission spectrum in photon counting experiments, defined as the rate of photon emission due to a single molecule, is[8]

$$\sigma_{\mathrm{em}}(\omega) = \frac{4\omega^3}{3\hbar c^3}\sum_{\mathrm{f}} |\langle\Psi_{\mathrm{i}}|\mu|\Psi_{\mathrm{f}}\rangle|^2\delta(E_{\mathrm{f}} - E_{\mathrm{i}} + \hbar\omega) \qquad (2)$$

The stick spectra in eqs 1 and 2 are usually convoluted with a Lorentzian or a Gaussian to simulate homogeneous or inhomogeneous broadening, respectively.

The intensity of a line of absorption or emission depends therefore on the transition dipole moment integral $\langle\Psi_i|\mu|\Psi_f\rangle$, where Ψ_i and Ψ_f are the molecular wave functions and μ is the electric dipole moment.

In the Born−Oppenheimer approximation, the wave function of each state can be expressed as a product of a nuclear ψ_n and an electronic ψ_e wave functions

$$\langle\Psi_{\mathrm{i}}|\mu|\Psi_{\mathrm{f}}\rangle = \langle\psi_{\mathrm{n}}\psi_{\mathrm{e}}|\mu|\psi'_{\mathrm{e}}\psi'_{\mathrm{n}}\rangle \tag{3}$$

The electric dipole moment can be separated into an electronic part μ_{e} and a nuclear part μ_{n}. By replacement of the electric dipole moment by these two components, the transition dipole moment integral can be divided into two terms

$$\langle\Psi_{\mathrm{i}}|\mu|\Psi_{\mathrm{f}}\rangle = \langle\psi_{\mathrm{n}}\psi_{\mathrm{e}}|\mu_{\mathrm{e}}|\psi'_{\mathrm{e}}\psi'_{\mathrm{n}}\rangle + \langle\psi_{\mathrm{n}}\psi_{\mathrm{e}}|\mu_{\mathrm{n}}|\psi'_{\mathrm{e}}\psi'_{\mathrm{n}}\rangle \tag{4}$$

Because of the orthogonality of the electronic wave functions of different electronic states, the second term of the rhs in eq 4 vanishes. As a consequence, the transition dipole moment integral depends on the nuclear wave functions and on the electronic transition moment $\mu_{\mathrm{if}} = \langle\psi_{\mathrm{e}}|\mu_{\mathrm{e}}|\psi'_{\mathrm{e}}\rangle$. Equation 4 can then be written

$$\langle\Psi_{\mathrm{i}}|\mu|\Psi_{\mathrm{f}}\rangle = \langle\psi_{\mathrm{n}}|\mu_{\mathrm{if}}|\psi'_{\mathrm{n}}\rangle \tag{5}$$

However, since no general analytical expression exists for the electronic transition dipole moment, this integral must be approximated. A first approach, proposed by Franck and Condon[16−18] is based on the assumption that molecular geometry does not change significantly during the transition, and therefore the electronic transition dipole moment can be treated as a constant. While this approximation is known to lead to very good results in many cases, it becomes not satisfactory for the dipole-forbidden or weakly allowed transitions. In these cases, one needs to improve the model by expanding the transition dipole moment in a Taylor series of the normal coordinates (either the set Q' of the final state or the set Q of the initial state). For computational reasons that will become clear later we choose the expansion on Q'

$$\mu_{\mathrm{if}}(Q') \cong \mu_{\mathrm{if}}(Q'_0) + \sum_{k=1}^{N} \frac{\partial\mu_{\mathrm{if}}}{\partial Q'_k} Q'_k + \frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N} \left(\frac{\partial^2\mu_{\mathrm{if}}}{\partial Q'_k\,\partial Q'_l}\right)_0 Q'_k Q'_l + \ldots \tag{6}$$

where Q'_0 refers to the equilibrium geometry of the final electronic state, while N is the number of normal modes.

In the current implementation we will consider the Taylor expansion up to the second derivatives. The zero-order, assuming that the transition dipole moment is unchanged during the transition, is the FC approximation. The Herzberg−Teller (HT) approximation corresponds to the development at the first order. In this case, we take into account a limited change in the structure during the transition.

Application of the Eckart conditions[28] allows minimization of the coupling between the rotational and vibrational motions of the nuclei in a molecule and as much as possible separate the nuclear wave function into rotational and vibrational contributions. Then, switching to the Dirac notation and assuming that the harmonic approximation can be used to represent the vibrational wave function of the initial and final states as eigenstates of the N-dimensional harmonic oscillator, i.e., by defining $|\psi_{\mathrm{n}}\rangle = |v\rangle$ and $|\psi'_{\mathrm{n}}\rangle = |v'\rangle$, it is possible to write eq 5 as

$$\langle\Psi_{\mathrm{i}}|\mu|\Psi_{\mathrm{f}}\rangle = \mu_{\mathrm{if}}(Q'_0)\langle v|v'\rangle + \sum_{k=1}^{N}\left(\frac{\partial\mu_{\mathrm{if}}}{\partial Q'_k}\right)_0 \langle v|Q'_k|v'\rangle + \frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left(\frac{\partial^2\mu_{\mathrm{if}}}{\partial Q'_k\,\partial Q'_l}\right)_0 \langle v|Q'_kQ'_l|v'\rangle \tag{7}$$

where μ_{if} has been replaced by its Taylor expansion given in eq 6.

Computation of the overlap integrals between initial and final vibrational states requires the use of a common coordinates set. Duschinsky proposed a solution to this problem by considering a linear transformation between the normal modes of the initial state and the final state[29]

$$Q' = JQ + K' \tag{8}$$

The Duschinsky matrix J describes the projection of the normal coordinates basis vectors of the final state on those of the initial state and represents the rotation of the normal modes upon the transition. The shift vector K' represents the displacements of the normal modes between the initial state and the final state structures.

Since J in principle is not diagonal, the calculation of the vibrational overlap integrals is not straightforward, and several methods were devised to compute the different terms of the rhs of eq 7. Schematically, we can divide them into two categories, analytical approaches[30,31] and recursive approaches.[32−34] In our case both methodologies have been employed, exploiting their respective advantages. Analytical methods can quickly and accurately compute the transition dipole moment integrals through ad hoc formulas, but the latter need to be generated beforehand and individually coded in the program to be truly efficient. As a consequence, the possible transitions, in the sense of a particular combination of quantum numbers for the initial state and final state, need to be predicted. This task is really cumbersome and in practice unfeasible when considering medium- or large-sized molecules. Moreover, it is difficult to design general-purpose programs based solely on this approach.

On the other side, recursive approaches provide methods to compute the overlap between given initial and final states through formulas that express it in terms of sums of integrals involving states with lower vibrational quantum numbers. Once the overlap between ground vibrational states has been calculated directly, that of any other transition can be obtained by recursively applying these formulas. While the derivation of the latter formulas has been a prerequisite for the calculation of the harmonic spectrum for a generic system, additional conceptual and technical difficulties remain to be faced with. First, in principle there is an infinite number of possible transitions, and therefore some limitations must be applied as will be described in section 2.2. Second, while the need to avoid redundant calculations would suggest to store each computed overlap integral, this strategy would easily fail leading to a saturation of the computer memory, and therefore efficient algorithms for their storage and/or partial recomputation must be figured out.

For many studied systems, the FC and HT approximations are sufficient to correctly describe both absorption and emission spectra, but for symmetry-forbidden transitions a correct reproduction of intensity of weak bands may require the inclusion of higher-order terms.[35] Thus, as a first attempt to deal with more general cases we will introduce also the expansion to the second order of eq 6.

Recursion formulas for the overlap integrals between vibrational states have been derived using several methods.[34] Briefly, using second quantization, the terms $\langle v|Q'|_k v'\rangle$ and $\langle v|Q'_k Q'_l|v'\rangle$ can be expanded in simpler terms, and then the second and third terms of rhs of eq 7 can be written as a sum of overlap integrals

$$\begin{aligned}\langle\Psi_i|\mu|\Psi_f\rangle &= \mu_{\mathrm{if}}\,(Q'_0)\,\langle v|v'\rangle + \\ &\sum_{k=1}^{N}\left(\frac{\partial\mu_{\mathrm{if}}}{\partial Q'_k}\right)_0\sqrt{\frac{\hbar}{2\omega'_k}}\left[\sqrt{v'_k}\,\langle v|v'-1'_k\rangle+\sqrt{v'_k+1}\,\langle v|v'+1'_k\rangle\right]+ \\ &\sum_{k=1}^{N}\left(\frac{\partial^2\mu_{\mathrm{if}}}{\partial Q_k'^2}\right)_0\frac{\hbar}{4\omega'_k}\left[\sqrt{v'_k(v'_k-1)}\,\langle v|v'-2'_k\rangle+\right. \\ &\left.(2v'_k+1)\,\langle v|v'\rangle+\sqrt{(v'_k+1)(v'_k+2)}\,\langle v|v'+2'_k\rangle\right]+ \\ &\sum_{k=1}^{N-1}\sum_{l=k+1}^{N}\left(\frac{\partial^2\mu_{\mathrm{if}}}{\partial Q'_k\partial Q'_1}\right)_0\frac{\hbar}{2\sqrt{\omega'_k\omega'_l}}\left[\sqrt{v'_kv'_l}\,\langle v|v'-1'_k-1'_l\rangle+\right. \\ &\sqrt{v'_k(v'_l+1)}\,\langle v|v'-1'_k+1'_l\rangle+\sqrt{(v'_k+1)v'_l}\,\langle v|v'+1'_k-1'_l\rangle+ \\ &\left.\sqrt{(v'_k+1)(v'_l+1)}\,\langle v|v'+1'_k+1'_l\rangle\right]\end{aligned}\tag{9}$$

where $|v' + x'_l\rangle$ is the vibrational state with all quantum numbers equal to $|v'\rangle$ apart from mode l which has x'_l quanta more. In a recursive approach, when computing $\langle v|v'\rangle$, $\langle v|v' + 1'_l\rangle$, $\langle v|v' + 2'_l\rangle$, $\langle v|v' + 1'_l - 1'_k\rangle$, and $\langle v|v' + 1'_l + 1'_k\rangle$ are not yet calculated and need to be expressed with respect to integrals involving states with lower quanta. In those electronic methods where the excited-state Hessian is obtained by numerical derivatives of the analytical gradient (with positive and negative increments), first derivatives as well as diagonal second derivatives of the transition dipole moment are automatically obtained. We have taken into account both of them in our spectrum calculation (i.e., the second and third terms in eq 9) in the spirit of using all the available information provided by the electronic computational model.

2.2. Computational Strategy. The recursive approach described in the previous section can be successfully applied to the computation of spectra of medium-to-large molecular systems. However, in many cases the number of overlap integrals that must be taken into account can become extremely large, with a consequent increase in the required computational times and memory usage.

Efficient computational strategies must be able to individuate a priori the relevant transitions among the infinite number of possible final states. A first simple scheme[22,23] selects the transitions on the basis of their energy, considering only those falling in a predetermined energy range (which is in principle arbitrary). Such an approach is not effective enough for large systems with broad spectra, and more robust methods need to be devised. An estimation procedure should work a priori, not require too many computations, and correctly choose the relevant transitions. Several strategies have been devised to decrease the number of required calculations, while retaining the accuracy of the resulting spectra.[24,25,36,37] Here, we use an a priori method called FCclasses,[25,37] which provides very accurate vibrationally resolved spectra of medium and large molecular systems with limited computational resources. According to this method, transitions are partitioned into classes C_n, depending on the number n of simultaneously excited normal modes of the final electronic state of the transition. The overlap integrals for single vibrations (class C_1) and combination of two normal modes (class C_2) are computed up to a chosen limit (it can be also very large, since computation is cheap; therefore any loss of accuracy in this step can be avoided). The probabilities of all these transitions are stored and then used in the computation of FC integrals for higher-order classes to obtain a priori estimates of the maximum quantum number that needs to be considered for each normal mode. To that end, for each class the allowed transitions are chosen iteratively on the ground of a minimum threshold for C_1 and C_2 probabilities, so that the number of overlap integrals to be computed stays approximatively below a user-defined limit ($N_I^{\max}$), that rules the calculation accuracy (see ref 25 for further details). As we will show in the following, the partition of transitions in "classes" C_n allows a breakthrough in the spectrum calculation, since this latter quickly converges with the increase of the class order n. From a technical point of view, such partition is particularly appealing for a high-efficiency implementation since it leads to a straightforward parallelization of the calculations. Furthermore, separation in classes allows for an easy control on the memory requirements for storage of the overlap integrals used in recursive procedures.

2.3. Convergence of the Spectra Computations. When using an a priori method to selectively compute transitions, a general issue to address is the evaluation of the convergence of the calculations and hence of the reliability of their outcome. In the calculation of vibrationally resolved spectra, this can be easily done on the ground of analytical sum rules, by comparing the actual computed spectrum intensity I^n_{tot} to the exact analytical limit I^a_{tot}. In the current approach spectrum convergence is always improved by increasing the limit for the number of integrals computed for each *class* of transitions: $N_I^{\max}$ (MAXINT). A higher number of allowed transitions obviously yields a better spectrum convergence but also directly increases the required computational time and memory usage.

Skipping prefactors and the dependence on the frequency, the intensity of a given transition is equal to $I(v, v') = |\langle v|\mu_{\mathrm{if}}(Q')|v'\rangle|^2$ and summing over all the possible final states one gets

$$I^a_{\mathrm{tot}} = \sum_{v'}|\langle v|\mu_{\mathrm{if}}(Q')|v'\rangle|^2 = \sum_{\rho=x,y,z}\langle v|\mu^2_{\mathrm{if}\rho}(Q')|v\rangle \tag{10}$$

where the superscript a indicates that the sum has been carried out analytically by exploiting the closure relation. In the limit of a complete inclusion of all the possible final states the numerical sum of the state-to-state intensities $I^n_{\mathrm{tot}} = \sum_{v'} I(v, v')$ must approach I^a_{tot}, and the ratio $C = I^n_{\mathrm{tot}}/I^a_{\mathrm{tot}}$ can be used as a control of the convergence, which is complete when $C = 1$. It is possible that, for large systems, a great number of transitions has to be considered to reach convergence of the spectrum intensity, and calculations although feasible can become computationally demanding. On the other side, usually the main scope is to correctly reproduce the spectrum line shape and assign the most important vibronic transitions. It has been shown[25,26,37] that the spectrum line shape usually converges much faster than C. The line shape convergence can be easily checked by comparison of results calculated with two different thresholds for $N_l^{\max}$.

For FC calculations eq 10 trivially gives $|\mu_{\mathrm{if}}|^2$ and the convergence only depends on the sum of FC factors $|\langle v|v'\rangle|^2$, which must tend to 1. For HT calculations it has been shown in ref 37 that the integral in eq 10 is more easily evaluated by expanding μ_{if} on the normal coordinates of the initial state Q. By consideration of the cartesian component ($\rho = x$, y or z)

$$\mu_{\mathrm{if}\rho}(Q) \simeq \mu_{\mathrm{if}\rho}(Q_0) + \sum_{k=1}^{N} \frac{\partial \mu_{\mathrm{if}\rho}}{\partial Q_k} Q_k = \mu_{\mathrm{if}\rho}(Q_0) + M_{1\rho}^T Q \quad (11)$$

where $M_{1\rho}$ is the vector of the dipole derivatives $\partial\mu_{\mathrm{if}\rho}/\partial Q$. Writing an analogous expression for the expansion as a function of Q'

$$\mu_{\mathrm{if}\rho}(Q') = \mu_{\mathrm{if}\rho}(Q'_0) + {M'_{1\rho}}^T Q' \quad (12)$$

and substituting eq 8 into eq 11 one gets

$$\mu_{\mathrm{if}\rho}(Q_0) = \mu_{\mathrm{if}\rho}(Q'_0) + {M'_{1\rho}}^T K' \quad (13)$$

$$M_{1\rho}^T = {M'_{1\rho}}^T J \quad (14)$$

These relations provide the data for the computation of $I_{\mathrm{tot}}^{\mathrm{a}}$

$$I_{\mathrm{tot}}^{\mathrm{a}} = |\mu_{\mathrm{if}}(Q_0)|^2 + \sum_{k=1}^{N} \frac{\hbar}{2\omega_k}(2v_k + 1) \sum_{\rho=x,y,z} M_{1\mathrm{k}\rho}^2 \quad (15)$$

An analogous procedure can be followed to compute $I_{\mathrm{tot}}^{\mathrm{a}}$ for an expansion of μ_{if} up to the second order. One has

$$\mu_{\mathrm{if}\rho}(Q) = \mu_{\mathrm{if}\rho}(Q_0) + M_{1\rho}^T Q + Q^T M_{2\rho} Q = \mu_{\mathrm{if}\rho}(Q'_0) + {M'_{1\rho}}^T Q' + Q'^T M'_{2\rho} Q' \quad (16)$$

where

$$\mu_{\mathrm{if}\rho}(Q_0) = \mu_{\mathrm{if}\rho}(Q'_0) + {M'_{1\rho}}^T K' + K'^T M'_{2\rho} K' \quad (17)$$

$$M_{1\rho}^T = M'^T_{1\rho} J + K'^T M'_{2\rho} J \quad (18)$$

$$M_{2\rho} = J^T M'_{2\rho} J \quad (19)$$

Now, by consideration of eq 10, expanding μ_{if} up to the second order and taking into account that terms in Q_k^n do not contribute by symmetry if n is odd, one gets

$$I_{\mathrm{tot}}^{\mathrm{a}} = |\mu_{\mathrm{if}}(Q_0)|^2 + \langle v|\{\sum_{k} \sum_{\rho=x,y,z} [(2\mu_{\mathrm{if}\rho}(Q_0) M_{2kk\rho} + M_{1k\rho}^2) Q_k^2 + M_{2kk\rho}^2 Q_k^4] + 2 \sum_{\substack{k,l \\ k>l}} \sum_{\rho=x,y,z} (M_{2kk\rho} M_{2ll\rho} + 2M_{2kl\rho}^2) Q_k^2 Q_l^2 \}|v\rangle \quad (20)$$

It can be easily shown that

$$\langle v|Q_k^2|v\rangle = \frac{\hbar}{2\omega_k}(2v_k + 1) \quad (21)$$

$$\langle v|{Q_k}^2 {Q_l}^2|v\rangle = \frac{\hbar^2}{4\omega_k\omega_l}(2v_k + 1)(2v_l + 1) \quad (22)$$

$$\langle v|Q_k^4|v\rangle = \frac{\hbar^2}{4\omega_k^2}(6v_k^2 + 6v_k + 3) \quad (23)$$

By substitution of eqs 21– 23 in eq 20, one finally obtains the analytical sum for a second order expansion of μ_{if}

$$I_{\mathrm{tot}}^{\mathrm{a}} = |\mu_{\mathrm{if}}(Q_0)|^2 + \sum_{k} \sum_{\rho=x,y,z} \frac{\hbar}{2\omega_k} \times [2\mu_{\mathrm{if}\rho}(Q_0) M_{2kk\rho} + M_{1k\rho}^2](2v_k + 1) + \sum_{k} \sum_{\rho=x,y,z} \frac{\hbar^2}{4\omega_k^2} M_{2kk\rho}^2 (6v_k^2 + 6v_k + 3) + 2 \sum_{\substack{k,l \\ k>l}} \sum_{\rho=x,y,z} \frac{\hbar^2}{4\omega_k\omega_l} (M_{2kk\rho} M_{2ll\rho} + 2M_{2kl\rho}^2)(2v_k + 1)(2v_l + 1) \quad (24)$$

In the current implementation we restrict ourselves to the diagonal terms of the second derivative of μ_{if}, thus the last term of eq 24 is not taken into account.

2.4. Anharmonic Corrections for the Ground and Excited Electronic States. As already stated, the methods outlined in previous sections are based on the harmonic approximation. However, improving the accuracy of simulated spectra requires going beyond harmonicity and taking into account anharmonic effects, couplings between modes, and vibrational or vibronic resonances. As a first step in this direction, some of us introduced[38] anharmonic corrections to the vibrational frequencies of both ground and excited electronic states. Here we briefly summarize a simple scheme to derive excited-state mode-specific scaling factors starting from the ground-state ones. These latter should be provided by the user, e.g., from perturbative anharmonic frequency calculations,[39] or from easily accessible ground-state experimental data. For each particular normal mode Q_k, the frequency scaling vector α is computed first, using the formula $\alpha(k) = \nu(k)/\omega(k)$ where ν is the anharmonic frequency and ω is the harmonic frequency. To proceed further, we shall assume that, if there is a one-to-one relation between the normal modes Q_k and Q'_k of the initial and final state, the scaling factors α_k and α'_k are equal. However, the normal modes are in general not coincident ($J \neq I$), and α cannot be transferred directly to scale the frequencies of the final state. In other words, the scaling vector must be adapted to the excited-state frequencies. In the case of small-amplitude vibrations, this can be obtained by expressing the normal modes of the excited state as linear combinations of the normal modes of the initial state by means of the Duschinsky transformation. The J_{kl} coefficients can now be applied to derive the relation between the initial (k) and final (l) state mode-specific anharmonicity scaling factors

$$\alpha'_l = \sum_{k}^{N} J_{lk}^2 \alpha_k \quad (25)$$

2.5. Implementation. The inclusion of vibronic computations in a general purpose electronic code (here Gaussian) requires an effective control of the memory usage. However, the storage of the overlap integrals for the recursion formulas is often the main bottleneck in FC calculations, and by use of a conventional method, keeping most integrals in an array is quickly unviable when dealing with large systems. Taking advantage of the capabilities of current computers, it is possible to consider a semidirect method, storing only a group of overlap integrals at a time. The computational costs induced by the necessary recalculations are generally on par with the times required to find a specific element in a large array. The main problem of a semidirect approach is to devise a consistent scheme to split the calculations of transition dipole moment integrals in groups requiring only a moderate need for recalculation in order to not hamper the efficiency of the spectrum generation. In our implementation, we chose to part the overlap integrals following a two-step procedure. First, they are naturally divided in classes as described in section 2.2. Second, each possible combination of n modes defines a set belonging to class C_n, and only the integrals calculated for a given set are stored at a time, thus dropping the storage requirements.

This partition technique also allows the implementation of several strategies to improve the efficiency of our calculations. First, it is not necessary to work on the full-dimensionality (N, the number of normal modes) Sharp and Rosenstock matrices (A, B, C, D, E),[30] and these latter, together with the diagonal matrices of reduced frequencies, can be replaced inside a given set by smaller matrices containing only the elements required by the calculations. The size of these new matrices, n for class C_n, is generally much lower than N speeding up the search for specific elements in these arrays. Additionally, the subdivision into well-defined sets allows the tailoring of the storage of the overlap integrals and their indexing. In fact, since the overlap integrals to be computed for a given set are known in advance, it is possible to choose a linear storage with a particular indexing function adapted to it, boasting the efficiency in retrieving a specific FC integral with respect to what would be allowed by a generic storage technique. Finally, different sets are independent, and therefore their calculation can be tackled in parallel, improving greatly the velocity of the spectrum simulation. Thanks to the double partition, two levels of parallelization can be pursued, either treating all the classes contemporaneously but their sets in serial or treating even these latter simultaneously.

In our computational strategy we calculate by default the intensities of the C_1 and C_2 transitions through analytical formulas. Nonetheless, a recursive calculation can be adopted also for these classes, if requested by the user. The present computational tool is designed for large systems where it may be necessary to take into account a large number of simultaneously excited modes to adequately simulate the spectra. Hence, a general method has been devised to let the order n of the highest class be set by the user.

The general integrated procedure to compute an absorption or emission spectrum is straightforward. As a first step the geometry structures of both electronic states and the frequencies of one of the two electronic states, by default the final one, have to be determined. These computations have to be performed beforehand and are stored in the internal checkpoint files. In this way all data needed in subsequent steps are extracted directly from internal files, and the accuracy of results is not arbitrarily truncated. The suite of modules dedicated to spectrum computations is integrated through a ad hoc subprogram, which proceeds to the treatment of the data available from previous calculations. Thus, the procedure to compute the vibrationally resolved spectrum extracts the stored data, and in a next step computes the frequencies for the second electronic state (by default the initial one) and the vibronic spectrum. When dealing with two structures computed independently, spurious effects might be introduced in the spectrum, if displacements due to translation and rotation are not carefully dealt with. While translation can be exactly removed by superposing the center of masses of the two equilibrium structures, rotational effects can be minimized by maximizing their mutual superposition. Our subprogram checks the superposition of the equilibrium structures and, if necessary, proceeds to their rotation and translation. All necessary data, such as frequencies and normal modes, are regenerated in this case.

The simplest computation of Franck−Condon spectrum requires the following data:

(1) Cartesian coordinates of the atoms.

(2) Masses of the atoms.

(3) Energy of the ground and excited states.

(4) Frequencies for the two electronic states involved in the transition.

(5) Normal modes for the two electronic states, expressed by the atom displacements.

All of which are available from frequency calculations performed for each electronic state. Additional data can be needed depending on the requested calculation. For HT calculations also the electronic transition dipole moment and its derivatives are required. For TD-DFT method, these data are directly available from the frequency calculations, since they are performed by numerical differentiation of the analytical TD-DFT energy gradients. If the electronic excited states are calculated with methods adopting analytical second derivatives by default, the numerical calculation of frequencies must be requested together with the HT calculation. Numerical derivatives of the transition dipole moment provide at no additional cost the diagonal terms of the second order of its Taylor expansion (see eq 6). Our method can take into account also these data, to utilize all the available information and maximize the accuracy of the computed spectra. Other data are also required to apply the anharmonic corrections. It is possible to provide anharmonic frequencies for both electronic states or derive anharmonic corrections for excited-state based on the ground-state data, as explained in section 2.4. In both cases proper sets of frequencies need to be included directly in the input for spectrum computation. In this way it is possible to apply anharmonic corrections by means of calculations performed beforehand (e.g., using the perturbative approach)[39] or making use of available (e.g., experimental) data. The spectrum convoluted with a Gaussian function [with half-width at half-maximum (SPECHWHM)

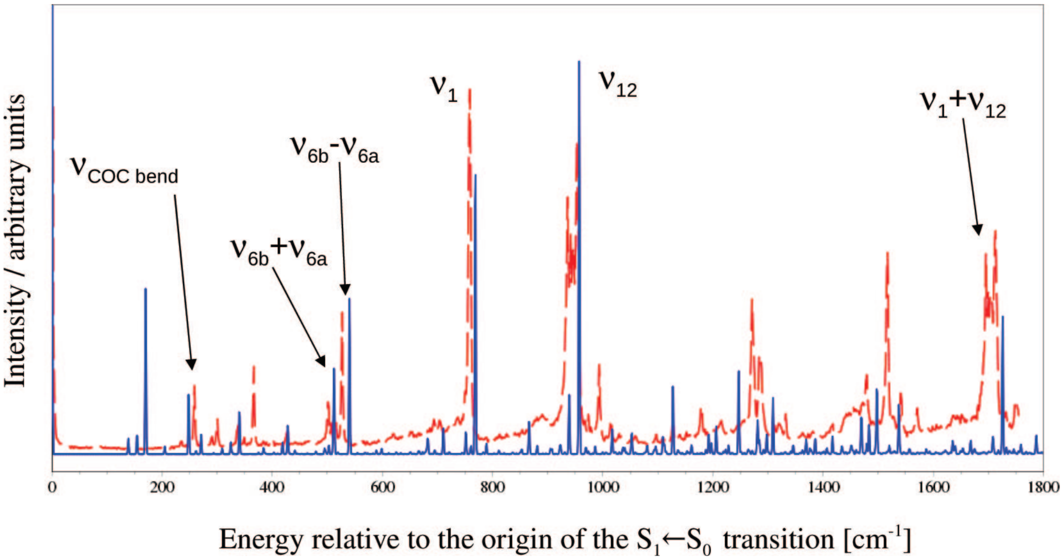


Figure 1. Computed TA (blue lines) and experimental REMPI[55] (red dashed lines) spectra of the $S_1 \leftarrow S_0$ transition of anisole along with the assignment of the most intense bands; see ref 38 for the details.

Table 1. Convergence of Spectra Computations for Adenine and Adenine@Si(100)[a]

	adenine		adenine@Si(100)	
class (n)	${}_{N}C_n$	progression	${}_{N}C_n$	progression
3	9.14×10^3	84.54%	4.27×10^7	87.31%
4	8.23×10^4	93.57%	6.75×10^9	94.82%
5	5.76×10^5	97.48%	8.54×10^{11}	97.37%
6	3.26×10^6	98.32%	8.98×10^{13}	97.88%
7	1.54×10^7	98.39%	8.08×10^{15}	97.93%

[a] For each class C_n the number of combinations of the n excited oscillators ${}_{N}C_n$ and spectrum progression are listed. The C_1 and C_2 transitions have been computed by analytical formulae allowing a maximum quantum number $v_i = 30$, and $v_1 = v_2 = 20$ (MaxC1 = 30, MaxC2 = 20) respectively. For the classes C_n, $n \geq 3$ the transitions to be computed have been selected setting the parameter $N_I^{\max}$ to 10^8 (the default value).

of 135 cm^{-1} by default] is generated as an outcome of vibronic calculations (Lorentzian broadening can be introduced as well). The assignment of vibronic transitions is available through the corresponding overlap integrals that are printed up to a chosen threshold (PRTINT), the transitions with intensity larger than 1% of $I(0, 0')$ are printed out as default. The convergence of the spectrum calculations and the contribution of the classes are also reported together with the requested matrices (e.g., J).

The parameters that rule spectra calculations can be divided into two groups, the first being related to the spectrum output as the convolution and its spectral range and grid, etc., which can be freely chosen in a manner appropriate for the problem under study. The second set of parameters is related to the accuracy of the computed spectrum. These are the maximum quantum number for each mode in C_1 and C_2 integrals (MAXC1 and MAXC2, respectively) and the maximum number of simultaneously excited modes MAXBANDS, which all determine the transitions to be computed, together with an additional parameter $N_I^{\max}$ (MAXINT). This latter limits an a priori and computationally cheap estimate of the number of integrals N_I to be computed for each class (see ref 25). This quantity is evaluated through the expression: $N_I = {}_{N}C_n \times \langle v'_{\max}\rangle^n$. The binomial coefficient represents the number of combinations of the n excited oscillators and $\langle v'_{\max}\rangle$ is the arithmetic mean of the N maximum quantum numbers $v'_{\max}$. In the following, for brevity, we will generically indicate $N_I^{\max}$ as the maximum number of integrals computed for each class. An internal check procedure has been devised to monitor computation convergence in order to prevent unnecessary calculations but also to warn for the cases when transitions to higher quanta need to be considered. Various aspects of spectrum convergence with respect to these parameters are discussed further in sections 4.2, 4.3, and 4.5.

3. Computational Chemistry Models

Full geometry optimizations and harmonic frequency calculations have been performed for all systems under study and in both final and initial electronic states. In some cases anharmonic perturbative[39] calculations have been performed for the ground electronic state, and subsequently used to account for anharmonicity in both S_0 and S_1 electronic states, as described in section 2.4. The computational chemistry methods have been chosen accordingly to the system under study in order to find a satisfactory balance between feasibility of calculations and accuracy of results. For anisole molecule the ground and excited states computations have been carried out using DFT and TD-DFT,[19] respectively, with the well-known B3LYP functional[40] and the 6-311+G(d,p) basis set. For SF_6 and its negative ion, computations have been performed at the MP2[41] level with aug-cc-pVTZ basis set.[42,43] For chlorophyll c2 both singlet ground and triplet excited electronic states have been calculated at the DFT level with the PBE0[44] functional and 6-31G(d) basis set. Solvent effects on the UV spectrum of acrolein have been introduced by a model where the solute molecule is treated explicitly by means of DFT with B3LYP functional and N07D double-ζ basis set,[45] and the continuum medium is described by the CPCM.[46] In the case of adenine molecule adsorbed on Si(100) surface the ONIOM[47] quantum me-

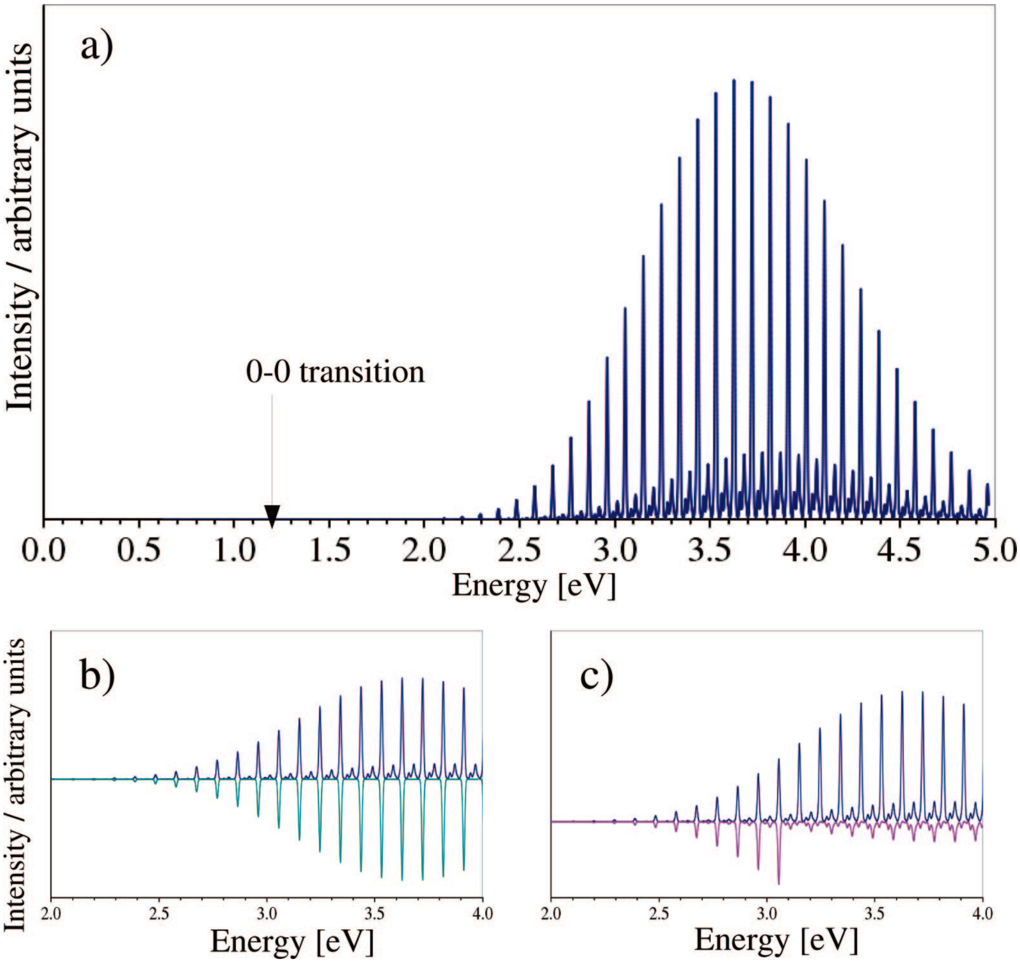


Figure 2. Theoretical electron photodetachment spectrum of SF_6^-. (a) Full spectrum in a range from 0 to 5 eV calculated within FC approximation with combinations between all modes considered; the energy of 0−0 transition is marked by an arrow. (b) Comparison between the full spectrum (upper panel) and a spectrum with coupling between modes excluded (lower panel). (c) Comparison between the full spectrum calculated with MAXC1 set to 100 (upper panel) vs the one with the maximum 20th overtone considered (lower panel).

chanical/molecular mechanical (QM/MM) scheme has been adopted with the Si(100) surface represented by a cluster of 119 silicon atoms. The QM part corresponding to the adenine molecule has been calculated at the B3LYP/6-31+G(d,p) level, while the cluster has been modeled by molecular mechanics using the UFF force field.[48] In the QM/MM calculation, the MM part has been polarized taking into account the interaction with the ground-state of the neutral molecule or its cation, respectively. Moreover in some cases the electronic energies of the initial and final electronic states have been refined at the coupled cluster level.[49−51] The CCSD and EOM-CCSD calculations have been performed with the MOLPRO[52] package. All other calculations have been performed with a locally modified version of the Gaussian suite of quantum chemistry programs.[53]

4. Applications

The integrated approach to compute vibrationally resolved optical spectra can be applied to a large variety of systems ranging from small molecules in the gas phase to macrosystems in condensed phases, whenever nonadiabatic couplings are negligible and harmonic approximation is reliable. The given examples of absorption spectrum of $S_1 \leftarrow S_0$ electronic transitions of anisole, photodetachment spectrum of SF_6^-, emission $T_1 \rightarrow S_0$ phosphorescence spectrum of chlorophyll c2, UV spectrum of acrolein in the gas phase and aqueous solution and a photoelectron spectra of adenine adsorbed on the Si(100) surface, are chosen to illustrate the flexibility of the present computational tool.

4.1. Vibrationally Resolved Optical Spectrum of Anisole. The recently published vibrationally resolved absorption spectrum of the $S_1 \leftarrow S_0$ electronic transitions of anisole[38] represents an example of the simulation accuracy achievable when good quality geometries and force fields for both electronic states are provided. The planarity of anisole in both electronic states has been determined by the high resolution laser-induced fluorescence (LIF) spectroscopy.[54] For anisole, methods based on the density functional theory and its time-dependent extension for electronic excited states [B3LYP/6-311+G(d,p) and TD-B3LYP/6-311+G(d,p)] have been applied to geometry optimizations and harmonic frequency calculations. The remarkable overall agreement between theoretical and experimental[54] rotational constants (average deviation of about 0.5%, for both electronic states) confirms good quality of the calculated geometry structures. The relative energy of the electronic states has been refined by EOM-CCSD/CCSD//aug-cc-pVDZ computations. The vibrational frequencies in the first excited electronic state have been corrected according to the ground-state experimental frequencies (EA) or to the calculated perturbative anharmonic frequencies[39] (TA). The spectrum has been computed with the default choice of parameters yielding 99.6% of the total spectrum intensity. The simulated vibronic profile convoluted with a fwhm of 2 cm^{-1} is compared to the highly accurate experimental data from resonance-enhanced multiphoton ionization (REMPI) spectrum[55] in Figure 1. On the whole, a very good agreement has been achieved (the root-mean-square deviation between computed and experimental bands is 15 cm^{-1}), as detailed in ref 38. To reproduce correctly the band intensities and the rich vibrational structure of the REMPI spectra, it has been necessary to account for changes in structure, vibrational frequencies, and normal modes between the involved electronic states. It is worthwhile highlighting that the remarkable overall agreement, also as far as band positions are concerned, has only been possible by correcting the frequencies for anharmonicity. The discrepancy between the absolute position of experimental and simulated spectra remains the main shortcoming of the purely theoretical approach: to achieve a good match between spectra, the energy of the electronic transition should be computed with an accuracy of ~10 cm^{-1}. Thus, even if DFT/TD-DFT computations are able to provide quite reasonable estimates of the relative energetics of the electronic states (within 0.2 eV) and despite the refinement based on coupled clusters calculations (0.05 eV), it was still necessary to compare spectra shifted to the 0−0 origin. Nevertheless, the remarkable agreement between theoretical and experimental spectra allowed for revision of some assignments of fundamental vibrations in the S_1 state of anisole. In particular, for many bands that had been assigned to S_1 fundamentals, consideration of the relative intensities has suggested instead a different interpretation as combinations or overtones.[38]

4.2. Photodetachment Spectrum of SF_6^-. SF_6^- represents an interesting case of a relatively small and highly symmetric system where accurate ab initio methodologies can be applied, giving possibility for an easy computation of accurate theoretical spectra which can be of great value for the interpretation of the best available experimental spectroscopic data in the gas phase. The photodetachment spectrum of SF_6^- is characterized by a broad progression, with the band maximum shifted by more than 2 eV from the 0−0 transition corresponding to the adiabatic electron affinity (AEA). These features have been attributed to the significant changes in the geometry between ionic and neutral species.[56,57] Indeed MP2/aug-cc-pVTZ calculations show a significant elongation of the S−F bond upon electron attachment (from 1.5750 to 1.7146 Å), while the molecule octahedral symmetry remains unchanged. Moreover, the theoretical [CCSD(T)/aug-cc-pVTZ] AEA of 1.06 eV is in very good agreement with the experimental value of 1.0 eV.[58] The fully theoretical photodetachment spectrum, calculated in an energy range from 0 to 5.0 eV within the FC approximation on the basis of the aforementioned ab initio results (see Figure 2) clearly resembles its recently measured experimental counterpart.[56] For the SF_6^- photodetachment spectrum most of the vibrational progression derives from excitation of the single totally symmetric S−F stretching mode (ν_1). Indeed, the computed spectrum shows a regular pattern of bands, the most intense corresponding to ν_1. Panel b of Figure 2 shows a comparison between the spectrum obtained with only the FC integrals from class C_1 and the complete spectrum, where couplings between all modes are also taken into account. In the latter case it is found that the weaker bands gain intensity from excitation of this totally symmetric mode, being related to the combinations between the overtones of ν_1 and doubly excited degenerate mode ν_4 or ν_5. As a consequence of the large changes in the S−F bond length, the most intense transitions are related to high overtones, which must therefore be considered to reproduce accurately the spectrum features. Our approach allows such a flexibility, through the change of the MAXC1 parameter to a user-defined value instead of the default value of 20. Indeed in the latter case only about 30% of the spectrum intensity has been achieved in comparison to 94% when all necessary excitations have been taken into account. The comparison of the spectra calculated with MAXC1 set to 100 and to 20 is shown on panel c) of Figure 2.

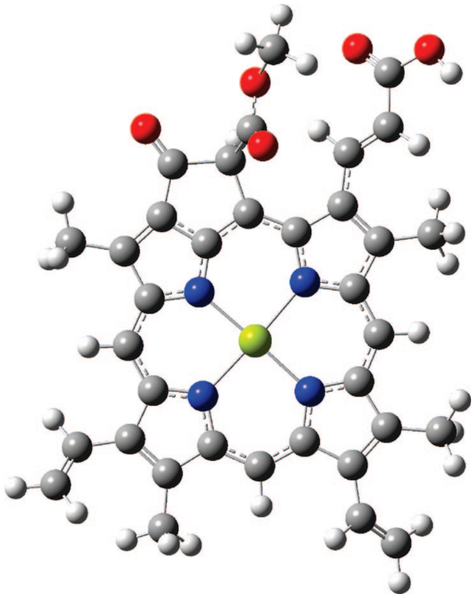


Figure 3. Structure of chlorophyll c2.

4.3. Phosphorescence Spectrum of Chlorophyll c2. Despite ongoing experimental and theoretical research, the understanding of the molecular mechanism of light harvesting in photosystem II is not yet satisfactory. Quantum chemical computations of optical properties combined with spectroscopic experiments can undoubtedly contribute to shed further light on this phenomenon.[59] The triplet states of chlorophylls are of particular interest due to their dual photodamage and photoprotective role in photosystem II.[60] Chlorophyll c2 (see Figure 3) is a large molecule with 73 atoms and 213 normal modes. Its $T_1 \rightarrow S_0$ phosphorescence spectrum has been chosen to demonstrate various aspects related to spectrum convergence and applicability of the integrated approach. More detailed studies of spectra of photosynthetic pigments are planned in future works. The

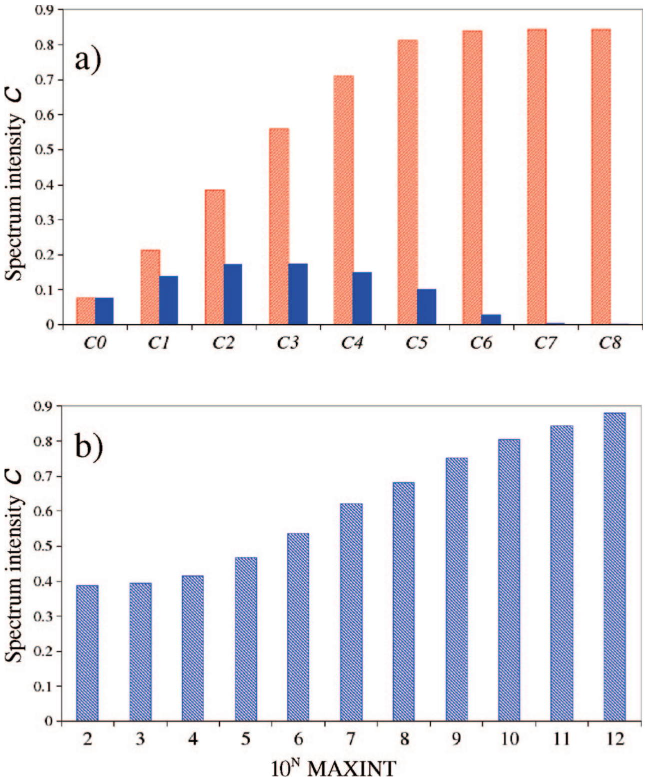


Figure 4. Convergence of the spectrum calculation for chlorophyll c2 with classes (upper panel) and with the maximum number of integrals MAXINT set for each class (lower panel). For spectrum convergence with classes the total convergence up to class C_n is shown as red blocks, while contribution of class C_n is shown in blue.

main factor which rules the accuracy of the calculations is the maximum number of integrals which are computed. Higher number of allowed transitions obviously improves the spectrum quality but at the same time strongly increases the required computational times and memory usage. For larger systems it becomes crucial to select only the most important transitions. In the present approach transitions are partitioned in classes, and the pool of C_1 and C_2 transitions is used to estimate a priori the maximum quantum number that has to be considered for each mode (see section 2.2) in order to keep the number of integrals computed for each class within the limit determined by MAXINT. This procedure has been proved to be very cost-effective.[25,26,37] Panel a of Figure 4 shows the spectrum convergence with classes calculated with MAXINT set to 10^{11}, while the spectrum convergence with increase of MAXINT (from 10^2 to 10^{12}) is depicted on panel b. As expected, a very small number of integrals is not sufficient and leads only to about half of the spectrum intensity. Moreover, even as many as 10^{12} integrals cannot provide the full convergence of spectrum intensity (C). On the other hand it is apparent that the contribution of classes higher than C_5 decreases steeply and that the difference in spectrum intensity calculated up to C_7 and up to C_6 is smaller than 1%, confirming spectrum convergence with respect to classes.

Nevertheless, in most cases the convergence of the spectrum line shape is much faster[25] than the full convergence of the total spectrum intensity. This fact is particularly encouraging for large systems like chlorophyll c2. Figure 5 compares spectrum line shapes calculated with MAXINT set to 10^2, 10^6, and 10^9. It is clear that the main spectral features are well reproduced even if total spectrum intensity is far from convergence. The spectra calculated with MAXINT = 10^9 or larger are identical on this scale. Thus, inspection of the spectrum line shape indicates that the most important transitions have been taken into account, and that reliable spectra have been computed already with $N_I^{\max}$ set to 10^9. The analysis of classes contributions to the total spectrum (Figure 6) shows that most of the spectrum bands are composed from classes up to C_4, with C_1 and C_2 influencing most the spectrum line shape. Contributions of the classes related to the simultaneous excitation of five and more modes are much flatter and of little importance for the spectrum line shape, although they are not negligible for the spectrum intensity. The present case of chlorophyll turns out to be much more challenging than that adopted as a benchmark by Dierksen et al.[24] and Jankowiak et al.,[27] a very large polycyclic aromatic hydrocarbon (PAH) derivative with 462 normal modes. Such PAH has a rather narrow photoelectron spectrum, and our method is able to converge it up to values

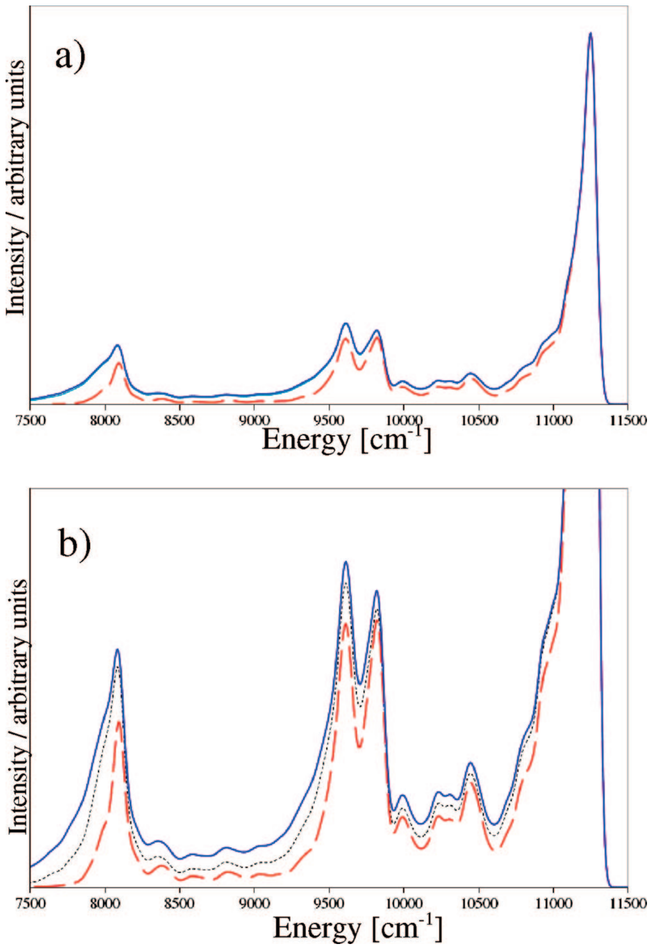


Figure 5. Convergence of the spectrum calculation for chlorophyll c2 with the threshold on the number of computed integrals. Comparison of spectrum shape calculated with MAXINT set to 10^2 (dashed red line) and 10^9 (solid blue line) is shown on upper panel, while the onset with spectra calculated with MAXINT set to 10^2, 10^6 (fine-dashed black line), and 10^9 is shown on lower panel.

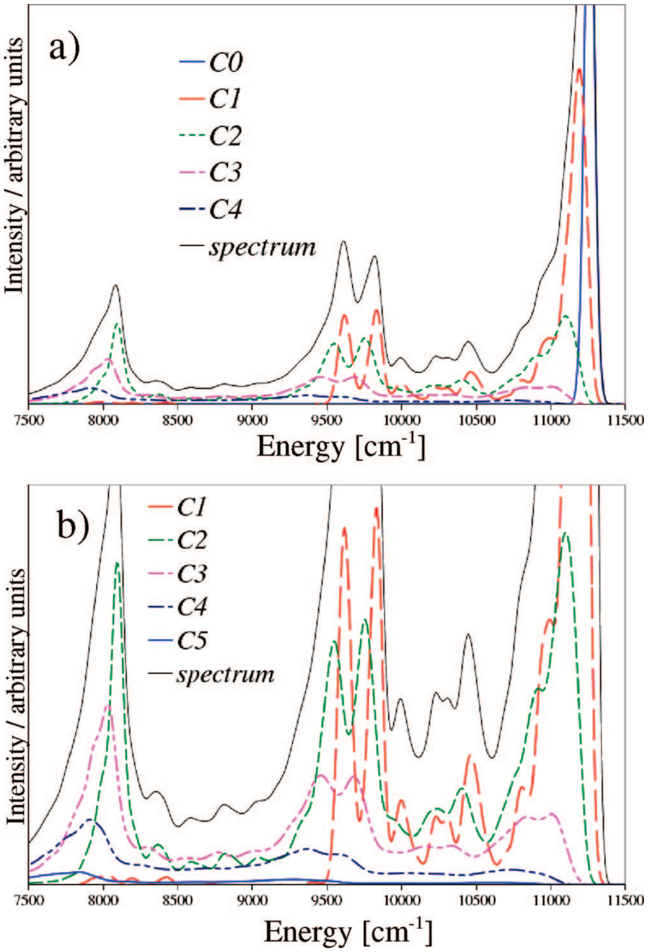


Figure 6. Convergence of the spectrum calculation for chlorophyll c2 with classes; contributions of specific classes are compared with total spectrum (see legend). Classes C_0–C_4 are shown on upper panel, while the onset with contribution of classes C_1–C_5 is shown on lower panel. Contribution of higher classes are not visible in this scale.

of C larger than 0.99. These tests show that our methodology can satisfactorily compute converged spectra also for large challenging systems. When the interest is focused on the high-energy wing of the spectrum (the one suffering of the largest relative error) as for instance for computation of nonradiative transition rates, a careful check of convergence in that energy region must be performed and purposely tailored methods may result more suitable.

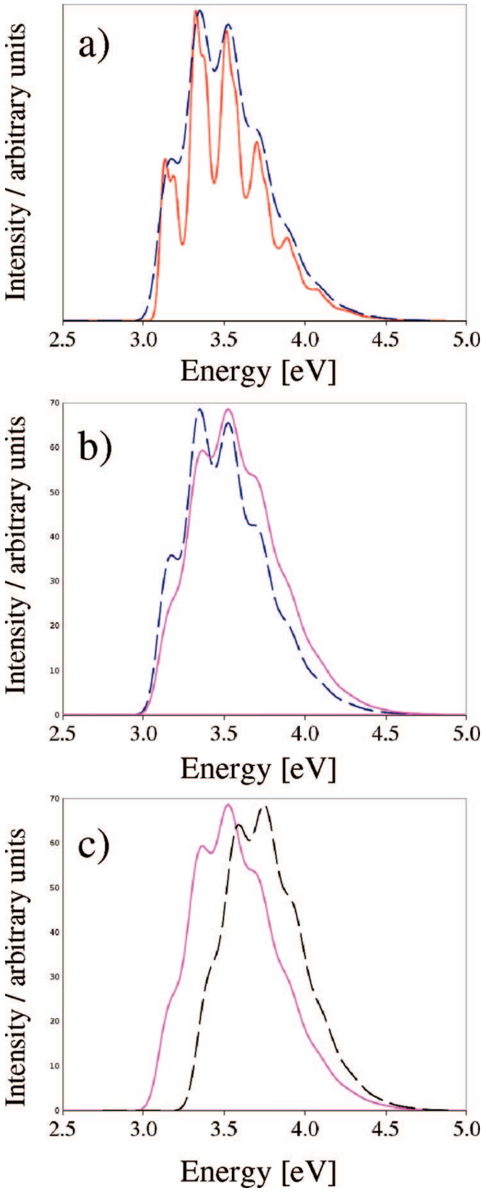


Figure 7. Theoretical absorption UV spectra of $n \rightarrow \pi^*$ electronic transition of acrolein. (a) Gas phase spectrum in a range from 2.5 to 5 eV calculated within FC approximation with fwhm of 500 (red line) and 1000 cm^{-1} (blue dashed line). (b) Comparison between gas-phase spectra calculated within either the FC (blue dashed line) or the FC-HT approximations (pink line). (c) Comparison between calculated spectra for acrolein in gas phase (pink line) and in water solution described by the CPCM model (black dashed line).

4.4. UV Spectrum of Acrolein in the Gas Phase and Aqueous Solution. The UV absorption spectrum of acrolein has attracted significant attention since this molecule exhibits two conjugated chromophores C=C and C=O, a common feature for many natural systems. In particular, a blueshift of the $n \rightarrow \pi^*$ transition of the C=O group has been observed in going from gas phase to aqueous solution.[61] Our integrated approach allows a straightforward computation of the gas phase and aqueous solution absorption spectra of acrolein, giving direct insights into the experimentally observed effect. The structures and frequencies of acrolein have been determined by DFT/TD-DFT computations with the B3LYP functional and N07D polarized double-ζ basis set, both in gas phase and in aqueous solution. The effect of water solvent has been included by means of the polarizable continuum model, where the solvent is represented by a homogeneous dielectric polarized by the solute, placed within a cavity built as an envelope of spheres centered on solute atoms.[46] The solvent is described in the nonequilibrium limit where only its fast (electronic) degrees of freedom are equilibrated with the excited-state charge density while the slow (nuclear) degrees of freedom remain equilibrated with the ground state. This assumption is sufficient to describe absorption spectrum in solution, because of the different time scales of the electronic and nuclear response components of the solvent reaction field.[25] To simulate the spectrum line-shape it is necessary to convolute the stick-spectrum with a Gaussian

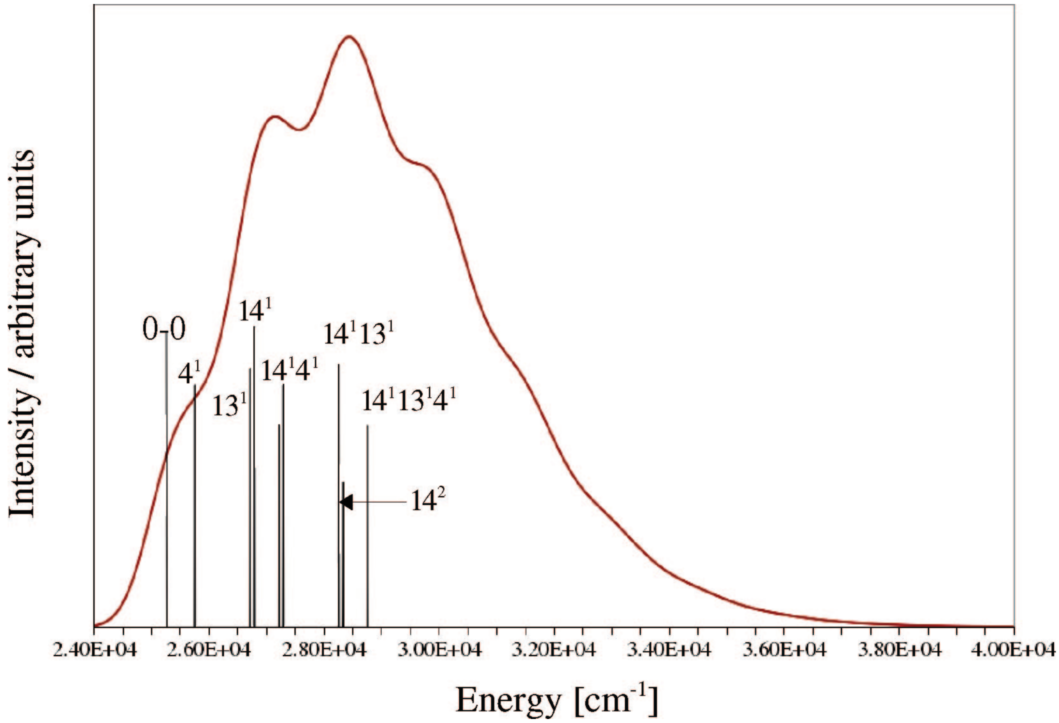


Figure 8. Assignment of the main bands of the theoretical absorption UV spectrum of $n \rightarrow \pi^*$ electronic transition of acrolein in gas phase. Solid line reports the spectrum in a range of 24000−40000 cm^{-1} calculated within FC-HT approximation with fwhm of 1000 cm^{-1}. The main stick bands are assigned as n^x where n is the excited normal mode and x its quantum number.

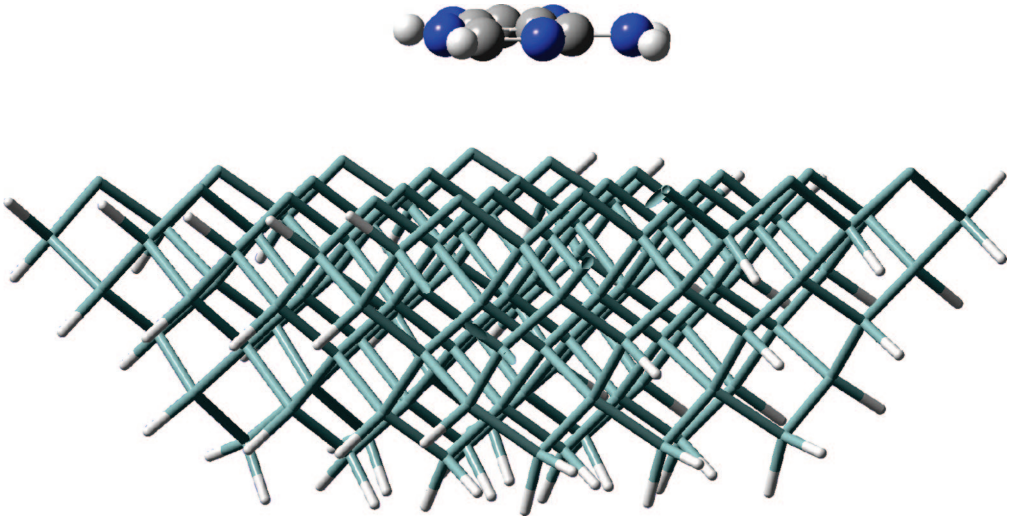


Figure 9. Adenine adsorbed on a cluster of 119 silicon atoms, modeling the Si(100) surface.

with an appropriate full-width at the half-maximum (fwhm): panel a of Figure 7 compares spectra calculated with the values of the fwhm set to 500 and 1000 cm^{-1}. For acrolein the latter choice better reproduces the broad structure of the experimental spectrum. In the present approach it is possible to improve the FC spectrum by considering changes of the transition dipole moment with the geometry. It is worth mentioning that in the present case inclusion of the HT term does not require any additional quantum mechanical computation, since the TD-DFT frequencies are calculated numerically giving direct access to the necessary derivatives of the transition dipole moment with respect to the normal coordinates of the excited electronic state. Inclusion of the HT term is particularly important for dipole forbidden or weakly allowed transitions where the FC approximation is unreliable. This is the case of the weakly dipole allowed $n \rightarrow \pi^*$ transition of acrolein ($\mu = 0.0463$ au), where the HT contribution indeed influences significantly the spectrum lineshape, as shown by the comparison of the FC and FC-HT spectra on panel b of Figure 7. Both the FC and the FC-HT spectra are fully converged (100.0%) to their respective limits (see section 2.3). The FC-HT spectra calculated in the gas phase and in the aqueous solution are compared in panel c of Figure 7. It is evident that not only the solvent shift is well reproduced by the theory but that also changes in the band shapes agree with the recent results obtained with a

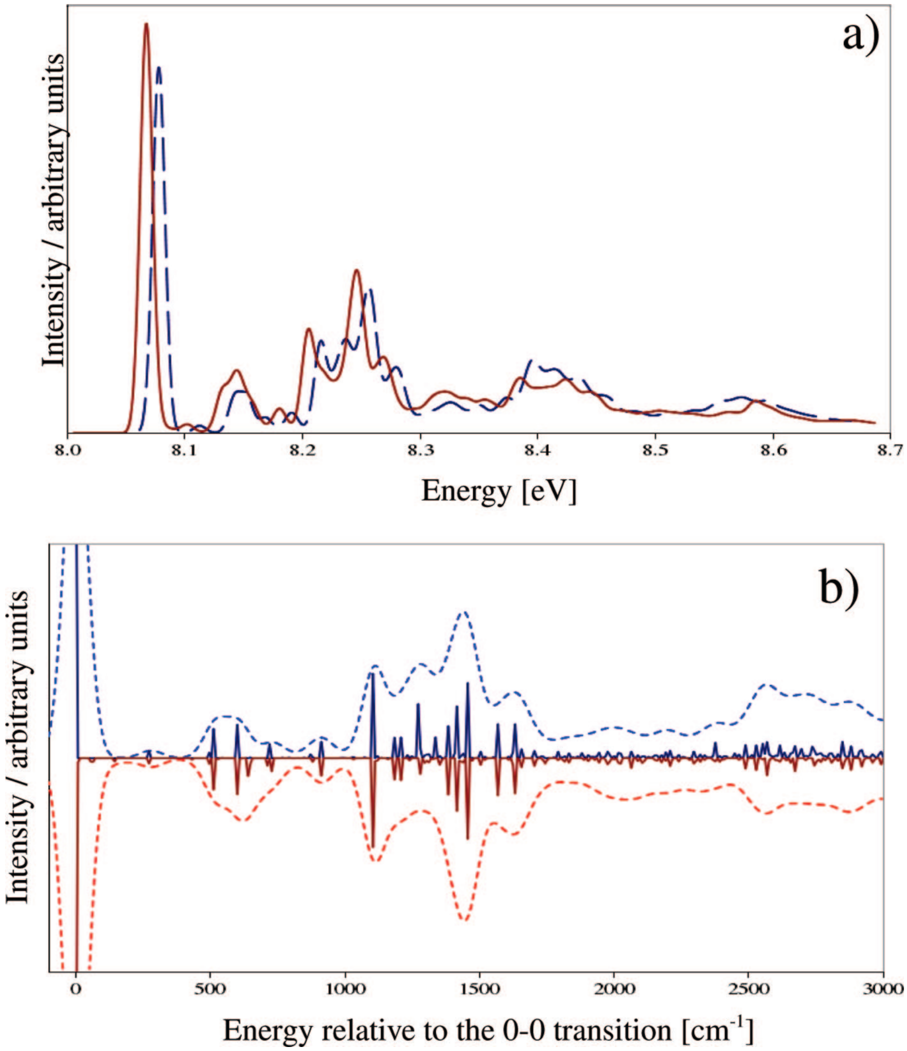


Figure 10. Comparison between the theoretical FC photoionization spectra in gas phase of isolated adenine (blue dashed line) and adenine adsorbed on a Si(100) surface (solid red line): (a) Spectra in an absolute energy range from 8.0 to 8.7 eV calculated within FC approximation with fwhm = 100 cm^{-1}. (b) Spectra shifted to the relative origins of the 0−0 electronic transitions, isolated molecule (upper panel), and adenine@Si(100) (lower panel); the stick bands show the most important transitions.

more accurate but computationally demanding time-domain approach.[61] Moreover, the present method works directly in the frequency domain, so that all the individual vibronic contributions to the total spectrum are computed and can be easily assigned, as shown in Figure 8. We foresee that the accessibility to an easy and straightforward method for the computations of vibrationally resolved spectra within the integrated approach here described, may lead to breakthroughs in the studies of UV−vis spectra in condensed phases.

4.5. Photoelectron Spectrum of Adenine Adsorbed on Si(100). Reliable computational studies of optical properties for large nanosystems in condensed phases can support the design of new materials relevant for optics, photonics, and sensoristics. Our approach is in line with such a demand as illustrated by simulation of the photoelectron spectrum of adenine adsorbed on Si(100) surface. The full valence photoelectron spectrum of adenine is composed from several overlapping excitations.[62] The present work is aimed to show feasibility of spectra simulations for nanosystems; thus only ionization from the highest occupied molecular orbital (HOMO) has been considered. The Si(100) surface has been modeled by a cluster of 119 silicon atoms, shown in Figure 9, resulting in a total system with 636 normal modes. For computation of geometry structures and frequencies the ONIOM QM/MM scheme has been adopted. The photoelectron spectra have been calculated for both isolated adenine molecule and adenine@Si(100), putting into evidence spectrum changes upon adsorption. Both spectra are plotted in the range of 8.0−8.7 eV roughly corresponding to the first band of valence shell photoelectron spectrum. Figure 10 shows the spectra both in the absolute energy scale and in a relative scale where the 0−0 transition is set to zero. It can be seen that adsorption on Si surface yields a small red shift of the excitation origin, while new vibronic transitions corresponding to intermolecular vibrations modulate the spectrum line shape. It is interesting to analyze the number of combinations for each class C_n for such a large system, which is directly related (see section 2.5) to the number of transitions to compute, and investigate the efficiency of the adopted selection procedure. Table 1 lists ${}_{N}C_n$ for isolated adenine and adenine@Si(100) and the spectrum intensity achieved with MAXINT set to the default value 10^8. It is noteworthy that in both cases, either an isolated molecule with 39 normal modes or a macrosystem with over 600, almost all spectrum intensity (about 98%) has been recovered at an equivalent computational cost. For the cluster, the default value of MAXINT is not sufficient to consider the whole initial pool even for only three simultaneously excited modes (C_3 class). This particular case shows the ability of the a priori strategy to select only the relevant transitions and discard the less probable ones.

Conclusions

A general approach for simulation of absorption and emission vibronic spectra has been implemented and applied to a variety of molecular systems, showing the high flexibility of the developed computational tool. The integration of all procedures within the same computational package allows for the fully automatic computation of vibrationally resolved optical spectra. Despite the fact that our computational scheme has been tailored for large systems, it can be utilized as well to generate high quality spectra for small systems, when nonadiabatic and anharmonic couplings are negligible, since it does not imply unnecessary approximations. Moreover, good quality spectra can be effectively computed even for large systems with hundreds of normal modes, paving the route to spectroscopic studies of systems of direct biological and/or technological interest.

It has been demonstrated that a very good agreement between the computed and experimental vibrationally resolved REMPI spectra can be achieved, when good accuracy geometries and force fields are determined for the ground and excited electronic states, for example, for the case of anisole. It is worthwhile highlighting that the remarkable overall agreement, also as far as band positions are concerned, has been achieved only through the correction of the frequencies for anharmonicity. The accuracy and effectiveness of our a priori procedure for the selection of the relevant transitions has been shown by comparison of the photoelectron spectra computed for isolated adenine and for adenine adsorbed on silicon surface. Despite significant difference in the systems size, in both cases the electronic transition is localized on adenine molecule. The implemented scheme indeed effectively chooses the relevant transitions, providing an equivalent computational cost for spectra computations for isolated molecule and adenine@Si(100). Spectrum quality has been discussed on the example of chlorophyll c2 showing that line shape converges much faster than spectrum intensity. The former, which provides all necessary information for most important transitions, is usually the actual interesting property. Thus, reliable computations for large systems can be performed with relatively small computational cost. Overall, we hope that the tool here presented to compute vibrationally resolved optical spectra, together with its integration into a computational chemistry package, will routinely allow detailed analyses of UV−vis spectra, improving their interpretation and understanding.

Acknowledgment. This work was supported by Italian MIUR. The large scale computer facilities of the VILLAGE network (http://village.unina.it) are kindly acknowledged. CINECA and the Wroclaw Centre for Networking and Supercomputing are acknowledged for access to the MOLPRO code and related computer time.

References

(1) Barone, V.; Polimeno, A. *Chem. Soc. Rev.* **2007**, *36*, 1724.

(2) Barone, V.; Improta, R.; Rega, N. *Acc. Chem. Res.* **2008**, *41*, 605.

(3) Clary, D. C. *Science* **2006**, *314*, 5797.

(4) Hellman, A.; Baerends, E. J.; Biczysko, M.; Bliggard, T.; Christensen, C. H.; Clary, D. C.; Dahl, S.; van Harrevelt, R.; Honkala, K.; Jonsson, H.; Kroes, G. J.; Luppi, M.; Manthe, U.; Norskov, J. K.; Olsen, R. A.; Rossmeisl, J.; Skulason, E.; Tautermann, C. S.; Varandas, A. J. C.; Vincent, J. K. *J. Phys. Chem. B* **2006**, *110*, 17719.

(5) Jacquemin, D.; Preat, J.; Wathelet, V.; Fontaine, M.; Perpete, E. *J. Am. Chem. Soc.* **2006**, *128*, 2072.

(6) Altoe, P.; Bernardi, F.; Garavelli, M.; Orlandi, G.; Negri, F. *J. Am. Chem. Soc.* **2005**, *127*, 3952.

(7) *Computational Molecular Spectroscopy*; Jensen, P., Bunker, P. R., Eds.; John Wiley & Sons: London, 2000.

(8) Lami, A.; Petrongolo, C.; Santoro, F. In *Conical Intersections, Electronic Structure, Dynamics and Spectroscopy*; Domcke, W., Yarkony, R., Koeppel, H., Eds.; World Scientific Publishing Co.: Singapore, 2004; pp 699−739.

(9) Grimme, S. *Rev. Comput. Chem.* **2004**, *20*, 153.

(10) Christiansen, O. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2942.

(11) Bowman, J.; Carter, S.; Huang, X. *Int. Rev. Phys. Chem.* **2003**, *22*, 533.

(12) Vaara, J. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5399.

(13) Brancato, G.; Barone, V.; Rega, N. *Theor. Chem. Acc.* **2007**, *117*, 1001.

(14) Barone, V.; Brustolon, M.; Cimino, P.; Polimeno, A.; Zerbetto, M.; Zoleo, A. *J. Am. Chem. Soc.* **2006**, *128*, 15865.

(15) Rega, N. *Theor. Chem. Acc.* **2006**, *116*, 344.

(16) Franck, J. *Trans. Faraday Soc.* **1926**, *21*, 536.

(17) Condon, E. *Phys. Rev.* **1926**, *28*, 1182.

(18) Condon, E. *Phys. Rev.* **1928**, *32*, 858.

(19) Scalmani, G.; Frish, M. J.; Menucci, B.; Tomasi, J.; Cammi, R.; Barone, V. *J. Chem. Phys.* **2006**, *124*, 094107.

(20) Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2004**, *121*, 12772.

(21) Köhn, A.; Hättig, C. *J. Chem. Phys.* **2003**, *119*, 5021.

(22) Kemper, M.; Van Dijk, J.; Buck, H. *Chem. Phys. Lett.* **1978**, *53*, 121.

(23) Berger, R.; Fischer, C.; Klessinger, M. *J. Phys. Chem. A* **1998**, *102*, 7157.

(24) Dierksen, M.; Grimme, S. *J. Chem. Phys.* **2005**, *122*, 244101.

(25) Santoro, F.; Improta, R.; Lami, A.; Bloino, J.; Barone, V. *J. Chem. Phys.* **2007**, *126*, 084509.

(26) Santoro, F.; Improta, R.; Lami, A.; Bloino, J.; Barone, V. *J. Chem. Phys.* **2007**, *126*, 184102.

(27) Jankowiak, H.-C.; Stuber, J. L.; Berger, R. *J. Chem. Phys.* **2007**, *127*, 234101.

(28) Eckart, C. *Phys. Rev.* **1935**, *47*, 552.

(29) Duschinsky, F. *Acta Physicochim. URSS* **1937**, *7*, 551.

(30) Sharp, T. E.; Rosenstock, H. M. *J. Chem. Phys.* **1964**, *41*, 3453.

(31) Chang, J.-L. *J. Chem. Phys.* **2008**, *128*, 174111.

(32) Manneback, C. *Physica* **1951**, *17*, 1001–1010.

(33) Doktorov, E. V.; Malkin, I. A.; Man'ko, V. I. *J. Mol. Spectrosc.* **1977**, *64*, 302.

(34) Ruhoff, P. T. *Chem. Phys.* **1994**, *186*, 355.

(35) Johnson, P. M.; Xu, H.; Sears, T. J. *J. Chem. Phys.* **2006**, *125*, 164330.

(36) Hazra, A.; Nooijen, M. *Int. J. Quantum Chem.* **2003**, *95*, 643.

(37) Santoro, F.; Improta, R.; Lami, A.; Bloino, J.; Barone, V. *J. Chem. Phys.* **2008**, *128*, 224311.

(38) Bloino, J.; Biczysko, M.; Crescenzi, O.; Barone, V. *J. Chem. Phys.* **2008**, *128*, 244105.

(39) Barone, V. *J. Chem. Phys.* **2005**, *122*, 014108.

(40) Becke, D. *J. Chem. Phys.* **1993**, *98*, 5648.

(41) Moeller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(42) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(43) Kendall, R.; Dunning Jr., T.; Harrison, R. *J. Chem. Phys.* **1992**, *96*, 6769.

(44) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.

(45) Barone, V.; Cimino, P. *Chem. Phys. Lett.* **2008**, *454*, 139.

(46) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.

(47) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419.

(48) Rappè, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

(49) Hampel, C.; Peterson, K.; Werner, H.-J. *Chem. Phys. Lett.* **1992**, *190*, 1.

(50) Deegan, M. J. O.; Knowles, P. J. *Chem. Phys. Lett.* **1994**, *227*, 321.

(51) Korona, T.; Werner, H.-J. *J. Chem. Phys.* **2003**, *118*, 300.

(52) 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; http://www.molpro.net.

(53) 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 Development Version*, revision G.01; Gaussian, Inc.: Wallingford, CT, 2006.

(54) Eisenhardt, C. G.; Pietraperzia, G.; Becucci, M. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1407.

(55) Hoffmann, L. J. H.; Marquardt, S.; Gemechu, A. S.; Baumgärtel, H. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2360.

(56) Bopp, J. C.; Roscioli, J. R.; Johnson, M. A.; Miller, T. M.; Viggiano, A. A.; Villano, S. M.; Wren, S. W.; Lineberger, W. C. *J. Phys. Chem. A* **2007**, *111*, 1214.

(57) Borrelli, R. *Chem. Phys. Lett.* **2007**, *445*, 84.

(58) Grimsrud, E. P.; Chowdhury, S.; Kebarle, P. *J. Chem. Phys.* **1985**, *83*, 1059.

(59) Vassiliev, S.; Bruce, D. *Photosynth. Res.* **2008**, *97*, 75.

(60) Noguchi, T. *Plant Cell Physiol.* **2002**, *43*, 1112.

(61) Brancato, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2006**, *125*, 164515.

(62) Trofimov, A. B.; Schirmer, J.; Kobychev, V. B.; Potts, A. W.; Holland, D. M. P.; Karlsson, L. *J. Phys. B* **2006**, *39*, 305.

CT8004744

J. Chem. Theory Comput. **2009,** *5,* 555–564

Journal of Chemical Theory and Computation

Benchmarks for Electronically Excited States: A Comparison of Noniterative and Iterative Triples Corrections in Linear Response Coupled Cluster Methods: CCSDR(3) versus CC3

Stephan P. A. Sauer,*,† Marko Schreiber,‡ Mario R. Silva-Junior,‡ and Walter Thiel‡

Department of Chemistry, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark, and Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim an der Ruhr, Germany

Received June 28, 2008

Abstract: CCSDR(3) calculations of vertical excitation energies are reported for a set of 24 molecules and 121 excited valence singlet states from a recently published benchmark of organic molecules. The same geometries (MP2/6−31G*) and basis set (TZVP) were employed as in our previous linear response CC2, CCSD, and CC3 calculations. The CCSDR(3) results are compared against the CCSD and CC3 results. Statistical evaluation of all CCSDR(3) excitation energies gives mean absolute deviations of 0.09 eV from CC3 and 0.30 eV from CCSD. For excited states, which are dominated by single excitations, the absolute mean deviation from CC3 is reduced to 0.02 eV and the maximum deviation is 0.09 eV. CCSDR(3) is thus a very cost-effective accurate alternative to CC3.

1. Introduction

In two recent studies[1,2] we have presented a benchmark set for the calculation of electronically excited states. This set comprises unsaturated aliphatic hydrocarbons (including polyenes and cyclic compounds), aromatic hydrocarbons and heterocycles, carbonyl compounds, and nucleobases. It consists of 28 medium size organic molecules with a total of 223 excited states (152 singlet and 71 triplet states) and is intended to cover the most important chromophores in organic photochemistry.

In the first study[1] calculations were performed with a series of linear response coupled cluster methods (CC2, CCSD, CC3)[3−15] and with multistate complete-active-space second-order perturbation theory (MS-CASPT2).[16−18] Based on these results and other high-level literature data, best theoretical estimates were chosen for the majority of the studied vertical excitation energies. The comparison of coupled cluster and multireference results showed that CC3 and CASPT2 excitation energies are in excellent agreement for states which are dominated by single excitations and that CC2 performs on average better than CCSD.

In the second study[2] we have investigated the performance of time-dependent density functional theory (TD-DFT) with three functionals (BP86, B3LYP, and BHLYP) and DFT-based multireference configuration interaction (DFT/MRCI) methods.

Møller−Plesset perturbation theory[19−29] and coupled cluster theory-based[3−15,26−30] response theory methods such as CC3[13−15] are suitable for states with low double excitation contributions. However, CC3 formally scales as N^7 with the number of orbitals N, and the high computational cost in the iterative treatment of the triple excitations in CC3 restricts its application to small systems and/or small basis sets. Looking for a computationally cheaper but comparably accurate approach, Christiansen and co-workers[31,32] have presented the CCSDR(3) method, in which a noniterative triples correction is added to the linear response CCSD excitation energy. In this respect, CCSDR(3) is analogous to the CCSD(T) method[33] which is so successful for ground-state energies. Both methods include fourth-order terms. In CCSDR(3) the reference singles and doubles amplitudes as well as the energies of single excitation dominated states

---

* Corresponding author e-mail: sauer@kiku.dk.

† University of Copenhagen.

‡ Max-Planck-Institut für Kohlenforschung.

10.1021/ct800256j CCC: $40.75

Table 1. Vertical Singlet Excitation Energies Δ*E* (eV)

molecule	state	CCSD (% R_1)^a ref 1	CCSDR(3) (Δ)^b this work	CC3 (% R_1)^a ref 1
ethene	1^1B_{1u} ($\pi\rightarrow\pi^*$)	8.51 (97.2)	8.36 (−0.01)	8.37 (96.9)
E-butadiene	1^1B_u ($\pi\rightarrow\pi^*$)	6.72 (95.5)	6.56 (−0.02)	6.58 (93.7)
	2^1A_g ($\pi\rightarrow\pi^*$)	7.42 (85.8)	6.95 (0.18)	6.77 (72.8)
all-*E*-hexatriene	1^1B_u ($\pi\rightarrow\pi^*$)	5.72 (95.0)	5.56 (−0.02)	5.58 (92.6)
	2^1A_g ($\pi\rightarrow\pi^*$)	6.61 (84.7)	6.04 (0.32)	5.72 (65.8)
all-*E*-octatetraene	2^1A_g ($\pi\rightarrow\pi^*$)	5.99 (85.4)	5.40 (0.43)	4.97 (62.9)
	1^1B_u ($\pi\rightarrow\pi^*$)	5.07 (94.7)	4.93 (−0.01)	4.94 (91.9)
	2^1B_u ($\pi\rightarrow\pi^*$)	6.89 (81.3)	6.91 (0.30)	6.06 (58.5)
	3^1A_g ($\pi\rightarrow\pi^*$)	6.98 (92.0)	6.72 (0.23)	6.50 (71.6)
	4^1A_g ($\pi\rightarrow\pi^*$)	7.05 (89.9)	7.01 (−0.05)	6.81 (92.1)
	3^1B_u ($\pi\rightarrow\pi^*$)	8.15 (94.6)	7.95 (0.03)	7.91 (91.9)
cyclopropene	1^1B_1 ($\sigma\rightarrow\pi^*$)	6.96 (94.5)	6.89 (−0.01)	6.90 (93.0)
	1^1B_2 ($\pi\rightarrow\pi^*$)	7.24 (96.3)	7.10 (0.00)	7.10 (95.5)
cyclopentadiene	1^1B_2 ($\pi\rightarrow\pi^*$)	5.87 (95.7)	5.72 (−0.01)	5.73 (94.3)
	2^1A_1 ($\pi\rightarrow\pi^*$)	7.05 (89.4)	6.76 (0.14)	6.61 (79.3)
	3^1A_1 ($\pi\rightarrow\pi^*$)	8.95 (95.8)	8.72 (0.02)	8.69 (93.1)
norbornadiene	1^1A_2 ($\pi\rightarrow\pi^*$)	5.80 (95.3)	5.65 (0.01)	5.64 (93.4)
	1^1B_2 ($\pi\rightarrow\pi^*$)	6.69 (94.5)	6.51 (0.02)	6.49 (91.9)
	2^1B_2 ($\pi\rightarrow\pi^*$)	7.87 (95.5)	7.65 (0.01)	7.64 (93.8)
	2^1A_2 ($\pi\rightarrow\pi^*$)	7.87 (95.0)	7.73 (0.02)	7.71 (93.0)
benzene	1^1B_{2u} ($\pi\rightarrow\pi^*$)	5.19 (90.5)	5.12 (0.04)	5.07 (85.8)
	1^1B_{1u} ($\pi\rightarrow\pi^*$)	6.74 (95.6)	6.70 (0.02)	6.68 (93.6)
	1^1E_{1u} ($\pi\rightarrow\pi^*$)	7.65 (94.5)	7.45 (0.00)	7.45 (92.2)
	2^1E_{2g} ($\pi\rightarrow\pi^*$)	9.21 (84.9)	8.71 (0.29)	8.43 (65.6)
naphthalene	1^1B_{3u} ($\pi\rightarrow\pi^*$)	4.41 (90.5)	4.34 (0.07)	4.27 (85.2)
	1^1B_{2u} ($\pi\rightarrow\pi^*$)	5.21 (94.3)	5.08 (0.05)	5.03 (90.6)
	2^1A_g ($\pi\rightarrow\pi^*$)	6.23 (90.1)	6.09 (0.11)	5.98 (82.2)
	1^1B_{1g} ($\pi\rightarrow\pi^*$)	6.53 (91.4)	6.26 (0.20)	6.07 (79.6)
	2^1B_{3u} ($\pi\rightarrow\pi^*$)	6.55 (93.9)	6.35 (0.02)	6.33 (90.7)
	2^1B_{1g} ($\pi\rightarrow\pi^*$)	6.97 (93.8)	6.81 (0.02)	6.79 (91.3)
	2^1B_{2u} ($\pi\rightarrow\pi^*$)	6.77 (93.8)	6.60 (0.03)	6.57 (90.5)
	3^1A_g ($\pi\rightarrow\pi^*$)	7.77 (88.4)	7.29 (0.39)	6.90 (70.0)
	3^1B_{2u} ($\pi\rightarrow\pi^*$)	8.77 (93.5)	8.53 (0.08)	8.44 (87.9)
	3^1B_{3u} ($\pi\rightarrow\pi^*$)	9.03 (84.1)	8.50 (0.38)	8.12 (58.7)
furan	1^1B_2 ($\pi\rightarrow\pi^*$)	6.80 (94.9)	6.64 (0.04)	6.60 (92.9)
	2^1A_1 ($\pi\rightarrow\pi^*$)	6.89 (90.8)	6.71 (0.09)	6.62 (84.9)
	3^1A_1 ($\pi\rightarrow\pi^*$)	8.83 (94.2)	8.57 (0.04)	8.53 (90.7)
pyrrole	2^1A_1 ($\pi\rightarrow\pi^*$)	6.61 (91.2)	6.47 (0.07)	6.40 (86.0)
	1^1B_2 ($\pi\rightarrow\pi^*$)	6.87 (94.2)	6.74 (0.03)	6.71 (91.6)
	3^1A_1 ($\pi\rightarrow\pi^*$)	8.44 (93.7)	8.20 (0.04)	8.17 (90.2)
imidazole	$1^1A''$ ($n\rightarrow\pi^*$)	7.01 (92.4)	6.87 (0.05)	6.82 (87.6)
	$2^1A'$ ($\pi\rightarrow\pi^*$)	6.80 (92.0)	6.64 (0.06)	6.58 (87.2)
	$3^1A'$ ($\pi\rightarrow\pi^*$)	7.27 (93.1)	7.15 (0.05)	7.10 (89.8)
	$2^1A''$ ($n\rightarrow\pi^*$)	8.15 (93.3)	7.98 (0.05)	7.93 (89.4)
	$4^1A'$ ($\pi\rightarrow\pi^*$)	8.70 (92.7)	8.49 (0.04)	8.45 (88.8)
pyridine	1^1B_2 ($\pi\rightarrow\pi^*$)	5.27 (90.6)	5.20 (0.05)	5.15 (85.9)
	1^1B_1 ($n\rightarrow\pi^*$)	5.25 (92.8)	5.12 (0.07)	5.05 (88.1)
	1^1A_2 ($n\rightarrow\pi^*$)	5.73 (92.4)	5.55 (0.05)	5.50 (87.7)
	2^1A_1 ($\pi\rightarrow\pi^*$)	6.94 (95.3)	6.88 (0.03)	6.85 (92.8)
	3^1A_1 ($\pi\rightarrow\pi^*$)	7.94 (94.2)	7.72 (0.01)	7.70 (91.5)
	2^1B_2 ($\pi\rightarrow\pi^*$)	7.81 (93.5)	7.61 (0.02)	7.59 (89.7)
	4^1A_1 ($\pi\rightarrow\pi^*$)	9.45 (89.5)	9.00 (0.33)	8.68 (74.1)
	3^1B_2 ($\pi\rightarrow\pi^*$)	9.64 (84.4)	9.09 (0.32)	8.77 (65.2)
pyrazine	1^1B_{3u} ($n\rightarrow\pi^*$)	4.42 (93.4)	4.31 (0.06)	4.24 (89.9)
	1^1A_u ($n\rightarrow\pi^*$)	5.29 (92.7)	5.11 (0.06)	5.05 (88.4)
	1^1B_{2u} ($\pi\rightarrow\pi^*$)	5.14 (90.8)	5.07 (0.05)	5.02 (86.2)
	1^1B_{2g} ($n\rightarrow\pi^*$)	6.02 (92.1)	5.86 (0.12)	5.74 (85.0)
	1^1B_{1g} ($n\rightarrow\pi^*$)	7.13 (90.8)	6.86 (0.11)	6.75 (83.8)
	1^1B_{1u} ($\pi\rightarrow\pi^*$)	7.18 (95.6)	7.10 (0.03)	7.07 (93.3)
	2^1B_{1u} ($\pi\rightarrow\pi^*$)	8.34 (93.9)	8.09 (0.03)	8.06 (90.9)
	2^1B_{2u} ($\pi\rightarrow\pi^*$)	8.29 (93.2)	8.08 (0.03)	8.05 (89.7)
	1^1B_{3g} ($\pi\rightarrow\pi^*$)	9.75 (83.5)	9.16 (0.39)	8.77 (61.1)
	2^1A_g ($\pi\rightarrow\pi^*$)	9.55 (89.1)	9.04 (0.35)	8.69 (74.2)
pyrimidine	1^1B_1 ($n\rightarrow\pi^*$)	4.70 (92.7)	4.56 (0.06)	4.50 (88.4)
	1^1A_2 ($n\rightarrow\pi^*$)	5.12 (92.6)	4.97 (0.05)	4.93 (88.2)
	1^1B_2 ($\pi\rightarrow\pi^*$)	5.49 (90.5)	5.42 (0.05)	5.36 (85.7)
	2^1A_1 ($\pi\rightarrow\pi^*$)	7.17 (94.8)	7.10 (0.04)	7.06 (92.2)
	2^1B_2 ($\pi\rightarrow\pi^*$)	8.24 (93.8)	8.02 (0.02)	8.01 (90.7)
	3^1A_1 ($\pi\rightarrow\pi^*$)	7.97 (93.5)	7.77 (0.03)	7.74 (89.7)
pyridazine	1^1B_1 ($n\rightarrow\pi^*$)	4.11 (93.1)	3.99 (0.07)	3.92 (89.0)
	1^1A_2 ($n\rightarrow\pi^*$)	4.76 (92.0)	4.57 (0.08)	4.49 (86.6)
	2^1A_1 ($\pi\rightarrow\pi^*$)	5.35 (90.2)	5.28 (0.06)	5.22 (85.2)
	2^1A_2 ($n\rightarrow\pi^*$)	6.00 (92.1)	5.84 (0.10)	5.74 (86.6)
	2^1B_1 ($n\rightarrow\pi^*$)	6.70 (92.0)	6.49 (0.08)	6.41 (86.6)
	1^1B_2 ($\pi\rightarrow\pi^*$)	7.09 (94.7)	6.99 (0.07)	6.93 (90.7)
	2^1B_2 ($\pi\rightarrow\pi^*$)	7.79 (93.8)	7.58 (0.04)	7.55 (90.2)
	3^1A_1 ($\pi\rightarrow\pi^*$)	8.11 (93.8)	7.86 (0.04)	7.82 (90.5)

Table 1. Continued

molecule	state	CCSD (% R_1)[a] ref 1	CCSDR(3) (Δ)[b] this work	CC3 (% R_1)[a] ref 1
s-triazine	$1^1A_1''$ $(n\rightarrow\pi^*)$	4.96 (92.3)	4.81 (0.04)	4.78 (88.0)
	$1^1A_2''$ $(n\rightarrow\pi^*)$	4.98 (92.5)	4.83 (0.07)	4.76 (88.0)
	$1^1E''$ $(n\rightarrow\pi^*)$	5.01 (92.5)	4.87 (0.05)	4.81 (88.1)
	$1^1A_2'$ $(\pi\rightarrow\pi^*)$	5.84 (90.2)	5.76 (0.06)	5.71 (85.1)
	$2^1A_1'$ $(\pi\rightarrow\pi^*)$	7.51 (93.7)	7.44 (0.03)	7.41 (90.8)
	$2^1E''$ $(n\rightarrow\pi^*)$	8.19 (90.9)	7.95 (0.15)	7.80 (88.1)
	$1^1E'$ $(\pi\rightarrow\pi^*)$	8.28 (93.7)	8.07 (0.02)	8.04 (88.8)
	$2^1E'$ $(\pi\rightarrow\pi^*)$	10.24 (91.2)	9.89 (0.46)	9.44 (74.3)
s-tetrazine	1^1B_{3u} $(n\rightarrow\pi^*)$	2.71 (93.2)	2.61 (0.08)	2.53 (89.6)
	1^1A_u $(\pi\rightarrow\pi^*)$	4.07 (92.2)	3.88 (0.08)	3.79 (87.5)
	1^1B_{1g} $(n\rightarrow\pi^*)$	5.32 (91.7)	5.15 (0.18)	4.97 (82.5)
	1^1B_{2u} $(\pi\rightarrow\pi^*)$	5.27 (90.0)	5.20 (0.08)	5.12 (84.6)
	1^1B_{2g} $(n\rightarrow\pi^*)$	5.70 (90.7)	5.51 (0.17)	5.34 (80.7)
	2^1A_u $(n\rightarrow\pi^*)$	5.70 (92.5)	5.56 (0.10)	5.46 (87.4)
	2^1B_{2g} $(n\rightarrow\pi^*)$	6.76 (90.1)	6.43 (0.20)	6.23 (79.2)
	2^1B_{1g} $(n\rightarrow\pi^*)$	7.25 (91.1)	6.98 (0.11)	6.87 (84.7)
	3^1B_{1g} $(n\rightarrow\pi^*)$	8.36 (86.9)	7.60 (0.52)	7.08 (63.2)
	2^1B_{3u} $(n\rightarrow\pi^*)$	6.99 (93.2)	6.77 (0.10)	6.67 (86.7)
	1^1B_{1u} $(\pi\rightarrow\pi^*)$	7.66 (94.9)	7.54 (0.09)	7.45 (91.0)
	2^1B_{1u} $(\pi\rightarrow\pi^*)$	8.06 (93.4)	7.83 (0.04)	7.79 (90.2)
	2^1B_{2u} $(\pi\rightarrow\pi^*)$	8.88 (93.2)	8.58 (0.07)	8.51 (87.7)
	2^1B_{3g} $(\pi\rightarrow\pi^*)$	9.44 (84.3)	8.86 (0.39)	8.47 (63.6)
formaldehyde	1^1A_2 $(n\rightarrow\pi^*)$	3.97 (93.4)	3.94 (0.00)	3.95 (91.2)
	1^1B_1 $(\sigma\rightarrow\pi^*)$	9.26 (93.4)	9.19 (0.00)	9.18 (90.9)
	2^1A_1 $(\pi\rightarrow\pi^*)$	10.54 (94.4)	10.43 (−0.02)	10.45 (91.3)
acetone	1^1A_2 $(n\rightarrow\pi^*)$	4.43 (93.4)	4.39 (0.00)	4.40 (90.8)
	1^1B_1 $(\sigma\rightarrow\pi^*)$	9.26 (93.8)	9.17 (0.01)	9.17 (91.5)
	2^1A_1 $(\pi\rightarrow\pi^*)$	9.87 (93.5)	9.66 (0.01)	9.65 (90.1)
p-benzoquinone	1^1A_u $(n\rightarrow\pi^*)$	3.19 (91.7)	3.01 (0.16)	2.85 (83.0)
	1^1B_{1g} $(n\rightarrow\pi^*)$	3.07 (92.0)	2.90 (0.15)	2.75 (84.1)
	1^1B_{3g} $(\pi\rightarrow\pi^*)$	4.93 (92.7)	4.69 (0.11)	4.59 (87.9)
	1^1B_{1u} $(\pi\rightarrow\pi^*)$	5.89 (92.5)	5.65 (0.03)	5.62 (88.4)
	1^1B_{3u} $(n\rightarrow\pi^*)$	6.55 (91.0)	6.09 (0.27)	5.82 (75.2)
	2^1B_{3g} $(\pi\rightarrow\pi^*)$	7.62 (91.0)	7.36 (0.08)	7.27 (83.8)
	2^1B_{1u} $(\pi\rightarrow\pi^*)$	8.47 (91.7)	8.10 (0.28)	7.82 (68.6)
formamide	$1^1A''$ $(n\rightarrow\pi^*)$	5.66 (93.6)	5.65 (−0.01)	5.65 (90.7)
	$2^1A'$ $(\pi\rightarrow\pi^*)$	8.52 (92.9)	8.30 (0.03)	8.27 (87.9)
	$3^1A'$ $(\pi\rightarrow\pi^*)$	11.34 (92.7)	11.06 (0.13)	10.93 (86.6)
acetamide	$1^1A''$ $(n\rightarrow\pi^*)$	5.71 (93.5)	5.69 (−0.01)	5.69 (90.6)
	$2^1A'$ $(\pi\rightarrow\pi^*)$	7.85 (92.8)	7.69 (0.02)	7.67 (89.1)
	$3^1A'$ $(\pi\rightarrow\pi^*)$	10.77 (93.0)	10.56 (0.06)	10.50 (88.7)
propanamide	$1^1\,A''$ $(n\rightarrow\pi^*)$	5.74 (93.6)	5.71 (−0.01)	5.72 (90.6)
	$2^1A'$ $(\pi\rightarrow\pi^*)$	7.80 (93.0)	7.64 (0.02)	7.62 (89.2)
	$3^1A'$ $(\pi\rightarrow\pi^*)$	10.34 (93.3)	10.13 (0.06)	10.06 (89.0)

[a] Weight of the single excitations in the coupled cluster calculations. [b] Difference between CCSDR(3) and CC3 results (in parenthesis).

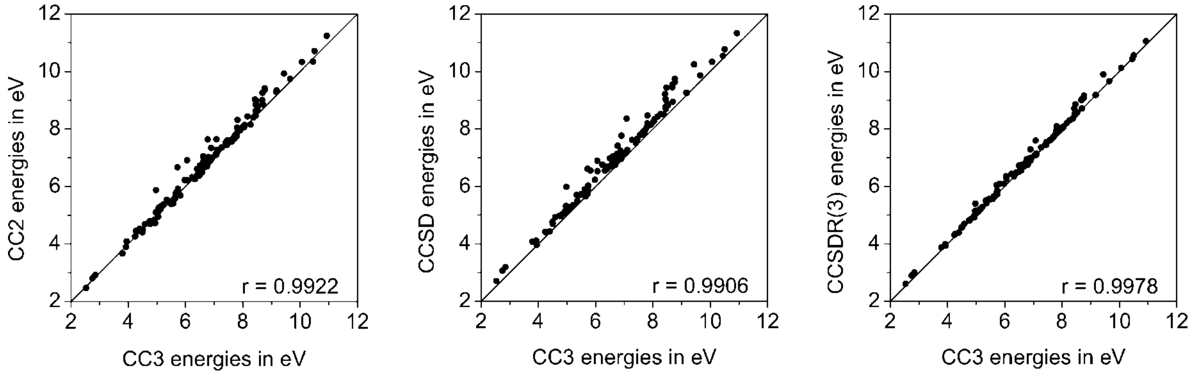


Figure 1. Correlation plots for all calculated singlet excited states: Coupled cluster vertical excitation energies.

are correct to third order whereas the reference triples amplitudes and the energies of double excitation dominated states are correct to second order exactly as in CC3.[31,32] While the computational scaling of CCSDR(3) is of the same order as in CC3, the N^7 step is noniterative in CCSDR(3), which implies a much smaller prefactor and therefore significant computational savings. Formally CCSDR(3) is based on a pseudoperturbation theory expansion of the CC3 eigenvalue problem and therefore bears some similarity to the CIS(D)[34] and RPA(D)[21,23] methods, which are based on

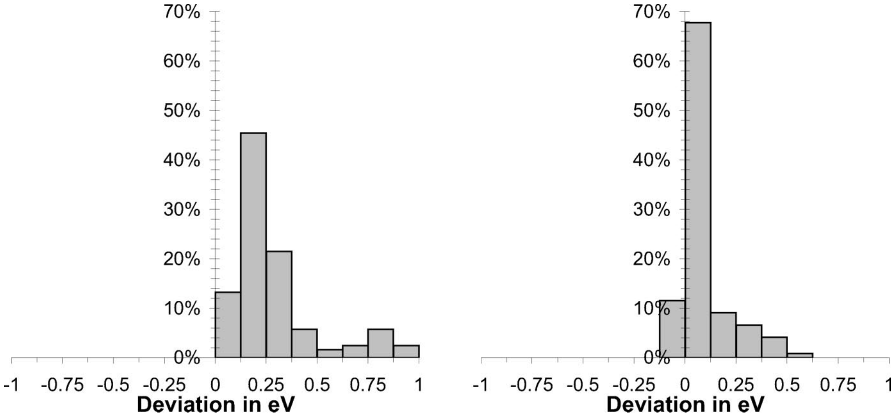


Figure 2. Histogram of the frequency of deviation (from CC3/TZVP in %) of all calculated CCSD/TZVP (left) and CCSDR(3)/TZVP (right) singlet excited states.

Table 2. Deviations in Excitation Energies of 121 Singlet Excited States with Respect to CC3/TZVP

	method		
	CC2[a]	CCSD[a]	CCSDR(3)
mean	0.13	0.30	0.09
abs mean	0.17	0.30	0.09
std dev	0.26	0.38	0.14
maximum	0.95	1.28	0.52

[a] CC2/TZVP and CCSD/TZVP results from ref 1.

the corresponding expansion of the CC2[3,4] or second-order polarization propagator approximation (SOPPA)[19,20,22,24,25] eigenvalue problem.

CCSDR(3) calculations have previously been carried out for a number of small[31,32,35−37] and medium size[36,38−45] molecules; however, a systematic comparison with CC3 and CCSD has not yet been published. In this article we present such a study for singlet excitation energies using our recently published benchmark set.[1] We have calculated CCSDR(3) excitation energies for 121 excited valence singlet states in 24 molecules, i.e. all the singlet states from the benchmark set for which CC3 results are available.[1] The new CCSDR(3) results are compared with the previously published CCSD and CC3 data. On the basis of a statistical evaluation of our results, we derive rules of thumb for the accuracy of the noniterative triples correction of CCSDR(3) relative to the iterative correction in CC3.

Although this study is only concerned with the question of how well CCSDR(3) with its noniterative triples correction is able to reproduce the results of CC3 calculations, we emphasize that CCSDR(3) is by no means the only method which includes noniterative triples corrections in the calculation of excitation energies. Several methods, EOM-CCSD(T), EOM-CCSD(T′), and EOM-CCSD($\bar{\mathrm{T}}$), have been proposed[46−49] that are based on the equation-of-motion coupled cluster approach (EOM-CC) developed by Bartlett and others.[50−63] Piecuch and co-workers have extended the completely renormalized coupled-cluster theory[64−66] to the calculation of excitation energies, by adding noniterative triples corrections to the EOM-CCSD energies in their CR-EOMCCSD(T), CR-EOMCCSD(T)$_\mathrm{L}$, and r-CR-EOMCCSD(T) treatments[65,67−71] which are based on the methods of moments of coupled cluster equations.[67,68,72−75] Finally one should also note in this context a recent approach[76] based on the EOM-CC(m)PT(n) methods,[77,78] the similarity transformed EOM-CC method (STEOM)[79−82] that implicitly includes triples excitations, the Fock space coupled cluster theory,[83] the SAC-CI approach of Nakatsuji,[84,85] and the spin-flip equation-of-motion coupled cluster method by Krylov and co-workers.[86−90]

The remaining parts of the paper are structured as follows. In Section 2 the details of the calculations are specified. In Section 3 the CCSDR(3) results are discussed in comparison with the previously published CCSD and CC3 results. Conclusions are drawn in section 4.

2. Computational Details

All calculations were carried out with the Dalton 2.0 program package.[91] The same MP2/6−31G* optimized geometries[92−95] and TZVP basis set[96] as in our previous studies were employed.[1,2]

As discussed before,[1,2] the TZVP basis set does not contain diffuse functions and might not be able to give a balanced description of excited states that are spatially extended and have (partial) Rydberg character. For the present study this is less important, since we are interested in a direct comparison of correlated ab initio methods in the calculation of valence excited states, which should be much less sensitive to the basis set than the absolute excitation energies. For example, the differences between the linear response CC3 and CCSD results deviate by only a few hundredths of 1 eV from the corresponding literature values[11,13,36,38−42,44,82,97−100] obtained with mostly larger basis sets. Our absolute CC3 excitation energies, on the other hand, tend to be slightly too large, normally by 0.02−0.15 eV and sometimes by up to 0.3 eV.

3. Results and Discussion

The CCSDR(3)/TZVP results for the 121 singlet valence states are given in Table 1 together with the previously published[1] CCSD and CC3 results. 83 of these states are of $\pi \rightarrow \pi^*$ type, 35 are of $n \rightarrow \pi^*$ type and three are of $\sigma \rightarrow \pi^*$ type. Figure 1 shows correlation plots between the CC2, CCSD, and CCSDR(3) results on one side and the CC3 results on the other side for all calculated states. Figure 2

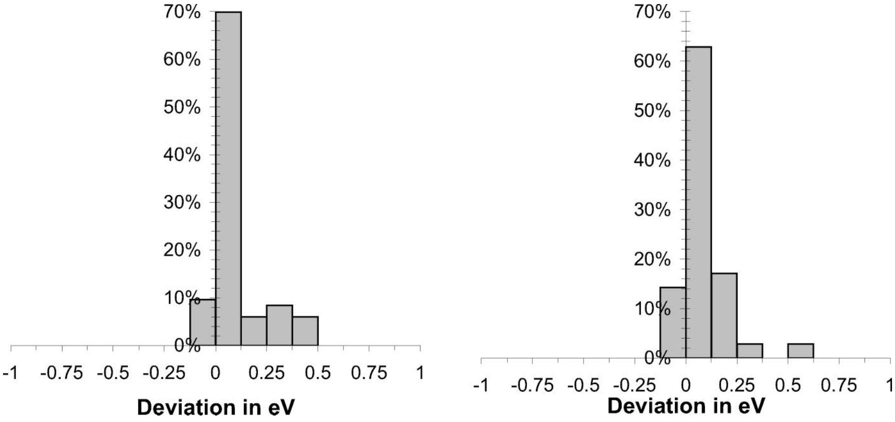


Figure 3. Histogram of the frequency of deviation (CCSDR(3)/TZVP vs CC3/TZVP, in %) of all calculated singlet $\pi \rightarrow \pi^*$ (left) and $n \rightarrow \pi^*$ (right) excited states.

Table 3. Deviations in Excitation Energies of 35 Singlet Excited $n \rightarrow \pi^*$ States with Respect to CC3/TZVP

	method		
	CC2[a]	CCSD[a]	CCSDR(3)
mean	0.04	0.28	0.10
abs mean	0.10	0.28	0.10
std dev	0.13	0.36	0.14
maximum	0.56	1.28	0.52

[a] CC2/TZVP and CCSD/TZVP results from ref 1.

Table 4. Deviations in Excitation Energies of 45 Singlet Excited States of the Benchmark Set with Respect to CC3/TZVP, for States Where the R_1 (CC3) Percentage Exceeds 90%

	method		
	CC2[a]	CCSD[a]	CCSDR(3)
mean	0.04	0.16	0.02
abs mean	0.09	0.16	0.02
std dev	0.11	0.18	0.03
maximum	0.27	0.30	0.09

[a] CC2/TZVP and CCSD/TZVP results from ref 1.

Table 5. Comparison of Relative CPU Times between CCSD, CCSDR(3), and CC3 Calculations

	molecule	
	benzene	naphthalene
basis functions	150	238
states	6[a]	3
S+D amplitudes	≈ 515 000[a]	3 353 467
relative CPU time		
CCSD	1	6
CCSDR(3)	22	132
CC3	1012	2706

[a] Six states in four different irreducible representations were calculated in the same run: one ^1A_g state with 517029 S+D amplitudes, two $^1B_{3u}$ states with 514210 S+D amplitudes, two $^1B_{2u}$ states with 515667 S+D amplitudes and one $^1B_{1g}$ state with 514202 S+D amplitudes.

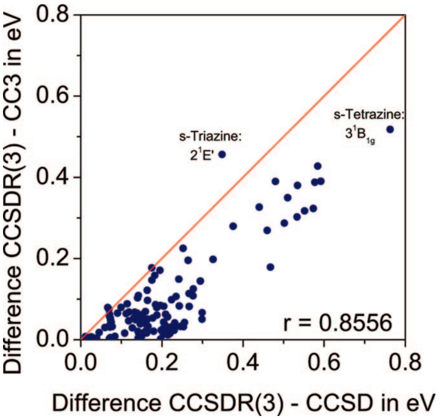


Figure 4. Correlation plot for all calculated singlet excited states: energy difference between CCSDR(3) and CC3 versus energy difference between CCSDR(3) and CCSD.

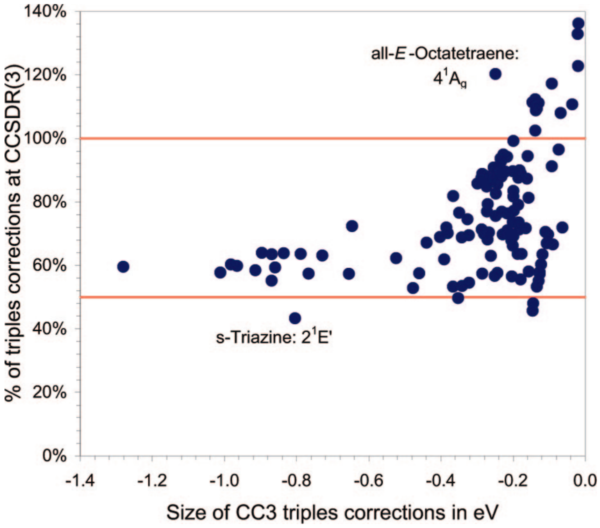


Figure 5. Correlation plot for all calculated singlet excited states: percentage of the triples correction obtained in the CCSDR(3) calculations versus the total CC3 triples correction.

presents histograms with the frequency of deviation from CC3 for the CCSD and CCSDR(3) results. It is obvious that compared with CCSD and CC2 the noniterative triples correction in CCSDR(3) leads to a much better agreement with the CC3 results over the whole range of energies. The mean deviation from CC3 (Table 2) drops from 0.30 eV in CCSD to 0.09 eV in CCSDR(3). The largest deviations from CC3 are found for the 3^1B_{1g} state of *s*-tetrazine (0.52 eV), the $2^1E'$ state of *s*-triazine (0.46 eV), and the 2^1A_g state of *all-E*-octatetraene (0.43 eV), which is in each case about half

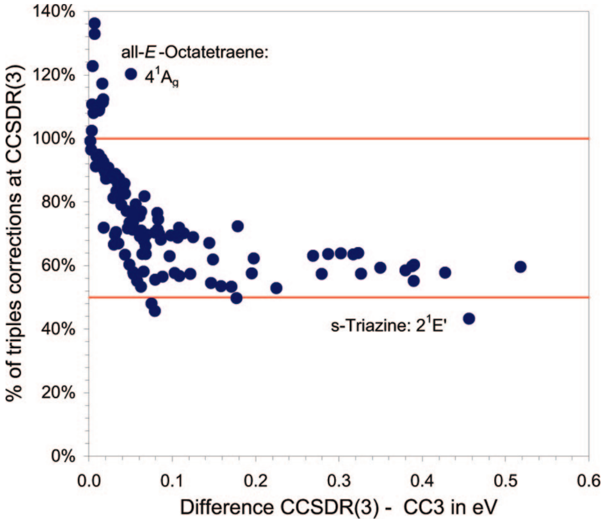


Figure 6. Correlation plot for all calculated singlet excited states: percentage of the triples correction obtained in the CCSDR(3) calculations versus the difference between CCSDR(3) and CC3.

the deviation found for CCSD. Correspondingly the standard deviation is also much smaller for CCSDR(3) than for CCSD as can be seen from Figure 2 and Table 2.

The significantly improved performance of CCSDR(3) is not restricted to a particular range of excited-state energies but holds for the whole benchmark set as illustrated in the correlation plots in Figure 1. For less than 10% of the states, CCSDR(3) predicts slightly smaller excitation energies than CC3, whereas the CCSD excitation energies in our benchmark set are always larger than the CC3 energies. We had previously found that CC2 gives excitation energies both larger and smaller than CC3 and consequently performs on average better than CCSD for our benchmark set.[1] However, compared with CCSDR(3), the spread of results is significantly larger in CC2 than in CCSDR(3). This can be seen in the correlation plots (Figure 1) and is also indicated by the standard deviations and maximum deviations in Table 2 which are about twice as large for CC2 than for CCSDR(3).

It is obvious from the histograms in Figure 3 and from the comparison of the statistical data for the $n \rightarrow \pi^*$ transitions alone (Table 3) against the data for all excited states (Table 2) that there is not much difference between the $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions. The largest outlier with respect to CC3, i.e., the 3^1B_{1g} state of *s*-tetrazine, is a $n \rightarrow \pi^*$ transition, whereas the two next largest outliers, the $2^1E'$ state of *s*-triazine and the 2^1A_g state of *all-E*-octatetraene, are $\pi \rightarrow \pi^*$ transitions.

When restricting the statistics to states with a CC3 single excitation weight larger than 90% (called % R_1 in Table 1), the agreement between CCSDR(3) and CC3 becomes almost perfect. The mean deviation is reduced to 0.02 eV, and the maximum deviation is also less than 0.1 eV (see Table 4).

Relative CPU times for two representative calculations of excited states in benzene and naphthalene are collected in Table 5. It is gratifying to see that the necessary CPU time is dramatically reduced in CCSDR(3) compared to CC3. Nevertheless, due to the formal N^7 scaling of CCSDR(3) these calculations are still considerably more expensive than the simpler CCSD calculations.

It would clearly be desirable to be able to estimate the remaining error of CCSDR(3) relative to CC3 based on the results of the CCSD and CCSDR(3) calculations alone. A correlation plot for the relevant energy differences (Figure 4) shows that the remaining errors in the CCSDR(3) results compared to CC3 are almost always smaller than the changes on going from CCSD to CCSDR(3) (prominent exception: the $2^1E'$ state of *s*-triazine).

Analyzing the performance of CCSDR(3) as a function of the single excitation weight in the CCSD calculations, one finds that for all states in our benchmark set with a CCSD single excitation weight smaller than 90% the difference between the CCSDR(3) and CC3 results is larger than 0.1 eV. This implies that the data from CCSD calculations allow us to pinpoint states, for which one very likely will encounter larger differences between CCSDR(3) and CC3. However, the opposite is not always true. There is one state in our benchmark set ($2^1E'$ in *s*-triazine), where the difference is 0.46 eV despite a single excitation weight of 91.2%, while there are two states in *p*-benzoquinone (1^1B_{3u} and 2^1B_{1u}) with a single excitation weight of 91−92% which differ by ≈0.27 eV and in total 15 states with single excitation weights of 91−93% and differences between CCSDR(3) and CC3 in the range between 0.1 and 0.2 eV.

Finally, it is important to know which fraction of the CC3 triples correction can be recovered in an CCSDR(3) calculation. This percentage is shown as a function of the size of the total CC3 triples correction in Figure 5 and as function of the remaining deviation from the CC3 results in Figure 6. It is obvious that with three exceptions (most prominent again the $2^1E'$ state of *s*-triazine) the CCSDR(3) triples correction gives at least 50% of the iterative CC3 triples correction, and on average it amounts to 78%. Even for the cases with large CC3 triple corrections (0.4 eV or more), CCSDR(3) yields about 60%. On the other hand, there are also some systems where CCSDR(3) overestimates the CC3 triples correction. However, with the exception of the 4^1A_g state of *all-E*-octatetraene, this happens only for states where the remaining difference between CCSDR(3) and CC3 is less than 0.02 eV and the total triples correction is less than 0.16 eV.

4. Conclusions

We have carried out CCSDR(3)/TZVP calculations of vertical excitation energies for 24 molecules and a total of 121 valence excited singlet states from a recently published benchmark set of organic molecules.[1] Statistical comparison of these data with the previously published linear response CC2, CCSD, and CC3 results shows that adding the noniterative triples corrections to the CCSD results leads to a substantial improvement over CC2 and CCSD for all states in this benchmark set.

Inclusion of the noniterative triples correction in CCSDR(3) often reproduces the iterative CC3 triples correction almost quantitatively, at dramatically reduced CPU times. For all states, which are dominated by single excitations (CC3 single excitation weight larger than 90%), the

CCSDR(3) results differ by at most 0.1 eV from the CC3 results. More important is that one can obtain a reasonable estimate of the accuracy of the CCSDR(3) results from the size of the triples correction in CCSDR(3) and the single excitation weight in the CCSD calculations. The remaining difference between CCSDR(3) and CC3 is in almost all cases smaller than the noniterative triples correction from CCSDR(3). CCSDR(3) normally gives at least 60% of the CC3 triples correction even if the latter is large (>0.4 eV). Furthermore one can easily identify states for which differences of more than 0.1 eV between CCSDR(3) and CC3 should be expected: for all states in our benchmark with a CCSD single excitation weight smaller than 90%, the CCSDR(3) results deviate from the CC3 results by 0.1 eV or more.

We conclude that CCSDR(3) may play a similar role for excited states as CCSD(T) does for ground states, and that an appropriate sequence of linear response-coupled cluster methods for the calculation of vertical excitation energies is CC2, CCSDR(3), CC3. However, since transition moments are not defined in CCSDR(3), the corresponding sequence for the calculation of oscillator strengths remains CC2, CCSD, CC3.

Acknowledgment. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 663, project C4) as well as the Danish Center for Scientific Computing, FNU, and the Carlsberg Foundation (SPAS). M.R.S.J. gladly acknowledges support by the Deutscher Akademischer Austausch Dienst.

References

(1) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.* **2008**, *128*, 134110.

(2) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. Benchmarks for electronically excited states: Time-dependent density functional theory and density functional theory based multireference configuration interaction. *J. Chem. Phys.* **2008**, *129*, 104103.

(3) Christiansen, O.; Koch, H.; Jørgensen, P. The second-order approximate coupled cluster singles and doubles model CC2. *Chem. Phys. Lett.* **1995**, *243*, 409–418.

(4) Hald, K.; Hättig, C.; Yeager, D. L.; Jørgensen, P. Linear response CC2 triplet excitation energies. *Chem. Phys. Lett.* **2000**, *328*, 291–301.

(5) Monkhorst, H. J. Calculation of Properties with the Coupled-Cluster Method. *Int. J. Quantum Chem. Symp.* **1977**, *11*, 421–432.

(6) Mukherjee, D.; Mukherjee, P. K. A response-function approach to the direct calculation of the transition-energy in a multiple-cluster expansion formalism. *Chem. Phys.* **1979**, *39*, 325–335.

(7) Ghosh, S.; Mukherjee, D.; Bhattacharyya, S. N. A spin-adapted linear response theory in a coupled-cluster framework for direct calculation of spin-allowed and spin-forbidden transition energies. *Chem. Phys. Lett.* **1982**, *72*, 161–176.

(8) Dalgaard, E.; Monkhorst, H. J. Some aspects of the time-dependent coupled-cluster approach to dynamic response functions. *Phys. Rev. A* **1983**, *28*, 1217–1222 .

(9) Koch, H.; Jørgensen, P. Coupled cluster response function. *J. Chem. Phys.* **1990**, *93*, 3333–3344.

(10) Datta, B.; Sen, P.; Mukherjee, D. Coupled-Cluster Based Linear Response Approach to Property Calculations: Dynamic Polarizability and Its Static Limit. *J. Phys. Chem.* **1995**, *99*, 6441–6451.

(11) Hald, K.; Hättig, C.; Jørgensen, P. Triplet excitation energies in the coupled cluster singles and doubles model using an explicit triplet spin coupled excitation space. *J. Chem. Phys.* **2000**, *113*, 7765–7772.

(12) Hald, K.; Hättig, C.; Olsen, J.; Jørgensen, P. CC3 triplet excitation energies using an explicit spin coupled excitation space. *J. Chem. Phys.* **2001**, *115*, 3545–3552.

(13) Christiansen, O.; Koch, H.; Jørgensen, P. Response functions in the CC3 iterative triple excitation model. *J. Chem. Phys.* **1995**, *103*, 7429–7441.

(14) Koch, H.; Christiansen, O.; Jørgensen, P.; Sanchez de Merás, A. M.; Helgaker, T. The CC3 model: An iterative coupled cluster approach including connected triples. *J. Chem. Phys.* **1997**, *106*, 1808–1818.

(15) Hald, K.; Jørgensen, P.; Olsen, J.; Jaszunski, M. An analysis and implementation of a general coupled cluster approach to excitation energies with application to the B2 molecule. *J. Chem. Phys.* **2001**, *115*, 671–679.

(16) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-Order Perturbation Theory with a CASSCF Reference Function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(17) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. Second-Order Perturbation Theory with a Complete Active Space Self-Consistent Field Reference Function. *J. Chem. Phys.* **1992**, *96*, 1218–1226.

(18) Roos, B. O.; Andersson, K.; Fülscher, M. P.; Malmqvist, P.-Å.; Serrano-Andrés, L.; Pierloot, K.; Merchán, M. Multiconfigurational Perturbation Theory: Applications in Electronic Spectroscopy In *Advances in Chemical Physics: New Methods in Computational Quantum Mechanics*; Prigogine, I., Rice, S. A., Eds.; John Wiley & Sons: New York, 1996; Vol. XCIII, pp 219−331.

(19) Nielsen, E. S.; Jørgensen, P.; Oddershede, J. Transition moments and dynamic polarizabilities in a second order polarization propagator approach. *J. Chem. Phys.* **1980**, *73*, 6238–6246.

(20) Packer, M. J.; Dalskov, E. K.; Enevoldsen, T.; Jensen, H. J. A.; Oddershede, J. A new implementation of the second order polarization propagator approximation (SOPPA): The excitation spectra of benzene and naphthalene. *J. Chem. Phys.* **1996**, *105*, 5886–5900.

(21) Christiansen, O.; Bak, K. L.; Koch, H.; Sauer, S. P. A. A Second-order doubles correction to excitation energies in the random phase approximation. *Chem. Phys. Lett.* **1998**, *284*, 47–62.

(22) Bak, K. L.; Koch, H.; Oddershede, J.; Christiansen, O.; Sauer, S. P. A. Atomic integral driven second order polarization propagator calculations of the excitation spectra of naphthalene and anthracene. *J. Chem. Phys.* **2000**, *112*, 4173–4185.

(23) Åstrand, P.-O.; Ramanujam, P. S.; Hvilsted, S.; Bak, K. L.; Sauer, S. P. A. Ab initio calculation of the electronic spectra of azobenzene dyes and its impact on the design of optical data storage materials. *J. Am. Chem. Soc.* **2000**, *122*, 3482–3487.

(24) Nielsen, C. B.; Sauer, S. P. A.; Mikkelsen, K. V. Response theory in the multipole reaction field model for equilibrium and nonequilibrium solvation: Exact theory and the second order polarization propagator approximation. *J. Chem. Phys.* **2003**, *119*, 3849–3870.

(25) Olsen, J.; Jørgensen, P.; Helgaker, T.; Oddershede, J. Quadratic response functions in a second-order polarization propagator framework. *J. Phys. Chem. A* **2005**, *109*, 11618–11628.

(26) Geertsen, J.; Oddershede, J. A coupled cluster polarization propagator method applied to CH^+. *J. Chem. Phys.* **1986**, *85*, 2112–2118.

(27) Geertsen, J.; Eriksen, S.; Oddershede, J. Some aspects of the coupled cluster based polarization propagator method. *Adv. Quantum Chem.* **1991**, *22*, 167–209.

(28) Sauer, S. P. A. Second Order Polarization Propagator Approximation with Coupled Cluster Singles and Doubles Amplitudes - SOPPA(CCSD): The Polarizability and Hyperpolarizability of Li-. *J. Phys. B: At. Mol. Opt. Phys.* **1997**, *30*, 3773–3780.

(29) Dalskov, E. K.; Sauer, S. P. A. Correlated, static and dynamic polarizabilities of small molecules. A comparison of four black box methods. *J. Phys. Chem. A* **1998**, *102*, 5269–5274.

(30) Sattelmeyer, K. W.; Stanton, J. F.; Olsen, J.; Gauss, J. A comparison of excited state properties for iterative approximate triples linear response coupled cluster methods. *Chem. Phys. Lett.* **2001**, *347*, 499–504.

(31) Christiansen, O.; Koch, H.; Jørgensen, P. Perturbative triple excitation corrections to coupled cluster singles and doubles excitation energies. *J. Chem. Phys.* **1996**, *105*, 1451–1459.

(32) Christiansen, O.; Koch, H.; Jørgensen, P.; Olsen, J. Excitation energies of H_2O, N_2 and C_2 in full configuration interaction and coupled cluster theory. *Chem. Phys. Lett.* **1996**, *256*, 185–194.

(33) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(34) Head-Gordon, M.; Rico, R. J.; Oumi, M.; Lee, T. J. A doubles correction to electronic excited states from configuration interaction in the space of single substitutions. *Chem. Phys. Lett.* **1994**, *219*, 21–29.

(35) Larsen, H.; Olsen, J.; Jørgensen, P.; Christiansen, O. Full configuration interaction benchmarking of coupled-cluster models for the lowest singlet energy surfaces of N_2. *J. Chem. Phys.* **2000**, *113*, 6677–6686.

(36) Cronstrand, P.; Christiansen, O.; Norman, P.; Ågren, H. Theoretical calculations of excited state absorption. *Phys. Chem. Chem. Phys.* **2000**, *2*, 5357–5363.

(37) Hättig, C. Structure optimizations for excited states with correlated second-order methods: CC2 and ADC(2). *Adv. Quantum Chem.* **2005**, *50*, 37–60.

(38) Christiansen, O.; Jørgensen, P. The Electronic Spectrum of Furan. *J. Am. Chem. Soc.* **1998**, *120*, 3423–3430.

(39) Christiansen, O.; Gauss, J.; Stanton, J. F.; Jørgensen, P. The electronic spectrum of pyrrole. *J. Chem. Phys.* **1999**, *111*, 525–537.

(40) Christiansen, O.; Koch, H.; Halkier, A.; Jørgensen, P.; Helgaker, T.; Sánchez de Merás, A. M. Large-scale calculations of excitation energies in coupled cluster theory: The singlet excited states of benzene*J. Chem. Phys.* **1996**, *105*, 6921−6939.

(41) Öhrn, A.; Christiansen, O. Electronic excitation energies of pyrimidine studied using coupled cluster response theory. *Phys. Chem. Chem. Phys.* **2001**, *3*, 730–740.

(42) Cronstrand, P.; Christiansen, O.; Norman, P.; Ågren, H. Ab initio modeling of excited state absorption of polyenes. *Phys. Chem. Chem. Phys.* **2001**, *3*, 2567–2575.

(43) Köhn, A.; Hättig, C. On the nature of the low-lying singlet states of 4-(dimethylamino)benzonitrile. *J. Am. Chem. Soc.* **2004**, *126*, 7399–7410.

(44) Osted, A.; Kongsted, J.; Christiansen, O. Theoretical Study of the Electronic Gas-Phase Spectrum of Glycine, Alanine, and Related Amines and Carboxylic Acids. *J. Phys. Chem. A* **2005**, *109*, 1430–1440.

(45) Pastore, M.; Angeli, C.; Cimiraglia, R. A multireference perturbation theory study on the vertical electronic spectrum of thiophene. *Theor. Chem. Acc.* **2007**, *118*, 35–46.

(46) Watts, J. D.; Bartlett, R. J. The inclusion of connected triple excitations in the equation-of-motion coupled-cluster method. *J. Chem. Phys.* **1994**, *101*, 3073–3078.

(47) Watts, J. D.; Bartlett, R. J. Economical triple excitation equation-of-motion coupled-cluster methods for excitation energies. *Chem. Phys. Lett.* **1995**, *233*, 81–87.

(48) Watts, J. D.; Gwaltney, S. R.; Bartlett, R. J. Coupled-cluster calculations of the excitation energies of ethylene, butadiene, and cyclopentadiene. *J. Chem. Phys.* **1996**, *105*, 6979–6988.

(49) Watts, J. D.; Bartlett, R. J. Equation-of-motion coupled-cluster calculations of excitation energies. The challenge of ozone. *Spectrochim. Acta A* **1999**, *55*, 495–507.

(50) Rowe, D. J. Equations-of-Motion Method and the Extended Shell Model. *Rev. Mod. Phys.* **1968**, *40*, 153–166.

(51) Sekino, H.; Bartlett, R. J. A linear response, coupled-cluster theory for excitation energy. *Int. J. Quantum Chem.: Quantum Chem. Symp.* **1984**, *18*, 255.

(52) Geertsen, J.; Rittby, M.; Bartlett, R. J. The Equation-of-Motion Coupled-Cluster Method: Excitation Energies of Be and CO. *Chem. Phys. Lett.* **1989**, *164*, 57–62.

(53) Stanton, J. F.; Bartlett, R. J. The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *J. Chem. Phys.* **1993**, *98*, 7029–7039.

(54) Comeau, D. C.; Bartlett, R. J. The equation-of-motion coupled-cluster method. Applications to open- and closed-shell reference states. *Chem. Phys. Lett.* **1993**, *207*, 414–423.

(55) Piecuch, P.; Bartlett, R. J. EOMXCC: A new coupled-cluster method for electronic excited states. *Adv. Quantum Chem.* **1999**, *34*, 295–380.

(56) Kowalski, K.; Piecuch, P. Excited-state potential energy curves of CH^+: a comparison of the EOMCCSDt and full EOMCCSDT results. *Chem. Phys. Lett.* **2001**, *347*, 237–246.

(57) Kowalski, K.; Piecuch, P. The active-space equation-of-motion coupled-cluster methods for excited electronic states: Full EOMCCSDt. *J. Chem. Phys.* **2001**, *115*, 643–651.

(58) Kucharski, S. A.; Włoch, M.; Musiał, M.; Bartlett, R. J. Coupled-cluster theory for excited electronic states: The full equation-of-motion coupled-cluster single, double, and triple excitation method. *J. Chem. Phys.* **2001**, *115*, 8263–8266.

(59) Hirata, S. Higher-order equation-of-motion coupled-cluster methods. *J. Chem. Phys.* **2004**, *121*, 51–59.

(60) Kállay, M.; Gauss, J. Calculation of excited-state properties using general coupled-cluster and configuration-interaction models. *J. Chem. Phys.* **2004**, *121*, 9257–9269.

(61) Bartlett, R. J.; Musial, M. Coupled-Cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291–352.

(62) Musiał, M.; Bartlett, R. J. Addition by subtraction in coupled cluster theory. II. Equation-of-motion coupled cluster method for excited, ionized, and electron-attached states based on the nCC ground state wave function. *J. Chem. Phys.* **2007**, *127*, 24106.

(63) Smith, C. E.; King, R. A.; Crawford, T. D. Coupled cluster methods including triple excitations for excited states of radicals. *J. Chem. Phys.* **2005**, *122*, 54110.

(64) Kowalski, K.; Piecuch, P. Renormalized CCSD(T) and CCSD(TQ) approaches: Dissociation of the N-2 triple bond. *J. Chem. Phys.* **2000**, *113*, 5644–5652.

(65) Kowalski, K.; Piecuch, P. New coupled-cluster methods with singles, doubles, and noniterative triples for high accuracy calculations of excited electronic states. *J. Chem. Phys.* **2004**, *120*, 1715–1738.

(66) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musiał, M. Efficient computer implementation of the renormalized coupled-cluster methods: The R-CCSD[T], R-CCSD(T), CR-CCSD[T], and CR-CCSD(T) approaches. *Comput. Phys. Commun.* **2002**, *149*, 71–96.

(67) Piecuch, P.; Kowalski, K.; Pimienta, I.; McGuire, M. Recent advances in electronic structure theory: Method of moments of coupled-cluster equations and renormalized coupled-cluster approaches. *Int. Rev. Phys. Chem.* **2002**, *21*, 527–655.

(68) Piecuch, P.; Kowalski, K.; Pimienta, I.; Fan, P.; Lodriguito, M.; McGuire, M.; Kucharski, S.; Kus, T.; Musiał, M. Method of moments of coupled-cluster equations: a new formalism for designing accurate electronic structure methods for ground and excited states. *Theor. Chim. Acta* **2004**, *112*, 349–393.

(69) Włoch, M.; Gour, J.; Kowalski, K.; Piecuch, P. Extension of renormalized coupled-cluster methods including triple excitations to excited electronic states of open-shell molecules. *J. Chem. Phys.* **2005**, *122*, 214107.

(70) Łoch, M. W.; Lodriguito, M. D.; Piecuch, P.; Gour, J. R. Two new classes of non-iterative coupled-cluster methods derived from the method of moments of coupled-cluster equations. *Mol. Phys.* **2006**, *104*, 2149–2172.

(71) Kowalski, K.; Valiev, M. Noniterative Corrections to Equation-of-Motion Coupled-Cluster Excited State Energies Based on the Reduced Method of Moments of Coupled Cluster Equations. *Int. J. Quantum Chem.* **2008**, *108*, 2178–2190.

(72) Kowalski, K.; Piecuch, P. The method of moments of coupled-cluster equations and the renormalized CCSD[T], CCSD(T), CCSD(TQ), and CCSDT(Q) approaches. *J. Chem. Phys.* **2000**, *113*, 18–35.

(73) Kowalski, K.; Piecuch, P. New type of noniterative energy corrections for excited electronic states: Extension of the method of moments of coupled-cluster equations to the equation-of-motion coupled-cluster formalism. *J. Chem. Phys.* **2001**, *115*, 2966–2978.

(74) Fan, P.-D.; Kowalski, K.; Piecuch, P. Non-iterative corrections to extended coupled-cluster energies employing the generalized method of moments of coupled-cluster equations. *Mol. Phys.* **2005**, *103*, 2191–2213.

(75) Piecuch, P.; Włoch, M. Renormalized coupled-cluster methods exploiting left eigenstates of the similarity-transformed Hamiltonian. *J. Chem. Phys.* **2005**, *123*, 224105.

(76) Shiozaki, T.; Hirao, K.; Hirata, S. Second- and third-order triples and quadruples corrections to coupled-cluster singles and doubles in the ground and excited states. *J. Chem. Phys.* **2007**, *126*, 244106.

(77) Hirata, S.; Nooijen, M.; Grabowski, I.; Bartlett, R. Perturbative corrections to coupled-cluster and equation-of-motion coupled-cluster energies: A determinantal analysis. *J. Chem. Phys.* **2001**, *114*, 3919–3928.

(78) Hirata, S.; Nooijen, M.; Grabowski, I.; Bartlett, R. Perturbative corrections to coupled-cluster and equation-of-motion coupled-cluster energies: A determinantal analysis (vol 114, pg 3919, 2001). *J. Chem. Phys.* **2001**, *115*, 3967–3968.

(79) Nooijen, M.; Bartlett, R. J. A new method for excited states: Similarity transformed equation-of-motion coupled-cluster theory. *J. Chem. Phys.* **1997**, *106*, 6441–6448.

(80) Nooijen, M.; Bartlett, R. Similarity transformed equation-of-motion coupled-cluster theory: Details, examples, and comparisons. *J. Chem. Phys.* **1997**, *107*, 6812–6830.

(81) Nooijen, M. Similarity transformed equation of motion coupled-cluster study of excited states of selected azabenzenes. *Spectrochim. Acta, Part A* **1999**, *55*, 539–559.

(82) Nooijen, M. Electronic excitation spectrum of *s*-tetrazine: An extended-STEOM-CCSD study. *J. Phys. Chem. A* **2000**, *104*, 4553–4561.

(83) Rittby, C. M. L.; Bartlett, R. J. Multireference coupled cluster theory in Fock space. *Theor. Chim. Acta* **1991**, *80*, 469–482.

(84) Nakatsuji, H. Cluster expansion of wavefunction - excited-states. *Chem. Phys. Lett.* **1978**, *59*, 362–364.

(85) Nakatsuji, H. Cluster expansion of the wavefunction - electron correlations in ground and excited-states by SAC (Symmetry-Adapted-Cluster) and SAC CI theories. *Chem. Phys. Lett.* **1979**, *67*, 329–333.

(86) Krylov, A. I. Size-consistent wave functions for bond-breaking: the equation-of-motion spin-flip model. *Chem. Phys. Lett.* **2001**, *338*, 42–46.

(87) Levchenko, S. V.; Krylov, A. I. Equation-of-motion spin-flip coupled-cluster model with single and double substitutions: theory and application to cyclobutadiene. *J. Chem. Phys.* **2004**, *120*, 175–185.

(88) Slipchenko, L. V.; Krylov, A. I. Spin-conserving and spin-flip equation-of-motion coupled-cluster method with triple excitations. *J. Chem. Phys.* **2005**, *123*, 84107.

(89) Krylov, A. I. The spin-flip equation-of-motion coupled-cluster electronic structure method for a description of excited states, bond breaking, diradicals, and triradicals. *Acc. Chem. Res.* **2006**, *39*, 83–91.

(90) Krylov, A. I. Equation-of-Motion Coupled-Cluster Methods for Open-Shell and Electronically Excited Species: The Hitchhiker's Guide to Fock Space. *Annu. Rev. Phys. Chem.* **2008**, *59*, 433–462.

(91) Dalton, a molecular electronic structure program, Release 2.0; http://www.kjemi.uio.no/software/dalton/dalton.html, 2005.

(92) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

(93) Pople, J. A.; Binkley, J. S.; Seeger, R. Theoretical Models Incorporating Electron Correlation. *Int. J. Quantum Chem. Symp* **1976**, *10*, 1–19.

(94) Frisch, M. J. et al. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.

(95) Hehre, W. J.; Radom, L.; v R. Schleyer, P.; Pople, J. A. *Ab initio molecular orbital theory*; Wiley: New York, 1986; pp 63−100.

(96) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577 .

(97) Christiansen, O.; Stanton, J. F.; Gauss, J. A coupled cluster study of the $1^1A_{1g}$and 1^1B_{2u} states of benzene. *J. Chem. Phys.* **1998**, *108*, 3987–4001.

(98) Christiansen, O.; Hättig, C.; Jørgensen, P. Ground and excited state polarizabilities and dipole transition properties of benzene from coupled cluster response theory. *Spectrochim. Acta, Part A* **1999**, *55*, 509–524.

(99) Hald, K.; Jørgensen, P.; Christiansen, O.; Koch, H. Implementation of electronic ground states and singlet and triplet excitation energies in coupled cluster theory with approximate triples corrections. *J. Chem. Phys.* **2002**, *116*, 5963–5970.

(100) Bomble, Y. J.; Sattelmeyer, K. W.; Stanton, J. F.; Gauss, J. On the vertical excitation energy of cyclopentadiene. *J. Chem. Phys.* **2004**, *121*, 5236–5240.

CT800256J

JCTC Journal of Chemical Theory and Computation

Franck−Condon Simulations including Anharmonicity of the $\tilde{A}^1A''-\tilde{X}^1A'$ Absorption and Single Vibronic Level Emission Spectra of HSiCl and DSiCl

Daniel W. K. Mok,*[,†] Edmond P. F. Lee,*[,†,‡] Foo-tim Chau,[†] and John M. Dyke[‡]

Department of Applied Biology and Chemical Technology, the Hong Kong Polytechnic University, Hung Hom, Hong Kong, and School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K.

Received November 28, 2008

Abstract: RCCSD(T) and/or CASSCF/MRCI calculations have been carried out on the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ states of HSiCl employing basis sets of up to the aug-cc-pV5Z quality. Contributions from core correlation and extrapolation to the complete basis set limit were included in determining the computed equilibrium geometrical parameters and relative electronic energy of these two states of HSiCl. Franck−Condon factors which include allowance for anharmonicity and Duschinsky rotation between these two states of HSiCl and DSiCl were calculated employing RCCSD(T) and CASSCF/MRCI potential energy functions, and were used to simulate the $\tilde{A}^1A'' \leftarrow \tilde{X}^1A'$ absorption and $\tilde{A}^1A'' \rightarrow \tilde{X}^1A'$ single vibronic level (SVL) emission spectra of HSiCl and DSiCl. Simulated absorption and experimental LIF spectra, and simulated and observed $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emission spectra, of HSiCl and DSiCl are in very good agreement. However, agreement between simulated and observed $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ and $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'$ SVL emission spectra of DSiCl is not as good. Preliminary calculations on low-lying excited states of HSiCl suggest that vibronic interaction between low-lying vibrational levels of the $\tilde{A}^1A''$ state and highly excited vibrational levels of the $\tilde{a}^3A''$ is possible. Such vibronic interaction may change the character of the low-lying vibrational levels of the $\tilde{A}^1A''$ state, which would lead to perturbation in the SVL emission spectra from these vibrational levels.

Introduction

In their continuing study of the spectroscopy of carbenes, silylenes, and germylenes, reactive intermediates important for chemical vapor deposition (CVD) processes in the semiconductor industry, the Clouthier group recorded laser induced fluorescence (LIF) spectra of HSiCl and DSiCl produced in an electric discharge of $SiHCl_3$ and $SiDCl_3$, respectively, diluted in Ar through a pulsed jet.[1] This work was performed over a decade ago. A few years later, they published single vibronic level (SVL) emission spectra of HSiCl and DSiCl.[2] In this SVL emission (or dispersed fluorescence) study, they employed data from harmonic force field and density functional theory (DFT) calculations using the B3LYP functional and the 6−311G(3df,3pd) basis set to simulate the $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emission spectrum of HSiCl, using computed Franck−Condon (FC) factors obtained within the harmonic oscillator model. Agreements between simulated and observed SVL emission spectra are reasonably good, particularly for the main $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'(0,v_2'',0)$ vibrational progression. However, for the weak vibrational structure, there are clearly noticeable discrepancies between simulated and observed SVL spectra. Specifically, in addition to the main $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'(0,v_2'',0)$ progression, simulated spectra show only one relatively weak $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'(1,v_2'',0)$ progression, but in the experimental SVL spectrum the only observed weak progression is the $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'(0,v_2'',1)$ progression (vide infra). The $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'(1,v_2'',0)$ progression was not observed in the experimental SVL emission spectrum as

---

* To whom correspondence should be addressed. E-mail: Daniel@polyu.edu.hk (D.W.K.M.); epl@soton.ac.uk (E.P.F.L.).

[†] The Hong Kong Polytechnic University.

[‡] University of Southampton.

10.1021/ct800513v CCC: $40.75

predicted by harmonic FC simulation. It was concluded that, “Obtaining reliable simulations of these spectra will be a stringent test of future Franck−Condon calculations which include vibrational anharmonicity.”[2] In this connection, we propose, in the present investigation, to carry out state-of-the-art *ab initio* calculations on the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ states of HSiCl and FC factor calculations between these two states, which include allowance for anharmonicity and Duschinsky rotation, aiming to obtain better agreement between theory and experiment.

For earlier spectroscopic and computational studies on HSiCl, Clouthier et al. has given thorough discussions previously,[1,2] and hence they will not be repeated here. After the above-mentioned LIF and SVL emission studies on HSiCl/DSiCl were published,[1,2] microwave spectra of $H^{28}Si^{35}Cl$, $H^{28}Si^{37}Cl$, $H^{29}Si^{35}Cl$, and $H^{30}Si^{35}Cl$ were measured at 14−15 GHz, and the effective rotational constants, chlorine nuclear quadrupole coupling constants, and nuclear spin-rotation constants of the four isotopomers were determined.[3] The molecular structure of HSiCl derived from this study is essentially identical to those reported in references 1 and 2. In addition, the ^{35}Cl nuclear quadrupole coupling constant in HSiCl was calculated at up to the MP3 level. A LIF and dispersed fluorescence study of HSiCl, produced in the UV photodissociation of 2-chloroethenylsilane, has also been published, and it was concluded that emission spectra of HSiCl could be useful to monitor the CVD process of 2-chloroethenylsilane.[4] On the computational front, a number of high-level *ab initio* studies have been carried out recently on the $\tilde{X}^1A'$ state of HSiCl.[5−7] The latest study reported CCSD(T) calculations of nuclear-spin-rotation and nuclear quadrupole coupling constants of some halocarbenes and halosilylenes, including HSiCl.[7] However, for the $\tilde{A}^1A''$ state of HSiCl, to our knowledge, the only available calculations are the above-mentioned B3LYP/6−311G(3df,3pd) calculations reported in ref 2.

Related closely to the present study, we have recently reported a combined *ab initio*/FC study on the SVL emission spectra of HSiF and DSiF.[8] In this work, very good agreement between simulated and observed spectra was obtained for the $\tilde{A}^1A''(1,0,0) \rightarrow \tilde{X}^1A'$ SVL emission of HSiF and the $\tilde{A}(0,0,0) \rightarrow \tilde{X}^1A'$ and $\tilde{A}(0,1,0) \rightarrow \tilde{X}^1A'$ SVL emissions of DSiF. However, discrepancies between simulated and observed spectra of the $\tilde{A}(0,1,0) \rightarrow \tilde{X}^1A'$ and (1,1,0) $\rightarrow \tilde{X}^1A'$ SVL emissions of HSiF have been found. It was concluded that they were most likely, partly, caused by experimental deficiencies and, partly, by inadequacies in the *ab initio* levels of theory employed in the calculation of the potential energy functions (PEFs). Nevertheless, on the basis of computed FC factors, minor revisions of previous vibrational assignments have been suggested (see ref 8 for details).

Computational Details

Ab Initio Calculations. RCCSD(T)[9,10] and CASSCF/MRCI[11,12] geometry optimization calculations were carried out on the $\tilde{X}^1A'$ state of HSiCl, while only CASSCF/MRCI calculations were performed on the lowest open-shell singlet $\tilde{A}^1A''$ state. For the closed-shell $\tilde{X}^1A'$ state, the computed T_1 diagnostics (<0.013 in all cases) obtained from RCCSD(T) calculations and the calculated CI coefficients (>0.92 for the main ground-state electronic configuration) from the computed CASSCF and MRCI wave functions indicate that multireference character is negligibly small for this state at the computed equilibrium geometries. Therefore, it can be concluded that the size-consistent, single-reference RCCSD(T) method should be the more suitable, and more reliable than the size-inconsistent MRCI method, for the $\tilde{X}^1A'$ state of HSiCl. Nevertheless, CASSCF/MRCI calculations were also carried out for the $\tilde{X}^1A'$ state to obtain relative electronic energies between the $\tilde{X}^1A'$ state and the $\tilde{A}^1A''$ state because the latter is an open-shell singlet state, which cannot be adequately treated by a single electronic configuration and hence requires a multireference method. It should be noted that all CASSCF/MRCI calculations carried out in the present study are single state (i.e., not average-state) calculations, with the CASSCF wave function (i.e., the CASSCF molecular orbitals and reference configurations to be used in subsequent MRCI calculations) optimized for the single state concerned (i.e., the $\tilde{X}^1A'$ state or the $\tilde{A}^1A''$ state). The CASSCF/MRCI wave functions thus obtained are optimal for the state concerned.

Correlation-consistent basis sets of aug-cc-pVQZ (AVQZ), aug-cc-pV5Z (AV5Z), and aug-cc-pwCVQZ (ACVQZ) qualities have been used in the present investigation. For H, the standard aug-cc-pVQZ and aug-cc-pV5Z basis sets[13] were used. With the AVQZ and AV5Z quality basis sets, the aug-cc-pV(Q+d)Z and aug-cc-pV(5+d)Z basis sets[14] were used for the second row elements, Si and Cl, and a full valence active space was employed in the CASSCF and MRCI calculations, while the default frozen core was employed in the RCCSD(T) calculations. Results obtained from calculations employing the AVQZ and AV5Z basis sets were used to estimate contributions on extrapolation to the complete basis set (CBS) limit to the optimized geometrical parameters and computed relative electronic energies. Two extrapolation techniques were used. First, the correction took half of the difference between the computed CCSD(T) or CASSCF/MRCI+D (D denoting the inclusion of Davidson corrections[15]) values (i.e., optimized geometrical parameters or computed relative electronic energies) obtained using the AVQZ and AV5Z basis sets.[16] Second, the extrapolation technique, which employed the $1/X^3$ formula[17] with the computed values (i.e., optimized geometrical parameters or computed relative electronic energies) obtained using the AVQZ and AV5Z basis sets, was used. With the ACVQZ quality basis sets, the aug-cc-pwCVQZ basis sets[18] were used for Si and Cl, and the RCCSD(T) and CASSCF/MRCI calculations that were carried out employed both the default frozen core, and also the core with only the Si $1s^2$ and Cl $1s^2$ electrons frozen. Results obtained from these calculations with the ACVQZ basis sets were used to estimate contributions from including the $2s^22p^6$ core electrons of Si and Cl in the correlation treatments to the optimized geometrical parameters and computed relative electronic energies. The estimated contributions of basis size extrapolation to the CBS limit and core electron correlation to the optimized geometrical parameters and computed relative electronic energies have been assumed to be additive.

All *ab initio* calculations carried out in the present study, including the energy scans for the potential energy surfaces to be described below, have employed the MOLPRO suite of programs.[19]

Potential Energy Functions, Variational Calculations of Anharmonic Vibrational Wave Functions, and Franck−Condon Factor Calculations. RCCSD(T) and CASSCF/MRCI+D potential energy functions (PEFs) for the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ states of HSiCl, respectively, were obtained by fitting the following polynomial to calculated *ab initio* total electronic energies {RCCSD(T)/AV5Z and CASSCF/MRCI+D/AV5Z energies, respectively}:

$$\mathbf{V} = \sum_{ijk} C_{ijk}(S_1)^i(S_2)^j(S_3)^k + V_{\text{eqm}} \qquad (1)$$

In the above expression of the PEF, S_2 is the bending coordinate of Carter and Handy,[20] $S_2 = \Delta\theta + \alpha\Delta\theta^2 + \beta\Delta\theta^3$, where $\Delta\theta$ is the displacement of the bond angle from the equilibrium value, $(\theta - \theta_e)$. S_1 and S_3 are the displacements of the HSi and SiCl bond lengths from the equilibrium values, $(r - r_e)$, respectively; 377 CCSD(T)/AV5Z energy points in the ranges of $1.25 \leq r(\text{HSi}) \leq 1.99$ Å, $70.0 \leq \theta(\text{HSiCl}) \leq 130.0°$, and $1.81 \leq r(\text{SiCl}) \leq 2.55$ Å scanned for the $\tilde{X}^1A'$ state and 559 CASSCF/MRCI+D/AV5Z energy points in the ranges of $1.29 \leq r(\text{HSi}) \leq 2.45$ Å, $75.0 \leq \theta(\text{HSiCl}) \leq 163.0°$, and $1.83 \leq r(\text{SiCl}) \leq 2.60$ Å scanned for the $\tilde{A}^1A''$ state were used in the fitting of the PEFs. The nonlinear least-squares fit procedure, NL2SOL,[21] was employed to obtain the C_{ijk} values (with the restriction of $i + j + k \geq 2$; vide infra), V_{eqm}, r_e, θ_e, α, and β from the computed single point energy data.

For both the $\tilde{X}^1A'$ and the $\tilde{A}^1A''$ states, variational calculations, which employed the rovibronic Hamiltonian of Watson[22] for a nonlinear molecule, were carried out to obtain the anharmonic vibrational wave functions and energies. The anharmonic vibrational wave functions were expressed as linear combinations of harmonic oscillator functions, $h(v_1, v_2, v_3)$, where v_1, v_2, and v_3 denote the quantum numbers of the harmonic basis functions for the HSi stretching, HSiCl bending, and SiCl stretching modes, respectively, as described previously.[23] Harmonic basis functions with vibrational quantum numbers up to $h(7, 16, 16)$ and a restriction of $v_1 + v_2 + v_3 \leq 16$ were included in the variational calculations of the $\tilde{X}^1A'$ state of both HSiCl and DSiCl. For the $\tilde{A}^1A''$ state, harmonic basis functions up to $h(9, 12, 12)$, with a restriction of $v_1 + v_2 + v_3 \leq 12$ were considered for HSiCl. For the $\tilde{A}^1A''$ state of DSiCl, harmonic basis functions up to $h(8, 12, 12)$ were included. FC factors were calculated employing computed anharmonic vibrational wave functions and with allowance for Duschinsky rotation, as described previously.[24] The best estimated computed geometrical parameters of the two states (vide infra) were used in the FC calculations. Employing these computed FC factors for spectral simulations gave the best theoretical simulated spectra.

In spectral simulations, vibrational components of the $\tilde{A}^1A'' \rightarrow \tilde{X}^1A'$ SVL emission spectra of HSiCl and DSiCl were simulated using Gaussian functions with a full-width-at-half-maximum (fwhm) of 10 cm^{-1}. The relative intensity of each vibrational component in a simulated SVL emission spectrum was expressed as the product of the corresponding computed anharmonic FC factor and a frequency factor of power four. For simulations of $\tilde{A}^1A'' \leftarrow \tilde{X}^1A'$ absorption spectra of HSiCl and DSiCl, the relative intensity of each vibrational component was expressed as the corresponding computed anharmonic FC factor multiplied by a frequency factor of power one, and a Gaussian line shape with a fwhm of 10 cm^{-1}. In all spectral simulations carried out, the experimental adiabatic relative electronic energies (including zero-point energy corrections) between the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ states {i.e., the T_0 values of the $\tilde{A}^1A''$ state of HSiCl (20717.769 cm^{-1}) or DSiCl (20773.431 cm^{-1}) obtained from the LIF spectra[1]} were used for the sake of making direct comparison between simulated and observed spectra easier. Similarly, the wavenumber scale in each simulated SVL emission spectrum is taken relative to the laser excitation line, giving a direct measure of the ground electronic state vibrational energy, as normally used in published experimental dispersed fluorescence spectra.[2]

Results and Discussion

The optimized geometrical parameters, computed vibrational frequencies, and relative electronic energies of the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ states of HSiCl obtained from the present investigation are summarized and compared with previously calculated or available experimental values in Tables 1−3. In Table 4, some of the computed anharmonic vibrational energies of the $\tilde{X}^1A'$ state of HSiCl and DSiCl obtained from variational calculations in the present work are compared with available experimental values from ref 2 (EPAPS Supporting Information[25]). In addition, some computed excitation energies from the $\tilde{X}^1A'(0,0,0)$ level to various vibrational levels of the $\tilde{A}^1A''$ state of HSiCl and DSiCl, corresponding computed FC factors and upper vibrational state assignments obtained from the FC factor calculations of the present study are compared with available experimental transition energies given in the LIF study of ref 1 in Tables 5 and 6. Calculated and experimental transition energies of some "hot" bands are also compared in these two tables. The fitted RCCSD(T)/AV5Z and CASSCF/MRCI+D/AV5Z PEFs of the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ state of HSiCl have root-mean-square (rms) deviations of 9.4 and 23.1 cm^{-1}, respectively, from the computed *ab initio* energy data points, and the C_{ijk} values of the polynomials are given in Table 7. The rms deviation of the PEF of the $\tilde{A}^1A''$ state is relatively large, when compared with that of the $\tilde{X}^1A'$ state. This is mainly because considerably more energy points and wider ranges of the geometrical parameters were employed in the fitting (cf., those of the $\tilde{X}^1A'$ state) to make sure that the PEF is adequate for simulating the full absorption spectra of HSiCl and DSiCl, yet to be recorded. The full lists of the computed FC factors between the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ states of HSiCl and DSiCl are available, upon request, from the authors.

Table 1. Computed Geometrical Parameters (in Å and deg) and Vibrational Frequencies {ω_1 (HSi), ω_2 (Bending), and ω_3 (SiCl) in cm^{-1}} of the $\tilde{X}^1A'$ State of HSiCl Obtained at Different Levels of Calculations

methods	r_e(HSi)	r_e(SiCl)	θ_e	ω_1	ω_2	ω_3
CAS/MRCI/AVQZ^a	1.5085	2.0716	95.918			
CAS/MRCI+D^b/AVQZ^a	1.5158	2.0756	95.462			
CCSD(T)/AVQZ^a	1.5184	2.0792	95.215	2054.1	818.6	524.2
CAS/MRCI+D^b/ACVQZ^c	1.5144	2.0461	97.479			
CCSD(T)/ACVQZ^c (fc)	1.5178	2.0778	95.233			
CCSD(T)/ACVQZ^c	1.5144	2.0715	95.183			
CAS/MRCI/AV5Z^d	1.5081	2.0693	95.932			
CAS/MRCI+D^b/AV5Z^d	1.5154	2.0730	95.480			
CCSD(T)/AV5Z^d	1.5180	2.0768	95.046			
CCSD(T)/AV5Z^d PEF	1.5181	2.0767	95.059	2051.0	820.9	527.1
PEF (ν's)				1973.1	805.5	521.8
PEF (ω's; DSi^{35}Cl)				1476.4	600.3	522.3
PEF (ν's; DSi^{35}Cl)				1436.5	592.2	517.1
CBS1^e	−0.0002	−0.0012	−0.085			
CBS2^f	−0.0004	−0.0025	−0.177			
Core1^g	−0.0040	−0.0077	−0.032			
Core2^h	−0.0034	−0.0063	−0.050			
(CBS+Core)$_{\text{lower}}{}^i$	1.5136	2.0666	94.819			
(CBS+Core)$_{\text{upper}}{}^j$	1.5144	2.0693	94.930			
best theoretical r_e values^k	1.514(1)	2.068(1)	94.9(1)			
B3LYP/6−311G(3df,3pd)^l	1.5256	2.0968	94.8			
QCISD/6−311+G(3df,3pd)^l	1.515	2.083	95.3			
CCSD(T)/cc-pCVTZ^m	1.5187	2.0808	95.449	2044.6	821.7	529.8
CCSD(T)/cc-pCVQZ^m	1.5138	2.0697	95.302			
empirical cc-pCVTZ^m (r_e, ν's)	1.5140	2.0724	94.66	1964.60	808.56	528.18
CCSD(T)/cc-pV(Q+d)Z^n	1.5180	2.0778				
CCSD(T)/cc-pwCVQZ^n	1.5140	2.0700				
CCSD(T)/cc-pCVQZ^n	1.5147	2.0700	95.303			
absorption^o (r_0, ν's)	1.561	2.064	102.8		808	522
LIF^p (r_e^z, ν's)	1.525(5)	2.067(3)	96.9(5)	1968.7(4)	805.9(2)	522.8(1)
DF^l (ν's)				1968.8	805.9	522.8
DF^l (r_z, ω^0's)	1.5275(10)	2.0747(2)	94.97(11)	2002.2(7)	809.5(2)	525.1(4)
DF^l (r_e^z, ω's)	1.515(2)	2.0700(3)	95.0(1)	2044(1)	822(1)	529(1)
DF^l (ν's; DSi^{35}Cl)				1434.4	592.3	518.1
DF^l (ω^0's; DSi^{35}Cl)				1452(1)	595.1(1)	521.2(5)
DF^l (ω's; DSi^{35}Cl)				1473.2(10)	601.3(10)	525.3(10)

[a] The aug-cc-pV(Q+d)Z basis set was used for Si and Cl. [b] MRCI energies plus Davidson corrections. [c] The aug-cc-pwCVQZ basis set was used for Si and Cl, and their $2s^22p^6$ electrons were correlated except when specified with (fc); see text. [d] The aug-cc-pV(5+d)Z basis set was used for Si and Cl. [e] The correction of extrapolation to the complete basis set (CBS) limit takes half of the difference between the computed CCSD(T) values obtained using the AVQZ and AV5Z basis sets; see text. [f] The correction of extrapolation to the complete basis set (CBS) limit employs the $1/X^3$formula using the computed CCSD(T) values obtained with the AVQZ and AV5Z basis sets; see text. [g] The correction of core electron correlation contribution to the computed geometrical parameters takes the difference between the computed CCSD(T) values obtained using the AVQZ (with the default frozen core) and ACVQZ {frozen only the Si and Cl $1s^2$ electrons} basis sets. [h] The correction of core electron correlation contribution to the computed geometrical parameters takes the difference between the computed CCSD(T) values obtained using the ACVQZ basis set with and without the $2s^22p^6$ electrons of Si and Cl being correlated in the CCSD(T) calculations. [i] Assuming that the basis set extrapolation and core electron correlation contributions are additive, this combination of the CBS and core corrections gives the lower limit of the combined contribution. [j] Assuming that the basis set extrapolation and core electron correlation contributions are additive, this combination of the CBS and core corrections gives the upper limit of the combined contribution. [k] The best theoretical estimates take the averaged value of the upper and lower limits (see footnotes i and j) and the uncertainties are the differences between the best estimates and the upper/lower limits. [l] From ref 2. [m] From ref 5. [n] From ref 6. [o] From ref 28. [p] From ref 1.

Optimized Geometrical Parameters, Computed Vibrational Frequencies, and Relative Electronic Energies. Considering the computed equilibrium bond lengths, r_e(HSi) and r_e(SiCl), and bond angle, θ_e(HSiCl), of the $\tilde{X}^1A'$ state of HSiCl, it can be seen from Table 1 that, RCCSD(T) values obtained previously (note that only relatively high level results are considered and included in Table 1) and from the present study using different basis sets or frozen cores agree very well. Specifically, the maximum differences in computed RCCSD(T) r_e(HSi), r_e(SiCl), and θ_e(HSiCl) values are smaller than 0.005 Å, 0.011 Å, and 0.5°, respectively. When the computed RCCSD(T) geometrical parameters of the $\tilde{X}^1A'$ state obtained using augmented basis sets in the present study are considered (previous calculations[5,6] did not use augmented basis sets; see Table 1), the ranges of the computed values are even narrower (within 0.0036 Å, 0.0077 Å, and 0.19°, respectively). However, computed CASSCF/MRCI and CASSCF/MRCI+D geometrical parameters of the $\tilde{X}^1A'$ state cover wider ranges of values. In particular, the computed CASSCF/MRCI+D/ACVQZ bond angle of 97.48° is larger than all other computed bond angles by ~2°. Because the multireference character is not important for the $\tilde{X}^1A'$ state, as mentioned above, the RCCSD(T) results should be more reliable than the CASSCF/MRCI+D results. Consequently, contributions from extrapolation to the CBS limit and core correlation were estimated based only on the RCCSD(T) results, as shown in Table 1 (see footnotes e to k of Table

Table 2. Computed Geometrical Parameters (in Å and deg) and Vibrational Frequencies {ω_1 (HSi), ω_2 (Bending), and ω_3 (SiCl) in cm^{-1}} of the $\tilde{A}^1A''$ states of HSiCl Obtained at Different Levels of Calculations

methods	r_e(HSi)	r_e(SiCl)	θ_e	ω_1	ω_2	ω_3
CAS/MRCI/AVQZ[a]	1.5110	2.0506	117.957			
CAS/MRCI+D[b]/AVQZ[a]	1.5090	2.0505	117.931			
CAS/MRCI+D/ACVQZ[c] (fc)	1.5083	2.0491	117.964			
CAS/MRCI+D[b]/ACVQZ[c]	1.5046	2.0451	118.137			
CAS/MRCI/AV5Z[d]	1.5107	2.0480	117.939			
CAS/MRCI+D[b]/AV5Z[d]	1.5086	2.0476	117.909			
CAS/MRCI+D[b]/AV5Z[d] PEF	1.5083	2.0476	117.722	1962.2	594.0	536.4
PEF (ν's)				1771.1	569.2	527.0
PEF (ω's; DSi^{35}Cl)				1413.0	426.9	545.0
PEF (ν's; DSi^{35}Cl)				1318.0	414.1	537.7
CBS1[e]	−0.0002	−0.0015	−0.011			
CBS2[f]	−0.0004	−0.0030	−0.023			
Core1[g]	−0.0044	−0.0054	+0.206			
Core2[h]	−0.0037	−0.0040	+0.173			
(CBS+Core)$_{lower}$[i]	1.5038	2.0392	118.059			
(CBS+Core)$_{upper}$[j]	1.5047	2.0422	118.104			
best theoretical r_e values[k]	1.504(1)	2.041(2)	118.1(1)			
B3LYP/6−311G(3df,3pd)[l]	1.5012	2.066	115.5			
absorption[m] (r_0, ν's)	1.499	2.047	116.1	1250; 1756	568	533
LIF[n] (r_s, ν's)	1.511(9)	2.051(11)	116.3(3)	1747.08	563.87	532.32
LIF[n] (r_z, ω^0's)	1.548(7)	2.045(1)	118.1(4)	1839.6(5)	567.6(2)	534.8(2)
LIF[n] (r_e^z, ω's)	1.532(8)	2.040(3)	118.1(5)	1956.4	585.8	540.9
LIF[n] (ν's; DSi^{35}Cl)				1300.799	408.649	543.169
LIF[n] (ω^0's; DSi^{35}Cl)				1351.0(3)	409.5(2)	546.1(2)
LIF[n] (ω's; DSi^{35}Cl)				1408.3	418.2	552.2

[a] The aug-cc-pV(Q+d)Z basis set was used for Si and Cl. [b] MRCI energies plus Davidson corrections. [c] The aug-cc-pwCVQZ basis set was used for Si and Cl, and their 2s^{2}2p^6 electrons were correlated except when specified with (fc); see text. [d] The aug-cc-pV(5+d)Z basis set was used for Si and Cl. [e] The correction of extrapolation to the complete basis set (CBS) limit takes half of the difference between the computed MRCI+D values obtained using the AVQZ and AV5Z basis sets; see text. [f] The correction of extrapolation to the complete basis set (CBS) limit employs the $1/X^3$ formula using the computed MRCI+D values obtained with the AVQZ and AV5Z basis sets; see text. [g] The correction of core electron correlation contribution to the computed geometrical parameters takes the difference between the computed MRCI+D values obtained using the AVQZ (with the default frozen core) and ACVQZ {frozen only the Si and Cl 1s^2 electrons} basis sets; see text. [h] The correction of core electron correlation contribution to the computed geometrical parameters takes the difference between the computed MRCI+D values obtained using the ACVQZ basis set with and without the 2s^{2}2p^6 electrons of Si and Cl being correlated in the CASSCF/MRCI+D calculations. [i] Assuming that the basis set extrapolation and core electron correlation contributions are additive, this combination of the CBS and core corrections gives the lower limit of the combined contribution. [j] Assuming that the basis set extrapolation and core electron correlation contributions are additive, this combination of the CBS and core corrections gives the upper limit of the combined contribution. [k] The best theoretical estimates take the averaged value of the upper and lower limits (see footnotes i and j) and the uncertainties are the differences between the best estimates and the upper/lower limits. [l] From ref 2. [m] From ref 28. [n] From ref 1.

1 for the methods employed to obtain the best theoretical estimates and their uncertainties). It can be seen that these contributions (Core1, Core2, CBS1, and CBS2 in Table 1) are generally quite small. The best estimated theoretical geometrical parameters of the $\tilde{X}^1A'$ state of HSiCl have r_e(HSi) = 1.514 ± 0.001 Å, r_e(SiCl) = 2.068 ± 0.001 Å, and θ_e = 94.9 ± 0.1°. These values agree very well with the "empirical cc-pCVTZ" values of 1.5140 Å, 2.0724 Å, and 94.66° (see Table 1) obtained previously[5] and also with the latest, estimated experimental r_e^z values of 1.515 ± 0.002 Å, 2.0700 ± 0.0003 Å, and 95.0 ± 0.1° reported in the dispersed fluorescence study.[2] The "empirical cc-pCVTZ" r_e bond lengths and bond angle (see Table 1) were derived from equilibrium rotational constants obtained based on r_0 rotational constants (A_0, B_0, and C_0 from ref 1) corrected for vibrational effects employing centrifugal distortion and vibration−rotation constants derived from calculated harmonic and cubic force fields at the CCSD(T)/cc-pCVTZ level (see ref 5 for details).

For the $\tilde{A}^1A''$ state of HSiCl, the computed CASSCF/MRCI and CASSCF/MRCI+D geometrical parameters obtained using different basis sets or frozen cores are very consistent (see Table 2), unlike what was obtained for the $\tilde{X}^1A'$ state, as mentioned above. Moreover, similar to the $\tilde{X}^1A'$ state, contributions from extrapolation to the CBS limit and core correlation based on the more reliable CASSCF/MRCI+D values (c.f., the CASSCF/MRCI values without Davidson corrections; see Table 2) are generally small. The best estimated theoretical geometrical parameters obtained (see footnotes e to k of Table 2) have r_e(HSi) = 1.504 ± 0.001 Å, r_e(SiCl) = 2.041 ± 0.002 Å, and θ_e= 118.1 ± 0.1°. When these values are compared with the only previously computed values of ref 2, the B3LYP bond angle of 115.5° is clearly too small (see Table 2). If the the best estimated theoretical values obtained here are compared with the experimental r_e^z (estimated equilibrium) values of r_e(HSi) = 1.532 ± 0.008 Å, r_e(SiCl) = 2.040 ± 0.003 Å, and θ_e= 118.1 ± 0.5° derived in the LIF study of ref 1, the agreement of within 0.001 Å and 0.1° for r_e(SiCl) and θ_e is very good. However, the agreement for r_e(HSi) is not as good, with the experimental r_e^z(HSi) value larger than the best estimated theoretical value by 0.028 Å. Of particular relevance to the present study, the estimated experimental r_e^z(HSi) values increase upon excitation from the $\tilde{X}^1A'$ to $\tilde{A}^1A''$ state (as derived from the dispersed fluorescence and LIF studies of refs 2 and 1, respectively; see Tables 1and 2), but the corresponding best estimated theoretical values obtained here decrease upon excitation. Nevertheless, the experimental uncertainty of 0.008 Å given in ref 1 for the estimated equilibrium r_e^z(HSi) value of the $\tilde{A}^1A''$ state of HSiCl is significantly larger than that estimated for r_e^z(SiCl). In addition, the experimentally derived r_0(HSi) values of the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ states of HSiCl (1.5214 ± 0.0008 Å and 1.505 ± 0.005 Å, respectively) reported in the LIF study[1] actually give a decrease in the r_0(HSi) bond length upon excitation. Similarly, the derived $r_0°$ and r_s HSi bond lengths also give decreases upon excitation (see ref 1 for details). It is only after including harmonic contributions, based on normal coordinate analysis and harmonic force fields, to obtain the r_z and r_e^z geometrical parameters that increases in the HSi bond length upon excitation are obtained. These increases are mainly because the r_z and r_e^z HSi bond lengths of the $\tilde{A}^1A''$ state become significantly larger than the r_0 and r_s values. It was noted in ref 1 that available data used in deriving the harmonic force fields were insufficient to determine some interaction force constants, which were assumed to be 0.0 in the fitting (see ref 1 for details). This may be the cause of the discrepancy between the best estimated theoretical r_e(HSi) bond length of the $\tilde{A}^1A''$ state of HSiCl obtained here and the experimental r_e^z value reported in ref 1. In any case, the best estimated theoretical r_e(HSi) value of the $\tilde{A}^1A''$ state of HSiCl agrees very well with the experimentally derived r_0 and r_s values of 1.505 ± 0.005 and 1.511 ± 0.009 Å, respectively, given in ref 1.

Regarding computed vibrational frequencies of the $\tilde{X}^1A'$ state of HSiCl, the differences between the computed harmonic values obtained using the AVQZ (numerical second derivative calculations) and AV5Z (variational calculations using the PEF) basis sets obtained in the present study are less than 4 cm^{-1} for all three vibrational modes (see Table 1). Comparing the computed fundamental frequencies of all three vibrational modes of the $\tilde{X}^1A'$ states of both HSiCl and DSiCl obtained here employing the RCCSD(T)/AV5Z PEF with the corresponding experimental values obtained from the dispersed fluorescence study,[2] the maximum discrepancy is less than 5 cm^{-1}. For the harmonic frequencies, the maximum difference between computed and experimentally derived values is 7 cm^{-1} for ω_1'', the HSi stretching mode, of HSiCl. Summing up, the agreement between theory and experiment is very good for the vibrational frequencies of the $\tilde{X}^1A'$ states of both HSiCl and DSiCl, suggesting that the RCCSD(T)/AV5Z PEF is highly reliable, particularly near the bottom of the potential energy well. For the $\tilde{A}^1A''$ state of HSiCl and DSiCl, the agreement between theory and experiment for the vibrational frequencies of the bending and SiCl stretching modes is within 10 cm^{-1} (see Table 2), which is reasonably good. For the HSi stretching mode, the computed ν_1' of HSiCl (1771.1 cm^{-1}; see Table 2) is larger than the experimental value (1747.08 cm^{-1}) by ~24 cm^{-1}, which is the largest discrepancy between computed and experimental vibrational frequencies of the $\tilde{A}^1A''$ states of HSiCl and DSiCl. Nevertheless, the computed and experimentally derived ω_1' values of the $\tilde{A}^1A''$ state of HSiCl agree to within 6 cm^{-1}. In addition, for the DSi stretching mode, the computed ν_1' value agree with the experimental value to within 18 cm^{-1}. It should be noted that the difference between the harmonic and fundamental frequencies of the HSi stretch mode is quite large (~200 cm^{-1} whether based on the computed or experimentally derived values; see Table 2), indicating large anharmonic effects. Employing linear combinations of harmonic basis functions for anharmonic vibrational wave functions may be inadequate for the HSi stretching mode of HSiCl.

Computed adiabatic electronic energies, T_e, of the $\tilde{A}^1A''$ state (relative to the $\tilde{X}^1A'$ state) of HSiCl obtained at different levels of calculation are summarized in Table 3. It can be seen that contributions from basis set extrapolation (CBS1 and CBS2 in Table 3) to the computed T_e values are very small (<0.0005 eV). However, core correlation contributions are significant (ca., −0.07 eV). The best estimated theoretical T_e value is 2.564 ± 0.067 eV (20683 ± 538 cm^{-1}). Including correction of zero-point vibrational energies of the two states involved (ΔZPE in Table 3), the best theoretical T_0 value is 2.54 ± 0.07 eV (20456 ± 560 cm^{-1}). It is pleasing that this value agrees with the experimental value of 2.5687 eV (20717.769 cm^{-1}) obtained from the LIF study[1] to within 0.03 eV (262 cm^{-1}). The best theoretical T_0 value of the $\tilde{A}^1A''$ state of DSiCl has also been estimated and given in Table 3, and it agrees with the experimental value also to within 0.03 eV (236 cm^{-1}). In summary, the *ab initio* results obtained in the present study agree very well with available experimental values.

Computed FC Factors and Simulated Absorption and SVL Emission Spectra. Before the simulated spectra are discussed, some computed and experimental vibrational energies of the $\tilde{X}^1A'$ state of HSiCl and DSiCl are compared in Table 4 (the form of the vibrational designations used follows those given in refs 2 and 25). It can be seen that agreements between computed and experimental energies are reasonably good, particularly for low-lying vibrational levels. Therefore, on energy grounds, all the vibrational assignments given in the dispersed fluorescence study[2] for the $\tilde{X}^1A'$ state of HSiCl and DSiCl are confirmed. Nevertheless, for a few vibrational levels, there are other vibrational levels, which are very close in energy according to our variational calculations of anharmonic vibrational wave functions of the $\tilde{X}^1A'$ states of HSiCl and DSiCl, and hence, alternative assignments are possible based on our computed anharmonic vibrational energies. These are the $1_1 2_7$ level of HSiCl and the 2_{11} and 2_{12} levels of DSiCl (as shown with two energy entries under "calcd" in Table 4). However, for these relatively high energy vibrational levels with high quantum numbers in the bending mode (υ_2''), the computed anharmonic vibrational wave functions are rather mixed with no single dominant harmonic basis functions. This is also the case for high vibrational levels of the $\tilde{A}^1A''$ state to be discussed below, particularly for HSiCl, where ν_2' and ν_3' are close in energy {difference 31.55 (experimental value) and 42.3 (computed value) cm^{-1}} and their normal modes show mixing of bending and SiCl stretching. For some of these levels, which are heavily mixed, the vibrational designations have been determined based on both the computed anharmonic vibrational wave functions and the expected vibrational spacing of a series, and they may be tentative.

Table 3. Computed Relative Electronic Energies in eV (cm^{-1}) between the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ States of HSiCl Obtained at Different Levels of Calculations

method[a]	T_e
CASSCF/MRCI/AVQZ[b]	2.6602 (21456.3)
CASSCF/MRCI+D/AVQZ[b]	2.6306 (21217.4)
CASSCF/MRCI+D/ACVQZ[b] (fc)	2.6291 (21205.3)
CASSCF/MRCI+D/AVCQZ[c]	2.5488 (20557.1)
CASSCF/MRCI+D/AVCQZ[b]	2.5628 (20670.2)
CASSCF/MRCI/AV5Z[b]	2.6533 (21400.3)
CASSCF/MRCI+D/AV5Z[b]	2.6311 (21220.9)
CBS1[d]	+0.00022 (+1.8)
CBS2[e]	+0.00047 (+3.7)
Core1[f]	−0.06782 (−547.0)
Core2[g]	−0.06634 (−535.1)
$(\mathrm{CBS+Core})_{\mathrm{lower}}$[h]	2.5635 (20675.7)
$(\mathrm{CBS+Core})_{\mathrm{upper}}$[i]	2.5652 (20689.5)
best theoretical T_e value[j]	2.564 ± 0.067 (20683 ± 538)
ΔZPE correction[k]	−0.02815 (−227.07)
best theoretical T_0 value[l]	2.54 ± 0.07 (20456 ± 560)
ΔZPE correction[k]	−0.01811 (−146.09)
best theoretical T_0 value[l] (DSiCl)	2.55 ± 0.07 (20537 ± 560)
absorption[m] T_0	2.5687 (20717.65)
LIF[n] T_0	2.5687 (20717.769)
LIF[n] T_0 (DSi^{35}Cl)	2.5756 (20773.431)

[a] For the basis set used, see footnotes of Table 1.

[b] At the corresponding optimized geometries.

[c] The $\tilde{X}^1A'$state CASSCF/MRCI+D/ACVQZ energy was computed at the $\tilde{X}^1A'$ CCSD(T)/AVQZ optimized geometry, because the computed θ_e value of the $\tilde{X}^1A'$ state obtained at the CASSCF/MRCI/ACVQZ level is significantly larger than values obtained with the RCCSD(T) method. It can be seen that this geometry effect on the computed T_e value is small (0.014 eV).

[d] The correction of extrapolation to the complete basis set (CBS) limit takes half of the difference between the computed MRCI+D values obtained using the AVQZ and AV5Z basis sets; see text.

[e] The correction of extrapolation to the complete basis set (CBS) limit employs the $1/X^3$ formula using the computed MRCI+D values obtained with the AVQZ and AV5Z basis sets; see text.

[f] The correction of core electron correlation takes the difference between the computed MRCI+D values obtained using the AVQZ (with the default frozen core) and ACVQZ {frozen only the Si and Cl $1s^2$ electrons} basis sets; see text.

[g] The correction of core electron correlation takes the difference between the computed MRCI+D values obtained using the ACVQZ basis set with and without the $2s^22p^6$ electrons of Si and Cl being correlated in the CASSCF/MRCI+D calculations.

[h] Assuming that the basis set extrapolation and core electron correlation contributions are additive, this combination of the CBS and core corrections gives the lower limit of the combined contribution.

[i] Assuming that the basis set extrapolation and core electron correlation contributions are additive, this combination of the CBS and core corrections gives the upper limit of the combined contribution.

[j] The best theoretical estimate takes the averaged value of the upper and lower limits (see footnotes h and i) and the uncertainty is the difference between the best estimate and the MRCI+D/AVQZ value (in view of the rather large core correlation contributions; see text).

[k] The zero-point vibrational energy correction has employed experimental fundamental vibrational frequencies of the two states from refs 1 and 2, respectively.

[l] $T_0 = T_e + \Delta\mathrm{ZPE}$.

[m] From ref 28.

[n] From ref 1.

Computed (E_{calcd} in Table 5) and experimental (E_{exptl}; from the LIF study of ref 1) excitation energies from the $\tilde{X}^1A'(0,0,0)$ level to the $\tilde{A}^1A''(v_1',v_2',v_3')$ levels of HSiCl and also excitation energies of some "hot" bands, with their corresponding computed FC factors (computed at a Boltzmann vibrational temperature of 300 K), are compared in Table 5. Some possible alternative assignments to those given in the LIF study are also given in Table 5, on the basis of the computed FC factors and anharmonic vibrational energies. These alternative assignments are mainly for transitions, where the discrepancies between the computed and experimental energies are larger than ~20 cm^{-1}, the experimental vibrational spacings differ significantly from what would be expected for the assigned progression or the computed FC factors with the assignments of ref 1 are relatively small. For example, for the $\tilde{A}^1A''(1,v_2',0)''$ $\tilde{X}^1A'(0,0,0)$ progression (with $v_2' = 1-4$; based on the assignments of the LIF study), the experimental transition energies are smaller than the corresponding computed energies by over 60−145 cm^{-1} (see Table 5). In addition, the experimental vibrational spacings of between ~498 and 463 cm^{-1} are significantly smaller than the corresponding computed values of between 538 and 498 cm^{-1}. The latter, larger computed vibrational spacings are reasonable for a progression involving the bending mode, which has a ν_2' value of 569.2 (computed value) or 563.87 (experimental value; see Table 2) cm^{-1}. In this connection, the experimental spacings seem to be too small for a vibrational progression involving ν_2'. Consequently, alternative assignments based on computed transition energies were found for these observed vibrational components. Most of the alternative assignments shown in Table 5 give better agreement between the computed and experimental transition energies, more reasonable vibrational spacings between the vibrational components concerned, or larger computed FC factors than those computed for the original assignments given in the LIF study.[1] Nevertheless, some alternative assignments also give possible weaker overlapping vibrational components (vide infra).

Similar to that shown in Table 5, computed (E_{calcd}) and experimental (E_{exptl}) excitation energies from the $\tilde{X}^1A'(0,0,0)$ level to various $\tilde{A}^1A''(v_1',v_2',v_3')$ levels of DSiCl and also excitation energies of some hot bands are compared in Table 6. Some possible alternative assignments are also included in the table, based on what has been discussed above for Table 5. For both HSiCl and DSiCl, it appears that most of the vibrational assignments of combination bands involving the HSi or DSi stretching mode given in ref 1 must be revised on the basis of the results of FC calculations carried out in the present investigation. In addition, hot bands arising from the excited vibrational level of the $\tilde{X}^1A'$ states of HSiCl and DSiCl with $v_1'' = 1$ (see Tables 5 and 6) have small computed FC factors at a Boltzmann vibrational temperature of 300 K, because ν_1'' has a relatively large value {1968.8 and 1434.4 cm^{-1} for HSiCl and DSiCl, respectively (experimental values; see Table 1); cf. the other two vibrational modes}. Consequently, alternative assignments with larger computed FC factors at similar transition energies were found for these observed vibrational components, assigned to hot bands arising from excited vibrational levels with $v_1'' = 1$ in ref 1. However, it should be noted that a Boltzmann vibrational temperature of 300 K has been assumed in the FC factor calculations. HSiCl or DSiCl molecules produced in an electric discharge of $SiHCl_3$ or $SiDCl_3$, respectively, diluted in Ar through a pulsed jet[1] may have a vibrational temperature higher than 300 K or may be excited preferentially in the HSi or DSi stretching modes (i.e., with a non-Boltzmann distribution). Because in their formation, there

Table 4. Computed Vibrational Energies (cm^{-1}) of the $\tilde{\mathrm{X}}^1\mathrm{A}'$ State of HSiCl and DSiCl and Their Assignments and the Corresponding Experimental Values[25] from the SVL Emission (DF, Dispersed Fluorescence) Study of Ref 2

	HSiCl		DSiCl	
assignment	DF	calcd	DF	Calcd
1_1	1968.8	1973.1	1434.4	1436.5
1_2	3870.1	3879.3	2832.9	2837.5
2_1	805.9	805.5	592.3	592.2
2_2	1607.7	1604.9	1183.6	1180.8
2_3	2400.7	2398.5	1769.0	1765.9
2_4	3190.9	3186.7	2354.4	2347.7
2_5	3977.5	3969.6	2932.6	2926.2
2_6	4755.5	4747.3	3509.2	3501.7
2_7	5526.4	5519.8	4082.7	4074.0
2_8	6292.6	6287.6	4652.2	4643.6
2_9			5217.7	5210.5
2_{10}			5781.6	5776.0
2_{11}			6338.0	6336.6; 6343.4
2_{12}			6893.3	6881.1; 6913.8
3_1	522.8	521.8	518.1	517.1
3_2	1037.9	1039.7	1031.2	1030.4
3_3	1553.5	1553.5		
3_4	2061.5	2063.4		
1_12_1	2758.7	2763.9	2020.3	2021.8
1_12_2	3543.6	3548.5	2603.4	2605.3
1_12_3	4324.9	4327.5	3184.3	3184.2
1_12_4	5100.9	5101.1	3759.0	3759.7
1_12_5	5865.5	5869.4	4333.6	4331.7
1_12_6	6628.6	6633.0	4901.6	4900.7
1_12_7	7384.4	7381.8; 7393.2	5466.6	5466.7
1_12_8			6029.6	6030.3
1_13_1	2491.4	2495.9	1952.1	1953.3
1_23_1	4390.4	4403.0		
2_13_1	1324.8	1323.9	1109.4	1107.0
2_23_1	2125.0	2119.9	1698.3	1693.2
2_33_1	2914.8	2910.4	2280.7	2276.1
2_43_1	3700.9	3695.7	2863.6	2855.6
2_53_1	4458.0	4475.8	3439.2	3432.1
2_63_1	5260.0	5250.9	4016.3	4005.5
2_73_1	6027.2	6021.4	4584.7	4576.0
2_83_1	6787.0	6788.9	5154.1	5143.9
2_93_1	7546.3	7557.8	5717.0	5710.6
$2_{10}3_1$			6279.4	6279.2
$2_{11}3_1$			6835.8	6834.1
2_13_2	1841.3	1838.2	1622.1	1617.8
2_23_2	2634.9	2631.0	2205.3	2201.8
2_33_2	3423.6	3418.3	2786.4	2782.4
2_43_2	4206.3	4200.6	3365.8	3359.8
2_53_2	4987.3	4978.0	-	
2_63_2	5757.3	5750.9	4512.3	4505.5
2_73_2	6522.3	6520.5	5081.1	5074.5
2_83_2			5647.6	5642.2
2_93_2			6211.7	6210.4
$2_{10}3_2$			6770.4	6773.9
2_13_3	2349.6	2348.6		
2_23_3	3141.8	3138.0		
2_33_3	3926.3	3922.3		
2_43_3	4706.0	4701.6	3864.8	3860.1
2_53_3	5484.3	5476.5	4438.2	4432.5
2_63_3			5008.1	5002.4
2_73_3			5574.8	5571.6
2_13_4	2855.2	2855.0		
2_23_4	3643.9	3641.1		
2_33_4	4425.8	4422.3		
2_43_4	5202.2	5199.0		
2_53_4	5974.2	5972.8		
$1_12_13_1$	3279.4	3283.0	2535.6	2537.1
$1_12_23_1$	4063.9	4064.3	3116.9	3117.2
$1_12_33_1$	4836.8	4840.1		
$1_12_43_1$	5608.2	5610.6	4271.9	4267.0
$1_12_53_1$	6374.0	6376.3	4837.2	4836.9
$1_12_63_1$	7134.0	7138.2		
$1_22_13_1$	5165.6	5174.5		
$1_22_23_1$	5931.9	5941.6		
$1_22_33_1$	6693.1	6704.1		
$1_22_43_1$	7449.6	7463.1		
$1_12_13_2$			3047.1	3047.5

is an increase in the HSi bond length from 1.464 Å {r_s(HSi) in $\mathrm{HSiCl_3}$}[26] to 1.515 Å in HSiCl (r_e^z value from ref 2).

The simulated absorption spectra of HSiCl and DSiCl at a Boltzmann vibrational temperature of 300 K are shown in Figures 1 and 2, respectively. The assignments of some major vibrational progressions and some hot bands are also indicated in these figures. It can be seen that the main simulated vibrational progression is $\tilde{\mathrm{A}}^1\mathrm{A}''(0,\upsilon_2',0) \leftarrow \tilde{\mathrm{X}}^1\mathrm{A}'(0,0,0)$, involving the bending mode in the upper state of both HSiCl and DSiCl, as expected from the major geometry change in the bond angle (see Tables 1 and 2) upon excitation. Based on the computed FC factors, some combination bands involving ν_2' and ν_3' should be excited, particularly for HSiCl (Figure 1). The full simulated absorption spectra of HSiCl and DSiCl cover excitation energies in the range between ~20 000 and 26 000 cm^{-1}. The band maxima are at 21 837.25 and 22 384.08 cm^{-1} (experimental values; Tables 5 and 6) corresponding to the $\tilde{\mathrm{A}}^1\mathrm{A}''(0,2,0) \leftarrow \tilde{\mathrm{X}}^1\mathrm{A}'(0,0,0)$ and $\tilde{\mathrm{A}}^1\mathrm{A}''(0,4,0) \leftarrow \tilde{\mathrm{X}}^1\mathrm{A}'(0,0,0)$ vibrational components for HSiCl and DSiCl, respectively. In addition to the major vibrational progressions shown in Figures 1 and 2, there are a large number of minor combination bands and hot bands according to the FC factor calculations (the lists are available upon request from the authors).

A portion of the simulated absorption spectrum of HSiCl is compared with the corresponding portion of the experimental LIF spectrum[1] in Figure 3 (top and bottom traces, respectively). Computed relative FC factors in the same spectral region are also shown as a bar diagram in Figure 3 (middle) to show some vibrational components, which are very close in energy and hence cannot be resolved in the simulated spectrum with a fwhm of 10 cm^{-1}. It can be seen from Figure 3 that the agreement between the simulated absorption and experimental LIF spectra is reasonably good, bearing in mind the spread of the rotational structure and uncorrected intensity (frequency-dependent laser intensity and detector sensitivity) of the experimental spectrum and that the published experimental LIF spectrum may consist of different fragment portions recorded under different experimental conditions. In any case, the assignments of the main vibrational structure given by ref 1 are confirmed by our spectral simulation. However, the weak hot bands assigned to $1_1^12_0^1$ and $1_1^12_0^2$ in the experimental LIF spectrum have very small computed relative FC factors (0.003 and 0.004, with the maximum value set to 100; see Table 5) and hence do not appear in the simulated spectrum with a Boltzmann vibrational temperature of 300 K. Instead, the 2_1^2 and $2_1^13_0^1$ hot bands are found in the simulated absorption spectrum to be roughly in the position of the weak structure assigned to the $1_1^12_0^1$ hot band in the experimental LIF spectrum and to have more intensity (see also Table 5). In addition, computed FC factors (the bar diagram in the middle of Figure 3) suggest that there are some weak hot bands, namely 3_1^1, 3_1^2, $2_0^13_1^1$, $2_0^13_1^2$, and $2_0^23_1^1$, underneath the main 0_0^0, 3_0^1, 2_0^1, $2_0^13_0^1$, and 2_0^2 vibrational components, respectively (see also Table 5).

Portions of the simulated absorption and experimental LIF spectra of DSiCl are compared in Figure 4 (bottom and top traces). Similar to HSiCl discussed above, the general

Table 5. Computed (E_{calcd}) and Experimental (E_{exptl}; from ref 1) Excitation Energies (in cm^{-1}) from the $\tilde{X}^1A'(0,0,0)$ Vibronic Level to the $\tilde{A}^1A''(v_1',v_2',v_3')$ Levels of HSiCl, Some "Hot" Bands with the Computed Franck−Condon (FC) Factors, and Possible Alternative Assignments Based on the Computed FC factors and Anharmonic Vibrational Energies (See Text)

LIF		calculation	
(v_1',v_2',v_3')	E_{exptl}	E_{calcd} [FCF]	alternative assignments
(0,0,0)	20717.769	20717.769^a [18.3]	20722.9 [1.5] 3_1^1
(1,0,0)	22464.85	22488.9 [0.1]	22422.1 [6.1] $2_0^4 3_1^0$; 22440.3 [0.4] $2_0^5 3_2^0$
(0,1,0)	21281.64	21287.0 [63.7]	21288.3 [3.7] $2_0^1 3_1^1$
(0,2,0)	21837.25	21848.0 [100.0]	21845.0 [3.1] $2_0^2 3_1^1$
(0,3,0)	22384.94	22400.4 [84.2]	
(0,4,0)	22925.27	22943.9 [33.5]	
(0,0,1)	21250.09	21244.8 [6.7]	21245. [0.9] 3_1^2
(0,0,2)	21777.89	21767.3 [1.2]	
(0,0,3)	22301.80	22285.4 [0.1]	
(0,0,4)	22819.76	22799.0 [0.005]	
(1,1,0)	22971.92	23032.9 [0.8]	23000.0 [1.0] (1,0,1); 22958.1 [3.7] $2_0^5 3_1^0$
(1,2,0)	23470.37	23571.2 [0.9]	23479.9 [4.2] (0,5,0)
(1,3,0)	23955.77	24064.7 [1.5]	23977.8 [0.8] (0,6,0)
(1,4,0)	24419.72	24564.2 [1.6]	24428.8 [0.2] (0,5,2)
(2,1,0)	24445.82	24534.0 [0.5]	24450.0 [0.1] (4,3,0)
(1,0,1)	23008.76	23000.0 [1.0]	
(1,0,2)	23547.89	23504,9 [2.7]	23535.0 [2.8] (1,1,1); 23542.9 [0.2] $1_0^1 2_0^3 3_1^0$
(1,0,3)	24081.38	24002.3 [4.7]	24064.7 [1.5] (1,3,0); 24069.6 [0.2] $1_0^1 2_0^1 3_1^3$
(2,0,1)	24527.72	24600.9 [0.03]	24520.0 [1.6] (1,1,3); 24519.8 [0.2] $1_0^1 3_1^5$
(0,1,1)	21808.75	21810.2 [30.1]	21807.6 [3.4] $2_0^1 3_1^2$
(0,2,1)	22357.40	22366.9 [57.7]	
(0,3,1)	22896.20	22913.3 [55.7]	
(0,4,1)	23427.25	23448.8 [18.6]	23421.9 [16.2] (0,3,2)
(0,1,2)	22331.55	22329.3 [6.3]	
(0,2,2)	22871.88	22882.0 [14.6]	
(0,3,2)	23399.93	23421.9 [16.2]	23393.5 [1.9] (0,2,3)
(0,1,3)	22849.84	22844.3 [0.6]	
(0,2,3)	23378.10	23393.5 [1.9]	23380.3 [0.3] $2_0^2 3_1^4$
(0,1,4)	23362.73	23355.4 [0.03]	
(1,1,1)	23508.38	23535.0 [2.8]	23504.9 [2.7] (1,0,2)
(1,2,1)	23999.81	24104.6 [0.3]	24002.3 [4.7] (1,0,3)
(1,3,1)	24479.14	24634.2 [0.03]	24489.4 [4.0] (1,0,4)
(1,1,2)	24039.56	24030.7 [3.9]	
1_1^1	20496.11	20515.8 [0.002]	20481.5 [2.0] 2_1^1
1_1^2	22023.31	22095.2 [0.0008]	22038.8 [0.015] $2_1^1 3_0^3$; 22031.5 [0.003] $2_1^1 3_1^4$
$1_1^1 2_0^1$	21003.21	21059.7 [0.003]	21004.7 [0.17] $2_1^1 3_0^1$; 21042.7 [1.0] 2_1^2
$1_1^1 2_0^2$	21501.55	21561.9 [0.004]	21523.8 [0.006] $2_1^1 3_0^2$
$1_1^1 2_0^3$	21986.95	22091.6 [0.0008]	21993.5 [0.0002] $2_1^0 3_0^4$
$1_1^1 3_0^1$	21040.03	21026.8 [0.003]	21042.5 [1.0] 2_1^2
$1_1^1 3_0^2$	21578.89	21561.9 [0.004]	21594.9 [0.006] 2_1^3; 21561.3 [0.003] $2_1^2 3_0^1$
$1_1^1 2_0^1 3_0^1$	21539.59	21531.7 [0.004]	21523.8 [0.006] $2_1^1 3_0^2$
$1_1^1 2_0^1 3_0^2$	22030.93	22029.1 [0.004]	21042.5 [1.0] 2_1^2; 21043.0 [0.03] $2_1^2 3_1^1$
2_1^0	19911.81	19912.3 [1.2]	
2_1^1	20475.66	20481.5 [2.0]	
$2_1^0 3_0^1$	20444.15	20439.3 [0.1]	20443.5 [0.02] $2_1^0 3_1^2$
$1_0^1 2_1^4$	23614.19	23828.8 [0.001]	23623.4 [0.08] $2_1^5 3_0^2$
$1_0^2 2_1^1$	23640.28	23728.5 [0.04]	23644.5 [0.07] $2_1^4 3_0^3$
$1_0^1 2_1^3 3_0^1$	23673.79	23683.9 [0.3]	
3_1^0	20194.95	20195.9 [0.03]	
$2_0^1 3_1^0$	20758.75	20765.2 [0.04]	
$2_0^2 3_1^0$	21314.45	21326.2 [1.1]	
$2_0^3 3_1^0$	21861.91	21878.6 [4.0]	
$2_0^4 3_1^0$	22402.45	22422.1 [6.0]	
3_1^1	20727.35	20722.9 [1.5]	
3_1^2	21255.11	21245.5 [0.9]	
3_1^4	22296.72	22277.2 [0.02]	
$1_0^1 2_0^1 3_1^0$	22449.06	22511.0 [0.008]	22422.1 [6.1] $2_0^4 3_1^0$
$1_0^1 2_0^2 3_1^0$	22947.41	23049.4 [0.08]	22958.1 [3.7] $2_0^5 3_1^0$
$1_0^1 3_1^1$	22485.95	22478.1 [0.004]	22488.9 [0.1] (1,0,0); 22489.6 [0.6] $2_0^5 3_1^1$
$1_0^1 3_1^2$	23024.83	22983.0 [0.04]	23032.9 [0.8] (1,1,0); 23049.4 [0.08] $1_0^1 2_0^2 3_1^0$
$2_0^1 3_1^1$	21285.93	21288.3 [3.7]	
$2_0^3 3_1^1$	22373.22	22391.4 [0.3]	
$1_0^1 2_0^1 3_1^1$	22985.56	23013.2 [0.01]	22971.8 [0.3] $2_0^6 3_2^0$; 22983.0 [0.04] $1_0^1 3_1^2$
$2_1^0 3_1^1$	19924.67	19920.9 [0.09]	19912.3 [1.2] 2_1^0

[a] Fixed to the experimental value.

agreement between theory and experiment is reasonably good, and most of the assignments of the main vibrational structure can be confirmed. However, the 3_0^2 and 1_0^1 vibrational components have very small computed FC factors (0.3 and 0.02, respectively; see Table 6) and hence are not observable in the simulated absorption spectrum (Figure 3 bottom trace). Nevertheless, such discrepancies between the simulated absorption and experimental spectra could be because the latter has not been corrected for frequency dependent intensity or because it consists of different portions

Table 6. Computed (E_{calcd}) and Experimental (E_{exptl}) Excitation Energies (in cm^{-1}) from the $\tilde{X}^1A'(0,0,0)$ Vibronic Level to the $\tilde{A}^1A''(v_1',v_2',v_3')$ Levels of DSiCl, Some "Hot" Bands, with the Computed Franck−Condon (FC) Factors, and Possible Alternative Assignments Based on the Computed FC factors and Anharmonic Vibrational Energies (See Text)

LIF		calculation	
(v_1',v_2',v_3')	E_{exptl}	E_{calcd} [FCF]	alternative assignments
(0,0,0)	20773.431	20773.431[a] [4.9]	
(1,0,0)	22074.23	22090.4 [0.02]	22061.2 [0.05] $2_1^2 3_0^2$
(0,1,0)	21182.08	21186.6 [23.7]	21203.9 [2.0] $2_0^1 3_1^1$
(0,2,0)	21587.19	21595.8 [57.6]	21609.9 [3.8] $2_0^2 3_1^1$
(0,3,0)	21988.07	22001.6 [90.5]	22012.3 [4.6] $2_0^3 3_1^1$
(0,4,0)	22384.08	22403.7 [100.0]	22411.0 [4.1] $2_0^4 3_1^1$
(0,5,0)	22774.51	22801.7 [78.8]	22775.6 [0.09] (0,1,3); 22805.8 [2.6] $2_0^5 3_1^1$
(0,6,0)	23158.20	23194.9 [42.3]	23156.6 [0.009] (1,0,2); 23197.6 [1.3] $2_0^6 3_1^1$
(0,7,0)	23533.85	23584.1 [13.3]	23545.5 [0.04] (1,1,2)
(0,8,0)	23899.33	23974.2 [1.1]	23881.2 [0.3] $2_0^9 3_1^0$; 23903.2 [0.06] $1_0^2 2_0^3 3_1^0$
(0,9,0)	24252.87	24369.4 [0.8]	24237.4 [0.07] (0,6,2); 24259.3 [0.01] $2_0^6 3_1^3$; 24263.3 [0.03] $1_0^2 2_0^4 3_1^0$
(0,10,0)	24591.74	24816.5 [0.3]	24571.4 [0.2] $1_0^1 2_1^0 3_0^8$
(0,0,1)	21316.60	21311.1 [2.0]	
(0,0,2)	21854.48	21844.1 [0.3]	21855.3 [0.08] 3_1^3
(1,1,0)	22464.33	22485.9 [0.3]	22460.2 [0.3] $2_1^2 3_0^3$
(1,2,0)	22849.09	22876.5 [1.4]	22855.8 [0.4] $2_1^2 3_0^4$
(1,3,0)	23228.62	23262.4 [3.9]	23194.9 [41.2] (0,6,0)
(1,4,0)	23601.73	23643.6 [6.4]	23584.1 [13.3] (0,7,0); 23591.4 [0.3] $2_0^7 3_1^1$
(1,5,0)	23966.83	24020.2 [6.6]	23974.2 [1.1] (0,8,0); 23951.1 [0.3] $1_0^1 2_1^5 3_0^1$
(1,6,0)	24321.72	24398.2 [2.4]	24309.3 [0.1] (1,3,2)
(1,7,0)	24663.59	24748.5 [0.8]	24686.8 [0.09] (1,4,2); 24646.5 [0.08] $1_0^1 2_0^8 3_1^0$
(2,1,0)	23641.52	23689.1 [0.00007]	23643.6 [6.4] (1,4,0)
(2,2,0)	24001.68	24056.6 [0.07]	24020.2 [6.6] (1,5,0)
(2,3,0)	24353.33	24420.3 [0.6]	24369.4 [0.8] (0,9,0); 24358.6 [0.1] $2_1^9 3_0^1$
(1,0,1)	22615.03	22625.9 [0.03]	22602.6 [3.0] 2_1^6
(0,1,1)	21722.07	21721.1 [8.7]	
(0,2,1)	22123.74	22127.0 [18.5]	
(0,3,1)	22521.05	22529.4 [24.1]	
(0,4,1)	22913.49	22928.1 [20.5]	
(0,5,1)	23299.85	23322.9 [11.2]	
(0,6,1)	23679.05	23714.7 [3.6]	23682.4 [0.001] (1,0,3); 23645.2 [0.1] $2_1^6 3_0^2$
(0,1,2)	22256.84	22250.7 [1.3]	
(0,2,2)	22655.04	22653.6 [2.2]	
(0,3,2)	23048.72	23052.4 [2.2]	
(0,4,2)	23437.17	23448.0 [1.3]	23448.2 [0.3] $2_0^4 3_1^3$
(1,1,1)	23001.42	23018.2 [0.2]	22991.8 [3.3] 2_1^7
(1,2,1)	23382.94	23405.4 [0.7]	23381.9 [1.5] 2_1^8
(1,3,1)	23758.63	23788.0 [1.3]	23777.2 [0.02] 2_1^9
(1,4,1)	24127.72	24166.4 [1.4]	24108.6 [0.5] (0,7,1)
(1,5,1)	24488.16	24543.3 [0.9]	24489.2 [0.04] $2_1^8 3_0^2$
(2,1,1)	24175.30	24219.1 [0.003]	24166.4 [1.8] (1,4,1); 24169.7 [0.2] $1_0^1 2_0^4 3_1^2$
(2,2,1)	24531.29	24583.5 [0.03]	24543.1 [0.9] (1,5,1);
(2,3,1)	24879.31	24943.3 [0.005]	24910.1 [0.1] (1,6,1)
(1,1,2)	23533.47	23545.5 [0.04]	
(1,2,2)	23911.09	23929.5 [0.1]	
(1,3,2)	24282.93	24309.3 [0.1]	
(1,4,2)	24647.50	24686.8 [0.09]	24646.5 [0.08] $1_0^1 2_0^8 3_1^0$
1_1^1	20640.08	20653.9 [0.006]	20669.4 [0.05] 2_1^1; 20614.1 [0.3] $2_1^1 3_1^1$
1_1^2	21841.02	21879.1 [0.0004]	21844.1 [0.3] (0,0,2); 21855.1 [0.08] 3_1^3
$1_1^1 2_0^1$	21029.73	21049.4 [0.03]	21020.1 [0.2] $2_1^2 3_1^1$
$1_1^1 2_0^2$	21414.9	21440.0 [0.07]	21409.3 [2.3] 2_1^3; 21422 [0.1] $2_1^3 3_1^1$
$1_1^1 2_0^3$	21794.18	21826.0 [0.1]	
$1_1^1 2_0^4$	22167.29	22207.1 [0.09]	22164.4 [0.04] $2_0^6 3_2^0$
$1_1^2 2_0^1$	22207.16	22252.6 [0.003]	22209.4 [0.8] 2_1^5; 22207.1 [0.09] $1_1^1 2_0^4$
$1_1^2 2_0^2$	22567.16	22620.1 [0.01]	22583.7 [0.04] $1_1^1 2_0^5$; 22582.7 [0.03] $2_1^2 3_0^3$
$1_1^2 2_0^3$	22919.06	22983.8 [0.03]	22928.1 [20.5] (0,4,1); 22930.9 [2.2] $2_0^4 3_1^2$
$1_1^1 2_0^2 3_0^1$	21948.52	21968.9 [0.01]	21937.2 [0.05] $2_1^3 3_0^1$
$1_1^1 2_0^3 3_0^1$	22324.00	22351.5 [0.02]	22335.9 [0.9] $2_1^4 3_0^1$
2_1^1	20589.41	20594.3 [3.5]	
2_1^2	20994.8	21003.6 [4.4]	
$2_0^2 3_1^0$	21069.1	21078.7 [0.004]	21069.9 [0.004] $2_2^1 3_0^2$
$2_0^3 3_1^0$	21470.0	21484.4 [0.4]	21472.6 [0.01] $2_2^2 3_0^2$
$2_0^2 3_1^1$	21605.65	21609.9 [3.8]	
$2_0^3 3_1^1$	22003.00	22012.3 [4.6]	
$2_0^1 3_1^2$	21728.62	21733.6 [1.3]	

[a] Fixed to the experimental value.

recorded under different experimental conditions, as mentioned above. Similar to HSiCl, computed FC factors suggest some contributions from weak hot bands to the main vibrational structure (see the bar diagram in the middle Figure 4), but they are so close in energy that they will not be resolved in the simulated spectrum with a fwhm of 10 cm^{-1}. Overall, it can be concluded that the agreement between the simulated absorption and experimental LIF spectra of HSiCl and DSiCl is good.

The simulated and experimental $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emission spectra of HSiCl are compared in Figure 5 (top and second from top, respectively). It can be seen that the

Table 7. RCCSD(T)/aug-cc-pV5Z and CASSCF/MRCI+D/aug-cc-pV5Z PEFs of the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ States of HSiCl {C_{ijk}; See Text and Eq *1*}

$(i,j,k)^a$	$\tilde{X}^1A'$	$\tilde{A}^1A''$
200	0.276297	0.252828
020	0.298425	0.312888
002	0.099921	0.049566
110	0.011694	−0.014379
101	−0.007743	0.014213
011	0.045098	0.014961
300	−0.425380	−0.494550
030	−0.453991	−0.507081
003	0.012296	−0.006902
210	0.014470	−0.005445
201	−0.000684	0.004654
120	−0.024499	−0.009130
021	−0.066495	−0.031292
102	−0.024798	−0.043138
012	−0.097468	−0.044741
111	−0.037525	−0.031976
400	0.440379	0.443721
040	0.483647	0.550709
004	0.027711	0.009272
310	−0.010392	0.006362
301	0.004372	−0.011545
130	0.022119	0.024504
031	0.035256	0.006376
103	−0.025911	0.010411
013	0.019163	0.013650
220	0.002746	0.021209
202	−0.033945	−0.034754
022	0.036571	0.012030
211	−0.003977	0.007443
121	0.050357	0.018099
112	0.024835	0.028248
050	−0.388735	−0.541210
060	0.155526	0.286202
005	0.040039	0.081228
006	0.048156	0.062328
500	−0.356954	−0.256333
600	0.159079	0.084583
β	−0.027306	−0.200886

[a] *i*, *j*, and *k* are subscripts (indices) in C_{ijk} {see eq 1} and with the restriction of $i + j + k \geq 2$.

Table 8. Computed Vertical Excitation Energies (T_v in eV) of Some Low-Lying Singlet and Triplet States of HSiCl from CASSCF/MRCI/AVQZ(spd,sp) Calculations^a at the RCCSD(T)/AVQZ Geometry of the $\tilde{X}^1A'$ State

states	configuration	CASSCF	C_{max}^b	MRCI	MRCI+D
$\tilde{X}^1A'$	$(13a')^2(3a'')^2$	0	0.928	0	0
$(1)^3A''$	$(13a')^1(4a'')^1$	1.48	0.937	1.649	1.684
$(1)^1A''$	$(13a')^1(4a'')^1$	2.88	0.911	2.863	2.823
$(1)^3A'$	$(13a')^1(14a')^1(3a'')^2$	4.52	0.930	4.651	4.656
$(2)^1A'$	$(13a')^1(14a')^1(3a'')^2$	5.83	0.881	5.720	5.614

[a] For the $\tilde{X}^1A'$, $(1)^3A'$, $(1)^1A''$, and $(1)^3A'$ states, the CASSCF calculations are average state calculations for four states. For the second $^1A'$ state, the CASSCF calculations are average state calculations for the $\tilde{X}^1A'$ and (2) $^1A'$ states. Only the s, p, and d functions were used for Si and Cl, and the s and p functions were used for H. [b] The largest computed CI coefficients in the MRCI calculations associated with the main electronic configuration (shown under configuration).

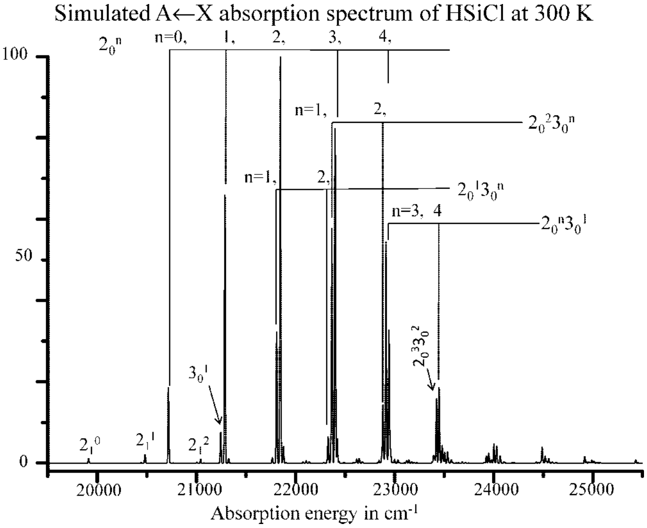


Figure 1. Simulated $\tilde{A}^1A'' \leftarrow \tilde{X}^1A'$ absorption spectrum of HSiCl at a Boltzmann vibrational temperature of 300 K with the assignments of some major vibrational progressions.

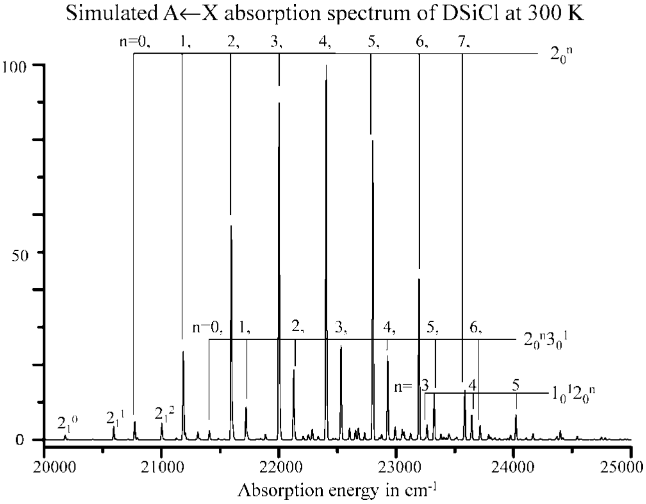


Figure 2. Simulated $\tilde{A}^1A'' \leftarrow \tilde{X}^1A'$ absorption spectrum of DSiCl at a Boltzmann vibrational temperature of 300 K with the assignments of some major vibrational progressions.

agreement is very good (bearing in mind that the intensity of the 0_0^0 vibrational component would be affected by the scattered excitation laser light). Specifically, in addition to the main $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'(0,v_2'',0)$ progression, the simulated spectrum shows only one weak progression, which is $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'(0,v_2'',1)$, in agreement with the experimental spectrum. Also shown in Figure 5 are the simulated spectra reported in ref 2 obtained employing the harmonic oscillator model (bottom and second from bottom traces in Figure 5; see ref 2 for details of the spectral simulations). Comparison between all the simulated and experimental spectra shows clearly that the simulated spectrum reported here, which includes anharmonicity, is superior to previously simulated spectra, which have ignored anharmonicity. Moreover, a more detailed comparison between the simulated spectrum obtained here (top trace in Figure 5) and the experimental SVL emission spectrum reported in ref 2 (second from top trace in Figure 5) suggests a gradual loss of intensity in the experimental spectrum toward low emission energy (i.e., larger displacement from the excitation line). Specifically, the strongest vibrational component in the simulated spectrum (top trace in Figure 5) is the $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'(0,3,0)$ component, while that in the experimental spectrum (second from top trace in Figure

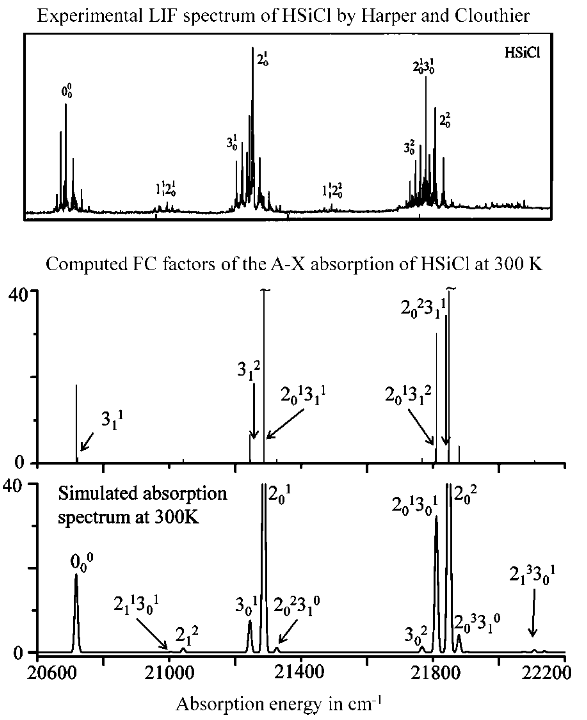


Figure 3. Computed Franck−Condon factors (the bar diagram in the middle), the simulated $\tilde{A}^1A'' \leftarrow \tilde{X}^1A'$ absorption spectrum (bottom trace) of HSiCl at a Boltzmann vibrational temperature of 300 K in the 20 600−22 200 cm^{-1} region with the assignments of some major vibrational components, and the corresponding portion of the experimental LIF spectrum of HSiCl (top trace) reported by Harper and Clouthier.[1]

5) is the $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'(0,2,0)$ component. In addition, for higher members of the $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'(0,v_2'',0)$ series, the relative intensities in the simulated spectrum are slightly stronger than those in the experimental spectrum. Assuming that the experimental SVL emission spectra recorded with the type of detector used in ref 2 (a red-sensitive photomultiplier) have not been corrected for wavelength dependence of detector sensitivity, which is expected to decrease toward low emission energy (see for example ref 27), it is anticipated that the simulated spectrum obtained here using the best *ab initio* geometries would match even better the observed spectrum if the experimental spectrum were corrected for wavelength dependent detector sensitivity. The comparison between the simulated and experimental $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emission spectra of DSiCl, as shown in Figure 6, yields very similar conclusions. First, the agreement between theory and experiment is very good, and would be better if the experimental spectra were corrected for wavelength dependence of detector sensitivity. Second, the excellent agreement between simulated and experimental $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emission spectra of HSiCl and DSiCl leads to the conclusions that the PEFs and the best theoretical geometrical parameters employed for the two states involved in the FC factor calculations are highly reliable.

Hostutler et al.[2] have also reported the experimental $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ and $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'$ SVL emission spectra of DSiCl, and they are compared with the corresponding simulated spectra in Figures 7 and 8, respectively.

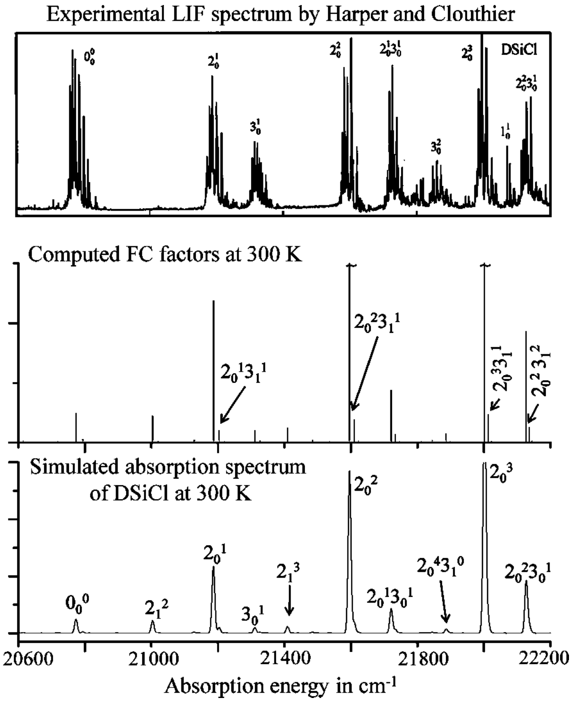


Figure 4. Computed Franck−Condon factors (the bar diagram in the middle), the simulated $\tilde{A}^1A'' \leftarrow \tilde{X}^1A'$ absorption spectrum (bottom trace) of DSiCl at a Boltzmann vibrational temperature of 300 K in the 20600−22200 cm^{-1} region with the assignments of some major vibrational components, and the corresponding portion of the experimental LIF spectrum of DSiCl (top trace) reported by Harper and Clouthier.[1]

For the $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'$ SVL emission, there are three major vibrational progressions, namely, $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'(0,v_2'',1)$, $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'(0,v_2'',0)$, and $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'(0,v_2'',2)$, in both the simulated and experimental spectra (Figure 8). The strongest progression is the $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'(0,v_2'',1)$ progression, with the $2_0 3_1$ and $2_1 3_1$ components being strongest in the series in both the simulated and experimental spectra. The agreement between the simulated and observed vibrational structures for this vibrational progression may be considered as acceptable. However, the overall agreement between the simulated and experimental vibrational structure of this SVL emission cannot be considered as good. For the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ SVL emission, the agreement between the simulated and experimental spectra for the relatively weak vibrational structure in the ~4000−6000 cm^{-1} region of the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ SVL emission spectrum can be considered as reasonably good (see Figure 7). However, the dominant vibrational structure of the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'(0,v_2'',0)$ series, for $v_2'' = 0-5$, in the 0−3000 cm^{-1} region of the simulated spectrum (top trace of Figure 7) increases in intensity and then decreases gradually, but this behavior is not displayed in the experimental spectrum (bottom trace of Figure 7). In the experimental spectrum, the 2_2^1 component is strong as in the simulated spectrum, but the 2_1^1 and 2_3^1 components are considerably weaker than those in the simulated spectrum. In addition, the 3_1 (or $2_0^1 3_1^0$) component of the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'(0,v_2'',1)$ series (i.e., the $2_n 3_1$ series, the notation used in ref 2; bottom trace of Figure 7) is very strong in the experimental spectrum, but the whole

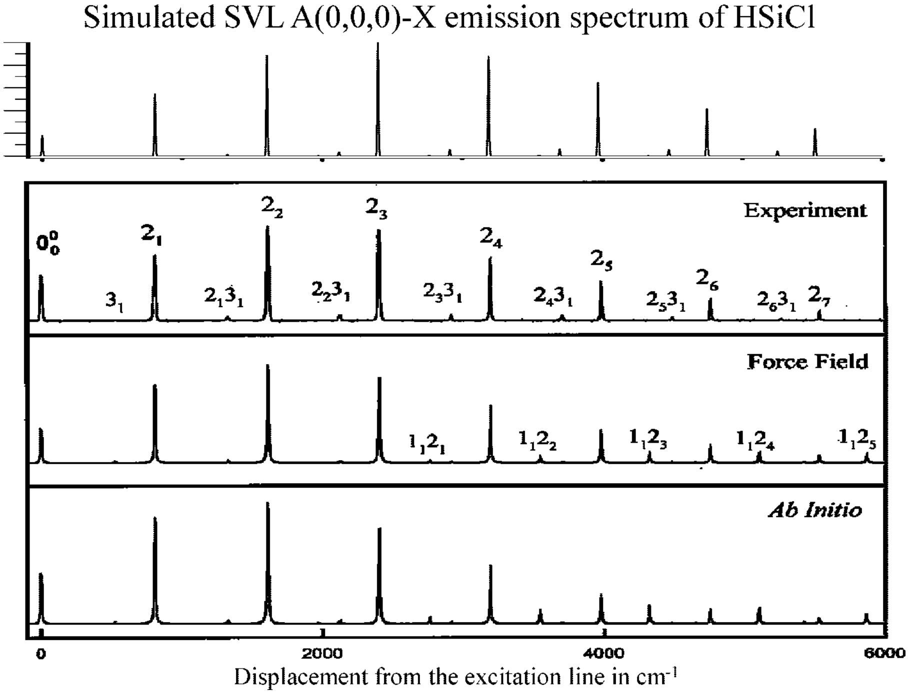


Figure 5. Simulated $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emission spectrum of HSiCl (top trace) and the corresponding experimental and simulated spectra reported by Hostutler et al.[2]

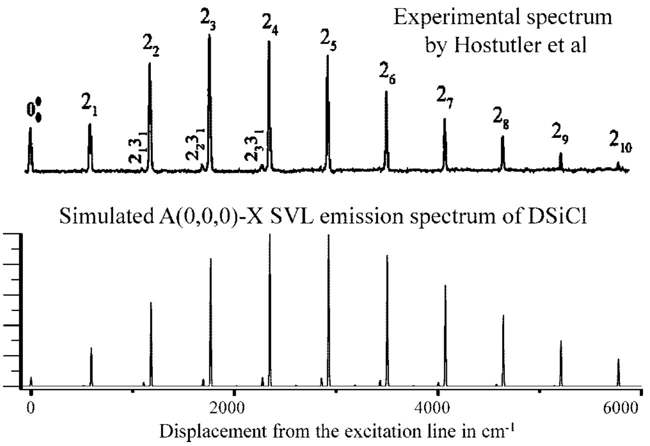


Figure 6. Simulated $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emission spectrum of DSiCl (bottom trace) and the corresponding experimental and simulated spectra reported by Hostutler et al.[2] (top trace).

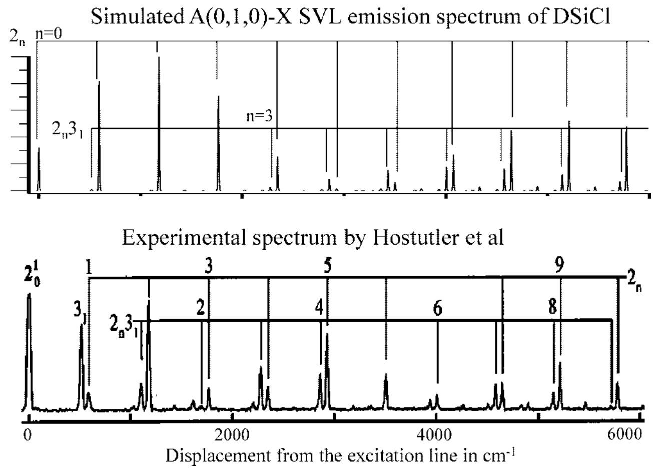


Figure 7. Simulated $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ SVL emission spectrum of DSiCl and the assignments of the two major vibrational progressions (top trace) and the corresponding experimental spectrum reported by Hostutler et al.[2] (bottom trace).

$\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'(0,v_2'',1)$ series is weak in the simulated spectrum. Nevertheless, it should be noted that the energy positions of all the observed vibrational components in the two experimental SVL emission spectra of DSiCl considered, namely, the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ and $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'$ spectra, are consistent with those of the $\tilde{X}^1A'$ state of DSiCl (see also Table 4 and discussion in Table 4 given above), suggesting that the emissions are to vibrational levels of the $\tilde{X}^1A'$ state of DSiCl, and the molecular carrier is DSiCl. The discrepancies between simulated and experimental structures of these two SVL emission spectra of DSiCl are in the relative intensities of some vibrational components. Nevertheless, in view of the fact that agreement is observed in vibrational component positions, but not in the relative intensities, it appears that, each of the two SVL emission spectra considered have major contributions from the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ and $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'$ SVL emissions, respectively. However, the comparison between the simulated and experimental vibrational structures, as shown in Figures 7 and 8, suggests that other emissions of DSiCl are involved in the experimental spectra, that is, other states of DSiCl are accessed, as well as the $\tilde{A}^1A''(0,1,0)$ and $\tilde{A}^1A''(0,2,1)$ levels, by the excitation laser energy used to record the experimental spectra.

Concluding Remarks

In summary, high-level *ab initio* calculations have been carried out on the $\tilde{X}^1A'$ and $\tilde{A}^1A''$ states of HSiCl, and FC factors including allowance for anharmonicity have been

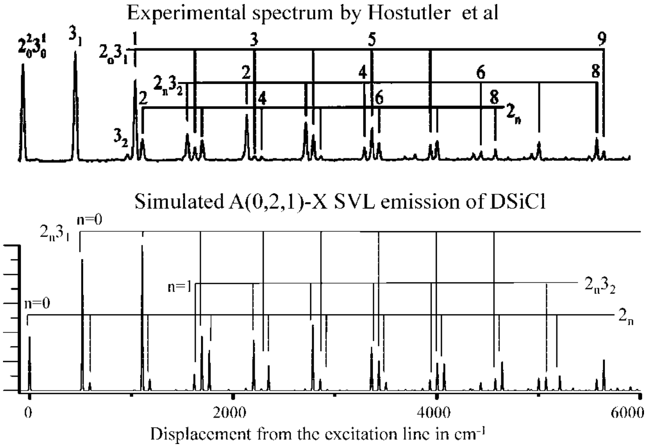


Figure 8. Simulated $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'$ SVL emission spectrum of DSiCl and the assignments of the three major vibrational progressions (bottom trace) and the corresponding experimental spectrum reported by Hostutler et al.[2] (top trace).

computed between these two states of HSiCl and DSiCl. Agreement between theory and experiment for the minimum-energy geometrical parameters, vibrational frequencies, relative electronic energy, and the $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emission spectra of HSiCl and DSiCl is very good. Simulated $\tilde{A}^1A'' \leftarrow \tilde{X}^1A'$ absorption spectra of HSiCl and DSiCl also agree reasonably well with corresponding experimental LIF spectra. Such good agreements between theory and experiment for the absorption/LIF and $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emission spectra of HSiCl and DSiCl suggest that the PEFs and the best estimated geometrical parameters of the two states of HSiCl employed in the FC calculations, and the computed anharmonic vibrational wave functions and energies of the two states of HSiCl and DSiCl are highly reliable. In fact, in view of the fact that the detector employed in recording the experimental $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emission spectra of HSiCl and DSiCl has a characteristic of discriminating against signals toward low emission energies and the recorded spectra have most likely not been corrected for frequency dependent sensitivity of the detector, the simulated spectra reported here are almost certainly more reliable than the experimental spectra.

However, for the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ and $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'$ SVL emissions of DSiCl, there are significant discrepancies between simulated and experimental spectra. Specifically, the relative intensities of the simulated and experimental vibrational structures do not match very well, although the energy positions of almost all observed vibrational components agree with the computed vibrational energies of the $\tilde{X}^1A'$ state, suggesting that the emissions are to the $\tilde{X}^1A'$ state of DSiCl. It has been mentioned above in the discussion on the simulated absorption spectra of DSiCl that some hot bands are computed to be underneath some vibrational components of the main $\tilde{A}^1A'' \leftarrow \tilde{X}^1A'(0,0,0)$ absorption band. For example, for the simulated vibrational component of DSiCl at a computed absorption energy of 21186.6 cm^{-1}, in addition to the $\tilde{A}^1A''(0,1,0)'' \leftarrow \tilde{X}^1A'(0,0,0)$ transition (see Table 6 and Figure 4), the $2_0{}^1 3_1{}^1$, $1_1{}^1 3_0{}^1$, $2_2{}^5 3_1{}^0$, and $1_0{}^1 2_1{}^2 3_2{}^0$ hot bands with the computed absorption energies of 21203.9, 21189.4, 21183.8, and 21183.3 cm^{-1} are also expected to contribute because of the proximity of their absorption energies (though only slightly according to the computed FC factors of ~2.0, 0.002, 0.002, and 0.00005, respectively). In this connection, the dispersed fluorescence spectrum recorded at an excitation energy of 21186.6 cm^{-1} (computed value; the experimental value is 21182.08 cm^{-1}) and assigned to the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ SVL emission may have contributions also from SVL emissions from the $\tilde{A}^1A''$ (0,1,1), (1,0,1), (0,5,0), and (1,2,0) levels (being accessed via hot band transitions), in addition to the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ SVL emission. However, including these SVL emissions arising from hot band transitions in the simulation of the experimental dispersed fluorescence spectrum still did not give a satisfactory agreement between theory and experiment. Although some observed vibrational features may be accounted for by parts of the vibrational structures of these SVL emissions arising from hot bands, the overall agreement cannot be considered as good. In fact, the agreement is far from "fingerprint identification", as for the $\tilde{A}^1A''(0,0,0) \rightarrow \tilde{X}^1A'$ SVL emissions of HSiCl and DSiCl. The main difficulty, particularly in the case of the dispersed fluorescence spectrum assigned to the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ SVL emission, is that the vibrational structures of the simulated SVL emission spectra always display a gradual change in the relative intensities over a number of vibrational components (i.e., a gradual increase followed by a gradual decrease over at least four vibrational components; see for example the simulated spectra in Figures 5−8). However, the experimental $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ SVL emission spectrum has the 3_1 and 2_2 vibrational components being very strong in the $2_n 3_1$ and 2_n series (see Figure 7 bottom trace).

To investigate this further, we have carried out preliminary calculations on some low-lying singlet and triplet states of HSiCl to see whether there are some other states nearby, which may contribute to the SVL emission spectra of HSiCl or DSiCl. Results of some CASSCF/MRCI calculations are summarized in Table 8. It can be seen that the $(1)^3A'$ and $(2)^1A'$ states are significantly higher in energy in the vertical excitation region (from the $\tilde{X}^1A'$ state) than the $\tilde{A}^1A''$ state, while the $(1)^3A''$ state is below the $\tilde{A}^1A''$ state. Further geometry optimization and vibrational frequency calculations on the lowest $^3A''$ state were carried out and gave $r_e(\mathrm{HSi}) = 1.485$ Å, $r(\mathrm{SiCl}) = 2.051$ Å, $\theta(\mathrm{HSiCl}) = 115.5°$, $\omega_1 = 2178$ cm^{-1}, $\omega_2 = 652$ cm^{-1}, and $\omega_3 = 552$ cm^{-1} at the RCCSD(T)/AVQZ level. The computed T_e and T_v values at this level are 1.484 and 1.702 eV, respectively. On the basis of these calculations, it can be concluded that the lowest $^3A''$ state is the ã state, and the T_v of the $\tilde{a}^3A''$ state is ~0.84 eV (6775 cm^{-1}) below the T_0 of the $\tilde{A}^1A''$ state. Only very high vibrational levels of the $\tilde{a}^3A''$ state may overlap with low vibrational levels of the $\tilde{A}^1A''$ state, and interaction between the $\tilde{A}$ and ã vibronic levels could occur via a spin−orbit mechanism. We have also estimated the barriers to linearity of the $\tilde{X}^1A'$ (a $^1\Sigma^+$ state in $C_{\infty V}$), $\tilde{a}^3A''$ (a $^3\Pi$ state with a $\sigma^1\pi^1$ open-shell configuration) and $\tilde{A}^1A''$ states (a $^1\Pi$ state with a $\sigma^1\pi^1$ open-shell configuration) of HSiCl by carrying out CASSCF/MRCI+D/AVQZ geometry optimization calculations with the bond angle fixed to 179.5°. They are ~3.11, 2.09, and 0.74 eV above their respective potential energy minima (the barriers to linearity of the ã and $\tilde{A}$ states are ∼3.37 and 3.57 eV above the potential energy minimum of the $\tilde{X}^1A'$ state). In DSiCl, the $\tilde{A}^1A''(0,1,0)$ and $\tilde{A}^1A''(0,2,1)$ levels lie below the barriers to linearity of the $\tilde{X}$ and ã states. Interaction between vibronic levels of the ã and $\tilde{A}$ states could lead to a change in vibrational character of their vibrational levels above the minimum of the $\tilde{A}^1A''$ state and a change in FC factors in the SVL spectra from those shown in the simulated spectra in Figures 7 and 8. Clearly, further experimental and computational investigations are required to clarify the discrepancies between theory and experiment for the $\tilde{A}^1A''(0,1,0) \rightarrow \tilde{X}^1A'$ and $\tilde{A}^1A''(0,2,1) \rightarrow \tilde{X}^1A'$ SVL emission spectra of DSiCl.

Acknowledgment. Financial support from the Research Grant Council (RGC) of the Hong Kong Special Administrative Region (HKSAR, Grant PolyU 5014/06P) is acknowledged for this work. The authors are also grateful to the provision of computational resources from the EPSRC (U.K.) National Service for Computational Chemistry Software (NSCCS).

References

(1) Harper, W. W.; Clouthier, D. J. *J. Chem. Phys.* **1997**, *106*, 9461.

(2) Hostutler, D. A.; Ndiege, N.; Clouthier, D. J.; Pauls, S. W. *J. Chem. Phys.* **2001**, *115*, 5485.

(3) Lin, W.; Novick, S. E.; Fukushima, M.; Jäger, W. *J. Phys. Chem. A* **2002**, *106*, 7703.

(4) de Nalda, R.; Mavromanolakis, A.; Couris, S.; Castillejo, M. *Chem. Phys. Lett.* **2000**, *316*, 499.

(5) Vázquez, J.; Stanton, J. F. *J. Phys. Chem. A* **2002**, *106*, 4429.

(6) Coriani, S.; Marchesan, D.; Gauss, J.; Hättig, C.; Jørgensen, P.; Helgaker, T. *J. Chem. Phys.* **2005**, *123*, 184107.

(7) Rizzo, A.; Puzzarini, C.; Coriani, S.; Gauss, J. *J. Chem. Phys.* **2006**, *124*, 064302.

(8) Mok, D. K. W.; Lee, E. P. F.; Chau, F.-T.; Dyke, J. M. *J. Chem. Phys.* **2004**, *120*, 1292.

(9) (a) Hampel, C.; Peterson, K.; Werner, H.-J. *Chem. Phys. Lett.* **1992**, *190*, 1. (b) Deegan, M. J. O.; Knowles, P. J. *Chem. Phys. Lett.* **1994**, *227*, 321.

(10) Knowles, P. J.; Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1993**, *99*, 5219; Erratum. *J. Chem. Phys.* **2000**,, *112*, 3106.

(11) Werner, H.-J. *Adv. Chem. Phys.* **1987**, *LXIX*, 1.

(12) Werner, H.-J.; Knowles, P. J. *J. Chem. Phys.* **1988**, *89*, 5803.

(13) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(14) Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, *114*, 9244.

(15) (a) Langhoff, S. R.; Davidson, E. R. *Int. J. Quantum Chem.* **1974**, *8*, 61. (b) Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Chem. Phys.* **1983**, *78*, 5682. (c) Simons, J. *J. Phys. Chem.* **1989**, *93*, 626.

(16) Lee, E. P. F.; Dyke, j. M.; Chau, F.-T.; Chow, W. K. *Chem. Phys. Lett.* **2003**, *376*, 465.

(17) Halkier, A.; Helgaker, T.; Klopper, W.; Jorgensen, P.; Csaszar, A. G. *Chem. Phys. Lett.* **1999**, *310*, 385.

(18) (a) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358. (b) Peterson, K. A.; Dunning, T. H., Jr. *J. Chem. Phys.* **2002**, *117*, 10548.

(19) 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, A package of ab initio programs*, 2008.

(20) Carter, S.; Handy, N. C. *J. Chem. Phys.* **1987**, *87*, 4294.

(21) Dennis, J. E. Jr.; Gay, D. M.; Welsh, R. E. *ACM Trans. Math. Software* **1981**, *7*, 348–369.

(22) Watson, J. K. G. *Mol. Phys.* **1970**, *19*, 465.

(23) Mok, D. K. W.; Lee, E. P. F.; Chau, F.-T.; Wang, D.-C.; Dyke, J. M. *J. Chem. Phys.* **2000**, *113*, 5791.

(24) Chau, F.-T.; Dyke, J. M.; Lee, E. P. F.; Mok, D. K. W. *J. Chem. Phys.* **2001**, *115*, 5816.

(25) EPAPS document No. E-JCPSA6-115-001137 from http://www.aip.org/pubservs/epaps.html (accessed Nov. 20, 2008) or from ftp.aip.org in the directory /epaps/; see the EPAPS homepage of *J. Chem. Phys.* and ref 2 for more information.

(26) Experimental geometry data from NIST Computational Chemistry Comparison and Benchmark Database (http://cccbdb.nist.gov/; accessed Nov. 26, 2008);Rotational constants from Smith, J. G. *J. Mol Spectrosc.* **1986**, *120*, 110.

(27) Liu, M.-L. Lee, C.-L. Bezant, A. Tarczay, G. Clark, R. J. Miller, T. A. Chang, B.-C. PCCP, 5, 1352, 2003.

(28) Herzberg, G.; Verma, R. D. *Can. J. Phys.* **1964**, *42*, 395.

CT800513V

JCTC Journal of Chemical Theory and Computation

Topological Analysis of the Interactions between Organic Molecules and Co(Ni)MoS Catalytic Active Phases

Emmanuel Krebs,[†,‡] Bernard Silvi,[*,†] and Pascal Raybaud[§]

Laboratoire de Chimie Théorique, Université Pierre et Marie Curie, (UMR-CNRS 7616), 4 Place Jussieu 75252-Paris cédex, France, IFP Direction Chimie et Physico-Chimie Appliquées, 1 & 4 Av. de Bois-Préau, 92852 Rueil-Malmaison cédex, France, and IFP Direction Catalyse et Séparation, Rond-point de l'échangeur de Solaize, BP3 69360 Solaize, France

Received August 19, 2008

Abstract: The adsorption modes of toluene, 2,3-dimethylbut-1-ene, and 2-methylthiophene on the edges of Co(Ni)MoS nanocrystallites has been investigated with the ELF topological approach of chemical bonding. The chemisorbed modes are characterized by the formation of bonding basins linking the substrate to the catalytic sites. The electronic rearrangements within the substrate are discussed. It is shown that a unique electronic descriptor, namely the metallic atomic basin contribution to the substrate ELF basins, provides a sizable correlation with the interaction energy.

1. Introduction

Hydrotreating reactions catalyzed by γ-alumina-supported Co(Ni)MoS catalysts are playing a key role in industrial refining in order to produce cleaner fuels.[1] The Co(Ni)MoS active phase has been widely studied by many experimental[2] and theoretical techniques[3] to provide an accurate description of the catalyst at an atomic scale. The active phase consists of MoS_2 nanocrystallites with a deformed hexagonal 2D-morphology and exposing two active edges, the so-called "S-"and "M-edges".[3,4] In order to increase the activity of the MoS_2 active phase, promoters (Co or Ni) are substituted to Mo atoms located at both edges of the crystallites. The edge energies and atomic structures depend on the sulfiding conditions and to the promoter coverage, and they have been investigated by "first principles" calculations based on periodic density functional theory in numerous previous works.[3−8] Furthermore, it has been shown that the 2D-morphology (and as a result the type of edge sites) depends on the nature and proportion of promoter (Co or Ni) at edges.[9] Depending on the reaction conditions, different typical proportions of the promoter can be stabilized at the edges, which induces different types of local arrangement of the metal and sulfur atoms. Hence, specific conditions can be found where the edges are fully covered by the promoters and other cases where the edges are only partially decorated by Co or Ni atoms, with some neighboring Mo atoms. These different local environments are strongly suspected to influence the catalytic reactivity[10] and the selectivity.[11] Many DFT works have also performed adsorption energy calculations of sulfur organic molecules including thiophene and dibenzothiophene[4,12−17] and nitrogen organic molecules.[3,18,19] A recent DFT study has focused on the determination of the most stable adsorption mode of 2,3-dimethylbut-1-ene and 2-methylthiophene on the Co(Ni)MoS edges.[11] These molecules represent two important classes of sulfur organic and olefin compounds present in hydrotreatment reactions of gasoline. This study has provided an interpretation of the selective hydrodesulfuration of 2-methylthiophene based on the adsorption energy calculated values. It is the subject of the present work to provide a more fundamental understanding based on the nature of the chemical bonding of these energy results. We will thus furnish a theoretical description of the electronic interaction between the adsorbed molecules and the active edges. In addition, aromatics are an important family of hydrocarbons molecules also present in hydrotreating reactions catalyzed by Co(Ni)MoS catalysts.[10,20] In order to achieve a coherent

---

[*] Corresponding author e-mail: silvi@lct.jussieu.fr.

[†] Université Pierre et Marie Curie.

[‡] IFP Direction Chimie et Physico-Chimie Appliquées.

[§] IFP Direction Catalyse et Séparation.

10.1021/ct800344r CCC: $40.75

comparison including this family of reactant, the present work will investigate the stable adsorption configuration of toluene on CoMoS edges and also provide an electronic analysis of its chemical bonding on the same edge sites.

The characterization of the bonding interactions between the adsorbed molecules and the active sites is a key for the *chemical* understanding of the catalytic process. The analysis of the bonding is usually carried out by either rather crude orbital-based methods or qualitative descriptions of electron density maps. Only few published works use more rigorous topological techniques such the AIM,[21] the topology of the electrostatic potential,[22] or the ELF approach.[23] Aray et al. have recently published interesting analysis carried out with AIM[24−27] and also with the electrostatic potential.[28−31]

To our knowledge, the topological ELF approach has already been used in surface science to characterize the adsorption of CO and CN^- on surfaces,[32] F centers on MgO surfaces,[33] and the adsorption of Pd atoms on MgO, Al_2O_3, and SiO_2 surfaces.[34]

This paper aims to provide a description of the bonding of adsorbed toluene, 2,3-dimethylbut-1-ene, and 2-methylthiophene molecules on promoted MoS_2 edges based on the ELF topological analysis in order to get a deeper insight on:

(1) the qualitative and quantitative aspects of bond formation and charge transfers between the catalyst and the substrate,

(2) the evolution of the bonds of the adsorbed species with respect to the uncomplexed molecules.

For this purpose we have selected a representative sample of four toluene, thirteen 2,3-dimethylbut-1-ene, and fourteen 2-methylthiophene adsorption modes on which the ELF topological analysis has been carried out.

2. Methodology

2.1. The Topological Analysis of Electron Distribution Functions. One of the goals of the topological analysis of electron distribution functions is to provide a partition of the geometrical space occupied by the chemical system of interest (molecule, aggregate, polymer, 1−3 D periodic system) into adjacent nonoverlapping volumes called basins. These basins are thought as corresponding to chemical entities such as atoms in molecules, atomic cores, bonds, or lone pairs. The partition is achieved with the help of a rigorous mathematical method, namely the Dynamical System Theory (see Abraham et al.),[35,36] applied to the gradient vector field of a quantum mechanically well defined local function of the electron distribution which carries the chemical information.

The choice of the local function is determined by the type of description of the bonding. The one electron density $\rho(r)$ has its regular maxima at nuclei where the Coulombic electron nucleus attractive potential tends to infinity. Therefore, one nuclear attractor and one basin is associated to each atom. As non-nuclear attractors are exceptions in molecular system at equilibrium,[37] the basins of the electron density attractors are called atomic basins and constitute the backbone of the AIM theory of R. F. W. Bader.[21] In order to recover a description close to Lewis's bonding picture, it is necessary to consider another local function able to locate the boundary between two regions where opposite spin pairs are localized. This can be achieved by the Electron Localization Function (ELF) which was originally designed by Becke and Edgecombe to identify "localized electronic groups in atomic and molecular systems".[38] It relies, through its kernel, on the laplacian of the conditional same spin pair probability scaled by the homogeneous electron gas kinetic energy:

$$\chi_\sigma(r) = \frac{D_\sigma(r)}{D_\sigma^0(r)} \tag{1}$$

in which

$$D_\sigma(r) = \tau_\sigma(r) - \frac{1}{4}\frac{(\rho_\sigma(r))^2}{\rho_\sigma(r)}$$

appears to be the difference of the actual definite positive kinetic energy $\tau_\sigma(r)$ and the von Weizsäcker kinetic energy functional,[39] whereas

$$D_\sigma^0(r) = \frac{3}{5}(6\pi^2)^{2/3}\rho_\sigma^{5/3}(r)$$

is the kinetic energy density of the homogeneous electron gas. This formulation led Savin to propose an interpretation of ELF in terms of the local excess kinetic energy because the Pauli repulsion enabled its calculation from Kohn−Sham orbitals.[40−42] Orbital-based interpretations of ELF have been proposed by Burdett[43] and more recently by Nalewajski et al.[44] who considered the nonadditive interorbital Fisher information. Another route pioneered by Dobson[45] explicitly considers the pair functions. It has been independently developed by Kohout et al.[46,47] and by one of us,[48] allowing the extension of ELF to correlated wave functions.[49]

The topological partition of the ELF gradient field[23,50] yields basins of attractors which can be thought as corresponding to atomic cores, bonds, and lone pairs. The core basins surround nuclei with atomic number $Z > 2$ and are labeled C(A) where A is the atomic symbol of the element. The valence basins are characterized by the number of atomic valence shells to which they participate, or in other words by the number of core basins with which they share a boundary. This number is called the synaptic order.[51] Thus, there are monosynaptic, disynaptic, trisynaptic basins, and so on. Monosynaptic basins, labeled V(A), correspond to the lone pairs of the Lewis model, and polysynaptic basins to the shared pairs of the Lewis model. In particular, disynaptic basins, labeled V(A, X) correspond to two-center bonds, trisynaptic basins, labeled V(A, X, Y) to three-center bonds, and so on. The valence shell of a molecule is the union of its valence basins. As hydrogen nuclei are located within the valence shell, they are counted as a formal core in the synaptic order because hydrogen atoms have a valence shell. For example, the valence basin accounting for a C−H bond is labeled V(C, H) and called protonated disynaptic. The valence shell of an atom, for example A, in a molecule is the union of the valence basins whose label lists contain the element symbol A. This description recovers Lewis's picture of the bonding[52,53] and provides very suggestive graphical representations of molecular systems. A quantitative analysis is further achieved by integrating the electron density and the pair functions over the volume of the basins yielding both basin populations:

$$\bar{N}(\Omega_A) = \int_{\Omega_A} \rho(r)\mathrm{d}r \qquad (2)$$

and the corresponding covariance matrix[54] which measures the quantum mechanical uncertainties of the electron distribution and supports a phenomenological interpretation in terms of weighted mesomeric structures.[55]

Combining the ELF and AIM approaches, Raub and Jansen[56] have introduced a bond polarity index defined as:

$$p_{XY} = \frac{\bar{N}[V(X, Y)|X] - \bar{N}[V(X, Y)|Y]}{\bar{N}[V(X, Y)|X] + \bar{N}[V(X, Y)|Y]} \qquad (3)$$

where $\bar{N}[V(X, Y)|X]$ denotes the contribution of the X atom to the population of the V(X, Y) basin, i.e., the integrated density over the intersection of the V(X, Y) ELF basin and the atomic basin of atom X.

In the context of the ELF analysis, the concept of domain is very important because it enables definition of chemical units within a system and to characterization of valence domains belonging to a given chemical unit. The sole gradient dynamical system mathematical properties do not provide the whole set of definitions necessary to describe the bonding in molecules, and therefore some other mathematically based approaches are required for this purpose. The topological concept of domain has been introduced in chemistry by P. Mezey in order to recognize functional groups within organic molecules.[57] Generalized to ELF isovalues this concept has proved to be an efficient "generator" of clear definitions. Any subset of the molecular space bounded by an external closed isosurface $\eta(r) = f$ is a domain. A f-localization domain is such a subset with the restriction that each point satisfies $\eta(r) > f$. A localization domain surrounds at least one attractor; in this case it is called irreducible. If it contains more than one attractor, it is reducible. An irreducible domain is a subset of a basin whereas a reducible one is the union of subsets of different basins. Except for atoms and linear molecules, the irreducible domains are always filled volumes whereas the reducible ones can be either filled volumes or hollowed volumes.

Upon the increase of the value of $\eta(r)$ defining the bounding isosurface, a reducible domain split into several domains each containing less attractors than the parent domain. The reduction of localization occurs at turning points which are critical points of index 1 located on the separatrix of two basins involved in the parent domain. Ordering these turning points (localization nodes) by increasing $\eta(r)$ enables tree-diagrams to be built reflecting the hierarchy of the basins. Three types of domains can be distinguished according to the nature of the attractors within them. A core domain contains the core attractor(s) of a given atom, a valence domain contains only valence attractors, and a composite domain contains both valence and core attractors. For any system there exists low values of $\eta(r) = f$ defining a unique composite parent domain. The successive reductions of localization will split this parent domain. Every child which is a composite domain corresponds to one or more chemical species. A chemical unit is the union of the basins of the last appearing composite domain of a branch provided it is a filled volume. This concept which has been successfully used for hydrogen bonding[58] enables characterization of physisorption vs chemisorption. In the former case the first reduction yields two composite domains corresponding to the substrate and to the surface, respectively, whereas in the latter case the parent domain splits into core domains encompassed by a hollowed volume.

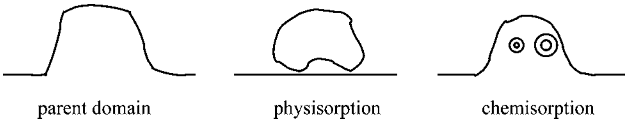


Figure 1. Bifurcation scheme of physisorption and chemisorption.

The case of adsorption is illustrated by Figure 1: the parent domain corresponding to a low ELF value contains both the substrate and the surface; when the isovalue defining the contour is raised, it splits either into two filled domains (physisorption) or becomes a hollowed volume encompassing the core domains (chemisorption). In the same manner as for the hydrogen bonding case where the ELF analysis is able to discriminate weak, medium, and strong H-bonds, several situations may occur for chemisorption according to the changes of the number of basins and of the synaptic order of some of them during the process. In a first case, very similar to physisorption, neither the number of basins nor the synaptic order varies. The topology of the substrate−surface system is just the addition of the topologies of the moieties which preserve their own valence shells, and the chemical interaction mostly consists of the increase of the electron density fluctuation of adjacent basins belonging to these valence shells, in other words an increase of the delocalization. The number of basins may remain constant whereas at least one synaptic order changes when one of the fragments behaves as a Lewis acid, in this case a monosynaptic basin (lone pair) of the other partner becomes disynaptic and therefore forms a dative bond. In the last case, the covalent chemisorption, the number of basins increases by the appearance of disynaptic basins between the substrate and the surface.

2.2. The ELF Analysis of Large Systems. In our group, the analysis of the ELF function is usually carried out with the TopMoD software[59] which has been initially designed to treat rather small-sized systems. In the first step, the program evaluates ELF over a three-dimensional regular grid whose recommended step size is on the order of 0.1−0.2 bohr. Larger step sizes may seriously downgrade the reliability of the results. In the second step, the program assigns the grid points to the different basins: from each grid point a trajectory is built until it leads to a region already assigned to an attractor basin. This procedure is robust if the whole set of attractors are located within the box. Because the computational cost increases as the cube of the box dimension, the ELF analysis of large systems becomes rapidly unfeasible. Moreover, the useful information provided by the calculation represents a small amount of the output. In order to overcome these difficulties we use a trick enabling us to safely carry out the analysis on a subset of atoms. For a given point, the contribution of the given localized basis function

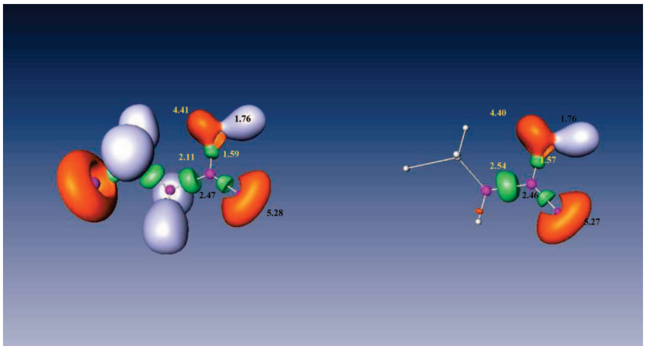


Figure 2. Localization domains and basin populations of the acid functional group of CH_2ClCH_2COOH. Left: full calculation; right: partial calculation. Color code: magenta = core, red = monosynaptic, blue = protonated disynaptic, green = disynaptic.

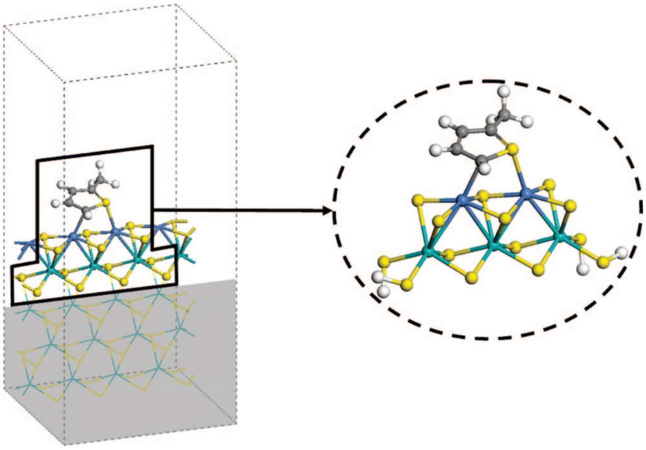


Figure 3. Example of the construction of a cluster (right) starting from a periodic supercell (left).

to both ELF and $\rho(r)$ dramatically depends upon the distance between the origin of the basis function and the point. Because of the exponential decay of the basis functions, contributions arising from those typically centered farther than 5 bohrs from the point are very small and thus can be neglected. Therefore, the basis functions centered on atoms which are not the nearest neighbors of those of the subset of interest can be legitimately discarded, yielding an auxiliary basis set to be actually used by the ELF analysis.

Figure 2 displays the localization domains and basin population of the functional group of CH_2ClCH_2COOH. The approximate calculation provides an almost identical aspect of the domains in the region of interest; moreover, the error on the populations is at most 0.02 e, the expected accuracy of the integration, except for the V(C,C) basin which belongs to the boundary zone.

2.3. Computational Methods. The ab initio calculations of molecules have been performed at the hybrid Hartree−Fock density functional B3LYP level[60−63] with the Gaussian 03 software.[64] The geometries and energetics of the adsorbed species have been optimized in periodic slab calculations carried out with the VASP plane waves periodic code.[65,66] The slab size is given in Figure 3. For more details on the slab model, the reader could refer to refs 9 and 11. The different edge structures for the CoMoS and NiMoS catalysts are reported in Figures 4 and 5. The energy values as used later are defined as:

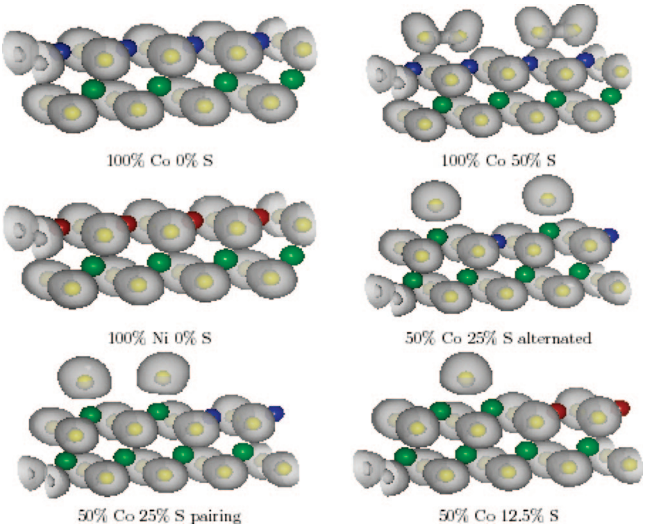


Figure 4. η = 0.7 Localization domains of the M edge structures after geometry optimization for the two promoter contents and the most stable sulfur coverage (yellow balls: S; green balls: Mo; blue balls: Co; brown ball: Ni).

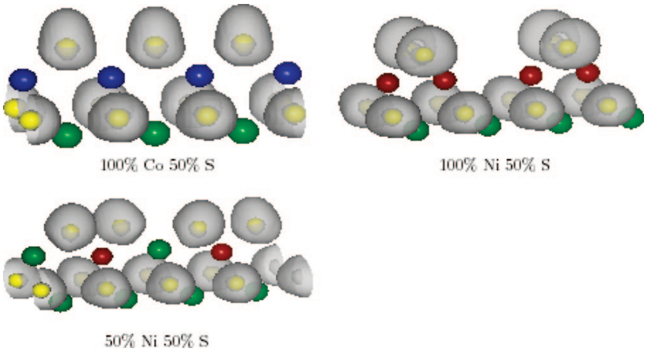


Figure 5. η = 0.7 Localization domains of the S edge structures after geometry optimization for the two promoter contents and the most stable sulfur coverage (yellow balls: S, green balls: Mo; blue balls: Co; brown ball: Ni).

$$\Delta E_{\mathrm{int}} = E(\mathrm{mol} + \mathrm{edge}) - E(\mathrm{edge}_{\mathrm{def}}) - E(\mathrm{mol}_{\mathrm{def}}) \quad (4)$$

$$\Delta E_{\mathrm{def}}(\mathrm{mol}) = E(\mathrm{mol}_{\mathrm{def}}) - E(\mathrm{mol}) \quad (5)$$

$$\Delta E_{\mathrm{def}}(\mathrm{edge}) = E(\mathrm{edge}_{\mathrm{def}}) - E(\mathrm{edge}) \quad (6)$$

$$\begin{aligned}\Delta E_{\mathrm{ads}} &= E(\mathrm{mol} + \mathrm{edge}) - E(\mathrm{edge}) - E(\mathrm{mol}) \\ &= \Delta E_{\mathrm{int}} + \Delta E_{\mathrm{def}}(\mathrm{edge}) + \Delta E_{\mathrm{def}}(\mathrm{mol}) \quad (7)\end{aligned}$$

where E(mol + edge) is the total energy of the molecule adsorbed on the edge, E(edge) the energy of the free edge (respectively, molecule for mol), and $E(\mathrm{edge}_{\mathrm{def}})$ the energy of the edge (respectively molecule for mol) deformed when the molecule is adsorbed on it. For adsorption energy calculations on the periodic slab, the GGA-PW91 functional[67] has been used. The cutoff energy is 500 eV, while the convergence criteria on forces is fixed at 0.05 eV/Å. For a more detailed description on the conditions for the periodic DFT calculations of 2,3-dimethylbut-1-ene and 2-methylthiophene adsorptions, the reader may also refer to ref 11

Representative clusters have been built from the periodic geometries (see Figure 3) and the B3LYP calculations performed with the standard 3−21G basis set.[68−73] The analysis of the ELF function has been carried out with the TopMoD program developed in the Laboratoire de Chimie Théorique de l'Université Pierre et Marie Curie.[59] The accuracy of the integrated densities is on the order of 0.02 e. The ELF isosurfaces have been visualized with the Amira 3.0 software.[74]

3. Results and Discussion

The adsorption modes on M- and S-edges are governed by the interactions between the metallic sites of the catalyst, which play the role of electrophilic centers, with the nucleophilic sites of the adsorbed molecules. Figures 4 and 5 display the ELF = 0.7 localization domains of some of the most important optimized M- and S-edges structures for the Co(Ni)MoS system calculated by VASP. In the M-edge the metal sites are above the average surface plane and therefore easily accessible by any nucleophilic part of the adsorbed molecule. However, at partial sulfur coverage the sulfur atoms may play a repulsive role and therefore partially hamper the chemisorption. On S-edges, the metal sites are much less accessible and the adsorption will therefore be much more difficult unless the shape of the nucleophilic center of the ad-molecule fits the available voids between sulfur atoms. Among other possibilities are the substitution of a sulfur of the edge by a heteroatom of the molecule and the displacement of sulfur atoms of the edge to create an accessible metal site.

Three types of nucleophilic sites can be encountered with respect to their dimensionality:

(1) polygonal sites correspond to electron delocalization within an aromatic ring; the adsorbed molecule is expected to lie in a plane parallel to the edge direction either on top of a metal site or in a bridged position between two,

(2) bond sites correspond to the interaction of a multiple bond with a metallic site; they are expected to be less subject to geometrical constraint,

(3) substitution sites in which an atom of the ad-molecule is involved in the coverage of the edge.

According to Pauling's electronegativity scale[75,76] Co should have a greater electrophilic character than Ni and Mo whereas in Allred and Rochow's scale[77] molydenum is the less electronegative. However, an indication of the electrophilic character of the metallic centers can be deduced from the ELF and QTAIM population analysis carried out in the present work. As a general rule, core basin populations are rather stable: the averaged core population of Co, Ni, and Mo are calculated to be 25.24 ± 0.05, 26.27 ± 0.06, and 39.72 ± 0.09, respectively, whereas larger standard deviations are calculated for the AIM atomic populations. According to these values, the stronger interactions are expected with molybdenum in agreement with Allred and Rochow's electronegativity.

Different types of bonding between the edge and the molecules are expected. They are named η_i-MM′ for both toluene (Figure 6) and 2-methylthiophene (Figure 7), where i is the number of atoms supposed to be linked to the metallic edge sites (M and M′ = Mo, Co, Ni). The sulfur atom of the thiophene molecule is always bonded to M. The nomenclature used for 2,3-dimethybut-1-ene is given in Figure 8.

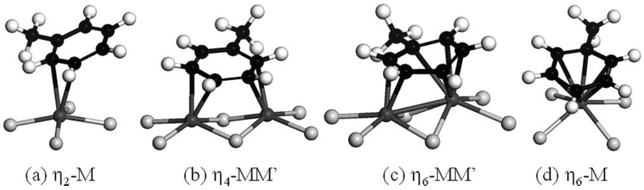
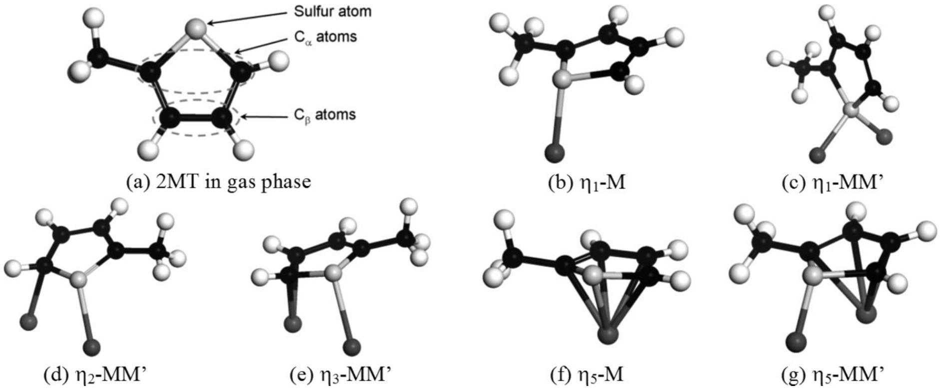


Figure 6. Views of toluene adsorption modes. (Black balls: C; dark gray balls: edge site (Mo, Co, or Ni); light gray balls: S).

3.1. Toluene Adsorption. It is first interesting to notice that the sum of the deformation energies (molecule+edge) is greater than 0.8 eV for the configurations strongly bound to the edge (η_i-MM′ for $i \geq 4$), whereas it is only about 0.42 eV for the less attached η_2-Co configuration. This is due to the very low molecule deformation energy for this last configuration compared to the others, while the deformation energy of the edge remains constant. Nevertheless, this configuration is the less favorably adsorbed one, considering interaction or adsorption energies.

The ELF localization domains of the free toluene molecule and of three of its adsorption modes are displayed in Figure 9 whereas the population analysis is reported in Table 1. The nucleophilic part of the toluene molecule is clearly delocalized over the aromatic ring, and therefore one can expect that the adsorption will take place preferentially on the M-edge with the molecular plane parallel to the edge direction. The "on top" adsorption over a Mo atom is expected to yield a six-coordinated complex (η_6) which could be characterized by the presence of six disynaptic basins between the metal and each carbon of the cycle. In fact there are only two disynpatic V(Mo, C) basins in the adsorbed species which involve the ipso and para carbons. The sum of the integrated densities over these basins is 1.18 e. It results from a density transfer from the $C_{ipso}-C_{ortho}$ and $C_{meta}-C_{para}$ bonds which lowers the population of the related disynaptic basins to ca. 2.3 e whereas the V(C_{ortho}, C_{meta}) population is slightly increased by 0.2 e with respect to the free molecule. Indeed, there is no destablizing loss of aromaticity because the delocalization remains unchanged for these latter bonds. However, the four other bonds are substantially elongated and therefore the distance between the metal and the ortho and meta carbon are slightly larger than those involving the ipso and para ones which is consistent with an effective η_2 structure. These bond elongations may be due to a strong electron delocalization between the Mo site and the toluene ring. In fact, the contribution of the Mo to the V(C, C) basins (0.23 e) has been found to be slightly larger than its contributions to the V(C, Mo) basins (0.19 e).

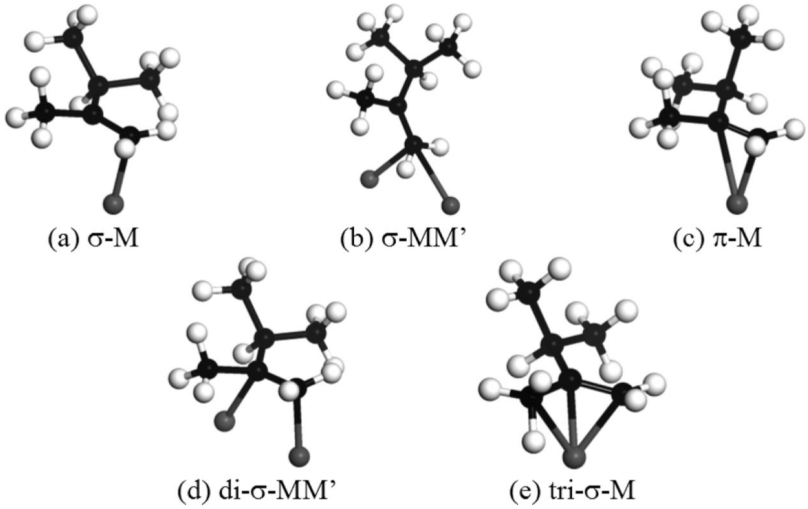
In the η_4 mode, two Co atoms are in bridging positions below the $C_{ortho}-C_{meta}$ and $C_{meta}-C_{para}$ bonds located on the same side with respect to the $C_{ipso}-C_{para}$ direction. Four V(Co, C) disynaptic basins are formed between the Co atoms



(a) 2MT in gas phase (b) η_1-M (c) η_1-MM'

(d) η_2-MM' (e) η_3-MM' (f) η_5-M (g) η_5-MM'

Figure 7. View of different modes of adsorption of 2-methylthiophene. (Black balls: C; dark gray balls: edge site (Mo, Co, or Ni); light gray balls: S).



(a) σ-M (b) σ-MM' (c) π-M

(d) di-σ-MM' (e) tri-σ-M

Figure 8. View of different modes of adsorption of 2,3-dimethylbut-ene. (Black balls: C; dark gray balls: edge site (Mo, Co, or Ni)).

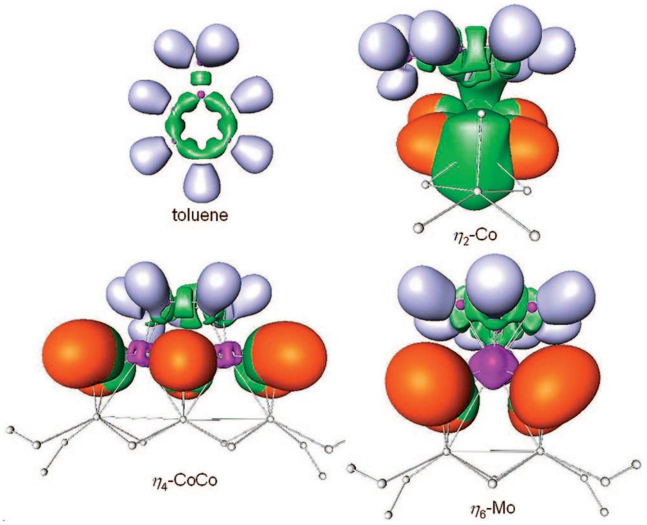


Figure 9. ELF = 0.7 localization domains of toluene showing the aromatic ring. ELF = 0.75 localization domains of the η_2, η_4, and η_6 adsorption modes.

and the nearest carbons, and therefore the η_4 bonding mode can be unambiguously assigned to this adsorption mode. The populations of the V(Co, C) basins range between 0.30 and 0.46 e and mostly come from all the C−C bonds of the cycle except the $C_{ipso}-C_{ortho}$ one which belong to the other half-plane. The basin populations are almost independent of the sulfur coverage, and the stability difference between the 0 and 12.5% edges should be due to the repulsion arising from the sulfur lone pairs. Similar to this, the Co contribution to the V(C, Co) and V(C,C) basins (0.33 e) is exactly the same for both edges.

In the η_2 mode, the Co atom is located below the $C_{ortho}-C_{meta}$ bond. The adsorption process yields disynaptic basins between Co and C_{ortho} and in the case of the 25% sulfur coverage investigated here an interaction of the second C_{ortho} with two linked S atoms because the related V(S, S) basin has a trisynaptic character. However, with respect to the increase of the number of basins due to adsorption, this structure should be named η_1 rather than η_2. Although the V(Co, C_{ortho}) population is low (0.36 e), the changes in the V(C, C) population are rather large; for example, the V(C_{ortho}, C_{meta}) basin loses 1.14 e with respect to the free molecule. Nevertheless, Co contributes to this bond up to 0.02 e over a total of 0.07 e. The calculated interaction energies are first consistent with the Allred and Rochow electronegativity scale and in second place with the number of disynaptic V(C, M) basins.

3.2. 2,3-Dimethylbut-1-ene Adsorption. The detailed results obtained on the adsorption modes and energies of

Table 1. Population Analysis, Deformation, Adsorption, and Interaction Energies (eV) of Toluene Molecule Adsorbed on M-Edge Sites[a]

		CoMoS M-edge: coverage (%M−%S)			
adsorp mode	free	50 (a)−12.5 η_6-Mo	100−0 η_4-CoCo	50 (p)−12.5 η_4-CoCo	50 (p)−25 η_2-Co
$\bar{N}$ [V(Mo, C_{ipso})]		0.53			
$\bar{N}$ [V(Co, C_{ortho})]			0.30	0.33	0.36
$\bar{N}$ [V(Co, C_{meta})]			0.45	0.46	
			0.30	0.39	
$\bar{N}$ [V(Co, C_{para})]			0.43	0.41	
$\bar{N}$ [V(Mo, C_{para})]		0.65			
$\bar{N}$ [V(S, S, C_{ortho})]					3.11
$\bar{N}$ [V(C_{ipso}, C_{ortho})]	2.81	2.35	2.35	2.32	2.44
			3.17	3.18	2.93
$\bar{N}$ [V(C_{ortho}, C_{meta})]	2.82	3.01	2.45	2.47	1.87
			2.33	2.27	2.49
$\bar{N}$ [V(C_{meta}, C_{para})]	2.79	2.28	2.12	2.11	2.60
			2.43	2.44	2.92
$\Sigma\bar{N}$ [V(C, M)]		1.18	1.48	1.59	0.36
$\Sigma\bar{N}$ [V(C, X)|M][b]		0.42	0.33	0.33	0.07
ΔE_{def} (mol)		0.20	0.45	0.49	0.06
ΔE_{def} (edge)		0.53	0.41	0.34	0.36
ΔE_{ads}		−1.73	−0.63	−0.65	0.00
ΔE_{int}		−2.46	−1.68	−1.48	−0.42

[a] Except for η_6-Mo, the toluene molecule loses the C_s symmetry. In the coverage entry, a means alternated, and p means pairing (see Figure 4). [b] X = C, M; M = Mo, Co.

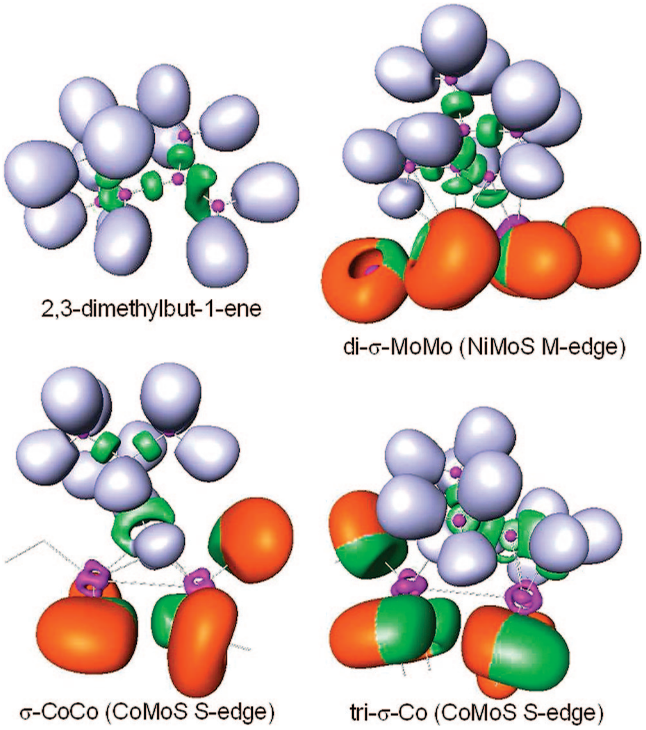


Figure 10. ELF = 0.7 localization domains of 2,3-dimethylbut-1-ene; ELF = 0.80 localization domains of the di-σ-MoMo, σ-CoCo (CoMoS S-edge), and tri-σ-Co adsorptions modes.

2,3-dimethylbut-1-ene (as well as 2-methylthiophene) are reported in ref 11. In what follows, we focus on the ELF analysis and the deformation energy values derived from the adsorption modes found in our previous study. The ELF localization domains of the free 2,3-dimethylbut-1-ene molecule and of three of its adsorption modes are displayed in Figure 10 whereas the population analysis is reported in Tables 2 and 3. The nucleophilic center is the terminal C═C double bond, the approach of which toward a metallic site is not hampered less by repulsions due to sulfur atoms than to those due to the aromatic ring of toluene. Therefore adsorptions on both M- and S-edges are thermodynamically possible.

3.2.1. Adsorption on the M-Edge. The two adsorption configurations on Mo sites (π-Mo and di-σ-MoMo) present significant energetic differences with respect to Co and Ni sites. Even if the molecule is strongly deformed (ΔE_{def} (mol) > +0.9 eV), the adsorption energies are the most exothermic ones (ΔE_{ads} < −1.5 eV) because of highly exothermic interaction energies (ΔE_{int} < −2.5 eV).

Adsorption configurations on Co sites are all presenting a sum of deformation energies lower or equal to +0.9 eV, whereas on Ni, it is less than +0.4 eV. As a consequence, even if adsorption energies on Co and Ni sites are in the same range (between 0 and −0.8 eV), interaction energies are always more exothermic on Co sites (between −0.98 and −1.41 eV) than on Ni sites (above −0.75 eV). In this case, we observe an anticorrelation between the interaction energies and the deformation energies of the molecules for all configurations.

The geometries of the adsorption modes on the CoMoS M-edge sites are all with the double bond "on top" of a metal atom. At low sulfur coverage, 12.5%, there are two disynaptic basins V(M, C(1)) and V(M, C(2)) linking the olefin to the metal M. The population of these basins is on the order of 1.2 with Mo and of 0.5 with Co in agreement with the respective electronegativities of these metals. The V(C(1), C(2)) population decreases from 3.73 in the free molecule to 2.50 in the adsorbed molecule. Such adsorption site should be referred as η_2 rather than as σ and π. At larger sulfur coverage, the adsorbed molecule is linked to the surface by only one disynaptic basin (η_1 mode), the population of which is always less than 0.8 e. The V(C(1), C(2)) population, ~2.72 e, is slightly larger than that of the η_2 mode.

On the NiMoS M-edge, the strongest adsorption occurs for a Ni coverage of 50% without sulfur, the C(1) and C(2) carbons are on top of two Mo atoms, and the olefin molecule is linked to the surface by a disynaptic V(Mo, C(2)) and a trisynaptic V(Mo, C(1), Mo′) basin both with populations larger than 1 e (1.3 and 1.64, respectively). Accordingly, the C(1)−C(2) bond loses its double bond character because its population is lowered to 1.84 e. This density transfers are expected to favor the hydrogenation of the olefin. When the Ni coverage is increased to 100%, the larger electronegativity of Ni hampers the formation of two basins and only the V(Ni, C(1)) basin with a rather small population of 0.64 e is observed. The C−Ni bond lengths are indeed larger (2.23 Å and 2.66 Å) than the C−Mo ones (from 2.12 Å to 2.37 Å). The population of V(C(1), C(2)), 2.70, is close to those calculated for the adsorption in the CoMoS structures. Finally, rather small sulfur coverage hampers the formation of any disynaptic basin. Nevertheless, a stable complex can be formed. Although this adsorption mode belongs to physisorption, the surface and the ad-molecule share the same valence shell because the ELF bifurcation value is 0.24. The

Table 2. Population Analysis, Deformation, Adsorption, and Interaction Energies (eV) of 2,3-Dimethylbut-1-ene Molecule Adsorbed on M-Edge Sites[a]

		CoMoS M-edge: coverage (%M−%S)				
adsorp mode	free	50 (a)−25 σ-Co	50 (a)−12.5 π-Mo	50 (p)−25 σ-Co	50 (p)−12.5 π-Co	100−25 π-Co
$\bar{N}$ [V(Co, C(1))]		0.79		0.79	0.65	0.66
$\bar{N}$ [V(Mo, C(1))]			1.26			
$\bar{N}$ [V(Co, C(2))]					0.36	
$\bar{N}$ [V(Mo, C(2))]			1.20			
$\bar{N}$ [V(C(1), C(2))]	3.73	2.72	2.50	2.71	2.50	2.73
$\Sigma\bar{N}$ [V(C, M)]		0.79	2.46	0.79	1.01	0.66
$\Sigma\bar{N}$ [V(C, X)|M][b]		0.15	0.48	0.16	0.24	0.25
ΔE_{def} (mol)		0.27	0.92	0.28	0.43	0.37
ΔE_{def} (edge)		0.63	0.13	0.34	0.26	0.24
ΔE_{ads}		−0.08	−1.54	−0.43	−0.72	−0.73
ΔE_{int}		−0.98	−2.59	−1.05	−1.41	−1.34

	NiMoS M-edge: coverage (%M−%S)		
adsorp mode	50 (p)−12.5 σ-Ni	50 (p)−0 di-σ-MoMo	100−0 σ-Ni
$\bar{N}$ [V(Ni, C(1))]			0.64
$\bar{N}$ [V(Mo, C(2))]		1.30	
$\bar{N}$ [V(Mo, C(1), Mo′)]		1.64	
$\bar{N}$ [V(C(1), C(2))]	3.56	1.84	2.70
$\Sigma\bar{N}$ [V(C, M)]	0.00	2.94	0.64
$\Sigma\bar{N}$ [V(C, X)|M][b]	0.08	0.63	0.14
ΔE_{def} (mol)	0.07	1.85	0.20
ΔE_{def} (edge)	0.06	0.25	0.15
ΔE_{ads}	−0.13	−1.94	−0.40
ΔE_{int}	−0.26	−4.04	−0.75

[a] In the coverage entry, a means alternated, and p means pairing (see Figure 4). [b] X = C, M; M = Mo, Co, Ni.

Table 3. Population Analysis, Deformation, Adsorption, and Interaction Energies (eV) of 2,3-Dimethylbut-1-ene Molecule Adsorbed on S-Edge Sites[a]

	CoMoS S-edge: coverage (%M−%S)		NiMoS S-edge: coverage (%M−%S)		
adsorp mode	100−50 σ-CoCo	100−37.5 tri-σ-Co	50 (a)-50 π-Mo	50 (a)-37.5 di-σ-NiMo	100−37.5 σ-NiNi
$\bar{N}$ [V(Co, C(1))]	0.99	0.87			
$\bar{N}$ [V(Mo, C(1))]			0.86		
$\bar{N}$ [V(Ni, C(1))]				1.34	0.66
$\bar{N}$ [V(Co, C(2))]		0.71			
$\bar{N}$ [V(Mo, C(2))]			0.69	0.91	
$\bar{N}$ [V(C(1), C(2))]	2.78	2.50	2.44	2.05	2.96
$\Sigma\bar{N}$ [V(C, M)]	0.99	1.58	1.55	2.25	0.66
$\Sigma\bar{N}$ [V(C, X)|M][b]	0.18	0.43	0.37	0.48	0.12
ΔE_{def} (mol)	0.16	0.37	0.54	1.37	0.10
ΔE_{def} (edge)	1.01	1.29	2.08	1.48	0.14
ΔE_{ads}	0.72	−0.51	0.81	0.16	−0.12
ΔE_{int}	−0.45	−2.18	−1.81	−2.69	−0.35

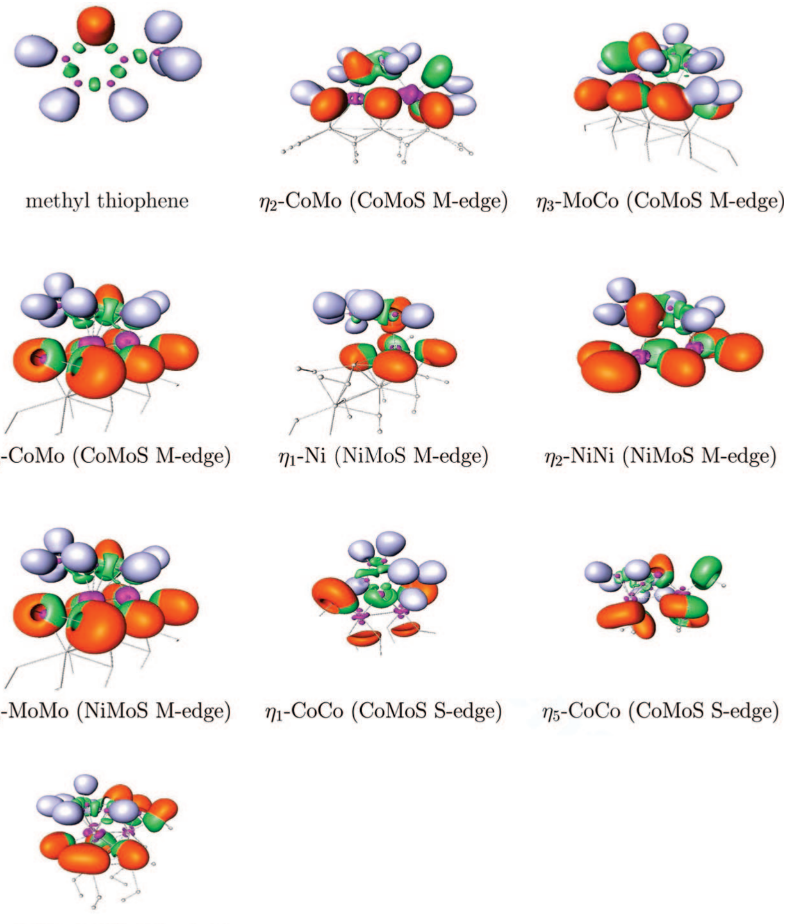
[a] In the coverage entry, a means alternated, and p means pairing (see Figure 5). [b] X = C, M; M = Mo, Co, Ni.

population of the double bond is noticeably perturbed because it is lowered by 0.17 e with respect to the free molecule.

As a general rule for equivalent adsorption modes, the Co and Ni contributions to the V(C, M) and V(C, C) basins have very similar values independent of the V(C, M) populations whereas the contribution of Mo is significantly larger.

3.2.2. Adsorption on the S-Edge. On this edge, we can distinguish the sites of adsorption of the molecule by comparing deformation energies: adsorption on pure Ni sites leads to a sum of deformation energies lower to +0.3 eV whereas it comprises between +1.1 and +1.7 eV on pure Co sites and above +2.6 eV on sites with at least one Mo atom. Hence, some resulting adsorption energies are endothermic, even if all interaction energies are exothermic. It is important to underline that the S-edge deformation energies are generally higher than those of the M-edge, because of the energy cost for the relaxation of S-atoms observed on the S-edge.

The olefin is linked to the S-edge by two disynaptic basins either when the sulfur coverage is less than 50% (CoMoS) or if a Mo site is involved (NiMoS). The thermodynamically favored adsorption, di-σ-NiMo, contains at the same time a molybdenum site and a sulfur coverage less than 50%. The double bond forms a bridge between the Ni and Mo sites; therefore, the two disynaptic basins connect different metal cores to different carbons, and their populations are 1.34 and 0.91 for V(Ni, C(1)) and V(Mo, C(2)), respectively. Consequently, the population of the V(C(1), C(2)) basins is lowered to a single bond value, i.e. 2.05 e. The two other



methyl thiophene

η_2-CoMo (CoMoS M-edge)

η_3-MoCo (CoMoS M-edge)

η_5-CoMo (CoMoS M-edge)

η_1-Ni (NiMoS M-edge)

η_2-NiNi (NiMoS M-edge)

η_5-MoMo (NiMoS M-edge)

η_1-CoCo (CoMoS S-edge)

η_5-CoCo (CoMoS S-edge)

η_5-NiMo (NiMoS S-edge)

Figure 11. ELF = 0.8 localization domains of 2-methylthiophene; ELF = 0.75 localization domains of the η_2-CoMo, η_3-MoCo, η_5-CoMo, η_1-Ni, η_2-NiNi, η_5-MoMo, η_1-CoCo η_5-CoCo, and η_5-NiMo adsorptions modes.

adsorption modes involving two disynaptic basins, the so-called tri-σ-Co and π-Mo, are electronically very similar: on the one hand, the basin population V(M, C(1)) and V(M, C(2)) are very close independent of the nature of the metal (0.87 and 0.86 for V(M, C(1)), 0.71 and 0.69 for V(M, C(2)); on the other hand, the V(C(1), C(2)) population has almost the same values as the other π modes on M-edge. The difference of the interaction energies is consistent with the metal contributions to the olefin basins.

The V(M, C(1)) populations, metal contributions to the olefin basins, and interaction energies of the σ-MM modes are in line with the electronegativities of Co and Ni.

3.3. 2-Methylthiophene Adsorption. Figure 11 displays the localization domains of free and adsorbed 2-methylthiophene whereas the ELF population analysis is reported in Tables 4, 5, and 6. The nucleophilic regions of 2-methylthiophene are, on the one hand, the heterocycle which has an aromatic character and, on the other hand, the sulfur lone pairs which can form a dative bond with the metallic sites.

3.3.1. Adsorption on the M-Edge. First, we observe that the sum of deformation energies depends on the type of mode, i.e. number of supposed interacting bonds between the molecule and the edge: the deformation energies increased with i in the η_i-mode. Consequently, a ranking is observed for interaction energies: greater than −0.6 eV for η_1 modes, between −0.6 and −1.3 eV for η_2 modes, around −1.8 eV for η_3 modes, and below −3.0 eV for η_5 modes. As for the olefin molecules, we observe an anticorrelation

Table 4. Population Analysis, Deformation, Adsorption, and Interaction Energies (eV) of 2-Methylthiophene Molecule Adsorbed on CoMoS M-Edge Sites[a]

		CoMoS M-edge: coverage (%M−%S)			
adsorp mode	free	50 (a)−25 η_2-CoMo	50 (a)−12.5 η_5-CoMo	50 (p)−25 η_3-MoCo	50 (p)−12.5 η_3-CoCo
$\bar{N}$ [V(Co, S)]		2.07	2.05		
$\bar{N}$ [V(Co′, S)]					2.26
$\bar{N}$ [V(Mo, S)]				1.98	
$\bar{N}$ [V(Co, C(5))]				1.06	1.28
$\bar{N}$ [V(Co, C(4))]				0.52	0.66
$\bar{N}$ [V(Mo, C(5))]		1.10	0.93		
$\bar{N}$ [V(Mo, C(4))]			0.45		
$\bar{N}$ [V(Mo, C(3))]			0.28		
$\bar{N}$ [V(Mo, C(2))]			1.08		
$\bar{N}$ [V(S, C(5))]	1.29	1.21	1.20	1.23	1.16
$\bar{N}$ [V(S, C(2))]	1.31	1.42	1.24	1.30	1.31
$\bar{N}$ [V(C(4), C(5))]	3.82	2.58	2.37	2.34	2.36
$\bar{N}$ [V(C(3), C(4))]	2.21	2.38	2.46	2.14	2.09
$\bar{N}$ [V(C(2), C(3))]	3.91	3.44	2.53	3.70	3.73
$\sum\bar{N}$ [V(C, M)]		1.10	2.74	1.58	1.94
$\sum\bar{N}$ [V(Y, X)|M][b]		0.21	0.53	0.34	0.47
ΔE_{def} (mol)		0.22	0.54	0.40	0.42
ΔE_{def} (edge)		0.49	0.47	0.43	0.27
ΔE_{ads}		−0.54	−2.31	−0.90	−1.19
ΔE_{int}		−1.25	−3.32	−1.74	−1.88

[a] In the coverage entry, a means alternated, and p means pairing (see Figure 4). [b] Y = C, S; X = C, S, M; M = Mo, Co.

Table 5. Population Analysis, Deformation, Adsorption, and Interaction Energies (eV) of 2-Methylthiophene Molecule Adsorbed on NiMoS M-Edge Sites[a]

	NiMoS M-edge: coverage (%M−%S)				
adsorp mode	50 (p)−12.5 η_2-MoNi	50 (p)−12.5 η_1-Ni	50 (p)−0 η_5-MoMo	100−0 η_2-NiNi	100−0 η_1-Ni
$\bar{N}$ [V(Ni, S)]		1.95		1.98	1.86
$\bar{N}$ [V(Mo, S)]	1.97		1.74		
$\bar{N}$ [V(Ni, C(5))]	0.87			0.85	
$\bar{N}$ [V(Mo, C(5))]			1.66		
$\bar{N}$ [V(Mo′, C(4))]			0.51		
$\bar{N}$ [V(Mo′, C(3))]			0.49		
$\bar{N}$ [V(Mo, C(2))]			1.44		
$\bar{N}$ [V(S, C(5))]	1.31	1.31	1.13	1.30	1.32
$\bar{N}$ [V(S, C(2))]	1.35	1.32	1.10	1.37	1.32
$\bar{N}$ [V(C(4), C(5))]	2.62	3.66	1.96	2.62	3.62
$\bar{N}$ [V(C(2), C(3))]	3.53	3.75	2.33	3.52	3.71
$\bar{N}$ [V(C(3), C(4))]	2.31	2.31	2.55	2.32	2.32
$\sum\bar{N}$ [V(C, M)]	0.87	0.0	4.10	0.85	0.00
$\sum\bar{N}$ [V(Y, X)|M][b]	0.22	0.08	0.70	0.24	0.11
ΔE_{def} (mol)	0.13	0.00	1.29	0.09	0.02
ΔE_{def} (edge)	0.27	0.05	0.26	0.19	0.11
ΔE_{ads}	−0.26	−0.23	−2.90	−0.45	−0.41
ΔE_{int}	−0.66	−0.28	−4.45	−0.73	−0.54

[a] In the coverage entry, a means alternated, and p means pairing (see Figure 4). [b] Y = C, S; X = C, S, M; M = Mo, Ni.

between the interaction energy and the deformation energy of the molecule. This trend explains why there is no direct link between adsorption energies and the bonding mode. Furthermore, deformation energies of the thiophene molecule are in general smaller than those of the olefin.

At low sulfur coverage, both CoMoS and NiMoS edges are able to adsorb 2-methylthiophene through an η_5 mode characterized by one V(M, S) basin and four V(Mo, C) ones. The population of the V(Mo, C) basins are on the order of 1 and 1.5 e for the CoMoS and NiMoS edges, respectively, whereas the two other V(Mo, C) basins have populations on the order of 0.5 e or less. Although the population of the V(M, S) basin is larger on the CoMoS than on the NiMoS edge, the interaction of the latter is larger in absolute value, indicating that the adsorption is controlled by the V(Mo, C) populations. An equalization of the V(C, C) basin populations of the cycle with respect to the free molecule is observed.

The η_3 adsorption modes occur on the CoMoS edge. They involve the sulfur and the C(4) and C(5) carbons. The interaction between Co and the former carbon is weaker than that between Co and the latter, the ratio of the populations of V(Co, C(5)) and V(Co, C(4)) being on the order of 2. The interaction being geometrically asymmetric with respect to S, only V(C(4), C(5)) undergoes a significant variation of its population.

The η_2 modes are electronically very similar, although the V(M, C(5)) population is larger for the CoMoS than for the NiMoS because it involves a Mo site in the CoMoS rather than a Ni site in the NiMoS. It is worth noting that the V(C(2), C(3)) is more perturbed than for η_3 modes.

The η_1 modes of NiMoS clearly belong to physisorption, as the bifurcation value of ELF is less than 0.2 for both.

3.3.2. Adsorption on the S-Edge. As for the M-edge, we can distinguish the adsorption modes by the sum of their deformation energies: for η_1 modes it is less than 1.5 eV whereas it is more than 2.0 eV for η_5 modes. Consequently, even if there is no direct link with adsorption energies, interaction energies of η_5 modes are always more exothermic than for η_1 modes (less than −2.5 eV vs more than −1.5 eV). Furthermore, for the 2-methylthiophene molecule as well as for the olefin, it appears that the deformation energy of the S-edge is always larger than that of the M-edge.

Two adsorption modes are possible on the S-edge: either the highly coordinated η_5 or the η_1 in which the 2-methylthiophene structure substitutes a sulfur edge atom. The bond properties of η_5-CoCo are very similar to those of the M-edge η_5-CoMo. The η_5-NiMo is in fact an η_4 mode involving a trisynaptic bridging V(Ni, S, Mo) basin. The equalization of the V(C, C) basin trend is weaker than for the true η_5 mode.

In the η_1 modes, the lone pairs of the sulfur atoms form two dative bonds with two neighboring metallic sites, the 2-methylthiophene sulfur occupying a "substitution" site. The populations within the heterocycle are very weakly perturbed and the contribution of the metallic sites is always larger than for the M-edge η_1 mode. This contribution as well as the interaction energy absolute value is larger for η_1-CoCo than for η_1-NiNi and η_1-MoNi. Similarly, the η_1-CoCo S−Co bonds (2.21 Å) are smaller than η_1-NiNi and η_1-MoNi S−Ni (2.32 Å to 3.01 Å) and S−Mo bonds (2.58 Å).

3.4. Correlation between Metallic Site Contributions and Interaction Energy. The calculation of the basin population provides a basis for the understanding of the electronic scheme during the adsorption of the three molecules (toluene, 2-methylthiophene, and 2,3-dimethylbut-1-ene) on both edges. As a first insight, it appears clear that even if the adsorption energy (defined according to eq 7) is the most representative for the thermodynamic process, this quantity cannot be correlated with any of the electronic descriptors provided by the basin population analysis. According to eq 7, the adsorption energy contains the contribution of three distinct terms: the interaction energy, and the deformation energies of the molecule and of the edge. According to the values reported in Table 1−6 and described in the previous section, it appears clear that the deformation energies cannot be neglected and strongly depend on the molecules, on the adsorption mode, and on the edge considered. Hence, it is impossible to find a direct correlation between ELF basin populations and the adsorption energy. In contrast, it can be reasonably expected that the V(X, M) basins shared between the atoms (X) of the adsorbed molecule and the metallic sites of the catalyst (M) contribute to the interaction energy. Therefore, a first descriptor, $\sum\bar{N}$[V(X, M)], is built by summing the populations of the V(X, M) basins. For toluene and 2,3-dimethylbut-1-ene, the values of this index are reported in the $\sum\bar{N}$[V(C, M)] entry of Tables 1−3 whereas for dimethylthiophene the V(M, S) populations have to be added to $\sum\bar{N}$[V(C, M)]. The plot of this sum as a function of the interaction energy (Figure 12) enables us to distinguish the thiophene molecule from the olefin and toluene molecules. The population of the V(S, M) basin is indeed a strong contributor to the sum of the population of the V(X, M) basins. This may qualitatively explain why the methylthiophene molecule is more strongly

Table 6. Population Analysis, Deformation, Adsorption, and Interaction Energies (eV) of 2-Methylthiophene Molecule Adsorbed on S-Edge Sites[a]

	CoMoS S-edge: coverage (%M−%S)		NiMoS S-edge: coverage (%M−%S)		
adsorp mode	100−50 η_1-CoCo	100−37.5 η_5-CoCo	50 (a)−50 η_1-MoNi	50 (a)−37.5 η_5-NiMo	100−37.5 η_1-NiNi
$\bar{N}$ [V(Co, S)]	2.28	2.26			
$\bar{N}$ [V(Co′, S)]	2.62				
$\bar{N}$ [V(Mo, S)]			2.15		
$\bar{N}$ [V(Ni, S)]			2.52		2.33
$\bar{N}$ [V(Ni′, S)]					2.44
$\bar{N}$ [V(Ni, S, Mo)]				4.65	
$\bar{N}$ [V(Co, C(5))]		0.96			
$\bar{N}$ [V(Co, C(2))]		0.98			
$\bar{N}$ [V(Co, C(4))]		0.47			
$\bar{N}$ [V(Co, C(3))]		0.47			
$\bar{N}$ [V(Mo, C(5))]				1.13	
$\bar{N}$ [V(Mo, C(2))]				1.02	
$\bar{N}$ [V(Mo, C(4))]				0.35	
$\bar{N}$ [V(S, C(5))]	1.22	1.20	1.27	1.22	1.27
$\bar{N}$ [V(S, C(2))]	1.22	1.16	1.28	1.24	1.27
$\bar{N}$ [V(C(4), C(5))]	3.69	2.40	3.68	2.40	3.67
$\bar{N}$ [V(C(2), C(3))]	3.77	2.53	3.75	2.56	3.76
$\bar{N}$ [V(C(3), C(4))]	2.13	2.25	2.20	2.78	2.17
$\Sigma\bar{N}$ [V(C, M)]	0.00	2.88	0.00	2.52	0.00
$\Sigma\bar{N}$ [V(Y, X)|M][b]	0.31	0.60	0.13	0.55	0.19
ΔE_{def} (mol)	0.04	0.73	0.03	0.22	0.03
ΔE_{def} (edge)	1.00	1.45	1.40	3.06	0.41
ΔE_{ads}	−0.24	−1.37	0.35	0.41	−0.59
ΔE_{int}	−1.28	−3.55	−1.08	−2.87	−1.03

[a] In the coverage entry, a means alternated, and p means pairing (see Figure 5). [b] Y = C, S; X = C, S, M; M = Mo, Co, Ni.

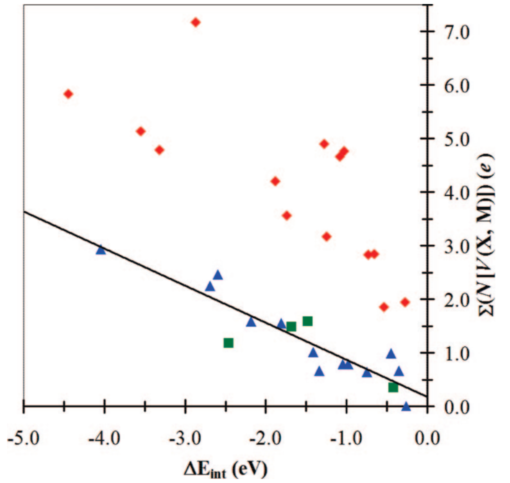


Figure 12. Sum of the populations $\Sigma\bar{N}$[V(X, M)] vs interaction energy. Green square: toluene; blue triangle: 2,3-dimethylbut-1-ene adsorption; red lozenge: 2-methylthiophene adsorption. (X = S, C; M = Mo, Co, Ni).

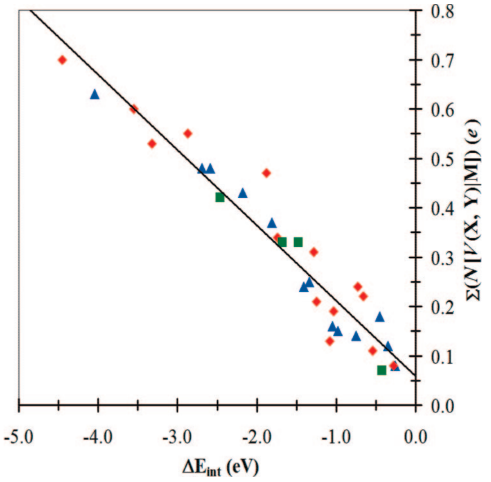


Figure 13. Sum of metal contributions $\Sigma\bar{N}$ [V(X,M) | M] vs interaction energy. Green square: toluene; blue triangle: 2,3-dimethylbut-1-ene adsorption; red lozenge: 2-methylthiophene adsorption. (X = S, C; M = Mo, Co, Ni).

stabilized. In contrast, for the olefin and the toluene, the unique contribution comes from the V(C, M) basins. However, as also revealed by Figure 12, the sum of the populations of V(X, M) basins cannot quantitatively unify the trend in interaction energies for the three molecules. For 2-methylthiophene, the ELF populations are overestimated because of the contribution of some nonbonding part in the V(S, M) population. A unique electronic descriptor valid for the whole sample of studied systems would be nevertheless very useful. In Figure 13, it appears that such a descriptor can be proposed if we consider the $\Sigma\bar{N}$[V(X, Y)|M] quantity. In this case, the value of the correlation coefficient, −0.96, validates this descriptor as relevant for the interaction energy.

The $\Sigma\bar{N}$[V(X, Y)|M] quantity (also reported in Tables 1−3) is calculated by adding the metal AIM atomic basin contributions to the V(C, M), V(C, S), V(C,C), and V(M, S) populations. In other words, it is the integrated density over the intersection of the V(C, M), V(C, S), V(C,C), and V(M, S) basins and the atomic basin of the metal. Hence, it clearly accounts for the electron density shared between the catalyst and the adsorbed molecule. The sum of the metal atomic basin contributions to the substrate ELF basins accounts for the partial covalent character of the interaction and, therefore, is found to be correlated with the electronic interaction energies of the molecules with metallic sites. This interesting correlation gives a rational explanation based on the ELF analysis of the interaction energy term contributing to the resulting adsorption energy. This evaluates the pure contribution of the chemical bonding between the adsorbed molecule and the metallic sites as also contained in the interaction energy, when the other effects such as deformation energies of the edges and molecules are subtracted.

4. Conclusions

The conclusions which can be drawn from the present work concern both technical and chemical aspects of computational chemistry. From a methodological viewpoint, our results demonstrate the feasibility of ELF analysis on rather large systems related to surface chemistry. On the one hand, the calculation of the ELF function and of the integrated densities with a cluster model in which the atomic positions have been optimized at a periodic level is a unique possibility until a periodic code is available for plane waves. Fortunately, this approach yields realistic results at least for the adsorbed molecule subset. On the second hand, the computational effort is significantly reduced by the atom selection procedure which has the additional advantage of simplifying the graphical representations.

The chemical pieces of information provided by the ELF analysis mostly concern the nature of the interactions in terms of bond formation between the catalyst and the adsorbed molecule. In the case of physisorption, such as for the 2-methylthiophene η_1 modes on the NiMoS M-edge, the topology of the catalyst + substrate is just the sum of their topologies. The chemisorption is characterized at least by the increase of the synaptic order of a monosynaptic basin, i.e. V(S) becomes V(S, M) in the 2-methylthiophene adsorption, or by the formation of new polysynaptic basins. The number of disynaptic basins linking the substrate to the metallic site may differ from that expected from a purely geometrical coordination as is the case for the η_6 adsorption mode of toluene for which only two disynaptic basins connect the ipso and para carbons to the molybdenum. The ELF analysis should be therefore helpful to improve the nomenclature used to name the adsorption modes, making it more precise. A proposal of the form η_n (M), $\eta_{n'}$ (M′),..., where the subscripts n, n',..., indicate the number of bonds made by the metallic sites M, M′,..., with the substrate improves the implicit description of the bonding properties of each site. In our opinion such nomenclature is less confusing than that used initially in this paper. The same π-Co name is used for two adsorption modes of the olefin on the CoMoS edge, giving rise to one and two V(C, M) basins; naming them η_2 (Co) and η_1 (Co) enables a better differentiation.

The ELF analysis also sheds light, in a chemical language, on the electronic rearrangements in the substrate, indicating which bonds are activated. For example, our results show that the 2-methylthiophene C–S bonds are weakened in the η_5 adsorption modes and more particularly in the η_5-MoMo one on the NiMoS M-edge which also corresponds to the largest absolute value of the interaction energy.

Finally, we propose an exhaustive analysis of the different energy components involved in the adsorption energy: the interaction energy and the deformation energy of the molecule and of the edge. In particular, because of non-negligible deformation energies of the molecule and of the edge, it can be better understood why the adsorption energy cannot correlate with the ELF electronic descriptor. In contrast, the interaction energy is the most representative of the electronic exchange taking place between the substrate and the adsorbate and thus exhibits a good correlation with the ELF descriptor.

References

(1) Prins, R. Hydrotreating Reactions. In *Handbook of Heterogeneous Catalysis*; Ertl, G.; Knözinger, H.; Weitkamp, J., Eds.; Wiley-VCH: Weinheim, Germany, 1997; vol. *4*, p 1908.

(2) Topsøe, H.; Clausen, B. S.; Massoth, F. E. Catalyst Characterization. In *Hydrotreating Catalysis Science and Technology*; Anderson, J. R., Boudart, M., Eds.; Springer-Verlag: Berlin, Germany, 1996; vol. *11*, p 29.

(3) Raybaud, P. Understanding and predicting improved sulfide catalysts: Insights from first principles modeling. *Appl. Catal. A* **2007**, *322*, 76.

(4) Paul, J.-F.; Cristol, S.; Payen, E. Computational studies of (mixed) sulfide hydrotreating catalysts. *Catal. Today* **2008**, *130*, 139.

(5) Raybaud, P.; Hafner, J.; Kresse, G.; Kasztelan, S.; Toulhoat, H. Structure, Energetics, and Electronic Properties of the Surface of a Promoted MoS_2 Catalyst: An ab Initio Local Density Functional Study. *J. Catal.* **2000**, *190*, 128.

(6) Byskov, L. S.; Nørskov, J. K.; Clausen, B. S.; Topsøe, H. DFT Calculations of Unpromoted and Promoted MoS_2-Based Hydrodesulfurization Catalysts. *J. Catal.* **1999**, *187*, 109.

(7) Schweiger, H.; Raybaud, P.; Toulhoat, H. Promoter Sensitive Shapes of Co(Ni)MoS Nanocatalysts in Sulfo-Reductive Conditions. *J. Catal.* **2002**, *212*, 33.

(8) Sun, M.; Nelson, A. E.; Adjaye, J. On the incorporation of nickel and cobalt into MoS_2-edge structures. *J. Catal.* **2004**, *226*, 32.

(9) Krebs, E.; Silvi, B.; Raybaud, P. Mixed sites and promoter segregation: A DFT study of the manifestation of Le Chatelier's principle for the Co(Ni)MoS active phase in reaction conditions. *Catal. Today* **2008**, *130*, 160.

(10) Gandubert, A. D.; Krebs, E.; Legens, C.; Costa, D.; Guillaume, D.; Raybaud, P. Optimal promoter edge decoration of CoMoS catalysts: A combined theoretical and experimental study. *Catal. Today* **2008**, *130*, 149.

(11) Krebs, E.; Silvi, B.; Daudin, A.; Raybaud, P. A DFT study of the origin of the HDS/HydO selectivity on Co(Ni)MoS active phases. *J. Catal.* **2008**, *260*, 276.

(12) Cristol, S.; Paul, J.-F.; Payen, E.; Bougeard, D.; Hutschka, F.; Clemendot, S. DBT derivatives adsorption over molybdenum sulfide catalysts: a theoretical study. *J. Catal.* **2004**, *224*, 138.

(13) Cristol, S.; Paul, J.-F.; Schovsbo, C.; Veilly, E.; Payen, E. DFT study of thiophene adsorption on molybdenum sulfide. *J. Catal.* **2006**, *239*, 145.

(14) Raybaud, P.; Hafner, J.; Kresse, G.; Toulhoat, H. Adsorption of Thiophene on the Catalytically Active Surface of MoS_2: An Ab Initio Local-Density-Functional Study. *Phys. Rev. Lett.* **1998**, *80*, 1481.

(15) Sun, M.; Nelson, A. E.; Adjaye, J. Adsorption Thermodynamics of Sulfur- and Nitrogen-containing Molecules on NiMoS: A DFT Study. *Catal. Lett.* **2006**, *109*, 133.

(16) Orita, H.; Uchida, K.; Itoh, N. A volcano-type relationship between the adsorption energy of thiophene on promoted MoS_2 cluster-model catalysts and the experimental HDS activity: ab initio density functional study. *Appl. Catal. A* **2004**, *258*, 115.

(17) Weber, T.; van Veen, J. R. A density functional theory study of the hydrodesulfurization reaction of dibenzothiophene to biphenyl on a single-layer NiMoS cluster. *Catal. Today* **2008**, *130*, 170.

(18) Sun, M.; Nelson, A. E.; Adjaye, J. Adsorption and hydrogenation of pyridine and pyrrole on NiMoS: an ab initio density-functional theory study. *J. Catal.* **2005**, *231*, 223.

(19) Sun, M.; Nelson, A. E.; Adjaye, J. First principles study of heavy oil organonitrogen adsorption on NiMoS hydrotreating catalysts. *Catal. Today* **2005**, *109*, 49.

(20) Guernalec, N.; Geantet, C.; Raybaud, P.; Cseri, T.; Aouine, M.; Vrinat, M. Dual Effect of H_2S on Volcano Curves in Hydrotreating Sulfide Catalysis. *Oil Gas Sci. Technol.* **2006**, *61*, 515.

(21) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford Univ.Press: Oxford, Great Britain, 1990.

(22) Gadre, S. R.; Shirsat, R. N. *Electrostatics of Atoms and Molecules*; Hyderabad Universities Press: Hyderabad, India, 2000.

(23) Silvi, B.; Savin, A. Classification of chemical bonds based on topological analysis of electron localization function. *Nature* **1994**, *371*, 683.

(24) Aray, Y.; Rodriguez, J.; Vega, D.; Rodriguez-Arias, E. N. Correlation of the Topology of the Electron Density of Pyrite-Type Transition Metal Sulfides with Their Catalytic Activity in Hydrodesulfurization. *Angew. Chem., Int. Ed.* **2000**, *39*, 3810.

(25) Aray, Y.; Rodriguez, J. Atoms in molecules theory for exploring the nature of the MoS_2 catalyst edge sites. *J. Mol. Catal. A: Chem.* **2007**, *265*, 32.

(26) Aray, Y.; Rodriguez, J.; Vega, D. Laplacian of the Electronic Charge Density and Heat of Adsorption of O_2 and CO Molecules on 3d Transition Metals. *J. Phys. Chem. B* **2000**, *104*, 5225.

(27) Aray, Y.; Rodriguez, J.; Vega, D.; Coll, S.; Rodriguez-Arias, E.; Rosillo, F. Adsorption of Thiophene on the RuS_2 (100) and (111) Surfaces: A Laplacian of the Electronic Charge Density Study. *J. Phys. Chem. B* **2002**, *106*, 13242.

(28) Aray, Y.; Rodriguez, J.; Coll, S.; Rodriguez-Arias, E.; Vega, D. Nature of the Lewis Acid Sites on Molybdenum and Ruthenium Sulfides: An Electrostatic Potential Study. *J. Phys. Chem. B* **2005**, *109*, 23564.

(29) Aray, Y.; Marquez, M.; Rodriguez, J.; Coll, S.; Simon-Manso, Y.; Gonzalez, C.; Weitz, D. Electrostatics for Exploring the Nature of Water Adsorption on the Laponite Sheets' Surface. *J. Phys. Chem. B* **2003**, *107*, 8946.

(30) Aray, Y.; Rodriguez, J.; Coll, D.; Gonzalez, C.; Marquez, M. Exploring the Nature of Wetting by Water of Surfaces of Alkane-Amidethiols Adsorbed on Gold Using the Electrostatic Potential Topology. *J. Phys. Chem. B* **2004**, *108*, 18942.

(31) Aray, Y.; Marquez, M.; Rodriguez, J.; Vega, D.; Simon-Manso, Y.; Coll, S.; Gonzalez, C.; Weitz, D. Electrostatics for Exploring the Nature of the Hydrogen Bonding in Polyethylene Oxide Hydration. *J. Phys. Chem. B* **2004**, *108*, 2418.

(32) Ample, F.; Curulla, D.; Fuster, F.; Clotet, A.; Ricart, J. M. Adsorption of CO and CN^- on transition metal surfaces: a comparative study of the bonding mechanism. *Surf. Sci.* **2002**, *497*, 139.

(33) Mori-Sánchez, P.; Recio, J. M.; Silvi, B.; Sousa, C.; Martín Pendás, A.; Luaña, V.; Illas, F. The rigorous characterization of MgO *F* centers as pseudoatoms. *Phys. Rev. B* **2002**, *66*, 075103.

(34) Gomes, J. R. B.; Illas, F.; Silvi, B. Topological analysis of the metal-support interaction: the case of Pd atoms on oxide surfaces. *Chem. Phys. Lett.* **2004**, *388*, 132.

(35) Abraham, R. H.; Shaw, C. D. *Dynamics: The Geometry of Behavior*; Addison Wesley: Redwood City, CA, 1992.

(36) Abraham, R. H.; Marsden, J. E. *Foundations of Mechanics*; Addison Wesley: Redwood City, CA., 1994.

(37) Martín Pendás, A.; Blanco, M. A.; Costales, A.; Mori Sánchez, P.; Luana, V. Non-nuclear Maxima of the Electron Density. *Phys. Rev. Lett.* **1999**, *83*, 1930.

(38) Becke, A. D.; Edgecombe, K. E. A simple mesure of electron localization in atomic and molecular systems. *J. Chem. Phys.* **1990**, *92*, 5397.

(39) von Weizsäcker, C. F. Zur Theorie der Kermassen. *Z. Phys.* **1935**, *96*, 431.

(40) Savin, A.; Becke, A. D.; Flad, J.; Nesper, R.; Preuss, H.; von Schnering, H. G. A New Look at Electron Localization. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 409.

(41) Savin, A.; Jepsen, O.; Flad, J.; Andersen, O. K.; Preuss, H.; von Schnering, H. G. Electron Localization in the Solid-State Structures of the Elements: the Diamond Structure. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 187.

(42) Savin, A.; Nesper, R.; Wengert, S.; Fässler, T. F. ELF: The Electron Localization Function. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 1809.

(43) Burdett, J. K.; McCormick, T. A. Electron Localization in Molecules and Solids: The Meaning of ELF. *J. Phys. Chem. A* **1998**, *102*, 6366.

(44) Nalewajski, R. F.; Koster, A. M.; Escalante, S. Electron Localization Function as Information Measure. *J. Phys. Chem. A* **2005**, *109*, 10038.

(45) Dobson, J. F. Interpretation of the Fermi hole curvature. *J. Chem. Phys.* **1991**, *94*, 4328.

(46) Kohout, M.; Pernal, K.; Wagner, F. R.; Grin, Y. Electron localizability indicator for correlated wavefunctions. I. Parallel spin pairs. *Theor. Chem. Acc.* **2004**, *112*, 453.

(47) Kohout, M.; Pernal, K.; Wagner, F. R.; Grin, Y. Electron localizability indicator for correlated wavefunctions. I. Antiparallel spin pairs. *Theor. Chem. Acc.* **2005**, *113*, 287.

(48) Silvi, B. The Spin Pair Compositions as Local Indicators of the Nature of the Bonding. *J. Phys. Chem. A* **2003**, *107*, 3081.

(49) Matito, E.; Silvi, B.; Duran, M.; Solà, M. Electron localization function at the correlated level. *J. Chem. Phys.* **2006**, *125*, 024301.

(50) Häussermann, U.; Wengert, S.; Nesper, R. Localization of Electrons in Intermetallic Phases Containing Aluminium. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 2069.

(51) Silvi, B. The synaptic order: a key concept to understand multicenter bonding. *J. Mol. Struct.* **2002**, *614*, 3.

(52) Lewis, G. N. The Atom and the Molecule. *J. Am. Chem. Soc.* **1916**, *38*, 762.

(53) Lewis, G. N. *Valence and the Structure of Atoms and Molecules*; Dover: New York, 1966.

(54) Silvi, B. How topological partitions of the electron distributions reveal delocalization. *Phys. Chem. Chem. Phys.* **2004**, *6*, 256.

(55) Poater, J.; Duran, M. Sol evaluation of electron delocalization in aromatic molecules by means of AIM and ELF topological approaches. *Chem. Rev.* **2005**, *105*, 3911.

(56) Raub, S.; Jansen, G. A quantitative measure of bond polarity from the electron localization function and the theory of atoms in molecules. *Theor. Chem. Acc.* **2001**, *106*, 223.

(57) Mezey, P. G. Quantum chemical shape: new density domain relations for the topology of molecular bodies, functional groups, and chemical bonding. *Can. J. Chem.* **1993**, *72*, 928.

(58) Alikhani, M. E.; Fuster, F.; Silvi, B. What can tell the topological analysis of ELF on hydrogen bonding. *Struct. Chem.* **2005**, *16*, 203.

(59) Noury, S.; Krokidis, X.; Fuster, F.; Silvi, B. Computational tools for the electron localization function topological analysis. *Comput. Chem.* **1999**, *23*, 597.

(60) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648.

(61) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev.* **1988**, *A38*, 3098.

(62) Lee, C.; Yang, Y.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev.* **1988**, *B37*, 785.

(63) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200.

(64) 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. 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.; G. Liu, A. L.; 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 A.1; Gaussian Inc., Pittsburgh PA, 2003.

(65) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15.

(66) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169.

(67) Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev.* **1992**, *B45*, 13244.

(68) Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.* **1980**, *102*, 939.

(69) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements. *J. Am. Chem. Soc.* **1982**, *104*, 2797.

(70) Pietro, W. J.; Francl, M. M.; Hehre, W. J.; Defrees, D. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods. 24. Supplemented small split-valence basis sets for second-row elements. *J. Am. Chem. Soc.* **1982**, *104*, 5039.

(71) Dobbs, K. D.; Hehre, W. J. Molecular orbital theory of the properties of inorganic and organometallic compounds. 4. Extended basis sets for third-and fourth-row, main-group elements. *J. Comput. Chem.* **1986**, *7*, 359.

(72) Dobbs, K. D.; Hehre, W. J. Molecular orbital theory of the properties of inorganic and organometallic compounds. 5. Extended basis sets for first-row transition metals. *J. Comput. Chem.* **1987**, *8*, 861.

(73) Dobbs, K. D.; Hehre, W. J. Molecular orbital theory of the properties of inorganic and organometallic compounds. 6. Extended basis sets for first-row transition metals. *J. Comput. Chem.* **1987**, *8*, 880.

(74) Amira 3.0; TGS, Template Graphics Sofware, Inc., San Diego, CA, 2002.

(75) 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.

(76) Pauling, L. *The Nature of the Chemical Bond*; Cornell University Press: Ithaca, NY, 1948.

(77) Allred, A. L.; Rochov, E. G. A Scale of Electronegativity Based on Electrostatic Force. *J. Inorg. Nucl. Chem.* **1958**, *5*, 264.

CT800344R

JCTC Journal of Chemical Theory and Computation

Valence–Bond Order (VBO): A New Approach to Modeling Reactive Potential Energy Surfaces for Complex Systems, Materials, and Nanoparticles

Meiyu Zhao,[†] Mark A. Iron,[†,§] Przemysław Staszewski,[†,‡] Nathan E. Schultz,[†]
Rosendo Valero,[†] and Donald G. Truhlar*[,†]

Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431, and Department of Theoretical Foundations of Biomedical Sciences and Medical Informatics, Collegium Medicum, Nicolaus Copernicus University, ul. Jagiellońska 13, 85-067 Bydgoszcz, Poland

Received October 28, 2008

Abstract: The extension of molecular mechanics to reactive systems, metals, and covalently bonded clusters with variable coordination numbers requires new functional forms beyond those popular for organic chemistry and biomolecules. Here we present a new scheme for reactive molecular mechanics, which is denoted as the valence–bond order model, for approximating reactive potential energy surfaces in large molecules, clusters, nanoparticles, solids, and other condensed-phase materials, especially those containing metals. The model is motivated by a moment approximation to tight binding molecular orbital theory, and we test how well one can approximate potential energy surfaces with a very simple functional form involving only interatomic distances with no explicit dependence on bond angles or dihedral angles. For large systems the computational requirements scale linearly with system size, and no diagonalizations or iterations are required; thus the method is well suited to large-scale simulations. The method is illustrated here by developing a force field for particles and solids composed of aluminum and hydrogen. The parameters were optimized against both interaction energies and relative interaction energies. The method performs well for pure aluminum clusters, nanoparticles, and bulk lattices and reasonably well for pure hydrogen clusters; the mean unsigned error per atom for the aluminum−hydrogen clusters is 0.1 eV/atom.

1. Introduction

The properties of molecules and materials result from internuclear (atomic) dynamics governed, in the Born−Oppenheimer approximation, by a potential energy surface that results from the system's electronic structure.[1] In principle, the electronic structure can be solved with high accuracy, and great progress has been made in accomplishing this for small molecules.[2,3] There are two general approaches for extending this progress to very large systems. The first is to derive simpler, yet sufficiently accurate, electronic structure methods. This includes theoretical advances in the description of electronic structure,[4−7] the development of semiempirical quantum mechanical electronic structure methods with reasonably reliable parametrizations,[8−13] and improved algorithms,[14−16] such as linear scaling methods.[17−19] The second general approach is to obtain analytic approximations to potential energy surfaces. If this is done for specific small systems,[20−22] additional work is needed to apply the treatment to extended systems. If this, however, can be done in terms of general parameters that are independent of the system size, then materials and large systems can be modeled without additional apparatus.

---

* To whom correspondence should be addressed. E-mail: truhlar@umn.edu.

[†] University of Minnesota.

[‡] Nicolaus Copernicus University.

[§] Current address: Computational Chemistry Unit, Department of Chemical Research Support, Weizmann Institute of Science, Rehovot, Israel 76100.

10.1021/ct8004535 CCC: $40.75

Examples of such an approach include recent potential functions for a variety of materials,[23−31] including Al nanoparticles.[30,31]

Although traditional analytical potentials often have all their parameters fit to experiment, a more modern approach is to use high-level calculations. We note first that potentials fit only to properties of bulk metals or bulk materials containing metal atoms are unlikely to be accurate for defects, surfaces, nanostructured materials, or nanoparticles, but systematic data for these latter types of features or systems are scarce. However, many energetic properties that are not experimentally readily accessible can be calculated with reasonable accuracy. Even when measured values are available, it is sometimes hard to say whether a given measured value should be used in a fit, especially if the possible experimental error is greater than the desired accuracy of the computational method being parametrized. One recent example involved the heats of formation of the oxides and hydroxides of alkali and alkaline earth metals where widely different experimental values were obtained.[32] One main obstacle, however, to using calculated rather than experimental data is the difficulty in carrying out benchmark electronic structure calculations on subsystems that are large enough to represent the full complexity of the materials.

Analytic potentials are sometimes called classical potentials because they can be evaluated without recourse to the Schrödinger equation or other quantum formalisms. This is, however, a deceptive name for potentials fit to quantum mechanical electronic structure calculations.[31] If the fit is carefully done and the functional forms properly represent the physics, these potentials can capture the full quantum mechanical behavior of the potential energy surface within the Born−Oppenheimer approximation.

Great progress has been made in developing general parameters for analytic potentials for hydrocarbons, carbohydrates, biopolymers, and diverse organic compounds,[33−42] where this approach is traditionally termed molecular mechanics. Most molecular mechanics force fields are expressed in terms of local valence coordinates, specifically stretches, bends, and torsions, plus interatomic distances of nonbonded atoms. Since most molecular mechanics force fields do not describe bond breaking or bond rearrangement, they are restricted to nonreactive systems. Furthermore, they are easiest to parametrize for systems where the geometries are reasonably rigid and bonding types can be readily classified and catalogued, whereas valences of metal atoms are more variable, and bending and torsion parameters at metal−atom centers seem to be less transferable because the compounds are more flexible and bond lengths more variable than in organic chemistry. Thus very little progress has been made in extending this kind of force field to organometallic and inorganometallic systems, although Landis and Deeth and co-workers have attempted to incorporate some of this complexity for modeling organometallic complexes,[43,44] and Goddard and co-workers have suggested that a central force formalism, with local perturbations based on bond angles and torsions, may provide a useful starting point for both reactive and metal-containing systems, and they use this approach in their ReaxFF reactive force field.[25,27,29]

The goal of the present work is to present a new type of analytic function for modeling reactive and metal-containing systems and to test it for nanoparticles composed of aluminum and hydrogen. In recent years, we have been interested in designing potentials that can be used to describe and/or predict processes that occur in aluminum nanoparticles.[30,31,45] We developed two reasonably accurate potentials,[31] called NP-A and NP-B, for pure aluminum; NP-B is less accurate than NP-A but is much less expensive to evaluate and has been widely applied.[46−51] However, neither the NP-A nor the NP-B functional form is well suited for straightforward extension to heteronuclear systems.

A potential capable of describing metal hydride materials in both nanoparticles and the bulk would be of general interest. Hydrogen is potentially an important component of a clean energy storage medium, yet its transport, handling, and utilization involve a number of safety and technical problems. One possible strategy is to use metal hydrides as a storage medium.[52−63] Therefore, a potential energy function capable of accurately describing such a system would be useful.

In the present article we present a model called the valence–bond order (VBO) model. It provides a physically motivated starting point for representing the potential energy surfaces of both homonuclear and heteronuclear metal-containing and reactive systems. In Supporting Information we also present an extension called VBO2 as an illustration of an attempt to obtain better quantitative performance with a functional form motivated by VBO but more general. We illustrate VBO and VBO2 for particles composed of aluminum and hydrogen and for bulk aluminum, but the approach used here is also applicable to other metal hydride systems and to other materials and nanoparticles, although for some such more general applications it may be necessary to extend the functional form, for example, to include explicit Coulomb interactions, attractive noncovalent interactions, more complicated bond order terms, explicit bond angle or screening (three-body) terms, and so forth. Before considering such extensions, it is useful to learn how well the simple VBO form without such enhancements can represent potential energy functions, and the present paper provides a first answer to that question.

Section 2 presents the new formalism, section 3 presents the parametrization, and section 4 gives the results and characterizes the accuracy achievable with the method. Section 5 gives further discussion, especially of the relationship to other approaches.

2. Theory

2.1. Background. The key physical idea underlying the present approach is that the quantum mechanical nature of bonding manifests itself in the concepts of valence, bonding orbitals, and steric repulsion, all of which have a quantum mechanical basis ultimately rooted in the Pauli principle. These concepts are utilized, for example, in the valence-shell electron-pair repulsion (VSEPR) model[64,65] and the molecular orbital aufbau principle.[65,66] The most straightforward way to incorporate the effects of bonding orbitals is to obtain the molecular orbitals by some variational or semiempirical method, which invariably involves diagonalizing a Hamiltonian matrix (such as a matrix representing a Fock or Kohn−Sham operator or some other effective one-electron Hamiltonian). Alternatively the quantum mechanical bonding effects can be included by valence bond theory,[1,65−71] which involves the diagonalization of a configuration interaction matrix. The difficulty in developing analytical potentials that do not require diagonalizations or an iterative process that is essentially equivalent to a diagonalization is in representing the saturation of bonding power (valence) of a given atom. Conventional molecular mechanics effectively accomplishes this by adopting a single-configuration valence bond formalism, where a different set of molecular mechanics parameters is associated with each bonding pattern, which in turn is associated with each valence bond configuration. This works well for nonreactive organic chemistry but requires a multiconfigurational extension[72,73] for reactions and is hard to extend to metals. Our approach presented here is different and is motivated instead by the second moment approximation to tight binding theory.

Tight-binding theory[9−11,13,74,75] may be justified[76−79] as an approximation to density functional theory, especially by using the noniterative formulation of Harris and Foulkes.[77−80] Tight-binding theory has been developed to simulate materials systems directly,[9−11,13,17,18,45,81−86] but it can also be further approximated to motivate analytic potentials. In particular, the closely related functional forms of the quasiatom theory,[87] the Gupta potential,[88] the embedded-atom model (EAM),[89] the modified EAM (MEAM),[90] the second-nearest-neighbor EAM (2NN-EAM),[91] the Finnis-Sinclair scheme,[92] the effective medium theory,[93] and the bond order potential[94] can all be motivated by the second moment approximation to tight binding theory.[95−105] These methods are closely related to each other, and they are also closely related to the Tersoff potential[106] and hence to the Brenner[23] and ReaxFF[25,27−29] potentials.[107] The relationship of EAM to the Pauling bond order has also been discussed.[108]

In the EAM and other closely related methods that we have mentioned, the energy is the sum of an attractive term that may be considered to approximate the cohesive band energy of tight-binding theory and a repulsive term representing mainly core−core repulsion. The cohesive term is proportional to the square root of the second moment of the density of states ($\sqrt{M}$), which may be approximated for each atom as the square root of a sum of two-body transfer integrals or the square root of the local coordination number. This idea is the basis for the functional form proposed here. In the basic version of the theory, to be called VBO (see below), unlike some earlier models, we retain only the simplest possible functional form, without added complications such as bond-angle terms and Coulombic terms. However, we generalize the square rooot, $\sqrt{M}$, to allow any fractional root, M^{ϕ}, where the fraction ϕ is positive and less than unity. This is motivated by a recognition that the value of one-half for ϕ is just one example of a more general principle of valence saturation, by which we mean that the bonding power of an atom for an additional ligand decreases as the number of ligands already bound increases. Other ways to motivate a value of ϕ other than one-half would be to consider the fourth root of the fourth moment (rather than the square root of the second moment) or to note that whereas the solution of a two-configuration bonding model, such as the London equation,[109] expresses valence saturation through a square root, a three-configuration model could involve a third root in the lowest eigenvalue.

In Supporting Information we present a second version, called VBO2, in which we include two cohesive terms for the homonuclear case and a third cohesive term for the heteronuclear case; we include explicit 3-body dependence for the heteronuclear case but still do not include explicit dependence on bond angles or coulomb forces. The VBO2 formulation involves two fractional powers rather than one; this may be justified either empirically (it sometimes gives more accurate results for a given number of parameters) or—more satisfactorily—by a two-band model.[110]

2.2. VBO. The new method presented here is called VBO. The first key element in the functional form is the bond order between atoms i and j, which is represented by a monotonically decreasing function of the distance r_{ij} between atoms i and j. A variety of such functions may be (and were) considered, but in the present work we present results only for the following cutoff decaying exponential function

$$b_{ij} = \begin{cases} N_{ij} \exp\left[\dfrac{-\gamma_{ij}}{1-(r_{ij}/\Delta_{ij})^{0.5}}\right] & \text{if } r_{ij} < \Delta_{ij} \\ 0 & \text{if } r_{ij} \geq \Delta_{ij} \end{cases} \tag{1}$$

where γ_{ij} and Δ_{ij} are parameters that depend on the atomic numbers of i and j. Notice that Δ_{ij} also plays the role of a cutoff value: as r_{ij} goes to Δ_{ij} from the left, b_{ij} goes to zero with an infinite number of continuous derivatives. N_{ij} is a normalization constant defined as

$$N_{ij} = \exp\left[\frac{\gamma_{ij}}{1-(R_{ij,0}/\Delta_{ij})^{0.5}}\right] \tag{2}$$

where $R_{ij,0}$ is the nominal bond distance of the ij dimer; the bond order of the dimer at this distance is defined to be unity. The factor N_{ij} and the constant $R_{ij,0}$ are included only for convenience of interpretation, and no energies depend on how the bond order is normalized because a different normalization would simply yield different optimized coefficients in expressions like eq 6.

The energy of the system is the sum of individual atomic contributions

$$E = \sum_i E_i \tag{3}$$

The form for E_i was inspired by the Morse potential[111] where the energy of a diatomic molecule is approximated as

$$E_{ij} = D_{ij}\{X_{ij}^2(r_{ij}) - 2X_{ij}(r_{ij})\} \tag{4}$$

where D_{ij} is a constant and

$$X_{ij} = \exp\left[-\alpha_{ij}(r_{ij} - R_{ij,0})\right] \tag{5}$$

and α_{ij} is another constant. Note that X_{ij} can be interpreted as the Pauling bond order.[112] In the VBO model, we replace

X_{ij} by b_{ij}, and we generalize eq 4 to a many-body system as follows

$$E_i = \sum_{j \neq i} (c_{i,1} b_{ij,1} + c_{i,2} b_{ij,2}) - V_i^{n_i} \tag{6}$$

where the valence V_i of atom i is defined by

$$V_i = \sum_{j \neq i} (c_{i,3} b_{ij,3} + c_{i,4} b_{ij,4}) \tag{7}$$

and b_{ij} has now become $b_{ij,p}$. Four different $b_{ij,p}$ functions with $p = 1, 2, 3, 4$, each with its own $\gamma_{ij,p}$ parameters, were used; $c_{i,p}$ (with $p = 1, 2, 3, 4$) and n_i are atomic parameters, whereas $\gamma_{ij,p}$ (with $p = 1, 2, 3, 4$) are parameters that depend on the atomic numbers of atoms i and j. All these parameters are positive by definition. Note that if $c_{i,2} = 0$, $c_{i,4} = 0$, and $n_i = 0.5$, then this functional has the same spirit as that used in the embedded atom[108] method but with a bond order function rather than an embedded density; furthermore, if one considers a diatomic molecule, uses eq 5 instead of eq 1 for the bond order, and sets $c_{i,2} = 0$, $c_{i,4} = 0$, $c_{i,3} = (2c_{i,1})^2$, and $n_i = 0.5$, one obtains a Morse[111] function. In eq 6, the last term is called the valence energy; this term is purely attractive, and the repulsive term is pairwise additive. The key physical constraint on the parameters of the VBO model is that n_i should be less than unity to account for valence saturation. But a key feature of the new method is that we will not constrain it to be exactly one-half, as in EAM.

The extension of the VBO model to a heteronuclear system such as a mixed aluminum−hydrogen system is more complicated than the way Hall combined two Morse potentials in order to map the potential energy surface for a collinear reaction.[113] Equation 4 is a quadratic function, and the simplest multinomial extension of such functions is to include cross-terms[113]

$$E_{\mathrm{ABC}} = D_{\mathrm{AB}}(X_{\mathrm{AB}}^2 - 2X_{\mathrm{AB}}) + D_{BC}(X_{\mathrm{BC}}^2 - 2X_{\mathrm{BC}}) + 2HX_{\mathrm{AB}}X_{\mathrm{BC}} \tag{8}$$

where H is a fitting parameter. But this cannot be applied to VBO because it does not show valence saturation in the same way. Therefore, we generalize VBO by replacing eq 7 by a more general valence expression, in particular

$$V_i = \sum_{j \neq i} (c_{i,3} \sigma_{ij,3} b_{ij,3} + c_{i,4} \sigma_{ij,4} b_{ij,4}) \tag{9}$$

In general the parameters depend on the atomic numbers of atoms i and j. When i and j have the same atomic numbers, we set $\sigma_{ij,3} = \sigma_{ij,4} = 1$ so that eq 9 reduces to eq 7 for homonuclear systems. In heteronuclear systems, $\sigma_{ij,p} \neq \sigma_{ji,p}$ if atoms i and j have different atomic numbers. The new parameters $\sigma_{ij,p}$ with $i \neq j$ allow the contribution of an Al atom to the valence of H to be different from the contribution of an H atom to the valence of Al. Similarly, $\gamma_{ij,p} \neq \gamma_{ji,p}$ for $p = 3$ or 4. However, the parameters $\gamma_{ij,p}$ for $p = 1$ or 2 are assumed invariant to permuting the order of the indices because these terms are pairwise additive. Equations 3, 6, and 9 define the VBO model.

2.3. Zero of Energy. The zero of energy is taken as the sum of the energies of the individual atoms at infinite separation. Thus all energies are interaction energies.

3. Determination of the Parameters

3.1. The Data Set. Three data sets, consisting of energies for specific geometries, were used: one each for pure aluminum, pure hydrogen, and mixed aluminum−hydrogen clusters.

For pure aluminum (Al_n clusters and nanoparticles), the data set, consisting of 808 structures calculated at either the PBE0[114]/6-311+G(3d2f)[115,116] (for $n \leq 13$) or PBE0[114]/MEC[117] (for $n > 13$) level, was taken from previous work.[31] The PBE0 functional was chosen based on comparison to high-level correlated wave function calculations on small clusters.[118] Data for bulk face-centered cubic (fcc), hexagonal close-packed (hcp), and body-centered cubic (bcc) quasi-spherical clusters data sets were also included, and they also were taken from previous work.[31]

For pure hydrogen, part of the data set was also taken from the literature: 80 H_2 points from the accurate calculations by Kołos and Wolniewicz,[119] 602 H_3 points (307 C_{2v}, 22 $D_{\infty h}$, 21 D_{3h}, and 252 $C_{\infty v}$) determined using the BKMP2 H_3 potential energy surface of Boothroyd et al.,[120] and 586 H_4 points (279 for parallel approach of two H_2 molecule, 307 for perpendicular approach of two H_2 molecules) determined using the BMKP H_4 potential energy surface of Boothroyd et al.[121]

In addition to these data for H_2, H_3, and H_4, we generated new data for H_6 and H_8 by performing high-level ab initio calculations at selected geometries. Because H atoms do not form extended systems, except for condensed-phase H_2, in contrast to aluminum atoms, which can form nanoparticle and bulk aluminum phases, pure hydrogen clusters with more than 8 atoms are not considered here.

Two types of H_6 structures were included in the data set. To construct the first type, the saddle point of the $H + H_2$ reaction was optimized at the UHF[122]/cc-pVTZ[123] level of theory, and two identical collinear H_3 fragments in this geometry were used to form an H_6 fragment, with the second H_3 fragment pointing toward the central atom of the first H_3 fragment. The six H_6 geometries included in the data set have closest distances between the two H_3 fragments varying between 0.7 and 1.2 Å. The second type of H_6 structure was constructed starting with a configuration analogous to that of the CH_5^+ cation, which on optimization leads to a structure of the type $H_4 + H_2$. In the six structures included in the data set, the distance between the H_4 and H_2 fragments was varied between 0.82 and 1.23 Å.

The H_8 data set is also composed of two types of structures. The first type has two square H_4 fragments situated side-by-side, where the H−H distances in the H_4 fragments were previously optimized. Then, six points were generated by varying the distance between the H_4 fragments from 0.46 and 0.80 Å. The second type of H_8 structure derives from an ethane-like configuration for the hydrogen atoms. The optimized structure has a long H−H central distance with each of the two H atoms involved being attached to three H atoms. Six points were added to the data set with the central distance ranging from 1.25 to 1.97 Å.

The electronic energies of the H_6 and H_8 structures were obtained from extrapolated multireference configuration interaction[124] (MRCI) calculations with a quadruples correction[125] (+Q). First, full-CI (FCI) calculations were performed for H_6 and H_8 with a double-ζ basis set with two goals: (1) to determine whether the singlet, triplet, or quintet is the lowest-energy spin state for the given structure, and (2) to establish a benchmark on which to base the choice for the reference space in the MRCI calculations. In this way, it was determined that the H_6 and H_8 structures have triplet and singlet ground states, respectively. To determine the complete active space self-consistent field[126,127] (CASSCF) reference state in the MRCI calculations, the H_3 and H_4 energies were calculated at the FCI level. Complete active spaces of 3 electrons in 6 orbitals, denoted (3,6), and 4 electrons in 6 orbitals, denoted (4,6), were chosen as the MRCI reference spaces for H_3 and H_4, respectively, and provided a maximum error of about 0.1 kcal/mol with respect to the FCI results. Thus, (6,12) and (8,12) reference spaces were chosen for the H_6 and H_8 calculations.

The final electronic energies for the H_6 and H_8 structures were obtained using the following form[128] of basis-set extrapolation

$$E_{\infty}^{\text{tot}} = \frac{3^{\alpha}}{3^{\alpha} - 2^{\alpha}} E_3^{\text{CASSCF}} - \frac{2^{\alpha}}{3^{\alpha} - 2^{\alpha}} E_2^{\text{CASSCF}} + \frac{3^{\beta}}{3^{\beta} - 2^{\beta}} E_3^{(\text{MRCI+Q})-\text{CASSCF}} - \frac{2^{\beta}}{3^{\beta} - 2^{\beta}} E_2^{(\text{MRCI+Q})-\text{CASSCF}} \quad (10)$$

where E_2 and E_3 are energies obtained using the cc-pVDZ and cc-pVTZ basis sets, respectively, and MRCI+Q denotes multireference CI with singles and doubles and Pople's correction for quadruples; the parameters α and β were taken to have the values[128] 3.4 and 2.4, respectively.

The pure hydrogen data set for H_2, H_3, and H_4 consists of 1268 data, and adding the H_6 and H_8 data increases this to 1292 data; the resulting data set is called the H1292 database. The data set of mixed aluminum−hydrogen particles consists of 906 Al_nH_m geometries ($n = 1-13$, $m = 1-12$) and energies calculated at the PBE0/6-311++G(3d2f,2p) level of theory and is called the AlH906 database.

We also considered bulk data for pure aluminum. The accurate cohesive energies for fcc, bcc, and hcp lattices with given lattice constants were estimated in ref 31, and those data are also used here. To calculate bulk cohesive energies for a given lattice constant from the model potentials, we used the procedure of ref 31 where the bulk cohesive energy is defined as the energy per atom required to atomize a particle, extrapolated to an infinitely large particle from finite-size fcc, bcc, and hcp quasispherical particles with given nearest-neighbor distances corresponding to the lattice constants for which accurate data are available.

All geometries and energies used in fitting are provided in Supporting Information.

3.2. The Optimization of Parameters. A microgenetic algorithm[129] was used to fit the parameters. In particular, we used the FORTRAN version 1.7a of Carroll's code,[130] locally modified with our own fitness function and designed to run in parallel using the message-passing interface (MPI).[131,132]

The fitness function used here is based on the fitness functions used in previous work[31,45] to parametrize analytical functions and tight-binding models. In each case, the data sets are divided into subgroups (with n_k particles in subgroup k) according to the size of the particles, i.e., the number of atoms N_i in each particle i and their stoichiometry. Thus, for example, although Al_4H and AlH_4 clusters both have $N_i = 5$; they are in different subgroups, but, as previously,[31] some subgroups for pure aluminum contain more than one N_i. Note that we use "particle" as a generic name for either a "cluster" (which is arbitrarily defined to have less than 20 Al atoms) or a "nanoparticle" (which is correspondingly defined to have 20 or more Al atoms).

Table 1. VBO Parameters for Pure Al and Pure H

parameter (units)	Al	H
$\gamma_{ij,1}$	0.3599	2.7292
$\gamma_{ij,2}$	4.2538	8.8100
$\gamma_{ij,3}$	0.5684	16.1462
$\gamma_{ij,4}$	2.9562	16.1330
$c_{i,1}$ (eV)	0.2248	2.0112
$c_{i,2}$ (eV)	0.2172	0.8084
$c_{i,3}$ (E_h^{1/n_i})^a	0.0133	0.00001
$c_{i,4}$ (E_h^{1/n_i})^a	0.0068	0.00001
n_i	0.7931	0.1536
$R_{ij,0}$ (Å)	2.7306^b	0.7414^b

[a] E_h denotes one hartree, that is, one atomic unit of energy. [b] As explained in the text, this is not a fitting parameter; it simply gives the bond order a convenient ("physical") normalization.

Table 2. VBO Parameters for Al−H Interactions

parameter	Al−H	parameter	Al−H
$\gamma_{HAl,1}$	7.4891	$\sigma_{HAl,4}$	0.0191
$\gamma_{HAl,2}$	1.8993	$\sigma_{AlH,3}$	2.9306
$\gamma_{HAl,3}$	13.5669	$\sigma_{AlH,4}$	0.00008
$\gamma_{HAl,4}$	13.6902	$c_{HAl,1}$ (eV)	0.3509
$\gamma_{AlH,3}$	1.6911	$c_{HAl,2}$ (eV)	0.5174
$\gamma_{AlH,4}$	2.8674	$R_{ij,0}{}^a$ (Å)	1.6637
$\sigma_{HAl,3}$	0.00001		

[a] As explained in the text, this is not a fitting parameter; it simply gives the bond order a physical normalzation.

Table 3. Mean Unsigned Errors (in eV/atom) for Aluminum

method (model)	$\varepsilon_{\text{bulk}}$	ε_{CE}	$\varepsilon_{\text{part}}$	$\varepsilon_{\text{cluster}}{}^a$	$\varepsilon_{\text{nano}}{}^b$
VBO	0.02	0.002	0.05	0.06	0.03
NP-A^c	0.02	0.002	0.05	0.06	0.03
NP-B^c	0.03	0.002	0.06	0.08	0.04

[a] MUE for cluster with 2−19 atoms. [b] MUE for nanoparticles with 20−177 atoms. [c] Reference 31.

Table 4. Bulk Lattice Constants (LC, in Ångstroms) and Bulk Cohesive Energies (E_c in eV/atom) for Aluminum

	fcc		bcc		hcp	
method (model)	LC	E_c	LC	E_c	LC	E_c
accurate^a	4.02	3.43	3.24	3.33	2.87	3.39
VBO	4.02	3.44	3.22	3.25	2.85	3.49
NP-A^a	4.01	3.43	3.22	3.34	2.84	3.42
NP-B^a	4.03	3.43	3.27	3.35	2.86	3.41

[a] Reference 31.

For pure Al, there are 11 groups with $N_i = 2$ in group $k = 1$, $N_i = 3$ in group $k = 2$, up to $N_i = 89-177$ in group $k = 11$.[31] The sum of the n_k for pure Al is 808. For hydrogen particles, there are five subgroups, each with a single value (2, 3, 4, 6, or 8) of N_i. For heteronuclear particles each subgroup k has a unique combination of N_i and stoichiometry. Thus for pure hydrogen and heteronuclear systems, we introduce the notation N_k for the N_i value of subgroup k.

We consider errors on a per-atom basis so that errors in the larger clusters do not dominate our considerations simply because of the size of the larger clusters. (The per-atom convention is standard in discussing cohesive energy of bulk materials.) We define a mean unsigned error (MUE) per atom for subgroup k as

$$\varepsilon_k = \frac{1}{2}\left(\frac{\sum_{i=1}^{n_k} w_i \Delta E_i^k}{\sum_{i=1}^{n_k} w_i} + \frac{\sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} w_i w_j \delta_{N_iN_j} \Delta\Delta E_{ij}^k}{\sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} w_i w_j \delta_{N_iN_j}}\right) \quad (11)$$

where w_i is the weight of particle i, and $\delta_{N_iN_j}$ is a Kronecker delta. For pure aluminum, the form of w_i is the same as eq 1 of ref 31. For pure hydrogen and mixed aluminum−hydrogen particles, $w_i = 1$. In eq 11, the quantity ΔE_i^k is the difference between the accurate (acc) and the VBO energies (each relative to separated atoms) for structure i of subgroup k

$$\Delta E_i^k = (E_i^{k,\mathrm{acc}} - E_i^{k,\mathrm{VBO}})/N_i \quad (12)$$

and $\Delta\Delta E_{ij}^k$ is defined by

$$\Delta\Delta E_{ij}^k = (\Delta E_i^k - \Delta E_j^k)/N_i \quad (13)$$

Thus ΔE_i^k is a measure of the error in absolute interatomic interaction energy on a per atom basis for a given geometry, whereas $\Delta\Delta E_{ij}^k$ is a measure, again on a per atom basis, of the accuracy of relative binding energies, that is, of the shape of the potential energy surface. The importance of including $\Delta\Delta E_{ij}^k$ was demonstrated previously.[13,30]

For pure Al we computed a mean unsigned error for all particles by

$$\varepsilon_{\mathrm{part}} = \left(\sum_{k=1}^{N_\mathrm{s}} \varepsilon_k\right)/N_\mathrm{s} \quad (14)$$

where N_s is the number of subgroups (11 for pure Al). For pure Al, we also compute $\varepsilon_{\mathrm{cluster}}$ in which eq 14 is applied only to the six subgroups with $N_i \leq 19$ and $\varepsilon_{\mathrm{nano}}$ in which eq 14 is applied only to the five subgroups with $N_i \geq 20$. For hydrogen we used eq 14 with $N_\mathrm{s} = 3$ (see below). For mixed aluminum−hydrogen particles, we used

$$\varepsilon_{\mathrm{part}} = \frac{\sum_k n_k^{\alpha} N_k^{\beta} \varepsilon_{N_k}}{\sum_k n_k^{\alpha} N_k^{\beta}} \quad (15)$$

with $\alpha = 0.4$ and $\beta = 0.6$.

Since a genetic algorithm is used to maximize a function, but our goal is to minimize the total error ε, the fitness function f is defined as

$$f = -\varepsilon \quad (16)$$

where ε is the error to be minimized. For hydrogen and heteronuclear hydrogen−aluminum particles, we set ε equal to $\varepsilon_{\mathrm{part}}$. For pure aluminum, we used[31]

$$\varepsilon = \frac{1}{4}(\varepsilon_{\mathrm{bulk}} + \varepsilon_{\mathrm{CE}}) + \frac{1}{2}\varepsilon_{\mathrm{part}} \quad (17)$$

where $\varepsilon_{\mathrm{bulk}}$ and $\varepsilon_{\mathrm{CE}}$ are, respectively, the errors in the bulk energies and in the FCC cohesive energies. In eq 17, $\varepsilon_{\mathrm{bulk}}$ is the average unsigned error over 12 energetic data for fcc, bcc, and hcp as in ref 31 (we considered four lattice constants for each crystal habit), whereas $\varepsilon_{\mathrm{CE}}$ is for the subset of four data four fcc.

3.3. Parameterization. In eq 1, the parameter Δ_{ij} acts as a cutoff distance for the bond order. In previous work[31] on pure Al, we used a similar cutoff, with values of 6.50 Å for our best potential, called NP-A, and 5.38 Å for a second potential, called NP-B, that was slightly less accurate but much more economical. The calculation times for large particles increase with increasing Δ_{ij} and the onset of linear scaling is pushed up to larger nanoparticles. Therefore, in all further work, Δ_{ij} was frozen somewhat arbitrarily at 6.88 Å for Al−Al and Al−H interactions and at 5.29 Å for H−H interactions.

A more commonly used approximation for the bond order is Pauling's approximation,[112] which is a simple exponential function as in eq 5. We found that replacing b_{ij} with a simple exponential did not significantly improve the results, and so we retained the cutoff bond order of eq 1 for the present parametrization.

There are nine free parameters in the VBO model for pure aluminum and nine for pure hydrogen. We first optimized the parameters for pure Al and pure H. VBO did not yield satisfactory results for H_6 and H_8, so we excluded them from the final optimization. The final values of the VBO parameters and $R_{ij,0}$ for Al−Al and H−H interactions are listed in Table 1. The optimized values of n_i are less than one; this physically corresponds to valence saturation as the bond order increases.

When optimizing the heteronuclear parameters, we froze the Al−Al and H−H parameters at the values already found for the pure systems. The number of heteronuclear parameters for Al−H systems is 12 in VBO. In the initial stages of optimization all parameters were optimized, and the importance of various parameters was examined. No terms were dropped, and the final heteronuclear parameters are given in Table 2.

4. Results

For pure aluminum, the mean unsigned errors for VBO are given in Table 3. Table 3 shows that for aluminum particles VBO performs as well as the much more expensive NP-A, which contains two three-body terms, and better than the less expensive embedded-atom-based NP-B, which includes

Table 5. MUEs (eV/atom) for Pure Hydrogen[a]

K	1	2	3	$\varepsilon_{\mathrm{part}}$
formula	H_2	H_3	H_4	
N_k	2	3	4	
n_k	80	602	586	
VBO	0.02	0.06	0.14	0.07

[a] In Tables 5 and 6, the last row gives the MUE per atom for the n_k data subgroup k, where subgroup k contains particles with N_k atoms, except in the last column where the last row gives ϵ_{part}.

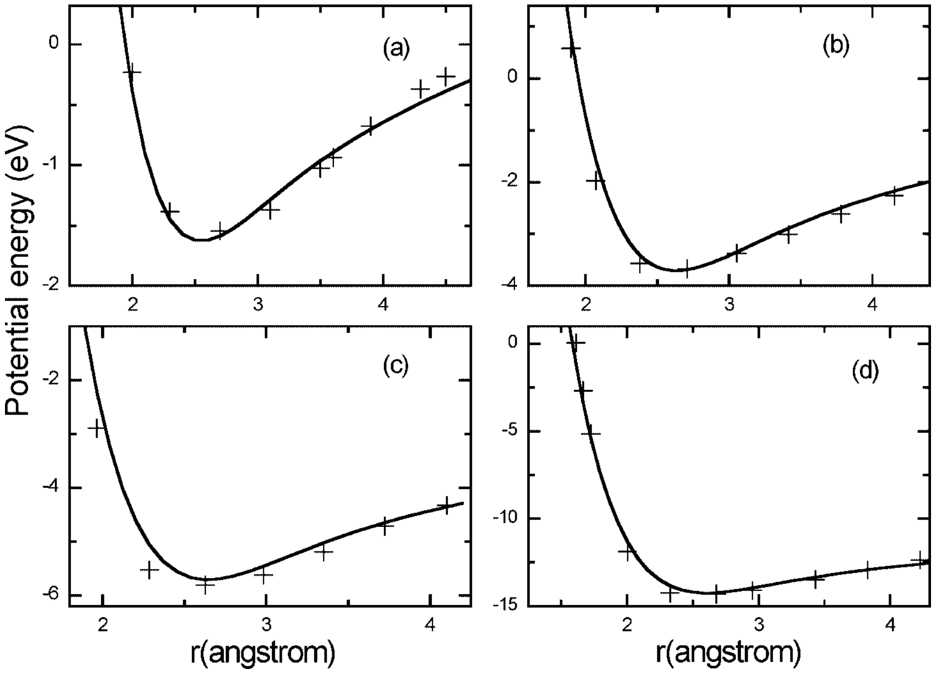


Figure 1. Potential energies of aluminum clusters as calculated by the VBO method (solid curves) as compared to PBE0/6-311+G(3d2f) reference data (crosses): (a) Al_2 dimer, (b) Al_3 for the sideways approach of Al to the dimer with a dimer bond length of 2.8635 Å, (c) Al_4 for the planar, short-bridge approach of Al to Al_3 (for this plot, Al_3 is an acute isosceles triangle with the short bond being the base of length 2.54724 Å and the height being 2.20598 Å, and the fourth Al approaches the short bond along the C_2 axis), and (d) Al_7 for the approach of Al to a 3-fold face of a regular octahedron with bond length 2.6634 Å. In all plots, r is the distance between the approaching Al atom and the closest Al atom in the approached subcluster.

Table 6. MUEs (eV/atom) for Heteronuclear Particles Composed of Aluminum and Hydrogen

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	ε_{part}
formula	AlH	AlH_2	AlH_3	AlH_4	AlH_5	AlH_6	AlH_{12}	Al_2H	Al_2H_2	Al_3H	Al_4H	$Al_{13}H$	$Al_{13}H_2$	$Al_{13}H_8$	$Al_{13}H_{12}$	
N_k	2	3	4	5	6	7	13	3	4	4	5	14	15	21	25	
n_k	37	108	54	3	3	3	3	304	3	74	180	125	3	3	3	
VBO	0.05	0.14	0.33	0.05	0.10	0.30	0.35	0.06	0.16	0.08	0.11	0.03	0.05	0.06	0.08	0.11

three-body effects through the use of an embedding function rather than by explicit three-body terms. Figure 1 shows four cuts through the potential surfaces for various sizes of clusters (these plots are similar to plots in previous papers[13,118] on Al clusters; the geometries used for these cuts are specified qualitatively in words in the figure captions and precisely by tables of geometries and energies in Supporting Information). The figure shows that VBO not only gives small mean unsigned errors, it also reproduces the shapes of the potential energy curves.

Table 4 shows that VBO also reproduces the bulk aluminum cohesive energy and lattice constants semiquantitatively.

Mean unsigned errors for pure hydrogen are given in Table 5. The average error for hydrogen is 0.07 eV/atom for VBO, which is the same order of magnitude as for aluminum, but a little larger.

MUEs for heteronuclear aluminum−hydrogen particles are given in Table 6. The number n_k of particles in each subgroup is also included in the table. The maximum MUE per atom is 0.35 eV, which is the MUE for the AlH_{12} cluster. The minimum MUE per atom is 0.03 eV/atom for the $Al_{13}H$ cluster. Most of the MUEs are in the range of 0.03−0.16 eV/atom. The overall error is 0.11 eV/atom, which is about twice as large as $\epsilon_{cluster}$ for pure Al. Figure 2 shows four cuts through the potential energy surfaces for mixed aluminum−hydrogen clusters (the geometries used in these cuts are specified qualitatively in words in the figure captions and precisely by tables of geometries and energies in Supporting Information); as for pure aluminum we see that the VBO model represents the qualitative dependence of the energy on structure quite well. The local maximum in Figure 2d corresponds to the passage of H through a triangular face between the exterior of the cluster and an interstitial site; VBO does quite well for reproducing this potential energy curve.

5. Discussion

Methods for calculating potential energy functions may be divided into those in which electrons are treated explicitly, for example, coupled cluster theory,[2,3] density functional theory,[4,5] or semiempirical molecular orbital theory[8−13] (including tight binding theory[9−11,13]), or those in which electrons are only implicit; a generic name for the latter is molecular mechanics.[33−42,133] As mentioned in the introduction, the molecular mechanics method usually yields analytic potential functions with general parameters designed to be transferable between systems (as opposed, for example, to potentials designed for a specific system, such as the water dimer[21] or the reaction of H with CH_4[22]). The set of general parameters is sometimes called

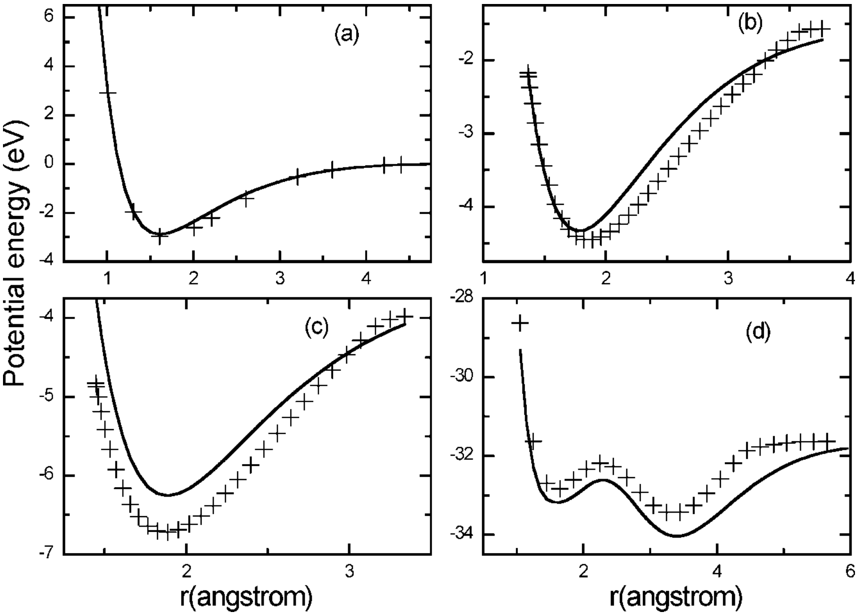


Figure 2. Potential energies of aluminum−hydrogen mixed clusters calculated by the VBO method (solid curves) as compared to PBE0/6-311++G (3d2f,2p) reference data (crosses): (a) AlH dimer, (b) Al_2H for the sideways approach of H to the Al_2 dimer with an Al−Al bond distance of 2.73 Å, (c) Al_3H for the 3-fold approach of H to an Al_3 triangle (for this plot, Al_3 is an equilateral triangle with the a bond length equal to 2.50696 Å, and the H approaches its center along the C_3 axis), and (d) approach of H to Al_{13} cluster (FCC structure) toward a 3-fold site but with C_1 symmetry. In all plots, *r* is the distance between the approaching H atom and an Al atom in the approached subcluster.

a force field,[37−42] although literally that term should refer to the gradient field of the potential energy function for a specific system.

As mentioned in the introduction, many molecular mechanics force fields (conventional molecular mechanics) are designed for treating only nonreactive systems or reactive systems for geometries in the neighborhoods of equilibrium geometries,[33−42,133] although there are some notable exceptions such as the Tersoff potential for systems composed of Si and C or Ge,[106] the Brenner potentials for hydrocarbons,[23] the Streitz−Mintmire potential for systems composed of Al and O,[24] and the ReaxFF force fields for a variety of materials, including hydrocarbons[25] and materials composed of Si and O,[27] Al and O,[28] or Mg and H.[29] These exceptions maybe called reactive molecular mechanics. VBO is designed as a form of reactive molecular mechanics, and in this article we presented a general parametrization for systems composed of Al and H.

A key difference between the reactive molecular mechanics force fields[23−29,107] and the conventional nonreactive ones[33−42,133] is the absence of typing in reactive force fields. In conventional molecular mechanics one specifies the type and bonding arrangement for each atom, e.g., an O might be typed as sp^3, and the user of the force field specifies the two atoms to which it is bonded. In reactive force fields, an atom interacts with all other atoms, except possibly for a smooth cutoff based solely on distance, and the parameters are the same for every Al, or every O, or every H independent of any perceived bonding pattern in a given geometry. Furthermore, the bonding interactions yield physically correct results when bonds break, as in a Morse potential,[111] as contrasted to harmonic or other polynomial interactions often used in conventional molecular mechanics.

For ionic or charged systems or systems whose bonds have large partial ionic character, one should include explicit Coulomb interactions in the functional form.[2,4,25,27−29,34,36−42] Al and H have different electronegativity values, but the difference corresponds to only about 10% partial ionic character[112a] (as compared to about 20% for Mg−H and 30% for Al−O). Therefore, at least for now, we omitted explicit Coulombic terms. Another key issue is whether to include explicit angle-dependent terms.[23,25,27,29,33,4,106,134] Similarly to deciding not to include Coulombic terms, at least for the present, we attempted here to see how well the VBO form can work without angle-dependent terms.

These choices led us to a very simple functional form, much simpler than any of those mentioned above, and the present results are encouraging in that this simple functional form already provides a qualitatively correct description of aluminum-hydrogen clusters with a wide range of compositions (wide range of Al-to-H ratios in the clusters). This kind of analytic potential energy function is therefore well suited to the largest simulations, such as long-time simulations of the dynamics of large nanoparticles, ceramics, or polycrystalline materials.[135,136] A second possible use for low-cost potential energy functions with general parameters is as a starting point for a more accurate potential energy function for a specific system or reaction; for example, sometimes one starts with a general parametrization and then introduces specific reaction parameters.[137−140] For a more specific example, one might reoptimize the general parameters for diffusion of a single hydrogen atom in bulk Al or to treat a system of two hydrogen atoms interacting with Al_{13} or with a specific crystal face of solid Al. By focusing on a less diverse set of structures, one can obtain a much more accurate potential energy surface.

One goal of the present work has been to test the accuracy attainable within a VBO scheme without explicit dependence on bond angles. The primitive VBO scheme employs one valence per atom and involves only radial two-body forces. The VBO2 scheme presented in Supporting Information involves two or three valences per atom and includes two-body homonuclear and heteronuclear radial forces and three-body heteronuclear forces; the latter implicitly bring in bond angle dependences. Directions for further study include three-body homonuclear forces, four-body forces, explicit dependence on bond angles or dihedral angles, and explicit accounting for Coulomb and dispersion interactions.

6. Conclusion

The VBO method has been proposed and developed for modeling a metal hydride system. The VBO model should be applicable to many other pure and binary materials as well, and there is no reason why it cannot be extended to ternary or other multicomponent systems. Here the model was used to develop a force field for nanoparticles composed of aluminum and hydrogen. The overall error is just 0.11 eV/atom on average for heteronuclear systems and 0.05−0.07 eV/atom for homonuclear systems. In Supporting Information we illustrate how the VBO functional form can be used as a starting point for functional forms with more parameters and flexibility. In particular, we present an extension called VBO2 and show that it is more accurate for pure hydrogen clusters and slightly more accurate for mixed aluminum−hydrogen clusters.

The computational expense of VBO scales linearly with system size, and diagonalization or iterations are not required; the method is therefore particularly useful for simulating large systems.

Acknowledgment. The authors are grateful to Ahren Jasper for helpful assistance and to Steven Mielke, Steve Valone, Akin Budi, Julian Gale, David Henry, and Irene Yarovsky for helpful comments. This work was supported in part by the National Science Foundation under Grant Nos. ITR-0428774 and CHE-0704974.

Supporting Information Available: Equations and results for the VBO2 model; geometries and energies (relative to separated atoms) for all data in the training set. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Slater, J. C. *Quantum Theory of Matter*, 2nd ed.; McGraw-Hill: New York, NY, 1968.

(2) Raghavachari, K.; Anderson, J. B. *J. Phys. Chem.* **1996**, *100*, 12960.

(3) Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.

(4) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974.

(5) Kohn, W. *Rev. Mod. Phys.* **1999**, *71*, 1253.

(6) Truhlar, D. G. In *Foundations of Molecular Modeling and Simulation*; Cummings, P. T., Westmoreland, P. R., Eds.; American Institute of Chemical Engineers: New York, NY, 2001; pp 71−83.

(7) Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6648.

(8) Zerner, M. C. *Rev. Comput. Chem.* **1991**, *2*, 313.

(9) Wang, C. Z.; Ho, K. M. *J. Comput.-Aided Mater. Des.* **1996**, *3*, 139.

(10) *Tight-Binding Approach to Computational Materials Science*; Turchi, P. E. A.; Gonis, A.; Colombo, L., Eds.; Materials Research Society: Warrendale, PA, 1998.

(11) Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; Kohler, C.; Amkreutz, M.; Sternberg, M.; Hajnal, Z.; Di Carlo, A.; Suhai, S. *J. Phys.: Condens. Matter* **2002**, *14*, 3015.

(12) Bredow, T.; Jug, K. *Theor. Chem. Acc.* **2005**, *113*, 1.

(13) (a) Staszewska, G.; Staszewski, P.; Schultz, N. E.; Truhlar, D. G. *Phys. Rev. B* **2005**, *71*, 045423. (b) Iron, M. A.; Heyden, A.; Staszewska, G.; Truhlar, D. G. *J. Chem. Theory Comp.* **2008**, *4*, 804.

(14) *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*; Grotendorst, J., Marx, D., Muramatsu, A., Eds.; Neumann Institute for Computing: Julich, 2002.

(15) *Multiscale Computational Methods in Chemistry and Physics*; Brandt, A.; Bernholc, J.; Binder, K., Eds.; IOS Press: Amsterdam, 2001.

(16) Hill, J.-R.; Subramanian, L.; Maiti, A. *Molecular Modeling Techniques in Material Sciences*; Taylor and Francis Group: Boca Raton, FL, 2003.

(17) Qiu, S. Y.; Wang, C. Z.; Ho, K. M.; Chan, C. T. *J. Phys.: Condens. Matter* **1994**, *6*, 9153.

(18) Goedecker, S.; Colombo, L. *Phys. Rev. Lett.* **1994**, *73*, 122.

(19) Nam, K.; Gao, J. L.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2.

(20) Truhlar, D. G.; Steckler, R.; Gordon, M. S. *Chem. Rev.* **1987**, *87*, 217.

(21) Huang, X. C.; Braams, B. J.; Bowman, J. M. *J. Phys. Chem. A* **2006**, *110*, 445.

(22) Albu, T. V.; Espinosa-Garcia, J.; Truhlar, D. G. *Chem. Rev.* **2007**, *107*, 5101.

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

(24) Streitz, F. H.; Mintmire, J. W. *Phys. Rev. B* **1994**, *50*, 11996.

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

(26) (a) Bazant, M. Z.; Kaxiras, E. *Phys. Rev. Lett.* **1996**, *77*, 4370. (b) Bazant, M. Z.; Kaxiras, E.; Justo, J. F. *Phys. Rev. B* **1997**, *56*, 8542. (c) Justo, J. F.; Bazant, M. Z.; Kaxiras, E.; Bulatov, V. V.; Yip, S. *Phys. Rev. B* **1998**, *58*, 2539. (d) Stuart, S. J.; Tuten, A. B.; Harrison, J. A. *J. Chem. Phys.* **2000**, *112*, 6472. (e) Marks, N. A. *Phys. Rev. B* **2001**, *63*, 035401. (f) Brenner, D. W.; Shenderova, O. A.; Harrison, J. A.; Stuart, S. J.; Ni, B.; Sinnott, S. B. *J. Phys.: Condens. Matter* **2002**, *14*, 7–83. (g) Brenner, D. W.; Shenderova, O. A.; Areshkin, D. A.; Schall, J. D.; Frankland, S. J. V. *Comput. Mod. Eng. Sci.* **2002**, *3*, 643. (h) Ni, B.; Lee, K.-H.; Sinnott, S. B. *J. Phys.: Condens. Matter* **2004**, *I*, 7261. (i) Shi, Y.; Brenner, D. W. *J. Chem. Phys.* **2007**, *127*, 134503.

(27) van Duin, A. C. T.; Strachan, A.; Stewman, S.; Zhang, Q. S.; Xu, X.; Goddard, W. A. *J. Phys. Chem. A* **2003**, *107*, 3803.

(28) Zhang, Q.; Cagin, T.; van Duin, A.; Goddard, W. A.; Qi, Y.; Hector, L. G. *Phys. Rev. B* **2004**, *69*, 045423.

(29) Cheung, S.; Deng, W. Q.; van Duin, A. C. T.; Goddard, W. A. *J. Phys. Chem. A* **2005**, *109*, 851.

(30) Jasper, A. W.; Staszewski, P.; Staszewska, G.; Schultz, N. E.; Truhlar, D. G. *J. Phys. Chem. B* **2004**, *108*, 8996.

(31) Jasper, A. W.; Schultz, N. E.; Truhlar, D. G. *J. Phys. Chem. B* **2005**, *109*, 3915.

(32) Sullivan, M. B.; Iron, M. A.; Redfern, P. C.; Martin, J. M. L.; Curtiss, L. A.; Radom, L. *J. Phys. Chem. A* **2003**, *107*, 5617.

(33) Bowen, J. P.; Allinger, N. L. *Rev. Comput. Chem.* **1991**, *2*, 81.

(34) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(35) Allinger, N. L.; Chen, K.; Lii, J.-H. *J. Comput. Chem.* **1996**, *17*, 642.

(36) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.

(37) MacKerrell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; 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, D.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(38) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730.

(39) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205.

(40) Ponder, J. W.; Case, D. A. *Adv. Prot. Chem.* **2003**, *66*, 27.

(41) MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584.

(42) Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665.

(43) Landis, C. R.; Root, D. M.; Cleveland, T. *Rev. Comput. Chem.* **1995**, *6*, 73.

(44) Deeth, R. J. *Coord. Chem. Rev.* **2001**, *212*, 11.

(45) Jasper, A. W.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 210.

(46) Bhatt, D.; Jasper, A. W.; Schultz, N. E.; Siepmann, J. I.; Truhlar, D. G. *J. Am. Chem. Soc.* **2006**, *128*, 4224.

(47) Bhatt, D.; Schultz, N. E.; Jasper, A. W.; Siepmann, J. I.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 26135.

(48) Li, Z. H.; Jasper, A. W.; Truhlar, D. G. *J. Am. Chem. Soc.* **2007**, *129*, 14899.

(49) Li, Z. H.; Bhatt, D.; Schultz, N. E.; Siepmann, J. I.; Truhlar, D. G. *J. Phys. Chem. C* **2007**, *111*, 16227.

(50) Li, Z. H.; Truhlar, D. G. *J. Phys. Chem. C* **2008**, *112*, 11109.

(51) Li, Z. H.; Truhlar, D. G. *J. Am. Chem. Soc.* **2008**, *130*, 12698.

(52) Züttel, A. *Naturwissenschaften* **2004**, *91*, 157.

(53) Pranevicius, L.; Milcius, D.; Pranevicius, L. L.; Thomas, G. *J. Alloys Compd.* **2004**, *373*, 9.

(54) Hayden, L. E.; Tverberg, J. C. *Am. Soc. Mech. Eng. PVP Div.* **2004**, *475*, 223.

(55) Sandrock, G.; Reilly, J.; Graetz, J.; Zhou, W. M.; Johnson, J.; Wegrzyn, J. *Appl. Phys., A* **2005**, *80*, 687.

(56) Yarovsky, I.; Goldberg, A. *Mol. Sim.* **2005**, *31*, 475.

(57) Milcius, D.; Pranevicius, L. L.; Templier, C. *J. Alloys Compd.* **2005**, *398*, 203.

(58) Fichtner, M. *Ann. Chim. Sci. Mater.* **2005**, *30*, 483.

(59) Wagemans, R. W. P.; van Lenthe, J. H.; de Jongh, P. E.; van Dillen, A. J.; de Jong, K. P. *J. Am. Chem. Soc.* **2005**, *127*, 16675.

(60) Graetz, J.; Reilly, J. J. *J. Alloys Compd.* **2006**, *424*, 262.

(61) Graetz, J.; Reilly, J. J. *Scipta Mater.* **2007**, *56*, 835.

(62) Wang, Y.; Yan, J.-A.; Chou, M. Y. *Phys. Rev. B* **2008**, *77*, 014101.

(63) Scheicher, R. H.; Kim, D. Y.; Lebegue, S.; Arnaud, B.; Alouani, M.; Ahuja, R. *Appl. Phys. Lett.* **2008**, *92*, 201903.

(64) Gillespie, R. J. *Chem. Soc. Rev.* **1992**, *21*, 59.

(65) Gilheany, D. G. *Chem. Rev.* **1994**, *94*, 1339.

(66) Albright, T. A.; Burdett, J. K.; Whangbo, M.-H. *Orbital Interactions in Chemistry*; Wiley: New York, NY, 1985.

(67) Hurley, A. C.; Lennard-Jones, J.; Pople, J. A. *Proc. R. Soc. London, Ser. A* **1953**, *220*, 446.

(68) Raff, L. M.; Stivers, L.; Porter, R. N.; Thompson, D. L.; Sims, L. B. *J. Chem. Phys.* **1970**, *52*, 3449.

(69) Hiberty, P. C. *THEOCHEM* **1998**, *451*, 237.

(70) Shaik, S.; Shurki, A. *Angew. Chem., Int. Ed.* **1999**, *38*, 586.

(71) Truhlar, D. G. *J. Comput. Chem.* **2007**, *28*, 73.

(72) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.

(73) Albu, T. V.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 8465.

(74) Slater, J. C.; Koster, G. F. *Phys. Rev.* **1954**, *94*, 1498.

(75) Hoffmann, R. *J. Chem. Phys.* **1963**, *39*, 1397.

(76) Sutton, A. P.; Finnis, M. W.; Pettifor, D. G.; Ohta, Y. *J. Phys. C* **1988**, *21*, 35.

(77) Foulkes, W. M. C.; Haydock, R. *Phys. Rev. B* **1989**, *39*, 12520.

(78) Porezag, D.; Frauenhein, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947.

(79) Horsfield, A. P. *Phys. Rev. B* **1997**, *56*, 6594.

(80) Harris, J. *Phys. Rev. B* **1985**, *31*, 1770.

(81) Sankey, O.; Niklewski, D. *J. Phys. Rev. B* **1989**, *40*, 3979.

(82) Ducastelle, F. In *Computer Simulation in Materials Science*; Meyer, M., Pontikis, V., Eds.; NATO ASI SeriesE 205: Dordrecht, 1991; p 223.

(83) Lathiotakis, N. N.; Andriotis, A. N.; Mennon, M.; Connolly, J. *J. Chem. Phys.* **1996**, *104*, 992.

(84) Sutton, A. P.; Godwin, P. D.; Horsfield, A. P. *MRS Bull.* **1996**, *21*, 42.

(85) (a) Mehl, M.; Papaconstantopoulos, D. A. *Phys. Rev. B* **1996**, *54*, 4519. (b) Papaconstantopoulos, D. A.; Mehl, M. J. *J. Phys.: Condens. Matter* **2003**, *15*, R413.

(86) (a) Wang, C. Z.; Ho, K. M. *Adv. Chem. Phys.* **1996**, *93*, 651. *Tight-Binding Approach to Computational Materials Science*; (b) Turchi, P. E. A., Gonis, A., Colombo, L., Eds.; Materials Research Society: Warendale, PA, 1998.

(87) Stott, M. J.; Zaremba, E. *Phys. Rev. B* **1980**, *22*, 1564.

(88) Gupta, R. P. *Phys. Rev. B* **1981**, *23*, 6265.

(89) Daw, M. S.; Baskes, M. I. *Phys. Rev. B* **1984**, *29*, 6443.

(90) (a) Baskes, M. *Phys. Rev. B* **1992**, *46*, 2727. (b) Lee, B.-J.; Shim, J.-H.; Baskes, M. I. *Phys. Rev. B* **2003**, *68*, 144112.

(91) (a) Lee, B.-J.; Baskes, M. I.; Kim, H.; Cho, Y. K. *Phys. Rev. B* **2001**, *64*, 184102. (b) Lee, B.-J.; Lee, J. W. *Comput. Coupl. Phase Diagr. Thermochem.* **2005**, *29*, 7.

(92) Finnis, M. W.; Sinclair, J. E. *Phil. Mag. A* **1984**, *50*, 45.

(93) Jacobsen, K. W.; Nørskov, J. K.; Puska, M. J. *Phys. Rev. B* **1987**, *35*, 7423.

(94) (a) Horfield, A. P.; Bratkovsky, A. M.; Fearn, M.; Pettifor, D. G.; Aoki, M. *Phys. Rev. B* **1996**, *53*, 12694. (b) Pettifor, D. G.; Oleinik, I. I. *Phys. Rev. B* **1999**, *59*, 8487. (c) Drautz, R.; Murdick, D. A.; Nguyen-Manh, D.; Zhou, X.; Wadley, H. N. G.; Pettifor, D. G. *Phys. Rev. B* **2005**, *72*, 144105. (d) Aoki, M.; Nguyen-Manh, D.; Pettifor, D. G.; Vitek, V. *Progr. Mat. Sci.* **2007**, *52*, 154.

(95) Stokbro, K.; Jacobsen, K. W. *Phys. Rev. B* **1993**, *47*, 4916.

(96) Polatoglou, H. M.; Methfessel, M.; Scheffler, M. *Phys. Rev. B* **1993**, *48*, 1877.

(97) Rey, C.; Gallego, L. J.; García-Rodeja, J.; Alonso, J. A.; Iñiguez, M. P. *Phys. Rev. B* **1993**, *48*, 8253.

(98) Canel, L. M.; Carlsson, A. E.; Fedders, P. A. *Phys. Rev. B* **1993**, *48*, 10739.

(99) Nastar, M.; Willaime, F. *Phys. Rev. B* **1995**, *51*, 6896.

(100) Foiles, S. M. *MRS Bull.* **1996**, *21*, 24.

(101) Kallinteris, G. C.; Papanicolaou, N. I.; Evangelakis, G. A.; Papaconstantopoulos, D. A. *Phys. Rev. B* **1997**, *55*, 2150.

(102) Garzón, I. L.; Michaelian, K.; Beltrán, M. R.; Posada-Amarillas, A.; Ordejón, P.; Artacho, E.; Sánchez-Portal, D.; Soler, J. M. *Eur. Phys. J. D* **1999**, *9*, 211.

(103) Berthier, F.; Legrand, B.; Tréglia, G. *Interface Sci.* **2000**, *8*, 55.

(104) (a) Karolewski, M. A. *Radiat. Eff. Defects Solids* **2001**, *153*, 239. (b) Mottet, C.; Tréglia, G.; Legrand, B. *Phys. Rev. B* **2002**, *66*, 045413.

(105) Pettifor, D. G.; Aoki, M. In *Equilibrium Structure and Properties of Surfaces and Interfaces*; Gonis, A., Stocks, G. M., Eds.; Plenum Press: New York, 1992; p 123.

(106) (a) Abel, G. C. *Phys. Rev. B* **1985**, *31*, 1985. (b) Tersoff, J. *Phys. Rev. B* **1989**, *39*, 5566.

(107) (a) Brenner, D. W. *Phys. Rev. Lett.* **1989**, *63*, 1022. (b) Brenner, D. W.; Shenderova, O. A.; Shall, J. D.; Areshkin, D. A.; Adiga, S.; Harrison, J. A.; Stuart, S. J. In *Handbook of Nanoscience, Engineering, and Technology*; Goddard, W. A., III; Brenner, D. W., Lyshevski, S. E., IafrateG. J., Eds.; CRC Press: Boca Raton, FL, 2003; p 24.

(108) Baskes, M. I. *Phys. Rev. Lett.* **1987**, *59*, 2666.

(109) (a) London, F. *Z. Elektrochem.* **1929**, *35*, 551. (b) Raff, L. M.; Stivers, L.; Porter, R. N.; Thompson, D. L.; Sims, L. B. *J. Chem. Phys.* **1970**, *52*, 3449. (c) Parr, C. A.; Truhlar, D. G. *J. Phys. Chem.* **1971**, *75*, 1844.

(110) Ackland, G. J.; Reed, S. K. *Phys. Rev. B* **2003**, *67*, 174108.

(111) Morse, P. M. *Phys. Rev.* **1929**, *34*, 57.

(112) (a) Pauling, L. *The Nature of the Chemical Bond*, 3rd. ed.; Cornell University Press: Ithaca, NY, 1960, p 98. (b) Johnston, H. S. *Adv. Chem. Phys.* **1960**, *3*, 131. (c) Lagana, A.; Aspuru, G. O. d.; Garcia, E. *J. Chem. Phys.* **1998**, *108*, 3886. (d) Lendvay, G. *THEOCHEM* **2000**, *501*−*502*, 389.

(113) Hall, G. G. *Theor. Chim. Acta* **1985**, *67*, 439.

(114) (a) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865. (b) Adamo, C.; Barone, V.*J. Chem. Phys.* **1999**, *110*, 6158.

(115) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(116) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639.

(117) Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 41.

(118) Schultz, N. E.; Staszewska, G.; Staszewski, P.; Truhlar, D. G. *J. Phys. Chem. B* **2004**, *108*, 4850.

(119) Kołos, W.; Wolniewicz, L. *J. Chem. Phys.* **1965**, *43*, 2429.

(120) Boothroyd, A. I.; Keogh, W. J.; Martin, P. G.; Peterson, M. R. *J. Chem. Phys.* **1996**, *104*, 7139.

(121) Boothroyd, A. I.; Martin, P. G.; Keogh, W. J.; Peterson, M. J. *J. Chem. Phys.* **2002**, *116*, 666.

(122) Pople, J. A.; Nesbet, R. K. *J. Chem. Phys.* **1954**, *22*, 571.

(123) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(124) Brown, F. B.; Shavitt, I.; Shepard, R. *Chem. Phys. Lett.* **1984**, *105*, 363.

(125) Pople, J. A.; Seeger, R.; Krishnan, R. *Int. J. Quantum Chem. Symp* **1977**, *11*, 149.

(126) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. *Chem. Phys.* **1980**, *48*, 157.

(127) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M.; Elbert, S. T. *Chem. Phys.* **1982**, *71*, 41.

(128) Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *294*, 45.

(129) Carroll, D. L. In *Developments in Theoretical and Applied Mechanics*; Wilson, H., Batara, R., Bert, C., Davis, A., Schapery, R., Stewart, D., Swinson, F., Eds.; School of Engineering, The University of Alabama: Tuscaloosa, AL, 1996; Vol. 17, p 411.

(130) Carroll, D. L. *GA-version 1.7a: FORTRAN Genetic Algorithm Driver*; CU Aerospace: Urbana, IL, 2001.

(131) Message Passing Interface Forum. *Document for a standard message passing interface, Technical Report No. CS-93-214*; University of Tennessee: Knoxville, TN, 1994.

(132) Gropp, W.; Lusk, E.; Skjellum, A. *Using MPI: Portable Parallel Programming with the Message-Passing Interface*; 2nd ed.; MIT Press: Cambridge, MA, 1999.

(133) Davidson, E. R. *Chem. Rev.* **1993**, *93*, 2337.

(134) Baskes, M. I.; Nelson, J. S.; Wright, A. F. *Phys. Rev. B* **1989**, *40*, 6085.

(135) Vashishta, P.; Kalia, R. K.; Nakano, A. *J. Phys. Chem. B* **2006**, *110*, 3727.

(136) Kadai, K.; Germann, T. C.; Lomdahl, P. S.; Albers, R. C.; Wark, J. S.; Higgenbotham, A.; Holian, B. L. *Phys. Rev. Lett.* **2007**, *98*, 135701.

(137) Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618.

(138) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231.

(139) Bash, P. A.; Ho, L. L. jr.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 3698.

(140) Layfield, J. P.; Owens, M. D.; Troya, D. *J. Chem. Phys.* **2008**, *128*, 194302.

CT8004535

J. Chem. Theory Comput. **2009,** *5,* 605–614

JCTC Journal of Chemical Theory and Computation

Toward an Accurate Density-Functional Tight-Binding Description of Zinc-Containing Compounds

Ney H. Moreira,* Grygoriy Dolgonos, Bálint Aradi, Andreia L. da Rosa, and Thomas Frauenheim

Bremen Center for Computational Materials Science, University of Bremen, Am Fallturm 1, 28359 Bremen, Germany

Received October 28, 2008

Abstract: An extended self-consistent charge density-functional tight-binding (SCC-DFTB) parametrization for Zn−X (X = H, C, N, O, S, and Zn) interactions has been derived. The performance of this new parametrization has been validated by calculating the structural and energetic properties of zinc solid phases such as bulk Zn, ZnO, and ZnS; ZnO surfaces and nanostructures; adsorption of small species (H, CO_2, and NH_3) on ZnO surfaces; and zinc-containing complexes mimicking the biological environment. Our results show that the derived parameters are universal and fully transferable, describing all the above-mentioned systems with accuracies comparable to those of first-principles DFT results.

1. Introduction

Zinc is the second most abundant transition metal in living organisms, playing an important role in many biological processes as a part of metalloenzymes or complexes with amino acids, nucleic acids, or proteins. Additionally, zinc oxide (ZnO) is a promising material for applications in electronics and optoelectronics, since it is a wide band gap semiconductor and has large excitonic binding energy. Recently, ZnO has been synthesized in a variety of nanostructures, including nanowires,[1−3] nanobelts,[4−6] nanorods,[7,8] and nanopilars,[9,10] which opens up the possibility for novel applications. Furthermore, the covalent attachment of functional organic linkers to ZnO surfaces can be used in biosensing applications.[11]

Zinc-containing systems have been widely investigated by first-principles methods.[12,13] Although such methods represent the state-of-the-art approach in materials science and solid-state simulations, they become prohibitive at treating a large number of atoms because of their high computational demands. In this context, one of the most promising approaches is the self-consistent charge density-functional tight-binding[14,15] (SCC-DFTB) method, which has been successfully applied to large-scale quantum-mechanical simulations in solid-state physics, chemistry, materials science, and biophysics.[16,17] The method is an approximation to the Kohn−Sham density-functional theory (DFT),[18] which combines reasonable accuracy and computational efficiency. However, the limited set of available parameters is still a drawback, and in some cases the transferability of parameters between solid and molecular environments is still a problem. For example, a recent SCC-DFTB parametrization for Zn has been successfully applied to investigate zinc-containing biological molecules,[19] but this set has been unable to model solid-phase zinc systems with acceptable accuracy.

It becomes evident that a new SCC-DFTB parametrization for zinc interactions, able to describe both solid-state and biological systems, would open up new possibilities for investigating Zn-containing materials. Further, it would reinforce the confidence in the transferability of SCC-DFTB parameters among different environments.

In this work, we present our recently derived SCC-DFTB parametrization for representing zinc and its interactions with hydrogen, carbon, nitrogen, oxygen, and sulfur. The parametrization has been validated by comparing SCC-DFTB results with literature data for solid-state zinc-containing systems (metallic Zn, ZnO in the wurzite (w), and ZnS in the zinc-blend (zb) structures), ZnO surfaces and nanostructures, adsorption of small species (H, CO_2, and NH_3) on ZnO, and models for zinc biological complexes. Our results have been found quantitatively comparable to those obtained by DFT calculations, demonstrating that the derived parameters

---

* Corresponding author e-mail: henrique.ney@bccms.uni-bremen.de.

10.1021/ct800455a CCC: $40.75

are reliable in predicting geometries and energetics of both solid and molecular systems, being fully transferable among different chemical environments.

2. Computational Methodology

2.1. SCC-DFTB Method. Since the foundations of the DFTB approach are described elsewhere[14,15] and have been also the subject of a recent review,[20] only a short description of the method will be given here.

The SCC-DFTB total energy expression (eq 1) is a second-order expansion of the DFT Kohn−Sham total energy with respect to charge density fluctuations (δn):

$$E_{\mathrm{tot}}^{\mathrm{DFTB}}=\sum_{i}^{\mathrm{occ}}n_i\langle\Psi_i|\hat{H}_0|\Psi_i\rangle+\frac{1}{2}\sum_{\alpha\beta}^{N}\gamma_{\alpha\beta}\Delta q_\alpha\Delta q_\beta+E_{\mathrm{rep}}=E_0[n_0]+E_2[n_0,\delta n]+E_{\mathrm{rep}} \quad (1)$$

In this expression, the zero-order term E_0 is the energy summation over all occupied eigenstates Ψ_i. The approximated non-SCC Hamiltonian ($\hat{H}_0$) is derived within the two-center approximation for an arbitrary reference charge density (n_0), introduced as the superposition of individual atomic charge densities. This Hamiltonian is also subjected to the frozen-core approximation (i.e., only the valence electrons are considered explicitly and 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 the charge density fluctuations δn (the first-order terms in this expansion vanish for any arbitrary n_0), approximated as atomic pointlike charges (Δq). Finally, the third term E_{rep} accounts for the "double-counting" terms and the ion−ion core interactions in a set of distance-dependent pairwise repulsive potentials, modeled as the difference between the SCC-DFTB electronic energy ($E_{\mathrm{Elec}}^{\mathrm{DFTB}} = E_0 + E_2$) and the total DFT energy for some reasonably chosen reference system:

$$E_{\mathrm{rep}}=\sum_{\alpha,\beta}U_{\alpha\beta}(R_{\alpha\beta}) \quad (2)$$

where

$$U_{\alpha\beta}(R_{\alpha\beta})=E_{\mathrm{Total}}^{\mathrm{DFT}}(R_{\alpha\beta})-E_{\mathrm{Elec}}^{\mathrm{DFTB}}(R_{\alpha\beta})$$

with $R_{\alpha\beta}$ representing the distance between the atoms α and β.

The wave functions (Ψ_i) are expanded as a linear combination of atomic orbitals:

$$\Psi_i=\sum_{\mu}c_\mu\phi_\mu \quad (3)$$

In the common DFTB framework, the atomic orbitals (ϕ) in eq 3 are constructed as linear combinations of Slater-type orbitals, obtained by solving self-consistently the modified Kohn−Sham equation (eq 4) for the spin-unpolarized free atom:

$$\left[\hat{T}+w^0+\left(\frac{r}{r_0}\right)^2\right]\phi_\mu(r)=\varepsilon_\mu\phi_\mu(r) \quad (4)$$

The modified Hamiltonian in eq 4 consists of a kinetic energy operator $\hat{T}$, the potential energy for the neutral atom w^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 energy band structures (see ref 20 for details).

Having defined the atomic orbitals (ϕ) and an initial set of expansion coefficients (c_μ), we estimated the atomic charge fluctuations (Δq) via Mulliken population analysis, and the DFTB total energy problem is solved self-consistently by using the secular eq 5

$$\sum_{i}c_{i\mu}(H_{\mu\nu}-\varepsilon_i S_{\mu\nu})=0 \quad \forall\,\mu,i \quad (5)$$

with

$$H_{\mu\nu}=H_{\mu\nu}^0+\frac{1}{2}S_{\mu\nu}\sum_{\xi}^{N}(\gamma_{\alpha\xi}+\gamma_{\beta\xi})\Delta q_\xi$$

$$H_{\mu\nu}^0=\langle\phi_\mu|\hat{H}|\phi_\nu\rangle \quad \forall\,\mu\in\alpha,\nu\in\beta$$

$$S_{\mu\nu}=\langle\phi_\alpha|\phi_\nu\rangle$$

In eq 5, the diagonal zero-order matrix elements $H_{\mu\nu}^0$ are taken as the eigenvalues obtained from the free-atom calculations, while the distance-dependent nondiagonal elements $H_{\mu\nu}^0$ and $S_{\mu\nu}$ are calculated within the two-center approximation and tabulated for all future calculations. This procedure avoids integral evaluations during the DFTB calculation, leading to a computational efficiency comparable to those of traditional semiempirical methods while retaining the accuracy comparable to those of DFT methods. The charge transfers among different atoms take into account their respective chemical hardnesses (Hubbard parameters), calculated as the first derivative of the total atomic energy with respect to the electronic occupation around the neutral-atom electronic density.

2.2. Parametrization Details. The parameters necessary to represent a system within the DFTB method include the Hubbard parameters for every chemical element, the $H_{\mu\nu}^0$ and $S_{\mu\nu}$ matrix elements, and the repulsive pairwise potentials for all interacting neighbors. Fortunately, all of these parameters are controlled by a few main quantities to be determined in the parametrization process, namely: (i) the reference input density (n_0), (ii) the wave function confinement radius (r_0), and (iii) the repulsive cutoffs, determining the distances where the repulsive pairwise potentials (E_{rep}) vanish.

As the parametrization reported here extends a previous well-established one,[21] all parameters not involving zinc atoms are assumed to be the same as in ref 19, including n_0 and r_0 for the H, C, N, O, and S atoms. For zinc, n_0 was confined into a 2.69 Å radius while r_0 was chosen to be 1.59 Å, because these values provided a good compromise among the description of geometries, cohesive properties, and electronic band structures for hexagonal closed-packed (hcp)-Zn and w-ZnO structures. The reference systems and cutoffs used to model the Zn−X (X = H, C, N, O, S, and Zn) pairwise repulsive potentials are summarized in Table 1.

The electronic DFTB parameters (i.e., Hubbard parameters, $H_{\mu\nu}^0$ and $S_{\mu\nu}$ matrix elements) were derived directly from DFT calculations, performed within the generalized gradient

Table 1. Parametrization Details of the Pairwise Repulsive Potentials for Zn−X Interactions (X = H, C, N, O, S, and Zn)[a]

interaction	reference system	repulsive cutoff (Å)	equilibrium Zn−X distance (Å)	
			SCC-DFTB	DFT
Zn−H	ZnH_2	1.63	1.63	1.54
Zn−C	$Zn(CH_3)_2$	2.01	1.97	1.95
Zn−N	$Zn(NH_3)_2$	2.10	1.92	1.95
Zn−O	zb-ZnO	2.23	2.00	1.98
Zn−S	$Zn(SH)_2$	2.40	2.21	2.17
Zn−Zn	fcc-Zn	2.75	2.79	2.74

[a] fcc = face-centered cubic.

approximation (GGA) and using the Perdew, Burke, and Ernzerhof (PBE)[22] exchange-correlation functional. All parameters derived here are available for use by the scientific community and can be downloaded from http://www.dftb.org/parameters/ (search for the set named "znorg-0-1").

2.3. Reference DFT Calculations. In the DFT reference calculations to derive the Zn−X repulsive potentials ($U_{\alpha\beta}$), we employed the GGA/PBE functional (for X = H, C, O, and Zn), as implemented in the SIESTA package[23] using a double-ζ plus polarization function (DZP) basis set and norm-conservative Troullier−Martins pseudo-potentials[24] (PP) for representing the valence and inner electrons, respectively. The Becke three-parameter, Lee, Yang, and Parr (B3LYP) hybrid functional[25,26] in conjunction with the 6-311G+(d, p) basis set, as implemented in the Gaussian03 package,[27] was used for X = N, S.

All GGA/PBE calculations were performed with periodic boundary conditions (PBC). For the fcc-Zn and zb-ZnO solid-state reference systems, the *k*-points were sampled with a (8 × 8 × 8) Monkhorst−Pack[28] (MP) grid. For molecular reference systems, the calculations were performed by employing large supercells, including a 25 Å vacuum region in all directions to isolate the molecules from their periodic replicas.

2.4. SCC-DFTB Calculations. The SCC-DFTB calculations were performed with the DFTB+ code.[29] Solid-state hcp-Zn, w-ZnO, and zb-ZnS properties were calculated using PBC and converged (8 × 8 × 4), (8 × 8 × 4), and (8 × 8 × 8) MP *k*-points, respectively. The calculations for ZnO surfaces (clean and with small adsorbates) were performed using suitably oriented supercells and PBC within the slab approach.[39] A vacuum region of 25 Å along the surface-normal direction and a (4 × 4 × 1) MP grid for the *k*-point sampling were employed. The [0001] ZnO nanostructures had their growth direction oriented along the *z*-axis in the supercell, with vacuum regions of 25 Å along the *x* and *y* directions, and were calculated using a (1 × 1 × 4) MP grid for *k*-point sampling. To mimic the zinc environment in biomolecules, we performed geometry optimizations of small zinc-containing complexes, as suggested by Elstner et al.[19] In all SCC-DFTB calculations, the atomic positions were relaxed until the forces in the system became smaller than 0.001 eV/Å.

Table 2. Selected hcp-Zn and w-ZnO Bulk Properties Calculated with SCC-DFTB and DFT Methods and Obtained from Experiments[h]

	E_{coh} (eV)	*a* (Å)	*c* (Å)	*V* (Å^3)	B_0 (GPa)	B_0'
			hcp-Zn			
SCC-DFTB[a]	2.45	2.71	5.04	32.20	114	5.67
PP-DZP/PBE[a]	1.92	2.68	5.00	31.24	73	5.93
PP-PW/PW91[b]		2.65	5.10	31.02	60	6.59
expt	1.36[c]	2.61[b]	4.91[b]	28.96[b]	60−80[b]	5.2−6.4[b]
			w-ZnO			
SCC-DFTB[a]	9.77	3.28	5.25	50.04	161	2.49
PP-DZP/PBE[a]	8.08	3.30	5.34	51.08	124	4.55
PP-PW/PBE[d]	8.98	3.34	5.30	51.40	171	
expt[e]	7.52	3.25	5.20	47.62	208	4
			zb-ZnS			
SCC-DFTB[a]	7.93	5.43		160.1	44.2	2.4
PP-PW/LDA[f]	7.22	5.35		153.3	82	4.6
PP-PW/PW91[g]		5.60		175.6	66.7	3.95
expt[f]	6.33	5.40		157.5	76.9	4.9

[a] GGA/PBE calculation performed with the SIESTA package; this work. [b] GGA/PW91 (Perdew−Wang 1991 exchange-correlation functional) calculation in ref 32. [c] Reference 33. [d] GGA/PBE calculation in ref 34. [e] Reference 35. [f] LDA calculation in ref 36. [g] GGA/PW91 (Perdew−Wang 1991 exchange-correlation functional) calculation in ref 37. [h] PW = plane-wave basis set.

3. Results and Discussion

3.1. Bulk Systems. The bulk hcp-Zn, w-ZnO, and zb-ZnS lattice parameters and elastic properties were determined by calculating energy-volume profiles in a ±15% range around the experimental equilibrium volumes and fitting the results to the Murnaghan equation of state. For hcp-Zn and w-ZnO structures, we fixed the experimental *c*/*a* ratios (1.630[30] and 1.602,[31] respectively).

As shown in Table 2, the DFTB parameters for all considered bulk systems are in fairly good agreement with DFT and experimental ones. The deviations in the cohesive energies are not surprising, since DFTB calculations usually overestimate this property,[38] while the error in the Zn bulk modulus is a consequence of strong zinc wave function compression applied to this element (~1.3 times larger than the Zn covalent radius), which shortens its dissociation bond distances and lowers its electronic energy wells. Nevertheless, this strong wave function compression ensures reasonable band structures for the solid materials, as exemplified in Figure 1 for the case of w-ZnO.

It should be mentioned that, because of the small basis set employed, the energy band structures calculated with DFTB are not in the same way affected by the band gap problem, as observed in local density approximation (LDA) and GGA calculations. Therefore, with no additional correction scheme being applied, the ZnO band gap obtained via DFTB (~4.1 eV) is closer to the experimental value (~3.3 eV) than the GGA-PBE result (~0.9 eV). It may be taken as an advantage to study electronic states introduced by defects or adsorbed species; however, it must be noted that the dispersions in the edge of the conduction band are considerably smaller than those found in GGA-PBE results and may lead to deviations in calculating transport properties.

Concerning the hcp-Zn and w-ZnO cohesive energies, the overestimation in principle could be corrected by applying

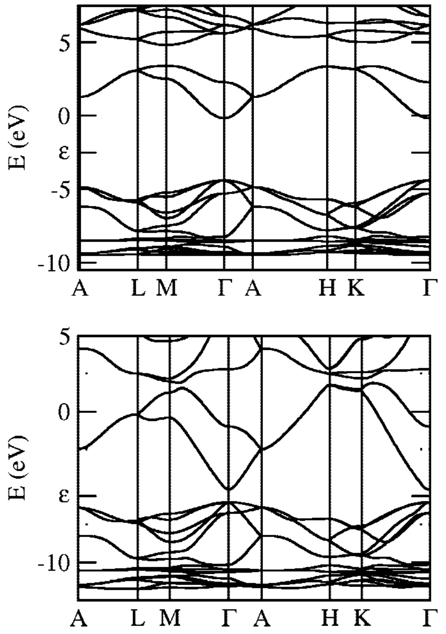


Figure 1. Electronic band structures for w-ZnO calculated with SCC-DFTB (top) and DFT (PP-DPZ/PBE) (bottom) methods. The DFT band structure was calculated using the SIESTA package. ε denotes the Fermi level for each case.

longer repulsive potentials, slightly sacrificing the geometrical parameters. However, we verified that this procedure leads to wrong relaxations for ZnO surfaces and nanostructures, where the outermost atoms move toward the vacuum region to avoid the artificially enhanced repulsion.

3.2. ZnO Surfaces. ZnO surfaces form a good probe set to test the new parameters, since they have the bulk symmetry broken along one direction. Additionally, they have been subject to several DFT studies in the literature.[39,40] The (10$\underline{1}$0), (1$\underline{2}$10), (0001/000$\underline{1}$), and depolarized dep-(0001) surfaces were constructed starting from the equilibrium DFTB w-ZnO structure. The surface unit cells for these four structures are represented in Figure 2. In all cases, our results for the surface relaxations are in good agreement with those reported in ref 40, as shown in Table 3.

The relaxations in the (10$\underline{1}$0) terminated slabs (Figure 3A) are only significant in the three outermost surface layers. The oxygen atoms remain close to their bulk positions, while the zinc atoms move inward in the top layer and slightly outward in the second layers. In comparison with the bulk values, the bond lengths between the top and the second layer, $d_{\mathrm{Zn-O}}$, are ~6% shortened while the bonds between the second and the third layers are ~3% larger. The O−Zn−O angle (α) changes from its 109° bulk value to 117° at the top surface layer, remaining unchanged in the inner layers. The (1$\underline{2}$10)-terminated slabs follow similar trends, with relaxations observed especially over the outermost Zn atoms, leading to a ZnO bond length ~4% shorter and strongly distorted bond angles at the surface.

The depolarized (0001) surface relaxation, as shown in Figure 4B, agrees well not only with the theoretical prediction derived by Claeyssens et al.,[40] but also with recent experimental data confirming the existence of such structure in ZnO ultrathin films.[41] This is a graphite-like structure, where both O and Zn atoms assume a planar sp^2 configuration, with all in-plane bonds by ~3% smaller than those in the w-ZnO, and with large interlayer distances (~2.34 Å). In this case, the relaxations do not differ significantly among different layers into the slab and are also not influenced by the slab

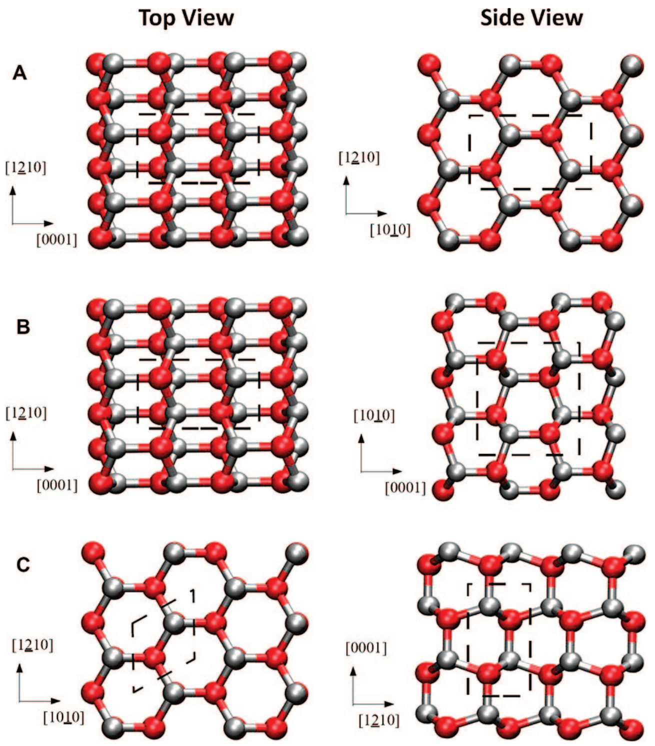


Figure 2. Top and side views of surface unit cells for (10$\underline{1}$0) (A), (1$\underline{2}$10) (B), and (0001/000$\underline{1}$) (C) surfaces. The dep-(0001) surface unit cell is similar to that used for its polar counterpart, but with planar ZnO sheets perpendicular to the [0001] direction.

Table 3. Comparison of Geometrical Relaxations in ZnO Surfaces Calculated with DFTB, for Selected Structural Parameters As Specified in Figures 3 and 4, with DFT (PP-PW/PW91) Results in Ref 40

		surface layers		inner layers	
surface	parameter	SCC-DFTB	PP-PW/PW91	SCC-DFTB	PP-PW/PW91
ZnO(10$\underline{1}$0)					
	$d_{\mathrm{Zn-O}}$ (Å)	1.88	1.85	2.01	1.99
	$d'_{\mathrm{Zn-O}}$ (Å)	2.05	2.06	2.01	1.99
	α (deg)	117	117	109	
	β (deg)	108		109	
ZnO(1$\underline{2}$10)					
	$d_{\mathrm{Zn-O}}$ (Å)	1.93	1.87	2.01	1.99
	$d_{\mathrm{O-Zn}}$ (Å)	2.00	1.96	2.01	1.99
	$d'_{\mathrm{Zn-O}}$ (Å)	2.05	2.06	2.01	
	α (deg)	118	117	109	
	β (deg)	97		109	
ZnO(0001/000$\underline{1}$)					
	$d_{\mathrm{Zn-O}}$ (Å)	1.94	1.92	1.99	1.97
	$d'_{\mathrm{Zn-O}}$ (Å)	2.08	2.15	2.05	2.08
	α (deg)	111	113	109	
	β (deg)	120	120	90	
ZnO (dep-0001)					
	$d_{\mathrm{Zn-O}}$ (Å)	1.90	1.93	1.90	
	$d'_{\mathrm{Zn-O}}$ (Å)	2.34	2.4	2.34	
	α (deg)	120	120	120	
	β (deg)	90		90	

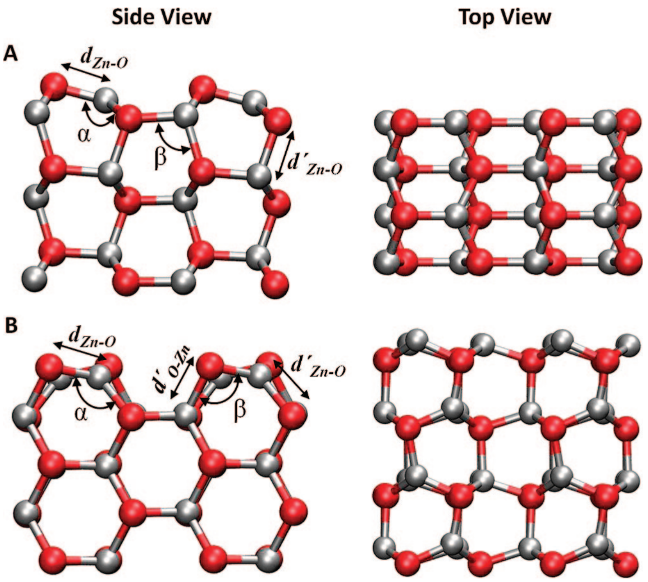


Figure 3. Side and top views of the first three bilayers of relaxed (10$\underline{1}$0) (A) and (1$\underline{2}$10) (B) ZnO surfaces calculated with SCC-DFTB.

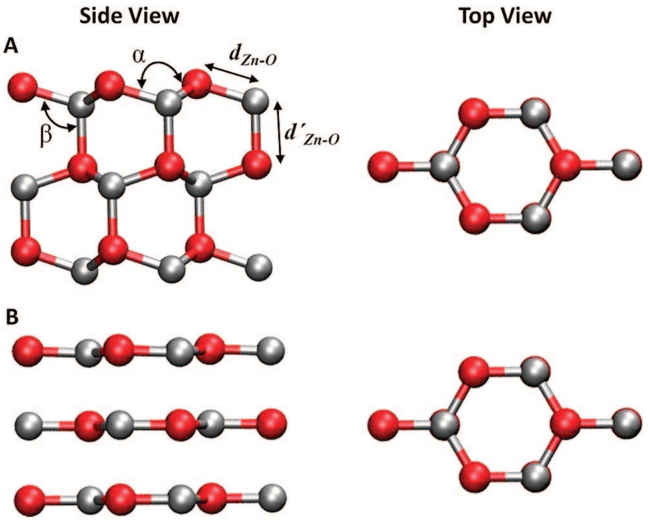


Figure 4. Side and top views of (A) the relaxed (0001/000$\underline{1}$) and (B) dep-(0001) ZnO surfaces calculated with SCC-DFTB.

thickness. For the polar (0001/000$\underline{1}$) surface, where the structure resembles the w-ZnO (Figure 4A) one, the relaxations extend through the whole slab and are more sensitive to variations in its thickness. Further, we also observed small charge transfers from the oxygen-terminated to the zinc-terminated surface (~0.3e per slab unit cell), which have usually been considered as a stabilizing mechanism for such surfaces.[42,43]

Since the polar (0001) and (000$\underline{1}$) surfaces are not equivalent, it is not possible to calculate their absolute formation energies. Therefore, we calculated the cleavage energy (the energy required to create two surfaces by cleaving the perfect crystal) for all calculated slabs to verify their relative stability as a function of the film thickness (Figure 5). The results are in excellent agreement with the DFT predictions in ref 40. As expected, the slabs with the polar surfaces are less stable than those (10$\underline{1}$0)- and (1$\underline{2}$10)-terminated. However, the (0001) depolarized structure is the most stable for the thinnest films, in accordance with the one experimentally found for ultrathin ZnO films,[41] but quickly rising in energy with increasing the slab thickness. It is necessary to remark that the phase transitions from the depolarized (0001) film to other configurations (in this study, at 5-, 6-, and 16-layer-thick slabs, respectively) take place at points slightly different from those found by Claeyssens et al. (namely, 9, 10, and 18 layers), but this does not affect the general conclusions drawn here.

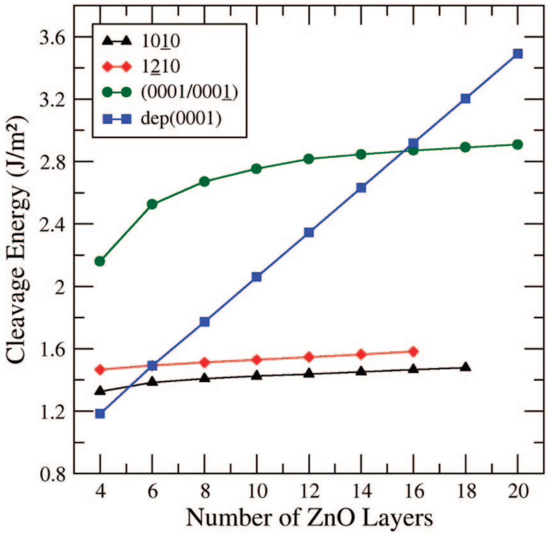


Figure 5. SCC-DFTB cleavage energy for (10$\underline{1}$0), (1$\underline{2}$10), (0001/000$\underline{1}$), and dep(0001) ZnO surfaces as a function of the slab thickness.

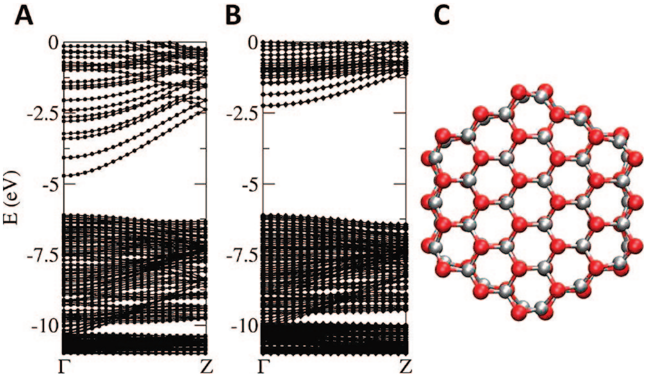


Figure 6. (A) DFT (PP-DPZ/PBE) and (B) SCC-DFTB electronic band structures for (C) the hexagonal [0001] nanowire with a diameter of 16.5 Å. The DFT calculation was performed using the SIESTA package.

3.3. ZnO Nanostructures. To extend the validation procedure to ZnO nanostructures, we investigated (10$\underline{1}$0)-faceted hexagonal nanowires with different diameters (3.7, 9.9, 16.5, and 23.0 Å, respectively). Figure 6 shows the relaxed cross section of one of the investigated nanowires and its band structure calculated with both DFT and DFTB. For all nanowires, the relaxations at the outermost layers follow the same general trends observed for the (10$\underline{1}$0) surfaces, being also in very good agreement with previous DFT results.[42] The inner atoms remain close to their positions in the bulk ZnO, as the O atoms at the surface, whereas the zinc atoms at the surface move inward, shrinking the distance

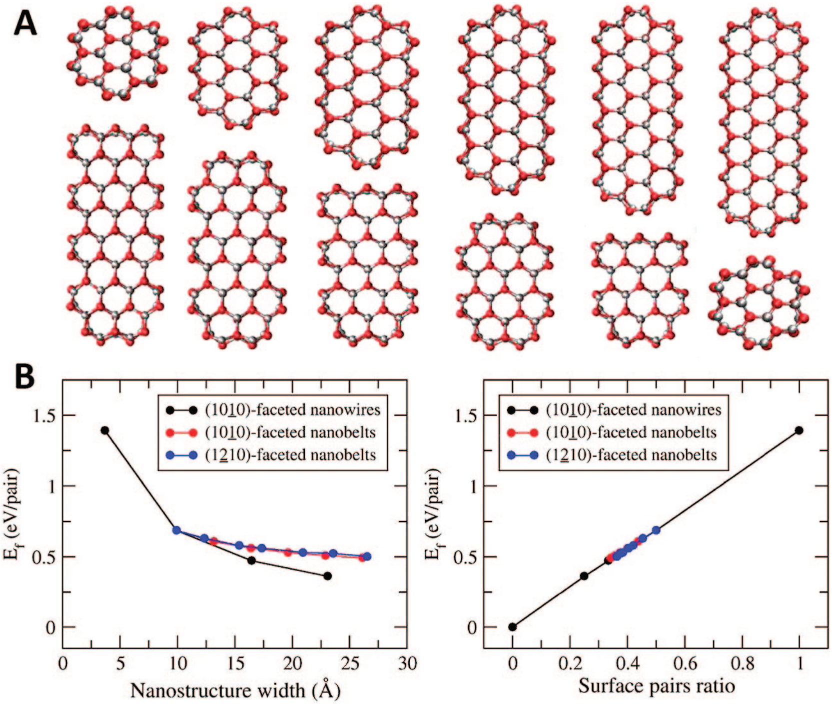


Figure 7. (A) Cross section of the relaxed structure for (10$\underline{1}$0)- (top) and (1$\underline{2}$10)-faceted (bottom) [0001] ZnO nanobelts with different widths across the dominating surface calculated with SCC-DFTB. (B) Formation energies of ZnO nanostructures grown along the [0001] direction, for hexagonal nanowires, (10$\underline{1}$0)- and (1$\underline{2}$10)-faceted nanobelts, as a function of the nanostructure width and surface pair ratio.

to their nearest inner Zn neighbors to 3.07 Å. The exception is the smallest nanowire, where the relaxations are more pronounced, with 2.86 Å as the minimal distance between Zn atoms. For these wires, the energy band structures along the Γ-Z direction in the Brillouin zone appeared in good agreement with DFT results,[42] with a direct band gap at the Γ-point, which shrinks from 4.1 to 3.6 eV as the nanowire diameter increases, indicating quantum confinement effects.

The relative stability of the nanowires is also in good agreement with plane waves calculations by Xu et al.,[12] who also demonstrated that for hexagonal nanowires the formation energy depends linearly on the ZnO surface pair ratio. To further validate the parameters, we calculated [0001]-oriented ZnO nanobelts and found a similar linear dependence, as shown in Figure 7. The nanobelts were found to have their relaxations consistent with those observed for their dominant surfaces, with their stabilities lying between that of the corresponding nanowire and that of the infinite thin film. It should be noted that the relative stability of the nanobelts increases with their width, as expected.

3.4. Adsorption of Small Molecules on ZnO Surfaces.

The results achieved in describing ZnO surfaces encouraged us to start validating our parametrization for the Zn−H, Zn−C, and Zn−N interactions by investigating the adsorption of small species (atomic H, NH_3, and CO_2) on the (10$\underline{1}$0) ZnO. We analyzed their geometrical configurations on the surface and also their adsorption energies, defined as $E_{ads} = (E_T - E_{ZnO-10\underline{1}0} - n\mu)/2$, where E_T is the ZnO−adsorbate complex total energy, $E_{ZnO-10\underline{1}0}$ is the energy of the bare slab, μ is the adsorbate chemical potential, and n is the number of adsorbed species. The factor $^1/_2$ is used because we have two equivalent surfaces.

Table 4. Comparison between SCC-DFTB and B3LYP/6-311+G(d,p) Results[43] for Hydrogen Monolayer Adsorption on ZnO (10$\underline{1}$0) Surface[a]

	SCC-DFTB	B3LYP
E_{ads} (eV)	−0.77	−0.62
r(Zn−H) (Å)	1.74	1.59
r(O−H) (Å)	0.99	0.99
θ(Zn−H) (deg)	31	41
θ(O−H) (deg)	28	38

[a] The angle θ represents the inclinations of the X−H bond (X = Zn or O) to the surface normal direction.

For the H-covered surfaces, we used the hydrogen chemical potentials equal to $^1/_2$ H_2 total energy. In accordance with ab initio investigations,[43−45] our results suggest that the Zn−H bond is less energetically favorable, as the half-monolayer coverage was found to be ~1 eV more stable than the monolayer. In both cases, the SCC-DFTB E_{ads} is slightly overestimated in comparison with first-principles results. The surface geometry for the monolayer coverage is also in reasonable agreement with the B3LYP predictions (Table 4).

For the CO_2-covered surfaces, the DFTB results (Figure 8) are in good agreement with recent DFT ones,[46] which show a tridentate surface carbonate (TSC) as the stable adsorbate species in the activation of CO_2 over ZnO. By using the CO_2 chemical potential as a variational parameter,[47] we obtained a phase-stability diagram similar to that proposed in ref 46. It should be noted that our E_{ads} values are overestimated by ~0.5 eV per CO_2 molecule in comparison to those from ref 46. This deviation is not surprising, since our model overestimates the Zn−O bond strength in comparison with those from DFT (Table 2). Therefore, the DFTB

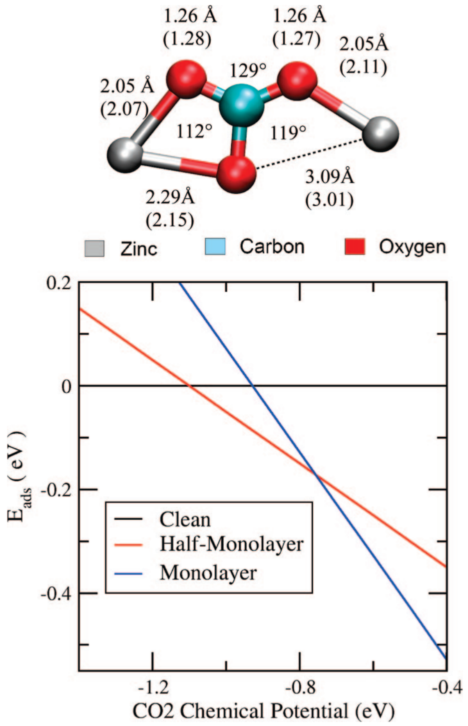


Figure 8. Top: Geometry of the TSC structure calculated with SCC-DFTB. The bond length values in the parentheses are DFT results taken from ref 46. The angles calculated with DFTB coincide exactly with those reported in ref 46. Bottom: Phase-stability diagram calculated with SCC-DFTB for different coverages of CO_2 on ZnO $(10\underline{1}0)$ surfaces.

adsorption energy related to the TSC structure is also overestimated in comparison with those from DFT, and the DFTB phase diagram is shifted toward lower CO_2 chemical potentials consequently. It is interesting to observe that the interaction with the CO_2 molecule strongly influences the surface geometry by pushing the Zn atoms outward and decreasing the angle α to 89° and increasing the ZnO dimer distance d_{Zn-O} to 2.29 Å.

ZnO has been also considered as a promising material for sensing NH_3 in gas phase,[48] but there are only a few theoretical studies on the absorption of ammonia on ZnO surfaces,[49] in which cluster models were employed to investigate single-molecule adsorption. A strong binding energy (~1.8 eV)[49] was reported for the H_3N-ZnO surface complex, which is not surprising given the basicity of NH_3 and the acidic character of zinc in ZnO. This strong binding energy and the small molecular volume of ammonia in comparison to CO_2 indicate that adsorption of NH_3 on ZnO$(10\underline{1}0)$ surfaces can also lead to the formation of self-assembled monolayers.

Our results show that ammonia binds to the ZnO surface through a covalent N−Zn bond (2.05 Å), thus reducing the angle α to 115° and increasing the ZnO dimer distance d_{Zn-O} to 1.96 Å. As expected, no ammonia dissociation was found. Instead, there is a typical hydrogen bond distance of 1.68 Å between ammonia hydrogen and ZnO oxygen. The NH_3 tilting angle with respect to the surface normal of 41° is in fairly good agreement with near-edge X-ray fine-structure spectroscopy reported by Kamada et al.[50] Additionally, our results (cf. Figure 9) agree very well with coupled-cluster calculations by Taft et al.[49]

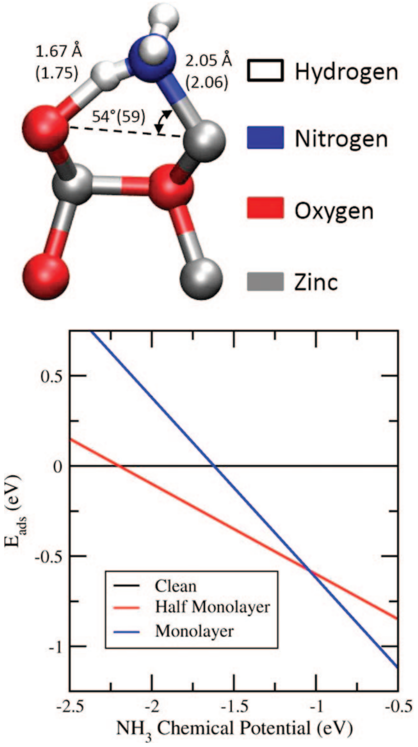


Figure 9. Top: Geometry of the H_3N-ZnO surface complex calculated with SCC-DFTB. The values in the parentheses are ab initio results taken from ref 49. Bottom: Phase-stability diagram calculated with SCC-DFTB for different coverages of NH_3 on ZnO $(10\underline{1}0)$ surfaces.

Table 5. Comparison between SCC-DFTB and B3LYP/6-311+G(d,p) Cohesive Energies and Equilibrium Geometries for Zinc Complexes Containing NH_3 and HS^- Ligands

	E_{coh} (eV)		bond length (Å) (Zn−X)		bond angles (deg) (X−Zn−X)	
species	SCC-DFTB	B3LYP	SCC-DFTB	B3LYP	SCC-DFTB	B3LYP
			Zn−N			
$[ZnNH_3]^{2+}$	5.36	5.98	1.93	1.97		
$[Zn(NH_3)_2]^{2+}$	11.86	10.70	1.92	1.95	180	180
$[Zn(NH_3)_3]^{2+}$	13.96	13.34	1.97	2.03	119.9	119.5
$[Zn(NH_3)_4]^{2+}$	16.75	15.3	2.02	2.09	109.5	109.4
			Zn−S			
$[Zn(SH)]^+$	18.06	18.35	2.03	2.18		
$Zn(SH)_2$	27.81	27.56	2.21	2.19	177.5	178.4

3.5. Modeling Zinc in Biological Systems. As a first step toward validating the parameters for Zn-containing biomolecules, we simulated small zinc complexes with NH_3 and SH^- ligands and found a reasonable agreement between our SCC-DFTB and B3LYP results (Table 5). We also reproduced recent B3LYP geometry parameters for cyclic ZnC_n ($n = 2-5$) clusters[51] and found the errors in the Zn−C and C−C bonds to be smaller than 0.05 Å in all cases.

By applying the procedure suggested by Elstner et al.,[19] we calculated three zinc-containing complexes to model zinc interactions with specific functional groups in proteins, using SH^-, $CH_2{=}NH$, and $HCOO^-$ ligands to represent the thiol

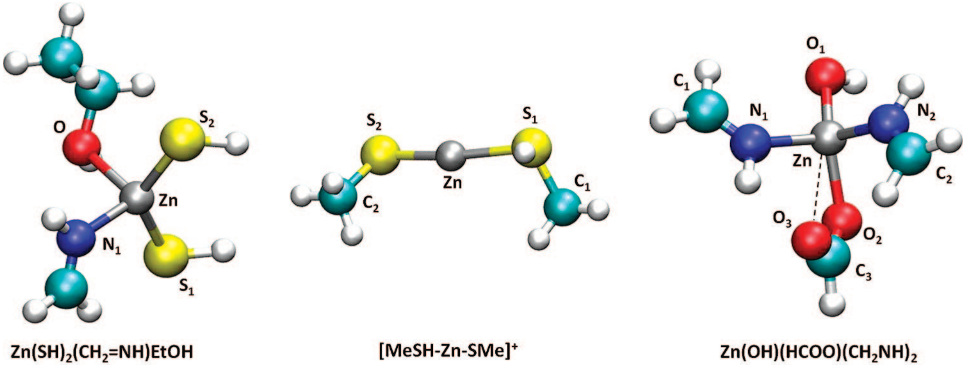


Figure 10. Optimized geometries of model zinc complexes calculated with SCC-DFTB.

Table 6. Comparison of Geometric Parameters of Model Zinc Complexes Calculated with B3LYP/6-311+G(d,p) and SCC-DFTB Methods[c]

complex	method B3LYP[a]	 SCC-DFTB[a]	 SCC-DFTB[b]
1: $Zn(SH)_2(CH_2{=}NH)EtOH$			
$r(Zn-S_1)$	2.261	2.258	2.301
$r(Zn-S_2)$	2.261	2.243	2.290
$r(Zn-O)$	2.272	2.288	2.158
$r(Zn-N)$	2.155	2.056	2.001
$\angle(S_1-Zn-S_2)$	143.5	140.3	134.0
$\angle(S_2-Zn-O)$	91.2	102.8	105.4
$\angle(S_1-Zn-N)$	106.9	104.1	105.4
2: $[MeSH-Zn-SMe]^+$			
$r(Zn-S_1)$	2.349	2.293	2.329
$r(Zn-S_2)$	2.152	2.123	2.169
$r(S_1-C)$	1.853	1.832	1.831
$\angle(C-S_1-Zn)$	104.9	108.0	104.9
$\angle(S_1-Zn-S_2)$	175.1	169.1	169.9
$\angle(Zn-S_2-C)$	105.8	110.2	105.5
3: $Zn(OH)(HCOO)(CH_2{=}NH)_2$			
$r(Zn-O_1)$	1.876	1.906	1.868
$r(Zn-O_2)$	1.972	2.080	2.023
$r(O_2-C_3)$	1.282	1.301	1.304
$r(C_3-O_3)$	1.237	1.241	1.251
$r(Zn-O_3)$	3.010	2.911	2.931
$r(N_1-C_1)$	1.270	1.267	1.262
$r(Zn-N_2)$	2.129	2.045	2.009
$\angle(O_1-Zn-O_2)$	131.6	109.1	125.8
$\angle(N_1-Zn-N_2)$	103.3	102.7	104.1
$\angle(N_2-Zn-O_1)$	100.3	104.0	109.8

[a] Reference 19. [b] Using current parametrization; this work. [c] Bond lengths are in angstroms, and angles are in degrees.

Table 7. Comparison of Formation Energies (in Kilocalories per Mole) of Model Zinc Complexes Calculated with B3LYP/6-311+G(d,p) and SCC-DFTB Methods

	method B3LYP	 SCC-DFTB[a]	 SCC-DFTB[b]
reaction 1[c]	−14.7[d]	−12.3	−20.1
reaction 2[e]	−50.9[f]		−53.2
reaction 3[g]	−159.4[d]	−180.9	−186.9

[a] Reference 19. [b] Using current parametrization; this work. [c] $(EtOH) + Zn(SH)_2(CH_2{=}NH) \rightarrow Zn(SH)_2(CH_2{=}NH)\ (EtOH)$. [d] B3LYP/6-311+G(d,p) data from ref 19 (see the supporting materials to the article in ref 19). [e] $CH_3SH + [Zn(SCH_3)]^+ \rightarrow [MeSH-Zn-SMe]^+$. [f] This work. [g] $OH^- + [Zn(HCOO)(CH_2{=}NH)_2]^+ \rightarrow Zn(OH)(HCOO)(CH_2{=}NH)_2$.

group in cysteine, the imidazole group in histidine, and the carboxylate group present in all amino acids, respectively. Figure 10 depicts the geometries of the investigated complexes, whereas Table 6 summarizes their geometrical characteristics calculated with B3LYP and two different SCC-DFTB parametrizations (i.e., the current one and that in ref 19). In comparison with DFT results, both DFTB parametrizations describe the Zn−X (X = S, N, O) reasonably well: Zn−S bond lengths in complex **1** were found to be longer by up to 2% compared to DFT ones, whereas the same bonds in complex **2** are described with an accuracy of 0.8% or better with the current parameters. Similar trends were found in the case of Zn−O bond lengths, with errors around 5 and 2.6% in comparison with DFT values for complexes **1** and **3**, respectively. The nonbonded $Zn-O_3$ distance in complex **3** is also in a good agreement with DFT ones (cf. Table 6). The current parametrization leads to underestimation of the Zn−N bond lengths, as exemplified by both complexes **1** and **3** (within 7 and 5.6% differences to DFT, respectively).

Also, the bond angles are reasonably described with both DFTB parametrization. Our S−Zn−S values in complexes **1** and **2**, for instance, agree within 6.6 and 3.3%, respectively, in comparison with DFT results. It should be noted that some bond angles disagree even up to 10° with DFT, which may indicate a relatively floppy bending potential-energy profile.

Table 7 lists formation energies of complexes under study calculated as the difference between the total energy of the complex and those of reactants. It can be inferred from these data that the current SCC-DFTB parametrization leads to an overbinding of ligands in zinc complexes with the largest error to DFT of 0.3 eV (37%) for the first reaction, and 0.1 eV (4.5%) and 1.2 eV (17.2%) for the second and third reaction, respectively. This finding is, however, in line with the general overbinding trend of the DFTB method, and one can safely conclude that the current parametrization is equally applicable to model zinc in biological environment and solid-state systems.

4. Conclusions

In this work, we presented a new SCC-DFTB parametrization for Zn−X interactions (X = H, C, N, O, S, and Zn), which demonstrated a reliable performance in representing the zinc-containing systems, including bulk phases (hcp-Zn, w-ZnO, and zb-ZnS), ZnO surfaces (clean and with adsorbates), ZnO nanostructures, and model zinc biomolecules. Our results indicate that this new set is universal, being transferable among several different chemical environments, and aims to be a promising computational approach for future in-depth explorations of the properties of complex zinc-containing systems (e.g., ZnO-based hybrid materials or active sites of enzymes).

Acknowledgment. This work was supported by the DFG Priority Program SPP-1165 and by the DIP-40100474. We thank Christof Köhler, Thomas Niehaus, and Simone Sanna for fruitful discussions.

References

(1) Law, M.; Greene, L. E.; Johnson, J. C.; Saykally, R.; Yang, P. *Nat. Mater.* **2005**, *4*, 455–459.

(2) Li, Y.; Meng, G. W.; Zhang, L. D.; Phillipp, F. *Appl. Phys. Lett.* **2000**, *76*, 2011–2013.

(3) Wang, X.; Song, J.; Li, P.; Ryou, D. R.; Dupuis, H. J.; Summers, C. J.; Wang, Z. L. *J. Am. Chem. Soc.* **2005**, *127*, 7920–7923.

(4) Hughes, W. L.; Wang, Z. L. *Appl. Phys. Lett.* **2005**, *86*, 043106.

(5) Wen, X. G.; Fang, Y. P.; Pang, Q.; Yang, C. L.; Wang, J. N.; Ge, W. K.; Wong, K. S.; Yang, S. H. *J. Phys. Chem. B* **2005**, *109*, 15303–15308.

(6) Pan, Z. W.; Dai, Z. R.; Wang, Z. L. *Science* **2001**, *291*, 1947–1949.

(7) Xu, X. Y.; Zhang, H. Z.; Zhao, Q.; Chen, Y. F.; Xu, J.; Yu, P. D. J. *J. Phys. Chem. B* **2005**, *109*, 1699–1702.

(8) Tian, Z. R.; Voigt, J. A.; Mackenzie, B.; Mcdermott, M. *J. Am. Chem. Soc.* **2002**, *124*, 12954–12955.

(9) Lui, R.; Vertegel, A. A.; Bohannan, E. W.; Sorenson, T. A.; Switzer, J. A. *Chem. Mater.* **2001**, *13*, 508–512.

(10) Tian, Z. R.; Voigt, J. A.; Mackenzie, B.; Mcdermott, M.; Rodrigues, M. A.; Konishi, H.; Xu, H. *Nat. Mater.* **2003**, *2*, 821–826.

(11) Taratula, O.; Galoppini, E.; Wang, D.; Chu, D.; Zhang, Z.; Chen, H.; Saraf, G.; Lu, Y. *J. Phys. Chem. B* **2006**, *110*, 6506–6515.

(12) Xu, H.; Zhang, R. Q.; Zhang, X.; Rosa, A. L.; Frauenheim, Th. *Nanotechnology* **2007**, *18*, 485713.

(13) Kohan, A. F.; Ceder, G.; Morgan, D.; Van der Walle, C. G. *Phys. Rev. B* **2000**, *61*, 15019–15027.

(14) Porezag, D.; Frauenheim, Th.; Köhler, Th.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947–12957.

(15) Seifert, G.; Porezag, D.; Frauenheim, Th. *Int. J. Quantum Chem.* **1996**, *58*, 185–192.

(16) Niehaus, Th.; Suhai, S.; DellaSala, F.; Lugli, P.; Elstner, M.; Seifert, G.; Frauenheim, Th. *Phys. Rev. B* **2001**, *63*, 085108.

(17) di Carlo, A.; Gheorghe, M.; Lugli, P.; Sternberg, M.; Seifert, G.; Frauenheim, Th. *Physica B* **2002**, *314*, 86–90.

(18) Hohenberg, P.; Kohn, W. *Phys. Rev. Lett.* **1964**, *136*, B864.

(19) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565–581.

(20) Seifert, G. *J. Phys. Chem. A* **2007**, *111*, 5609–5613.

(21) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, Th.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.

(22) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(23) Soler, J. M.; Artacho, E.; Gale, J. D.; García, A.; Junquera, J.; Ordejón, P.; Sánchez-Portal, D. *J. Phys.: Condens. Matter* **2002**, *14*, 2745–2779.

(24) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993–2006.

(25) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1993**, *37*, 785–789.

(26) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053–1062.

(27) 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.04; Gaussian, Inc.: Wallingford, CT, 2004.

(28) Monkhorst, H. J.; Pack, J. D. *Phys. Rev. B* **1976**, *13*, 5188–5192.

(29) Aradi, B.; Hourahine, B.; Frauenheim, Th. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.

(30) Steinberg, D. J. *J. Phys. Chem. Solids* **1982**, *43*, 1173–1175.

(31) Desgreniers, S. *Phys. Rev. B* **1998**, *58*, 14102–14105.

(32) Ghosh, G.; Delsante, S.; Borzone, G.; Asta, M.; Ferro, R. *Acta Mater.* **2006**, *54*, 4977–4997.

(33) Weiss, R. J.; Mazzone, G. *J. Appl. Crystallogr.* **1981**, *14*, 401–416.

(34) Li, C.; Guo, W.; Kong, Y.; Gao, H. *Phys. Rev. B* **2007**, *76*, 035322.

(35) Jaffe, J. E.; Snyder, J. A.; Lin, Z.; Hess, A. C. *Phys. Rev. B* **2000**, *62*, 1600–1665.

(36) Martins, L. J.; Troullier, N. *Phys. Rev. B* **1991**, *43*, 2213–2217.

(37) Sahraoui, F. A.; Zerroug, S.; Louail, S.; Maouche, D. *Mater. Lett.* **2007**, *67*, 1978–1981.

(38) Koskinen, P.; Hakkinen, H.; Seifert, G.; Sanna, S.; Frauenheim, Th.; Moseler, M. *New J. Phys.* **2006**, *8*, 9.

(39) Meyer, B.; Marx, D. *Phys. Rev. B* **2003**, *67*, 035403.

(40) Claeyssens, F.; Freeman, C. L.; Allan, N. L.; Sun, Y.; Ashfold, N. R.; Harding, J. H. *J. Mater. Chem.* **2005**, *15*, 139–148.

(41) Tusche, C.; Meyerheim, H. L.; Kirschner, J. *Phys. Rev. Lett.* **2007**, *99*, 026102.

(42) Fan, W.; Xu, H.; Rosa, A. L.; Frauenheim, Th.; Zhang, R. Q. *Phys. Rev. B* **2007**, *76*, 073302.

(43) Wander, A.; Harrison, N. M. *J. Phys. Chem. B* **2001**, *105*, 6191–6193.

(44) Nakatsuji, H.; Fukunishi, Y. *Int. J. Quantum Chem.* **1992**, *42*, 1101–1114.

(45) Zapol, P.; Jaffe, J. B.; Hess, A. C. *Surf. Sci.* **1999**, *422*, 1–7.

(46) Wang, Y.; Kovacik, R.; Meyer, B.; Kotsis, K.; Stodt, D.; Staemmler, V.; Qiu, H.; Traeger, F.; Langenberg, D.; Muhler, M.; Wöll, C. *Angew. Chem., Int. Ed.* **2007**, *46*, 5624–5627.

(47) Meyer, B.; Rabaa, H.; Marx, D. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1513–1520.

(48) Law, J. B. K.; Thong, J. T. L. *Nanotechnology* **2008**, *19*, 205502.

(49) Martins, J. B. L.; Longo, E.; Salmon, O. D. R.; Espinoza, V. A. A.; Taft, C. A. *Chem. Phys. Lett.* **2004**, *400*, 481–486.

(50) Ozawa, K.; Hasegawa, T.; Edamoto, K.; Takahashi, K.; Kamada, M. *J. Phys. Chem. B* **2002**, *106*, 9380–9386.

(51) Barrientos, C.; Retondo, P.; Largo, A. *J. Chem. Theory Comput.* **2007**, *3*, 657–664.

CT800455A

J. Chem. Theory Comput. **2009,** *5,* 615–626

JCTC Journal of Chemical Theory and Computation

Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins

Andreas Kukol*

School of Life Sciences, University of Hertfordshire, Hatfield AL10 9AB, United Kingdom

Received August 21, 2008

Abstract: United-atom force fields for molecular dynamics (MD) simulations provide a higher computational efficiency, especially in lipid membrane simulations, with little sacrifice in accuracy, when compared to all-atom force fields. Excellent united-atom lipid models are available, but in combination with depreciated protein force fields. In this work, a united-atom model of the lipid 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine has been built with standard parameters of the force field GROMOS96 53a6 that reproduces the experimental area per lipid of a lipid bilayer within 3% accuracy to a value of 0.623 ± 0.011 nm^2 without the assumption of a constant surface area or the inclusion of surface pressure. In addition, the lateral self-diffusion constant and deuterium order parameters of the acyl chains are in agreement with experimental data. Furthermore, models for 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-glycerol (POPG) result in areas per lipid of 0.625 nm^2 (DMPC), 0.693 nm^2 (POPC), and 0.700 nm^2 (POPG) from 40 ns MD simulations. Experimental lateral self-diffusion coefficients are reproduced satisfactorily by the simulation. The lipid models can form the basis for molecular dynamics simulations of membrane proteins with current and future versions of united-atom protein force fields.

Introduction

Proteins associated with lipid membranes constitute approximately 30% of genomes[2] and are estimated to form more than half of all drug targets.[3] However, because of well-known experimental problems in membrane protein structure determination, the number of atomic-resolution structures of soluble proteins far exceeds the number of membrane protein structures. Therefore, membrane proteins have been a popular target for computational modeling[4] as well as model building based on limited experimental data.[5] Additionally, the slowly growing repository of high-resolution membrane protein structures[6] forms a resource for atomic-level studies of function and conformational change. As membrane protein simulations with explicit solvents require in excess of 30 000 atoms, molecular dynamics (MD) simulations based on classical mechanics are usually applied.[7,8] A key ingredient of MD simulations is the force field composed of a set of mathematical functions and parameters that describe all of the bonded and nonbonded interactions between particles. Widely used biomolecular force fields are Amber,[9] CHARMM,[10] OPLS-AA,[11] and GROMOS,[12] the first three of which are all-atom force fields, whereas GROMOS is a united-atom force field that subsumes nonpolar hydrogen atoms into their adjacent carbon atom. While the accuracy of the force fields for protein modeling is comparable,[13] the united-atom approach poses a distinct advantage for membrane systems, reducing the number of particles by up to 60%; for example, a united-atom model of 1,2-dipalmitoyl-*sn*-glycero3-phosphocholine (DPPC) is composed of 50 particles, while the corresponding all-atom model consists of 130 particles. Accurate force field parameters and lipid models are available for the all-atom CHARMM force field.[14,15] Additionally, lipid models for the all-atom General Amber Force Field (GAFF) have been developed recently[16,17] that required the incorporation of surface tension in order to achieve the experimental lipid surface area. For the CHARMM force field, a united-atom model of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) is available[18] that was

---

* Author e-mail: a.kukol@herts.ac.uk.

10.1021/ct8003468 CCC: $40.75

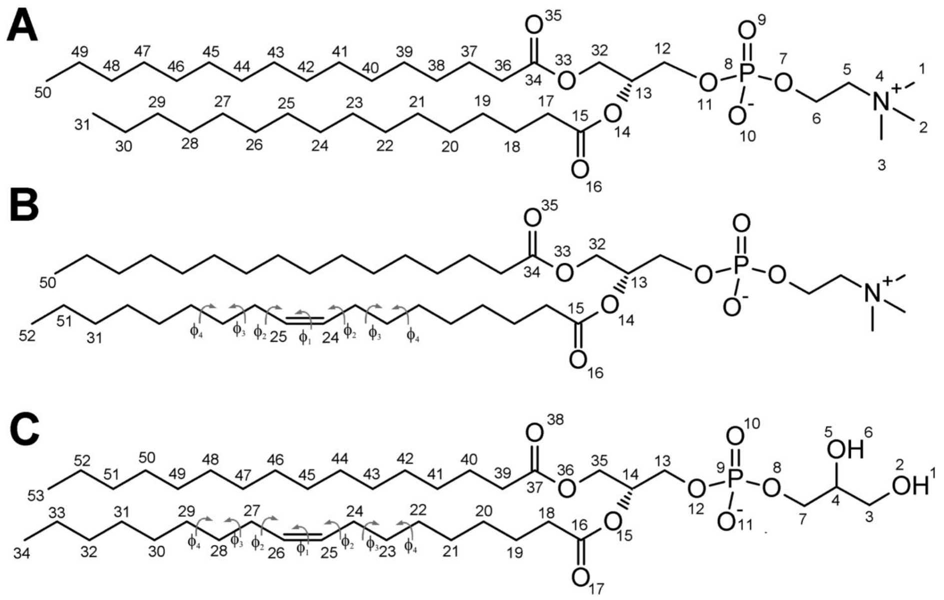


Figure 1. Definition of the atom numbers and torsion angles for the lipid models DPPC (A), POPC (B), and POPG (C).

used with constant area simulations in order to reproduce the experimental area per lipid ratio. Most united atom models of lipid molecules are based on the parameters developed by Berger et al.[19] within the united atom version of the OPLS force field.[20] Lipid models based on the Berger et al. parameters have been transferred to the united-atom GROMOS87 force field and used successfully for many lipid membrane MD simulations and membrane proteins. The united atom models for various lipids have successfully reproduced the most widely cited experimental data of area per lipid molecule within 2% accuracy as well as self-diffusion constants and order parameters for lipid bilayers without inclusion of surface tension or constant area simulations. Since GROMOS87,[21] the GROMOS force field has undergone significant revisions, with GROMOS96 in 1996[22] up to the latest versions 53a5 and 53a6 of the GROMOS96 force field.[12] The model of the DPPC lipid included in the latest 53a6 and previous versions of the force field distribution has unfortunately failed to reproduce the experimental parameters of DPPC membranes to satisfactory accuracy, as shown by various studies[23,24] as well as in this report. This has led to the paradoxical situation where many recent united-atom membrane protein simulations have used the GROMOS87 force field with lipid parameters based on Berger et al.[25−28] The problem with this approach is that MD simulations reproduce lipid bilayer attributes very well, while the membrane protein at the focus of interest may not be treated with comparable accuracy to other modern protein force fields. Some studies used a combination of the united-atom Berger et al. parameters and the all-atom OPLS force field, but the combination of different force fields is not straightforward and requires considerable care.[29] Ideally, lipid models should be developed with the same atom types, bonded and nonbonded parameters that are applied to the protein.

In order to support molecular dynamics simulations focused on membrane proteins, this work reports the development and evaluation of models of DPPC, 1,2-dimyristoyl- *sn*-glycero-3-phosphocholine (DMPC), POPC, and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol (POPG) with GROMOS96 53a6 force field parameters that reproduce the experimental area/lipid ratio to a sufficient accuracy, in cases where experimental data for comparison is available. The lipid bilayer coordinates after 40 ns of MD simulations as well as the lipid models (topologies) are available as Supporting Information.

Methods

Initial Structures. Initial coordinates of a 128-lipid DPPC bilayer hydrated with 3655 water molecules based on 2 ns equilibration,[30] a 128-lipid DMPC bilayer[31] with 3655 water molecules, and a POPC bilayer[32] with 2460 water molecules have been obtained from P. Tielman's Web site (http://moose.bio.ucalgary.ca). The initial coordinates of a 128-molecule POPG bilayer neutralized with 128 Na^+ counterions and hydrated with 3527 water molecules[33] have been obtained from M. Karttunen's Web site (http://www.apmaths.uwo.ca/∼mkarttu). The racemic POPG bilayer was composed of equal numbers of L-POPG and D-POPG.

Molecular Topologies. For DPPC simulations, three different models (topologies) were investigated on the basis of the topology included in the GROMOS96 53a6 force field distribution. DPPC1 uses the original topology file without any alterations. DPPC2 has been modified with the partial charges of the lipid headgroup due to Chiu et al.,[34] with a subdivision into four charge groups as suggested by Chandrasekhar et al.[24] (model C in that publication). DPPC3 uses the partial charges model of DPPC2 with a different ester-carbonyl carbon atom type, “CH0” instead of “C” as suggested previously.[24] This resulted in a van der Waals radius for the carbonyl-ester of 0.664 nm for atom type “CH0”, as opposed to 0.336 nm for atom type “C”. The atom numbering scheme of all lipid models is shown in Figure 1.

The DMPC model was based on DPPC3, with two united-atom CH2 groups less in each aliphatic carbon chain.

The POPC1 model adopted the same partial charges distribution for the lipid headgroup and torsion potentials as

Table 1. POPC Atom and Bond Parameters for the C=C Double Bond[a]

atom 1	atom 2	atom 3	atom 4	code	remark
24				CR1	
25				CR1	
24	25			gb_10	double bond
23	24	25		ga_27	
24	25	26		ga_27	
23	24	25	26	gi_1	improper dihedral

[a] For the definition of atom numbers, see Figure 1b. The meaning of the codes is given in the publication describing the GROMOS96 53a6 force field.[12]

Table 2. Atom Types and Charges for the POPG Head Group[a]

atom no.	code	charge
1	H	0.4170
2	OA	−0.5740
3	CH2	0.1570
4	CH1	0.1570
5	OA	−0.5740
6	H	0.4170
7	CH2	0.4000
8	OE	−0.8000
9	P	1.7000
10	OM	−0.8000
11	OM	−0.8000
12	OE	−0.7000
13	CH2	0.4000
14	CH1	0.3000
15	OE	−0.7000
16	CH0	0.7000
17	O	−0.7000

[a] For the definition of atom numbers, see Figure 1c. The meaning of the codes is given in the publication describing the GROMOS96 53a6 force field.[12]

Table 3. Bond and Bond Angle Parameters for the POPG Head Group[a]

atom 1	atom 2	atom 3	code
1	2		gb_1
2	3		gb_13
3	4		gb_27
4	5		gb_13
5	6		gb_1
4	7		gb_27
7	8		gb_18
8	9		gb_28
9	10		gb_24
9	11		gb_24
9	12		gb_28
12	13		gb_18
14	15		gb_27
15	16		gb_10
16	17		gb_5
1	2	3	ga_12
2	3	4	ga_13
3	4	5	ga_13
4	5	6	ga_12
5	4	7	ga_13
3	4	7	ga_13
4	7	8	ga_13
4	7	8	ga_15
7	8	9	ga_26
8	9	10	ga_14
8	9	11	ga_14
8	9	12	ga_5
9	12	13	ga_26
10	9	11	ga_29
10	9	12	ga_14
11	9	12	ga_14
12	13	14	ga_15
13	14	15	ga_13
13	14	35	ga_13
14	15	16	ga_22
14	35	36	ga_15
15	14	35	ga_13
15	16	17	ga_31
15	16	18	ga_16
16	18	19	ga_15

[a] For the definition of atom numbers, see Figure 1c. The meaning of the codes is given in the publication describing the GROMOS96 53a6 force field.[12]

DPPC3, while different bonded parameters were used in order to model the double bond, as shown in Table 1. The model POPC2 contains additional torsion potentials in the vicinity of the double bond (shown in Figure 1B), which have been developed previously by Bachar et al.[35] on the basis of ab initio calculations. The potentials included are

$$V(\varphi_2) = -5.865 + 7.470[1 + \cos(\varphi_2 - 180°)] + 3.99[1 + \cos(2\varphi_2)] + 1.1[1 + \cos(3\varphi_2 - 180°)]$$

$$V(\varphi_3) = 3.35[1 + \cos(\varphi_3)] - 1.66[1 + \cos(2\varphi_3 - 180°)] + 7.333[1 + \cos(3\varphi_3)]$$

For computational efficiency, the torsion potential $V(\phi_1)$ has been left unchanged as a standard Gromos96[12] improper torsion potential, and $V(\phi_4)$ is described as Gromos96 dihedral type 34. It has been shown that the dihedral angle distributions for ϕ_1 and ϕ_4 with Bachar et al. potentials are similar to the situation where standard potentials are used.[36]

The POPG model provided by Karttunen et al.[33] (for the two steroisomers D-POPG and L-POPG) has been translated into the GROMOS96 53a6 topology, while the partial charges for the glycerol headgroup have been retained. Atom types and charges are shown in Table 2. Bond and bond angle parameters are shown in Table 3, and torsion potientials for the lipid headgroup are shown in Table 4. The carbon chains of POPG1 have been modeled in the same way as POPC1, while for POPG2, the Bachar et al.[35] torsion potentials have been included as described above for POPC2. A POPG3 model has been generated that is indentical to POPG1, apart from the atom type "C" for the ester carbonyl atom, as used before in DPPC1 and DPPC2. The topologies are available as Supporting Information.

ErbB2 Transmembrane Domain Simulation. The starting coordinates of the ErbB2 transmembrane domain sequence GCPAEQRASPLTSIISAVVGILLVVVLGVVFGILIKRRQQKIRK were obtained from the RCSB Protein Data Bank[37] (PDB ID: 2JWA). This structure was originally obtained by NMR spectroscopy from the peptide in 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine/DMPC bicelles at 313 K.[38] For the MD simulation, the peptide was inserted into a pre-equilibrated DMPC lipid bilayer using a series of scaling and energy minimization steps, as described by Kandt et al.[39] Chloride ions were added at positions of most favorable electrostatic potential in order to neutralize the system. The resulting peptide/lipid/ions/water system contained 103 DMPC molecules, 12 chloride ions, and 5749 water molecules. MD simulations were carried out with the

Table 4. Torsion Potentials for the POPG Head Group[a]

atom 1	atom 2	atom 3	atom 4	code
1	2	3	4	gd_23
6	5	4	7	gd_23
5	4	7	8	gd_18
2	3	4	5	gd_18
2	3	4	7	gd_33
2	3	4	7	gd_17
3	4	5	6	gd_23
3	4	7	8	gd_33
3	4	7	8	gd_17
4	7	8	9	gd_29
7	8	9	12	gd_20
7	8	9	12	gd_27
8	9	12	13	gd_20
8	9	12	13	gd_27
9	12	13	14	gd_29
12	13	14	15	gd_34
12	13	14	35	gd_34
12	13	14	35	gd_17
13	14	35	36	gd_34
13	14	35	36	gd_17
13	14	15	16	gd_29
14	35	36	37	gd_29
14	15	16	18	gd_13
15	14	35	36	gd_18
15	16	18	19	gd_40

[a] For the definition of the atom numbers, see Figure 1c. The meaning of the codes is given in the publication describing the GROMOS96 53a6 force field.[12]

Table 5. Overview of Simulations Performed[a]

model	features
DPPC1	original model included in GROMOS96 53a6 force field
DPPC2	based on DPPC1, different partial charges of the headgroup atoms
DPPC3	based on DPPC2, larger carbonyl-ester atoms, atom type "CH0"
DMPC	same as DPPC3 with shorter acyl chains
POPC1	based on DPPC3, with unsaturated chain
POPC2	based on POPC1, changed torsion potentials next to double bond
POPG1	based on POPC1, with glycerol headgroup
POPG2	based on POPC2, with glycerol headgroup
POPG3	based on POPC1, smaller carbonyl-ester type "C" (data not shown)
ErbB2/DMPC	ErbB2 transmembrane peptide with GROMACS force field and Berger lipids
ErbB2/**DMPC**	ErbB2 transmembrane peptide and lipids with GROMOS96 53a6 force field

[a] Models in boldface are recommended and included in the Supporting Information.

GROMACS force field based on GROMOS87 and Berger et al.[19] lipid parameters and, in addition, with the GROMOS96 53a6 force field and the DMPC model developed in the present study. The lipid and water molecules were allowed to equilibrate around the peptide during a 10 ns MD simulation at 314 K with position restraints applied to the peptide atoms. The position restraint MD simulation was followed by unrestrained MD simulations over 20 ns. All simulations performed are summarized in Table 5.

Simulation Parameters. All simulations were performed with the GROMACS simulation package,[40−42] version 3.3.2. The NpT ensemble and periodic boundary conditions were used, while the temperature was kept constant with the Berendsen thermostat[43] at 325 K for all DPPC simulations, at 314 K for DMPC, and at 298 K for POPC and POPG, with a coupling time constant of 0.1 ps. We chose the temperature such that it was at least 10 K above the gel to liquid-crystalline phase transition temperature. The phase transition temperature of POPC and POPG is below 273 K; therefore, the standard thermodynamic temperature of 298 K was chosen. Lipid and water molecules were coupled separately to the thermostat; in the case of POPG, water molecules and 128 sodium ions were jointly coupled to the thermostat. Semianisotropic Berendsen pressure coupling[43] was applied with separate coupling to the *z* direction (the bilayer normal) and the *xy* plane with a coupling time constant of 2.0 ps in order to maintain a constant pressure of 1.0 bar. For Lennard-Jones interactions, a cutoff at 1.4 nm was applied, while electrostatic interactions were treated with the particle mesh Ewald (PME) method[44,45] and a real-space cutoff of 0.9 nm. A larger than usual Lennard-Jones cutoff was used in order to increase the accuracy of the van der Waals interactions. Electrostatic interactions were treated with the PME method, which does not introduce artificial ordering like cutoff methods.[46] The real-space cutoff of 0.9 nm used with the PME method is merely a numerical device in order to separate the direct- and reciprocal-space sums. Long-range dispersion corrections for energy and pressure were applied. All systems were first subjected to 1000 steps of energy minimization, followed by a 40 ns MD simulation with a time step of 2 fs. The lipid molecule bonds were constrained with the LINCS algorithm,[47] while the water molecules were constrained with the SETTLE algorithm.[48] All simulations were carried out on a dual-processor dual-core AMD Opteron 2.66 GHz workstation (DNUK, Ashton-under-Lyne, United Kingdom).

Data Analysis. The area per lipid molecule was calculated from the lateral *x* and *y* dimensions of the simulation box divided by the number of lipid molecules in one leaflet of the bilayer. The reported average area per lipid was taken over the time from 20 to 40 ns.

Lateral diffusion coefficients for lipid molecules were calculated from the mean square displacement (MSD) of the center of mass $\langle (r(t+\tau)\, r(\tau))^2 \rangle$, where $\langle\rangle$ refers to the average taken over all starting times τ and $r(t)$ to the position of the center of mass at time t. The lateral diffusion coefficient is then given by the Einstein relation MSD $= 4D_{\mathrm{lat}}t$ in two dimensions for long times t and is averaged over 128 lipid molecules. In order to obtain the long-range diffusion coefficients, the MSD was fitted between 20 and 38 ns. Calculations have been performed with the g_msd program of the GROMACS suite. Diffusion constants were not further corrected for random motions of the lipid monolayer, as it was found to be insignificant under periodic boundary conditions.

The deuterium order parameter S_{CD} for the carbon tails is calculated from the elements of the order parameter tensor $S_{xx} = 1/2\langle(3\cos^2\alpha_{\mathrm{i}} - 1)/2\rangle$ and S_{yy} as $S_{\mathrm{CD}} = 2/3S_{xx} + 1/3S_{yy}$. The angle α_i is the angle between the molecular axis given by the carbon atoms C_{i-1} and C_{i+1} and the lipid bilayer normal; the average is taken over the time of 20−40 ns and for all lipid molecules. Calculations have been performed with the g_order program of the GROMACS suite.

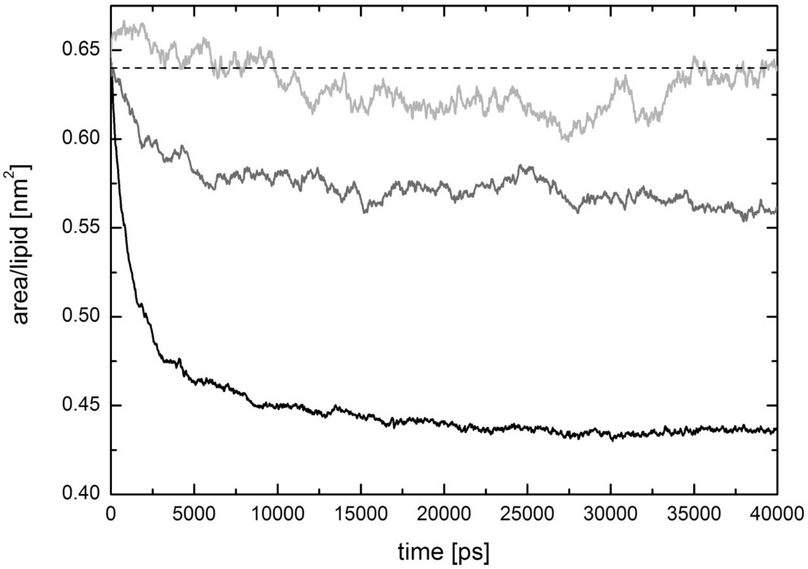


Figure 2. Area per lipid ratio dependent on MD simulation time for a 128-molecule lipid bilayer obtained with various lipid models: DPPC1 (black curve), DPPC2 (dark gray), and DPPC3 (light gray). The lipid models are explained in the text. The dashed line indicates the currently accepted experimental value for the area per lipid ratio.[50]

The radial distribution function of water hydrogens averaged between 20 and 40 ns around the ester carbonyl atoms of the *sn1* and *sn2* chains was calculated with the g_rdf program of the GROMACS suite.

The angles between the P−N vector of the lipid headgroups and the z axis were calculated with a Perl script developed in-house.

Results

DPPC. Lipid bilayers with three different DPPC models have been subjected to 40 ns MD simulation, and the time evolution of the area per lipid molecule has been calculated. The area per lipid molecule is the most widely used parameter for the characterization of lipid bilayers, as it is related to various other properties of the membrane like the lateral diffusion, membrane elasticity, or permeation. Figure 2 shows the area per lipid dependent on time for three different DPPC models. From the curves in Figure 2, the equilibration time of the simulation has been estimated qualitatively as less than 20 ns; thus, the period between 20 and 40 ns has been used to calculate average properties. For DPPC1, the original model included in the GROMOS force field distribution, the area per lipid reached an average of 0.436 ± 0.002 nm^2. Model DPPC2, which has alternative headgroup charges according to Chiu et al.,[34] reached an average area per lipid of 0.568 ± 0.007 nm^2, while the change of the ester carbonyl atom type from "C" to "CH0" in model DPPC3 led to an average area per lipid of 0.623 ± 0.011 nm^2, which is within a range of 3% of the experimental value of around 0.64 nm^2.[49,50]

The deuterium order parameters $|S_{CD}|$ dependent on the carbon atom number along the hydrocarbon tails show excellent agreement with experimental data[51] (Figure 3). There is some deviation from the experimental results for carbon-2 of the *sn1* chain, which corresponds to atom number 36 in Figure 1A.

The interactions between lipid and water molecules have been analyzed by calculating the radial distribution function of water hydrogens around the carbonyl-ester atoms for the *sn1* and *sn2* chains of DPPC (Figure 4). It is obvious from prominent peaks of the radial distribution function that water effectively solvates the carbonyl-oxygen atoms of both chains in the DPPC3 model (solid line). The more water-exposed carbonyl atom of the *sn2* chain achieves a stronger solvation, as indicated by the higher peak amplitude of the radial distribution function. In clear contrast, the DPPC1 model (dotted line) shows no significant solvation by water hydrogen atoms.

The molecular basis of these results can bee seen in the snapshot of the simulation trajectory of a lipid bilayer constructed with DPPC1 (Figure 5A) and that constructed with DPPC3 (Figure 5B), both taken at 40 ns. The DPPC3 bilayer shows water penetrating into the lipid headgoup region, and the lipid acyl chains seem to be in a clearly disordered liquid-crystalline state, while acyl chains are highly ordered in the case of DPPC1 (Figure 5A). The same liquid-crystalline state is adopted by DPPC2 (not shown).

The angle between the lipid headgroup dipole moment as defined by the P−N vector and the z axis is 90° ± 2 for DPPC1, 91° ± 3 for DPPC2, and 90° ± 2 for DPPC3, averaged over all lipid molecules during the last 20 ns of the simulation. This is similar to the starting structure obtained from a previous simulation,[30] which yields an angle of 89° ± 2 between the P−N vector and the z axis averaged over 128 lipid molecules.

The lateral diffusion of lipid molecules is characterized by fast fluctuations in the limited space provided by surrounding lipids as well as by long-range diffusion throughout the lipid bilayer leaflet. The long-range diffusion coefficient for DPPC3 given in Table 6 shows good agreement with experimental data obtained from the literature.

DPMPC, POPC, and POPG. Lipid models (topologies) for DMPC, POPC, and POPG are not provided with the

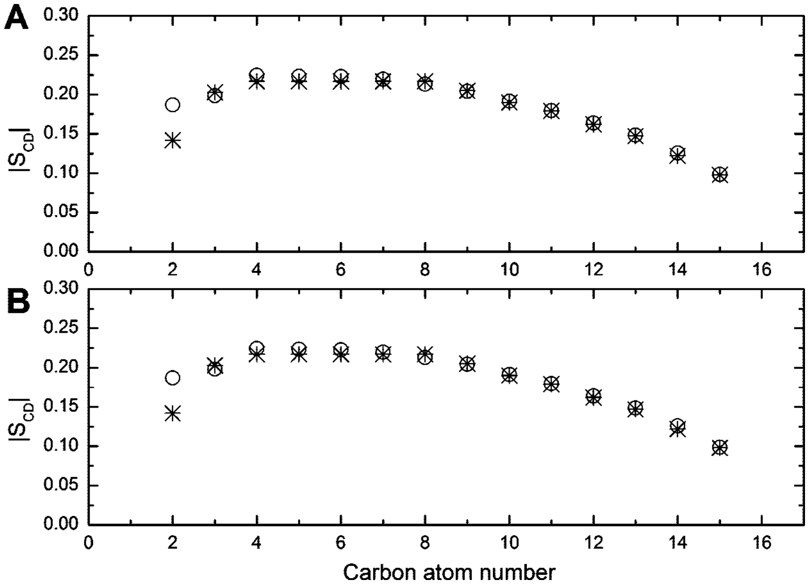


Figure 3. Comparison of the deuterium order parameters along the carbon atoms of the lipid acyl tails from simulation (○) and those from experimental results (*). (A) The *sn1* chain. (B) The *sn2* chain. The experimental values are from the *sn2* chain in both parts.

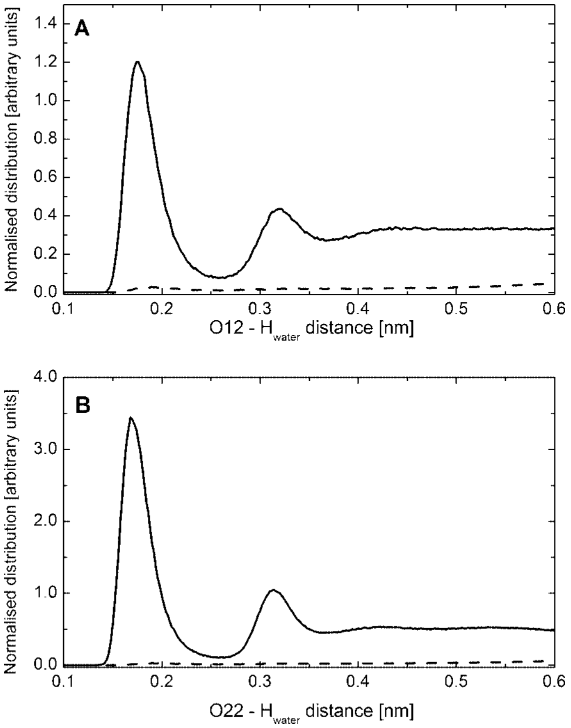


Figure 4. The radial distribution function of water hydrogens around the carbonyl-ester atoms for the *sn1* (A) and *sn2* chains (B) of DPPC. The radial distribution function has been normalized with respect to density and volume. The solid line represents the data obtained from simulation with the DPPC3 model, while the dotted line has been obtained with the DPPC1 model. The average between 20 and 40 ns has been taken.

GROMOS force field distribution. On the basis of the DPPC3 model, the DMPC model has been created by reducing the number of carbon atoms from 16 to 14 in each acyl chain and leaving all other parameters unchanged. The change of area per lipid ratio with simulation time shown in Figure 6 reveals that the bilayer is equilibrated almost immediately, because the DMPC bilayer based on Berger et al.[19] parameters has been used as the initial structure. The area per lipid averaged over the last 20 ns of the trajectory reaches a value of 0.625 ± 0.005 nm^2, and the lipid lateral diffusion constant is $5.7 \times 10^{-8} \pm 0.3$ cm^2 s^{-1} (Table 6).

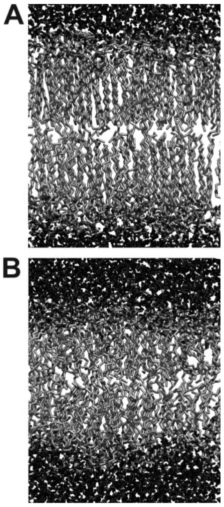


Figure 5. Snapshot of the MD simulation trajectory at 40 ns with the lipid models DPPC1 (A) and DPPC3 (B). The figure was created with VMD.[1]

Two models of POPC have been evaluated. The POPC1 model is based on DPPC3, while different atom types, bond types, and torsion potentials (Table 1) have been used in order to model the double bond. In POPC2, the torsion potentials for bonds adjacent to the double bond (ϕ_2 and ϕ_3 in Figure 1B) have been replaced with potentials developed by Bachar et al.[35] The area per lipid ratio equilibrates for both models after 5 ns and reaches an average value of 0.654 ± 0.008 nm^2 for POPC1 and 0.693 ± 0.005 nm^2 for POPC2, which is closer to the experimental value of 0.683 nm^2 (Table 6). The lateral diffusion constant of POPC1 is $3.0 \times 10^{-8} \pm$

Table 6. Comparison of Area Per Lipid, A_{sim}, and Lateral Self Diffusion Coefficients D_{sim} Averaged from Trajectories between 20 and 40 ns for Various Lipid Models to Experimental Data (A_{exp}, D_{exp})

lipid	A_{sim} [nm^2]	A_{exp} [nm^2]	D_{sim} [10^{-8} cm^2 s^{-1}]	D_{exp} [10^{-8} cm^2 s^{-1}]
DPPC1	0.436 ± 0.002	0.64[a]	3.0 ± 0.3	9.7
DPPC2	0.568 ± 0.007	0.64[a]	8.2 ± 0.6	9.7
DPPC3	0.623 ± 0.011	0.64[a]	6.8 ± 0.4	9.7[d]
DMPC	0.625 ± 0.005	0.606[b]	5.7 ± 0.3	14.3[e]
POPC1	0.654 ± 0.008	0.683[c]	3.0 ± 0.1	8.87
POPC2	0.693 ± 0.005	0.683[c]	7.4 ± 0.3	8.87[e]
POPG1	0.635 ± 0.008	N/A[f]	7.0 ± 0.2	3.0[g]
POPG2	0.700 ± 0.007	N/A[f]	7.1 ± 0.5	3.0[g]

[a] Reference 50. [b] Reference 57. [c] Reference 64. [d] Reference 65 (at 333 K). [e] Reference 52. [f] 0.64 nm^2 for PG prepared from egg Lecithin,[59] 0.66 nm^2 for a DPPC/DMPG (50:50) monolayer.[66] [g] Reference 67 (obtained from NBD-PG in a supported planar bilayer composed of 22% POPC and 76.5% POPG).

0.1 cm^2 s^{-1} and is 7.4 × 10^{-8} ± 0.3 cm^2 s^{-1} for POPC2, which is closer to the experimental value of 8.9 × 10^{-8} cm^2 s^{-1}.[52]

The POPG1 model uses the same parameters for the acyl chains as those of POPC1, while the parameters of the headgroup are shown in Tables 2−4. The charges of the headgroup atoms are the same as those used by Zhao et al.[33] The POPG2 model includes the torsion potentials for bonds adjacent to the double bond analogous to POPC2. The area per lipid ratio after 20 ns reaches an average value of 0.635 ± 0.008 nm^2 for POPG1 and 0.700 ± 0.007 nm^2 for POPG2, while the lateral diffusion coefficient is 7.0 × 10^{-8} ± 0.2 cm^2 s^{-1} for POPG1 and 7.1 × 10^{-8} ± 0.5 cm^2 s^{-1} for POPG2.

In general, the area per lipid values of DPPC3, DMPC, and POPC2 averaged from the trajectories between 20 and 40 ns show excellent agreement with experimental results (Table 6). The lateral diffusion constants are of the right order of magnitude, when compared to experimental values (Table 6), although it should be noted that the experimental conditions were sometimes different from the simulation, for example, using fluorescently labeled POPG molecules in order to measure the diffusion constant.

Since the area per lipid for POPG simulations is different from those of other simulation studies (see Discussion), an analysis of the sodium ion distribution has been conducted. The distribution of sodium ions with respect to the bilayer normal (z axis) is shown in Figure 7A for the POPG2 bilayer. Sodium ions penetrate deeply into the lipid headgroup region; the peak values of the sodium atom density are reached at a position deeper into the lipid bilayer than the peak values of the phosphorus atom density.

Sodium atoms are found hexagonally coordinated with the involvement of one or two lipid molecules, while the free coordination sites are occupied by water molecules (Figure 7B). In the bulk water, sodium ions are coordinated hexagonally and pentagonally (Figure 7B). The ester-carbonyl oxygens of the lipid molecules participate in sodium coordination as well as the oxygen atoms of glycerol hydroxyl groups or oxygen atoms of the phosphate ester group (P−O−C). There is no example of the participation of terminal oxygens of the phosphate groups in sodium coordination. Similar findings have been reported by Zhao et al.,[33] albeit with a much smaller area per lipid ratio.

Transmembrane Peptide Simulation. Since the ultimate aim was to use the lipid models for united-atom membrane protein simulations, the ErbB2 transmembrane domain has been subjected to 20 ns MD simulation using the established GROMOS87/Berger lipids force field combination as well as the GROMOS96 53a6 force field. Visual analysis of the trajectory showed that both simulations maintained the secondary structure and the integrity of the transmembrane helix dimer. Structural fluctuations occurred mainly in the unstructured regions outside the lipid bilayer. The root-mean-square deviation (rmsd) of the protein backbone with respect to the starting structure shown in Figure 8 increased over the first 1000 ps to a value in the region of 0.33 nm. Both force fields behaved initially identically until 7000 ps, where the rmsd values began to diverge. The backbone rmsd averaged between 10 and 20 ns was 0.31 nm ± 0.02 for the GROMOS96 53a6 force field and 0.42 nm ± 0.02 for the GROMOS87 force field. A similar divergence was observed for MD trajectories with different random initial atom velocities.

Discussion

The development and evaluation of various lipid models with standard GROMOS96 53a6 force field parameters has been shown. For DPPC, the most studied lipid molecule, the important experimental area per lipid ratio has been reproduced with the DPPC3 model to a reasonable accuracy of 3% by simulations. The initial improvement of the area per lipid ratio seen in Figure 2 on going from model DPPC1 to DPPC2 is caused by the increased penetration of water into the lipid bilayer headgroup region due to the higher partial charges on the headgroup atoms in the DPPC2 model, as reported before.[23] The increased water penetration can be seen clearly from the comparison of the radial distribution functions of water around lipid carbonyl-ester (Figure 4) between DPPC1 and DPPC3 (the same for DPPC2, not shown). Additionally the dipole−dipole interaction between the lipid headgroups may affect the area per lipid, as discussed by Wohlert and Edholm.[55] The in-plane component of the headgroup dipole contributes to attractive forces, while the components perpendicular to the membrane plane repel each other. The angle between the P−N vector and the z axis was found to be around 90° in all cases, which would contribute to attractive forces. The dipole moment of the headgroup is similar for DPPC1 and DPPC2, because the sum of the charges on the phosphate group (−1) and the choline group (+1) is the same for both charge schemes. It is due to the larger atomic charges of the DPPC2 headgroup that water penetrated into the lipid headgroup region and thus overcame parts of the attractive forces. Additionally, the use of the PME scheme for the treatment of electrostatics renders the system less vulnerable to electrostatic artifacts introduced by the cutoff scheme.[55]

The further increase of the area per lipid in the DPPC3 model is caused by the larger radius of the ester carbon of

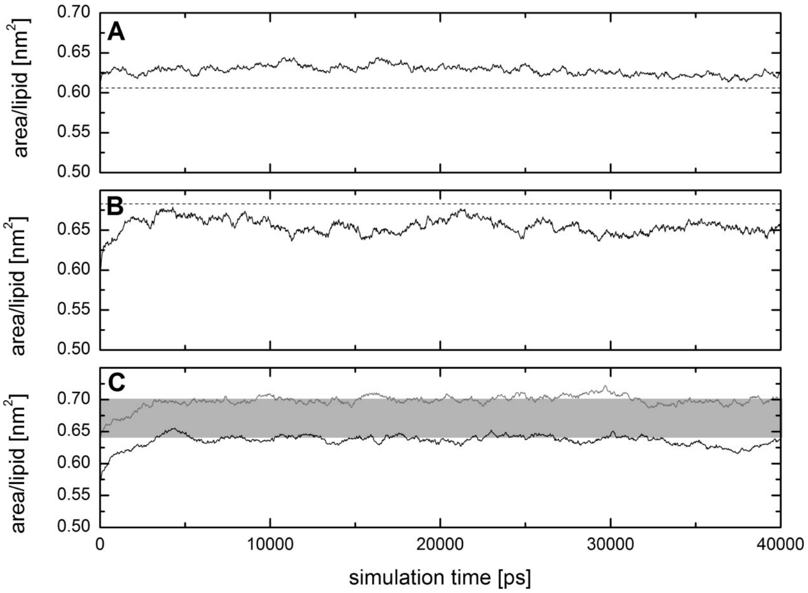


Figure 6. Area per lipid ratio dependent on MD simulation time for a 128-molecule lipid bilayer composed of (A) DMPC, (B) POPC2, and (C) POPG1 (dark curve) and POPG2 (gray curve). The dashed lines indicate the experimental value, and the gray shaded area indicates a probable range of experimental area per lipid ratios for POPG.

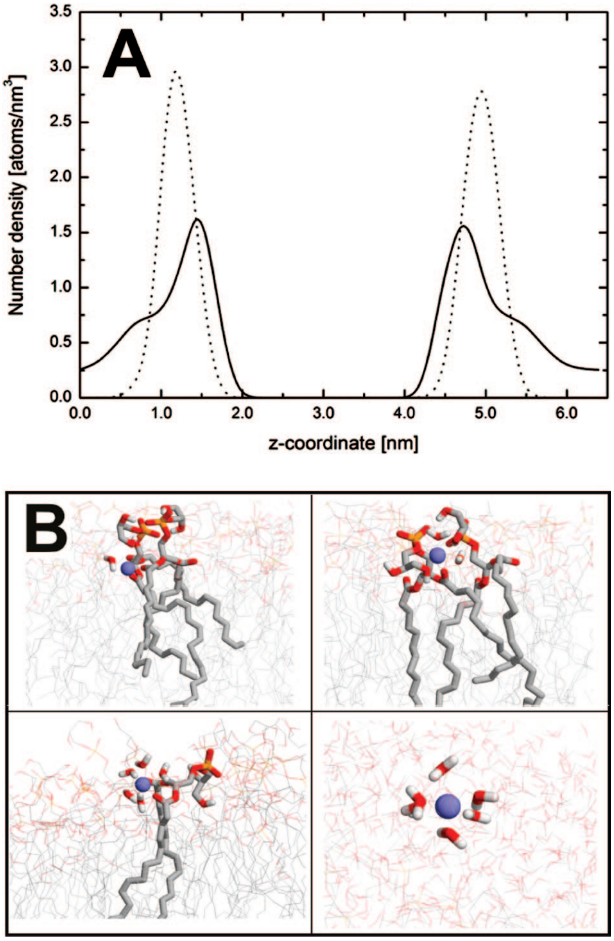


Figure 7. Characteristic features of the POPG bilayer. Panel A shows the atom number density of sodium (solid line) and phosphorus (dotted line) dependent on the *z* coordinate, the normal to the bilayer plane. Panel B shows snapshots from the MD trajectory taken at 38 ns of various modes of sodium interactions with lipid and water molecules. Panel B was created with Rasmol.[53,54]

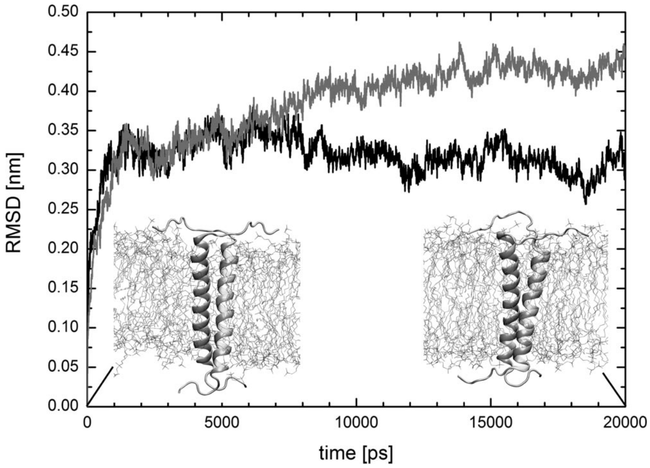


Figure 8. Root-mean-square deviation (rmsd) of the backbone atoms with respect to the experimental structure for the MD simulation of the ErbB2 transmembrane peptide dimer in a DMPC lipid membrane with the GROMOS96 53a6 force field (black curve) and the GROMACS force field (gray curve). The insets show the experimental structure (left) and a snapshot of the MD trajectory taken at 20 000 ps from the GROMOS96 simulation (right).

0.664 nm in the DPPC3 model compared to the 0.336 nm chosen in the DPPC1 and DPPC2 models. The molecular picture of Figure 5 reveals that additionally the disorder of the acyl chains is markedly increased for the DPPC3 model, as would be expected for a liquid-crystalline phase. The addition of further torsion potentials to the DPPC3 model did not yield any improvement in the area per lipid ratio, but a slight decrease to 0.61 nm^2/lipid was seen (data not shown). Furthermore, the experimental deuterium order parameters and the diffusion coefficients are well reproduced by the DPPC3 model (Figure 3 and Table 6). It can be concluded that the DPPC3 model with an area per lipid of 0.623 ± 0.011 nm^2 achieves high accuracy compared to experimental data. The results are in line with other studies based on the Berger et al.[19] model, which achieve even higher accuracy, for example, 0.645 nm^2 in a study by Patra et al.[46] or 0.655 nm^2 in another recent study.[56] The advantage of the lipid model presented here is that it is fully compatible with the recent GROMOS96 53a6 force field. Therefore, we recommend the DPPC3 lipid model for united-atom membrane protein MD simulations.

The lipid bilayer based on the DMPC model showed an equally satisfactory reproduction of experimental parameters.

The increased area per lipid ratio of the POPC1 lipid bilayer is most likely caused by the increased lipid bilayer packing disorder introduced by the cis double bond in the lipid acyl chain. The inclusion of more realistic torsion potentials for the bonds adjacent to the double bond in POPC2 further increases the area per lipid to 0.693 ± 0.005 nm^2, which is within 1.5% of the experimental result of 0.683 nm^2.[57] This is the best reproduction of the area per lipid ratio for POPC in the literature so far, to our knowledge. The all-atom lipid models using the CHARMM force field required constant area simulations or the application of surface pressure,[58] similar to a recent DOPC model using GAFF.[17] A recent simulation of 128 POPC[33] using the united-atom parameters of Berger et al.[19] achieved 0.658 nm^2.

The area per lipid for the POPG1 model reaches 0.635 and 0.700 nm^2 for POPG2. On the basis of the improved agreement between simulation and experimental results of the area per lipid for POPC2 due to the inclusion of modified torsion potentials, POPG2, which includes the same torsion potentials as POPC2, is chosen as the preferred model. There are no experimental data of the area per lipid ratio for pure POPG lipid membranes in the liquid-crystalline state available, to our knowledge. Phosphatidyl-glycerol of various chain lengths and compositions prepared from egg lecithin yielded an area per lipid of 0.64 nm^2 using X-ray diffraction[59] (extracted from Figure 1 in that publication). A DPPC/DMPG (50:50) lipid monolayer gave a value of 0.66 nm^2 from neutron reflection data. These values most likely underestimate the area per lipid of a pure POPG lipid bilayer, since an unsaturated acyl chain is expected to increase the surface area of a lipid molecule, as can be seen from the comparison between DPPC and POPC. DMPG used in the monolayer study has completely saturated acyl chains, and PG lipids prepared from egg lecithin contain a large amout of saturated chains. A more recent X-ray study of the negatively charged phospholipid 1,2-dioleoyl-*sn*-glycero-3-phospho-L-serine in the fluid state obtained a surface area per lipid of 0.653 nm^2 in the fluid state and in the absence of osmotic pressure.[60] Until further experimental data become available, the POPG2 model with a lipid surface area of 0.700 nm^2 agrees more with the experimental data mentioned above than the POPG1 model.

It is interesting to note that most other MD simulation studies so far report much smaller area per lipid ratios for POPG bilayers. For example, a study by Zhao et al.[33] reports a value of 0.530 nm^2, and Elmore[61] reports a value of 0.561 nm^2. Another simulation study of a POPE/POPG (75:25) mixture[62] obtains an area per lipid ratio for POPG of 0.628 nm^2 from a decomposition of the area into POPG and POPE contributions using a Voroni tessellation method. This result is not directly comparable to a pure POPG system, since the POPG area is most likely influenced by the presence of POPE. Other simulations of lipids with negatively charged headgroups obtain a similar small area per lipid ratio; for example, for POPS, a value of 0.55 nm^2 has been found.[63] The smaller lipid surface area of charged lipids in comparison to neutral lipids is explained by the binding of counterions that not only compensate for the electrostatic repulsion of the negatively charged headgroups but lead to a further condensation of the bilayer area. Of note is the deviation of the area per lipid in our study (POPG1, 0.635 nm^2) from the previous study by Zhao et al.[33] (0.530 nm^2), which used a similar POPG model with the same partial charge distribution. In the same study, Zhao et al. obtained an average area per lipid of 0.658 nm^2 for POPC, which is identical with our result for POPC1, taking into account the error margin. In order to rationalize these differences, further simulations of POPG1 have been conducted, where the van der Waals radius of the ester-carbonyl atoms has been changed from 0.664 to 0.336 nm (“CH0” to “C” atom type). The “C” atom type has been used in the DPPC1 and DPPC2 models, which led to an unsatisfactory reproduction of the area per lipid. For POPG, this resulted in an average area per lipid of 0.573 ± 0.004 nm^2, which is similar to simulation results reported by other groups. However, in light of available experimental data, as discussed above, an area per lipid of 0.700 nm^2 seems more accurate. The likely reason for the larger area per lipid in our work is the van der Waals repulsion between sodium atoms in the lipid headgroup region and lipid atoms. There is the possiblity that previous MD simulations of anionic lipid bilayers underestimated the surface area per lipid, because they did not model the space requirements of the lipid-bound counterions adequately. Future experimental studies of pure POPG and POPS lipid bilayers are required in order to provide a final answer.

The new DMPC/GROMOS96 53a6 lipid model has been compared to the Berger et al.[19]/GROMOS87 lipid model in membrane protein simulations. During a 20 ns simulation, the lipid/force field combination developed in this work showed better performance in maintaining the experimental structure, as revealed by the rmsd deviation. The experimental conditions for ErbB2 stucture determination of DHPC/DMPC bicelles are similar but not identical to a pure DMPC bilayer. Therefore, deviation from the experimental structure as a measure of force field quality has to be considered with caution. Unfortunately, most membrane protein structures have been obtained at conditions far different from a liquid-crystalline lipid bilayer, for example, from crystals in X-ray crystallography or from detergent micelles in NMR spectroscopy.

In conclusion, united-atom lipid models for DPPC, DMPC, POPC, and POPG have been developed that accurately reproduce experimental data and should enable researchers to carry out united-atom membrane protein simulations with the latest 53a6 version of the GROMOS96 force field, thus reducing the computational costs as compared to all-atom force fields without sacrificing significantly the accuracy of the simulation.

Acknowledgment. This work was supported by the School of Life Sciences at the University of Hertfordshire, United Kingdom.

Supporting Information Available: The atomic coordinates of the bilayer systems and topology files of the lipid molecules are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33–38; 27−28.

(2) Wallin, E.; von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **1998**, *7* (4), 1029–1038.

(3) Terstappen, G. C.; Reggiani, A. In silico research in drug discovery. *Trends Pharmacol. Sci.* **2001**, *22* (1), 23–26.

(4) Sperotto, M. M.; May, S.; Baumgaertner, A. Modelling of proteins in membranes. *Chem. Phys. Lipids* **2006**, *141* (1−2), 2–29.

(5) Kukol, A. Site-specific IR spectroscopy and molecular modelling combined towards solving transmembrane protein structure. *Spectrosc. Int. J.* **2005**, *19* (1), 1–16.

(6) White, S. H. Membrane proteins of known 3D structure. http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html (accessed Dec 15, 2008). University of California at Irvine.

(7) Biggin, P. C.; Bond, P. J. Molecular dynamics simulations of membrane proteins. *Methods Mol. Biol.* **2008**, *443*, 147–60.

(8) Lindahl, E. R. Molecular dynamics simulations. *Methods Mol. Biol.* **2008**, *443*, 3–23.

(9) 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 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.

(10) 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* (18), 3586–3616.

(11) 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* (45), 11225–11236.

(12) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. 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* (13), 1656–1676.

(13) Guvench, O.; MacKerell, A. D. Comparison of protein forcefields for molecular dynamics simulations. *Methods Mol. Biol.* **2008**, *443*, 63–88.

(14) Feller, S. E.; MacKerell, A. D. An improved empirical potential energy function for molecular simulations of phospholipids. *J. Phys. Chem. B* **2000**, *104* (31), 7510–7515.

(15) Feller, S. E.; Gawrisch, K.; MacKerell, A. D. Polyunsaturated fatty acids in lipid bilayers: Intrinsic and environmental contributions to their unique physical properties. *J. Am. Chem. Soc.* **2002**, *124* (2), 318–326.

(16) Rosso, L.; Gould, I. R. Structure and dynamics of phospholipid bilayers using recently developed general all-atom force fields. *J. Comput. Chem.* **2008**, *29* (1), 24–37.

(17) Siu, S. W. I.; Vacha, R.; Jungwirth, P.; Bockmann, R. A. Biomolecular simulations of membranes: Physical properties from different force fields. *J. Chem. Phys.* **2008**, *128* (12), 125103–12115.

(18) Henin, J.; Shinoda, W.; Klein, M. L. United-atom acyl chains for CHARMM phospholipids. *J. Phys. Chem. B* **2008**, *112* (23), 7008–7015.

(19) Berger, O.; Edholm, O.; Jahnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* **1997**, *72* (5), 2002–2013.

(20) Jorgensen, W. L.; Tiradorives, J. The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657–1666.

(21) Van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual*; BIOMOS B.V.: Groningen, The Netherlands, 1987.

(22) Daura, X.; Mark, A. E.; Van Gunsteren, W. F. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.* **1997**, *19* (5), 535–547.

(23) Chandrasekhar, I.; Bakowies, D.; Glattli, A.; Hunenberger, P.; Pereira, C.; Van Gunsteren, W. F. Molecular dynamics simulation of lipid bilayers with GROMOS96: Application of surface tension. *Mol. Simul.* **2005**, *31* (8), 543–548.

(24) Chandrasekhar, I.; Kastenholz, M.; Lins, R. D.; Oostenbrink, C.; Schuler, L. D.; Tieleman, D. P.; van Gunsteren, W. F. A consistent potential energy parameter set for lipids: dipalmitoylphosphatidylcholine as a benchmark of the GROMOS96 45A3 force field. *Eur. Biophys. J.* **2003**, *32* (1), 67–77.

(25) Beevers, A. J.; Kukol, A. Systematic molecular dynamics searching in a lipid bilayer: application to the glycophorin A and oncogenic ErbB-2 transmembrane domains. *J. Mol. Graphics Modell.* **2006**, *25* (2), 226–33.

(26) Liu, X.; Xu, Y.; Li, H.; Wang, X.; Jiang, H.; Barrantes, F. J. Mechanics of channel gating of the nicotinic acetylcholine receptor. *PLoS Comput. Biol.* **2008**, *4* (1), e19.

(27) Cuthbertson, J. M.; Bond, P. J.; Sansom, M. S. P. Transmembrane helix-helix interactions: Comparative simulations of the glycophorin A dimer. *Biochemistry* **2006**, *45* (48), 14298–14310.

(28) Haider, S.; Khalid, S.; Tucker, S. J.; Ashcroft, F. M.; Sansom, M. S. P. Molecular dynamics simulations of inwardly rectifying (Kir) potassium channels: A comparative study. *Biochemistry* **2007**, *46* (12), 3643–3652.

(29) Tieleman, D. P.; MacCallum, J. L.; Ash, W. L.; Kandt, C.; Xu, Z. T.; Monticelli, L. Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid-protein interactions, side chain transfer free energies and model proteins. *J. Phys.: Condens. Matter.* **2006**, *18* (28), S1221–S1234.

(30) Tieleman, D. P.; Berendsen, H. J. C. Molecular dynamics simulations of a fully hydrated dipalmitoyl phosphatidylcholine bilayer with different macroscopic boundary condi-

tions and parameters. *J. Chem. Phys.* **1996**, *105* (11), 4871–4880.

(31) Tieleman, D. P.; Berendsen, H. J.; Sansom, M. S. An alamethicin channel in a lipid bilayer: molecular dynamics simulations. *Biophys. J.* **1999**, *76* (4), 1757–1769.

(32) Tieleman, D. P.; Berendsen, H. J. A molecular dynamics study of the pores formed by Escherichia coli OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys. J.* **1998**, *74* (6), 2786–801.

(33) Zhao, W.; Rog, T.; Gurtovenko, A. A.; Vattulainen, I.; Karttunen, M. Atomic-scale structure and electrostatics of anionic palmitoyloleoylphosphatidylglycerol lipid bilayers with Na+ counterions. *Biophys. J.* **2007**, *92* (4), 1114–1124.

(34) Chiu, S.-W.; Clark, M.; Subramaniam, B. S.; Scott, H. L.; Jacobson, E. Incorporation of surface tension into molecular dynamics simulation of an interface: a fluid phase lipid bilayer membrane. *Biophys. J.* **1995**, *69* (10), 1230–1245.

(35) Bachar, M.; Brunelle, P.; Tieleman, D. P.; Rauk, A. Molecular dynamics simulation of a polyunsaturated lipid bilayer susceptible to lipid peroxidation. *J. Phys. Chem. B* **2004**, *108* (22), 7170–7179.

(36) Martinez-Seara, H.; Rog, T.; Karttunen, M.; Reigada, R.; Vattulainen, I. Influence of cis double-bond parametrization on lipid membrane properties: How seemingly insignificant details in force-field change even qualitative trends. *J. Chem. Phys.* **2008**, *12910*.

(37) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–42.

(38) Bocharov, E. V.; Mineev, K. S.; Volynsky, P. E.; Ermolyuk, Y. S.; Tkach, E. N.; Sobol, A. G.; Chupin, V. V.; Kirpichnikov, M. P.; Efremov, R. G.; Arseniev, A. S. Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. *J. Biol. Chem.* **2008**, *283* (11), 6950–6.

(39) Kandt, C.; Ash, W. L.; Tieleman, D. P. Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **2007**, *41*, 475–488.

(40) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(41) Lindahl, E.; Hess, B.; van der Spoel, D. Gromacs 3.0: A package for molecular dynamics simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317.

(42) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26* (16), 1701–18.

(43) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.

(44) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(45) 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* (19), 8577–8593.

(46) Patra, M.; Karttunen, M.; Hyvonen, M. T.; Falck, E.; Lindqvist, P.; Vattulainen, I. Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions. *Biophys. J.* **2003**, *84* (6), 3636–3645.

(47) 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* (12), 1463–1472.

(48) Miyamoto, S.; Kollman, P. A. Settle - An Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952–962.

(49) Nagle, J. F.; Tristram-Nagle, S. Lipid bilayer structure. *Curr. Opin. Struct. Biol.* **2000**, *10* (4), 474–480.

(50) Kucerka, N.; Tristram-Nagle, S.; Nagle, J. F. Closer look at structure of fully hydrated fluid phase DPPC bilayers. *Biophys. J.* **2006**, *90* (11), L83–L85.

(51) Douliez, J. P.; Leonard, A.; Dufourc, E. J. Restatement of Order Parameters in Biomembranes - Calculation of C-C Bond Order Parameters from C-D Quadrupolar Splittings. *Biophys. J.* **1995**, *68* (5), 1727–1739.

(52) Filippov, A.; Oradd, G.; Lindblom, G. Influence of cholesterol and water content on phospholipid lateral diffusion in bilayers. *Langmuir* **2003**, *19* (16), 6397–6400.

(53) Bernstein, H. J. Recent changes to RasMol, recombining the variants. *Trends Biochem. Sci.* **2000**, *25* (9), 453–5.

(54) Sayle, R. A.; Milner-White, E. J. RASMOL: Biomolecular graphics for all. *Trends Biochem. Sci.* **1995**, *20* (9), 374.

(55) Wohlert, J.; Edholm, O. The range and shielding of dipole-dipole interactions in phospholipid bilayers. *Biophys. J.* **2004**, *87* (10), 2433–2445.

(56) Patra, M.; Salonen, E.; Terama, E.; Vattulainen, I.; Faller, R.; Lee, B. W.; Holopainen, J.; Karttunen, M. Under the influence of alcohol: the effect of ethanol and methanol on lipid bilayers. *Biophys. J.* **2006**, *90* (4), 1121–35.

(57) Kucerka, N.; Liu, Y. F.; Chu, N. J.; Petrache, H. I.; Tristram-Nagle, S. T.; Nagle, J. F. Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using X-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophys. J.* **2005**, *88* (4), 2626–2637.

(58) Feller, S. E.; Pastor, R. W. Constant surface tension simulations of lipid bilayers: The sensitivity of surface areas and compressibilities. *J. Chem. Phys.* **1999**, *111* (3), 1281–1287.

(59) Cowley, A. C.; Fuller, N. L.; Rand, R. P.; Parsegian, V. A. Measurement of Repulsive Forces between Charged Phospholipid Bilayers. *Biochemistry* **1978**, *17* (15), 3163−3168.

(60) Petrache, H. I.; Tristram-Nagle, S.; Gawrisch, K.; Harries, D.; Parsegian, V. A.; Nagle, J. F. Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt. *Biophys. J.* **2004**, *86* (3), 1574–1586.

(61) Elmore, D. E. Molecular dynamics simulation of a phosphatidylglycerol membrane. *FEBS Lett.* **2006**, *580* (1), 144–148.

(62) Murzyn, K.; Rog, T.; Pasenkiewicz-Gierula, M. Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane. *Biophys. J.* **2005**, *88* (2), 1091–1103.

(63) Mukhopadhyay, P.; Monticelli, L.; Tieleman, D. P. Molecular dynamics simulation of a palmitoyl-oleoyl phosphatidylserine bilayer with Na+ counterions and NaCl. *Biophys. J.* **2004**, *86*, 1601–1609.

(64) Kucerka, N.; Tristram-Nagle, S.; Nagle, J. F. Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *J. Membr. Biol.* **2005**, *208* (3), 193–202.

(65) Konig, S.; Pfeiffer, W.; Bayerl, T.; Richter, D.; Sackmann, E. Molecular-Dynamics of Lipid Bilayers Studied by Incoher-

ent Quasi-Elastic Neutron-Scattering. *J. Phys. Ii.* **1992**, *2* (8), 1589–1615.

(66) Maierhofer, A. P.; Bucknall, D. G.; Bayerl, T. M. Modulation of cytochrome C coupling to anionic lipid monolayers by a change of the phase state: a combined neutron and infrared reflection study. *Biophys. J.* **2000**, *79* (3), 1428–1437.

(67) Gilmanshin, R.; Creutz, C. E.; Tamm, L. K. Annexin-Iv Reduces the Rate of Lateral Lipid Diffusion and Changes the Fluid-Phase Structure of the Lipid Bilayer When It Binds to Negatively Charged Membranes in the Presence of Calcium. *Biochemistry* **1994**, *33* (27), 8225–8232.

CT8003468

JCTC Journal of Chemical Theory and Computation

Application of Multiplexed Replica Exchange Molecular Dynamics to the UNRES Force Field: Tests with α and $\alpha+\beta$ Proteins

Cezary Czaplewski,[†,‡] Sebastian Kalinowski,[‡] Adam Liwo,[†] and Harold A. Scheraga*[,†]

Baker Laboratory of Chemisty and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, and, Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland

Received September 23, 2008

Abstract: The replica exchange (RE) method is increasingly used to improve sampling in molecular dynamics (MD) simulations of biomolecular systems. Recently, we implemented the united-residue UNRES force field for mesoscopic MD. Initial results from UNRES MD simulations show that we are able to simulate folding events that take place in a microsecond or even a millisecond time scale. To speed up the search further, we applied the multiplexing replica exchange molecular dynamics (MREMD) method. The multiplexed variant (MREMD) of the RE method, developed by Rhee and Pande, differs from the original RE method in that several trajectories are run at a given temperature. Each set of trajectories run at a different temperature constitutes a layer. Exchanges are attempted not only within a single layer but also between layers. The code has been parallelized and scales up to 4000 processors. We present a comparison of canonical MD, REMD, and MREMD simulations of protein folding with the UNRES force-field. We demonstrate that the multiplexed procedure increases the power of replica exchange MD considerably and convergence of the thermodynamic quantities is achieved much faster.

1. Introduction

The dynamics of proteins and protein folding plays a fundamental role in biological processes, such as enzymatic reactions, signal transduction, immunological processes, and cell motility, and also in malignant processes, such as cancer and amyloid formation.[1] Recent advancement of single-molecule studies[2] facilitates the experimental investigation of the folding pathways of some proteins, but generally, experimental studies of the mechanisms of protein folding are restricted to techniques that provide only indirect and fragmentary information, leaving a wide room to interpretation. Therefore, simulation techniques are used widely to study the dynamics and mechanisms of protein folding. The development of new methods of simulation of mechanisms of protein folding and comparison of theoretical and experimental characteristics of protein folding is of crucial importance for advancing our understanding of biological systems.

Because of the complexity of the systems, all-atom studies of protein folding are mostly restricted to unfolding simulations, starting from the experimental structure, except for small proteins.[3] However, even for small proteins, only a few trajectories can be run, which does not make it readily possible to compare the results with those of experiments that provide ensemble-averaged properties. To carry out large-scale simulations, one has to resort to reduced models of proteins; most of such simulations are carried out with the use of Gō-like[4,5] or related[6,7] potentials, which are biased toward the native structure, or model potentials,[8] which can reproduce general features of protein folding. Small-scale motions are also studied by using elastic-network models.[9] This is because most of the existing realistic and general coarse-grained potentials perform well (in particular, they

---

* To whom correspondence should be addressed. E-mail: has5@cornell.edu.

[†] Cornell University.

[‡] University of Gdańsk.

10.1021/ct800397z CCC: $40.75

can predict the native structure of a protein) when used in connection with information extracted from protein databases, which is not acceptable when studying protein dynamics. In turn, the reason for the deficiency of the coarse-grained potentials lies in their derivation either by analogy to all-atom force fields or from database statistics; neither method offers a clear way of first-principle systematic derivation of the force field.

The coarse-grained UNRES (united-residue) force field[10−15] developed in our laboratory has been derived from first principles[13] as a cluster-cumulant[16] expansion of the restricted free energy (or potential of mean force) of a polypeptide chain immersed in water, where secondary degrees of freedom have been integrated out. This approach enabled us to introduce the essential multibody terms in a systematic and database-independent way. The optimization of the parameters of the force field was performed with a novel method,[17−19] which makes use of the hierarchical structure of the protein energy landscape. The optimized force field was applied to search for the global energy minimum, and predicted complete structures or large portions of structure of proteins in CASP blind test experiments without ancillary information from protein structural databases.[20] Recently, we implemented the united-residue UNRES force field for mesoscopic molecular dynamics (MD).[21−24] Initial results from UNRES MD simulations show that we are able to simulate folding events which take place in a microsecond or even a millisecond time scale.

Methods such as canonical MD or Metropolis Monte Carlo (MC) can be used for estimating thermodynamical properties, as well as for a global search, but in practice, they easily become trapped and thus are not effective methods for studying rough free-energy landscapes of proteins. Efficient conformational sampling algorithms are an essential component of methods for studying protein structure and dynamics. One of the most effective sampling methods, the replica exchange method (RE, also known as exchange MC[25] or parallel tempering[26]), was initially developed to improve sampling in glassy systems in statistical physics. However, following Hansmann's use of the method in simulations of a simple peptide, Met-enkephalin[26] and Sugita and Okamoto's formulation of an MD version of the algorithm,[27] the RE method has been applied extensively in biomolecular simulations.

The replica exchange MD (REMD) method combines the idea of simulated annealing MD and MC methods and is one of the generalized-ensemble algorithms that perform a random walk in energy space because of a free random walk in temperature space. In the REMD method, n replica systems, each in the canonical ensemble and each at a different temperature, are simulated. At given intervals, swaps or exchanges of the configurational variables between systems are accepted with the Metropolis criterion. This is equivalent to exchanges of temperatures because the set of n replica systems can be treated as the set of n continuous MD trajectories of varying temperatures or the set of n canonical ensembles at particular temperatures with structures from all trajectories sorted by temperature. In this paper, we investigate the use of UNRES in multiplexed REMD (MREMD) introduced by Rhee and Pande.[28] In MREMD, to enhance sampling, the replicas are multiplexed with a number of independent molecular dynamics runs at each temperature. Exchanges of configurations between random replicas of neighboring temperatures are tried as in REMD, but there is a larger number of such pairs in MREMD than in REMD. In MREMD, it can be considered that there are several layers of replicas, each of which has all different temperature levels and is equivalent to a single REMD simulation. Exchanges between replicas in different layers are tried, as well as exchanges between replicas in the same layer.

The replica exchange method was the subject of a recent review,[29] which discussed both the history of the method and its application to various physicochemical simulations. The great potential of the RE method was also recognized in a review of sampling methods for molecular simulation.[30] The efficiency of replica exchange for canonical sampling of biomolecules was questioned by Zuckerman and Lyman,[31,32] but they were concerned only with canonical sampling at a fixed temperature and did not consider that efficient conformational searching is also necessary for proper canonical sampling. Also, in an erratum,[32] they agreed that RE is more promising than suggested in their original letter. Recently, we compared three generalized-ensemble algorithms for molecular simulations, namely, a replica exchange method (RE), a replica exchange multicanonical method (REMUCA), and a replica exchange multicanonical method with replica exchange (REMUCAREM) in both MC and MD versions, to determine the thermodynamic characteristics of the UNRES force field for efficient sampling at various temperatures.[33] Of those, the REMD method, especially in its multiplexed version (MREMD), turned out to be the most efficient. Despite having been shown to be very effective on some model systems, applications of MREMD in complex systems, such as those for the simulation of protein folding, have not been tested in detail. Here, we present a comparison of canonical MD, REMD, and MREMD simulations of protein folding with the UNRES force-field.

2. Methods

UNRES Force Field. In the UNRES model,[10−15] a polypeptide chain is represented by a sequence of α-carbon (C^{α}) atoms linked by virtual bonds with attached united side chains (SC) and united peptide groups (p). Each united peptide group is located in the middle between two consecutive α-carbons. Only these united peptide groups and the united side chains serve as interaction sites, the α-carbons serving only to define the chain geometry. Comparison of all-atom and UNRES models of polypeptide chain is shown in Figure 1.

The UNRES force field has been derived as a restricted free energy (RFE) function of an all-atom polypeptide chain plus the surrounding solvent, where the all-atom energy function is averaged over the degrees of freedom that are lost when passing from the all-atom to the simplified system.[12,13] The RFE is further decomposed into factors coming from interactions within and between a given number

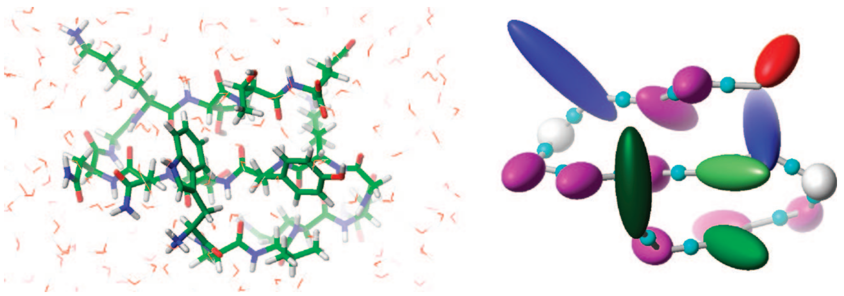


Figure 1. Illustration of all-atom and united-residue (UNRES) models of a polypeptide chain. In the all-atom model (left), different atom types have different colors, the polypeptide chain is shown using thick sticks, and surrounding water molecules are shown as thin sticks. The UNRES model (right) for a polypeptide chain has only the following centers of interaction: the united peptide groups represented by small cyan spheres and the united side chains represented by ellipsoids with different sizes and colors for different types of amino acids. The surrounding water is treated implicitly in the potential of mean force between these centers of interaction.

of united interaction sites.[13] Expansion of the factors into generalized Kubo cumulants[16] enables us to derive approximate analytical expressions for the respective terms,[12,13] including the multibody or correlation terms, which are derived in other force fields from structural databases or on a heuristic basis.[34]

The energy of the virtual-bond chain is expressed by eq 1.

$$U = w_{\mathrm{SC}}\sum_{i<j} U_{\mathrm{SC}_i\mathrm{SC}_j} + w_{\mathrm{SCp}}\sum_{i\neq j} U_{\mathrm{SC}_i\mathrm{p}_j} + w_{\mathrm{pp}}^{\mathrm{VDW}}\sum_{i<j-1} U_{\mathrm{p}_i\mathrm{p}_j}^{\mathrm{VDW}} + w_{\mathrm{pp}}^{\mathrm{el}}\sum_{i<j-1} U_{\mathrm{p}_i\mathrm{p}_j}^{\mathrm{el}} + w_{\mathrm{tor}}\sum_{i} U_{\mathrm{tor}}(\gamma_i) + w_{\mathrm{tord}}\sum_{i} U_{\mathrm{tord}}(\gamma_i, \gamma_{i+1}) + w_{\mathrm{b}}\sum_{i} U_{\mathrm{b}}(\theta_i) + w_{\mathrm{rot}}\sum_{i} U_{\mathrm{rot}}(\alpha_{\mathrm{SC}_i}, \beta_{\mathrm{SC}_i}, \theta_i) + \sum_{m=3}^{4} w_{\mathrm{corr}}^{(m)} U_{\mathrm{corr}}^{(m)} + w_{\mathrm{bond}}\sum_{i=1}^{n\mathrm{bond}} U_{\mathrm{bond}}(d_i) + w_{\mathrm{SS}}\sum_{i} U_{\mathrm{SS};i} \quad (1)$$

The term $U_{\mathrm{SC}_i\mathrm{SC}_j}$ represents the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains, which implicitly contains the contributions from the interactions of the side chain with the solvent. The term $U_{\mathrm{SC}_i\mathrm{p}_j}$ denotes the excluded-volume potential of the side-chain-peptide-group interactions. The peptide-group interaction potential is split into two parts: the Lennard-Jones interaction energy between peptide-group centers ($U_{\mathrm{p}_i\mathrm{p}_j}^{\mathrm{VDW}}$) and the average electrostatic energy between peptide-group dipoles ($U_{\mathrm{p}_i\mathrm{p}_j}^{\mathrm{el}}$); the second of these terms accounts for the tendency to form backbone hydrogen bonds between peptide groups p_i and p_j. The terms U_{tor}, U_{tord}, U_{b}, and U_{rot} are the virtual-bond-dihedral angle torsional terms, virtual-bond-dihedral angle double-torsional terms, virtual-bond angle bending terms, and side-chain rotamer terms, respectively; these terms account for the local propensities of the polypeptide chain. The terms $U_{\mathrm{corr}}^{(m)}$ represent correlation or multibody contributions from the coupling between backbone-local and backbone-electrostatic interactions. The multibody terms are indispensable for reproduction of regular α-helical and β-sheet structures. The terms $U_{\mathrm{bond}}(d_i)$, where d_i is the length of the ith virtual bond and $n$bond is the number of virtual bonds, are simple harmonic potentials of virtual-bond distortion,[22] and the terms U_{SS} describe the energetics of disulfide bonds.[35,36] The w's are the weights of the energy terms, and they were determined (together with the parameters within each cumulant term and the well depths of the side-chain pairwise interaction potential $U_{\mathrm{SC}_i\mathrm{SC}_j}$) by hierarchical optimization[19] of the potential-energy function. In this work, we used the version of the UNRES force field referred to as 4P,[19] which was parametrized in our earlier work[19] simultaneously on the training proteins 1GAB (α), 1E0L (β), 1E0G (α+β), and 1IGD (α+β), using the conformational space annealing (CSA) method[37] to generate the decoy sets. Although the 4P force field was parametrized to find the native structures of proteins as global minima of the potential energy, it performs quite well in folding proteins in molecular dynamics simulations.[21] On the other hand, because the optimization procedure was focused on producing as large an energy gradient with increasing native-likeness as possible,[17] the folding temperatures turned out to be unphysically high[21,23] and the heat capacity curves contain multiple peaks,[33] reflecting the fact that formation of small structural elements such as individual α-helices precedes packing into tertiary structure. This also means that the 4P UNRES force field produces rough energy landscapes, but such a feature is useful for performing a hard test of a conformational-search method. We note at this point that, in our recent work,[38] we have reported preliminary versions of the UNRES force field parametrized for canonical simulations.

Replica-Exchange Algorithm. In the REMD method,[27] M canonical MD simulations are carried out simultaneously, each one at a different temperature. Initially the temperatures increase with the sequential number of the simulation (trajectory). After every m steps, an exchange of temperatures between neighboring trajectories ($j = i + 1$) is attempted, the decision about the exchange being made based on the Metropolis criterion, which is expressed by eq 2

$$\Delta = (\beta_j - \beta_i)[U(\mathbf{X}_j) - U(\mathbf{X}_i)] \quad (2)$$

where $\beta_i = 1/RT_i$, T_i being the absolute temperature corresponding to the ith trajectory, and $\mathbf{X}_i$ denotes the variables of the UNRES conformation of the ith trajectory at the attempted exchange point. If $\Delta \leq 0$, T_i and T_j are exchanged, otherwise the exchange is performed with probability $\exp(-\Delta)$.

The multiplexed variant of the RE method (MREMD) developed by Rhee and Pande[28] differs from the original

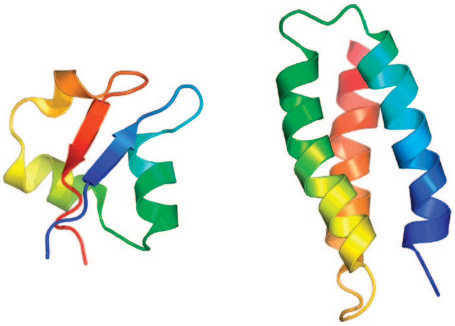


Figure 2. Structures of 1E0G (left) and 1LQ7 (right) proteins shown as ribbon models. Ribbons are in rainbow colors starting from blue on the N-terminus to red on the C-terminus.

RE method in that several trajectories are run at a given temperature. Each set of trajectories ran at a different temperature constitutes a layer. Exchanges are attempted not only within a single layer but also between layers.

The original REMD and MREMD algorithms require synchronization every time the exchange of temperatures between trajectories is attempted. Exchange of temperatures is performed by the master task, and the algorithm requires that all replicas communicate with the master task. All replicas have to perform the same number of steps between communications exactly. Such synchronization means that the master task and all replicas have to wait for the slowest replica. This deteriorates the performance of the algorithm beyond 256 processors heavily. We eliminated synchronization by allowing exchanges every time at which the master task has performed a given number of steps rather than requiring that all tasks have reached this point. Synchronization is also eliminated in a serial replica exchange algorithm (SREM).[39] SREM eliminates not only synchronization but also all communication/exchanges between replicas; each replica changes temperature not by direct exchange with neighbors but based on potential energy distributions. Recently we have implemented SREM with UNRES.[40] SREM reproduces the results of REM and is more efficient in terms of wall-clock time and scales better on distributed-memory machines. Unfortunately SREM can be applied only to the temperature-independent, but not to the temperature-dependent, UNRES force field.[40] Optimizing input/output operations is also important for parallel performance. UNRES MREMD and canonical MD simulations can use two modes of input/output. In the first mode, all processors read and write all files independently: input files with all parameters in the beginning of the run and several output files at defined intervals. In the second mode, only the master processor writes a text file with messages and one binary trajectory file with conformations collected from all processors. Only the second mode minimizes input/output operations and leads to high efficiency on massively parallel systems without local hard drives such as Cray XT3 or IBM BlueGene. An additional cache array on each processor is used for storage of calculated conformations before sending them to the master processor, which facilitates less synchronization between processors. The Europort Data Compression XDRF library is used for writing compressed binary trajectory files.[41]

WHAM. The weighted histogram analysis method (WHAM)[42] was used to extract maximum information from all replicas to evaluate thermodynamic quantities at any temperature. For a replica exchange simulation with M replicas at M distinct temperatures, a set of M energy histograms $N_m(E)$ is obtained. The densities of states $[n(E)]$ are then obtained self-consistently from the following WHAM equations:

$$n(E)=\frac{\sum_{m=1}^{M} g_m^{-1} N_m(E)}{\sum_{m=1}^{M} g_m^{-1} n_m \exp(f_m-\beta_m E)} \tag{3}$$

$$\exp(-f_m)=\sum_{E} n(E)\exp(-\beta_m E) \tag{4}$$

where $N_m(E)$ is the histogram at temperature T_m, $\beta_m = 1/(RT_m)$ is the inverse temperature, n_m is the total number of samples in the mth replica, $g_m = 1 + 2\tau_m$, and τ_m is the integrated autocorrelation time at temperature T_m. In biomolecular systems, g_m is approximately constant[42] and, therefore, can be canceled in eq 3. The WHAM eqs 3 and 4 are evaluated self-consistently, and the resulting densities of states are used to evaluate the expectation value of any observable A in eq 5

$$\langle A\rangle_T=\frac{\sum_{E} A(E)n(E)\exp(-\beta E)}{\sum_{E} n(E)\exp(-\beta E)} \tag{5}$$

3. Results and Discussion

The native structures of two proteins investigated in this work: the *Escherichia coli* Mltd Lysm domain (an α+β protein, 48 residues, 1E0G)[43] and de novo designed protein (an α protein, 67 residues, 1LQ7)[44] are shown in Figure 2. The 1E0G protein was one of four proteins used simultaneously, together with three others to optimize the set of UNRES energy parameters, designated as the 4P force field,[19] used in the present work.

Simulation of M replicas in REMD, rather than one canonical MD trajectory, requires on the order of M times more computational effort if each trajectory is simulated for the same length as the single canonical MD trajectory. To make comparison simpler, the total length of simulations should be the same. Many shorter MD simulations appear to be more efficient and provide more insight than a single longer simulation with the same overall length of time.[45,46] The main reason is nonergodicity of MD simulations, which especially for a rugged energy landscape are easily trapped in the region close to the starting conformation. Running shorter MD simulations provides a greater chance to explore different regions of conformational space than running a single long simulation. This is illustrated by the comparison of plots of energy versus rmsd from the native structure for the 1E0G protein generated using independent canonical MD simulations with different numbers of trajectories and the

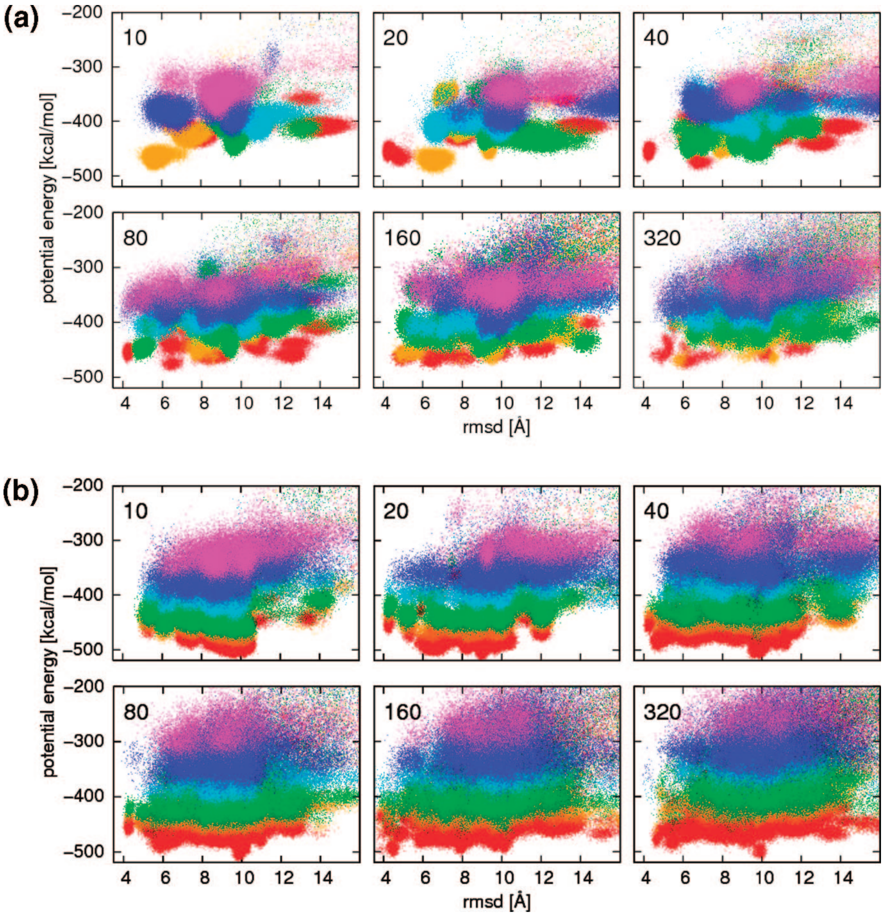


Figure 3. Plots of energy vs rmsd from the native structure for the 1E0G protein generated using (a) independent canonical MD simulations with different numbers of trajectories and (b) multiplexed-replica exchange MD simulations with different numbers of replicas. The number of trajectories (replicas) used in the simulation presented in each graph is shown in the top-left corner of each graph. The total length of the simulation is the same for all subgraphs, that is, simulation with a larger number of trajectories is proportionally shorter. Different colors from the rainbow spectrum represent different temperatures: red for 300 K, orange for 350 K, green for 400 K, cyan for 500 K, blue for 600 K, and magenta for 700 K.

same total length (80 mln steps) of simulations shown in Figure 3a. Simulations at six different temperatures were used, and scatter plots for each temperature are represented by different colors of points. The upper left panel with only 10 trajectories (8 mln steps each) shows that each trajectory visited only small regions of conformational space as marked by limited changes of rmsd and energy. Canonical sampling for this set of MD simulations did not converge because simulations performed with lower temperature (300 K, red points) did not visit the lowest-energy regions; the lowest energy was achieved in simulations at intermediate temperature (350 K, orange points). It should be noted that the important region close to the native structure, with rmsd around 4 Å, is not visited at all in this set of simulations. Adding more trajectories and reducing their length to keep the total simulation-length the same improves the sampling up to certain point. The native region around 4 Å rmsd is visited by at least one trajectory for simulations with 20, 40, and 80 trajectories but not for 160 and 320 trajectories, which are too short (in the 320 case, each trajectory has only 0.25 mln steps).

The MREMD simulation is more efficient than a set of independent canonical MD simulations: for the same length and the number of replicas (trajectories), it covers a larger portion of conformational space, as illustrated on plots of energy versus rmsd in Figure 3b, in comparison with Figure 3a. This improvement is especially visible for low-energy regions: all MREMD simulations reached lower energies for low-temperature replicas in comparison with the independent MD simulations. The convergence of each canonical distribution to proper energy regions for each temperature is much better, as shown by the proper order of color on each plots: lower temperature replicas always sample lower energy regions, which was not always the case for the set of independent MD simulations. Just as running several MD trajectories is more efficient than one long MD trajectory, adding more trajectories increases the effectiveness of MREMD compared to REMD. Because the total simulation length is constant, it is not always feasible to decrease the necessary simulation time by simultaneously performing a multiple number of replicas because thermodynamically acceptable results cannot be expected within a very short simulation time for each replica. This is especially important for larger proteins and proteins containing β structures, which fold more slowly.

A set of MREMD simulations with different numbers of replicas but the same total simulation time for 1LQ7 is shown in Figure 4. The shortest MREMD simulation (0.25 million

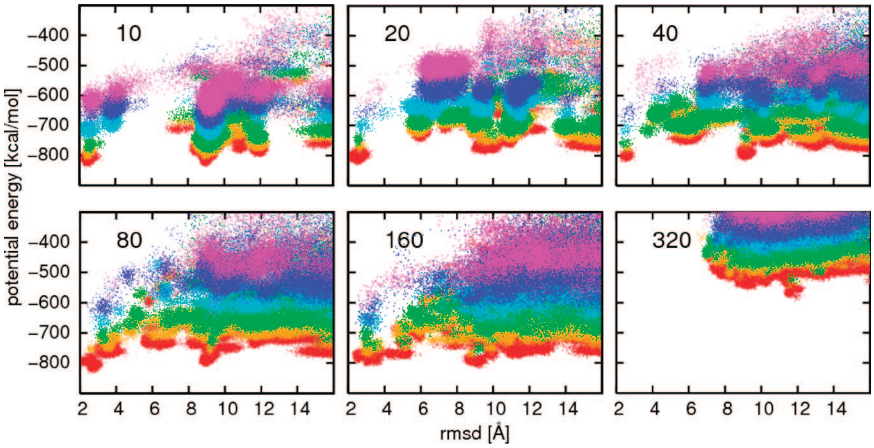


Figure 4. Plots of energy vs rmsd from the native structure for the 1LQ7 protein generated using multiplexed-replica exchange MD simulations with different numbers of replicas. The number of replicas is shown in top-left corner of each graph. The total length of the simulation is the same for all graphs. Different colors from the rainbow spectrum represent different temperatures: red for 300 K, orange for 350 K, green for 400 K, cyan for 500 K, blue for 600 K, and magenta for 700 K.

steps) with the largest number of replicas (320) did not visit low energy and low rmsd regions of conformational space at all. But an intermediate number of replicas and length of each replica's trajectory shows that, up to a certain point, adding more replicas and reducing each of their lengths also helps in this case: the region of conformational space with rmsd around 6 Å was not visited in simulations with 10 and 20 replicas but is occupied for 40, 80, and 160 replicas.

The observation that protein folding has a series of early conformational steps that lead to lag phases at the beginning of the folding kinetics and limit the use of very short simulation data was made by Fersht.[47] His criticism was directed toward the use of very short simulations in distributed computing for calculating kinetic data, but the presence of these lag phases is important also for MREMD simulations. It can bias short simulations; therefore, MREMD with many replicas but too short simulation time for each of them can lead to a false conformational landscape and wrong thermodynamic characteristics of folding. Convergence of simulations should always be checked carefully by monitoring convergence of sensitive thermodynamical variables such as heat of capacity profiles.

To compare the efficiency of REMD and MREMD in calculations of thermodynamical characteristics of protein folding, additional simulations for 1E0G and 1LQ7 were performed using 30 temperatures spanning the range 200−1800 K (the folding temperature of the 4P force field is very high, around 1200 K, because this force field was optimized using decoys generated by a global optimization method, not by canonical sampling). The REMD simulations were performed with one replica per temperature and 300 million steps for 1E0G and 48 million steps for 1LQ7, and MREMD simulations were performed with multiplexing of eight replicas per temperature with the same length of simulation (300 million steps for 1E0G and 48 millions steps for 1LQ7). To determine the added value of multiplexing, for 1LQ7, we carried out eight additional REMD simulations independent of each other at 30 temperatures each (i.e., with the same temperatures and total number of trajectories as the multiplexed simulation of that protein); to determine the added value of replica exchange, we carried out two series of 30 and 240 independent canonical MD simulations, respectively, at the same temperatures as in REMD/MREMD simulations with 1 or 8 trajectories per temperature, respectively. Such control runs were not carried out for 1E0G because of the slow convergence of the canonical simulations. In this set of simulations, all replicas were started from extended structures. To investigate the dependence on the starting conditions, a second set of simulations, REMD (300 million steps for 1E0G and 48 millions steps for 1LQ7) and MREMD with multiplexing of 8 replicas per temperature (300 million steps for 1E0G and 48 millions steps for 1LQ7) were performed with all replicas started from native structures. Conformations collected during simulations every 2000 steps were used in WHAM to calculate temperature profiles of heat capacity and average rmsd from the native structure.

The convergence of the heat capacity curve with increased simulation length, calculated for the 1E0G protein by WHAM using consecutive windows of length 20 mln steps for REMD and 2.5 mln steps for MREMD, is shown in Figure 5 (the length of a window was chosen to select the same number of conformations from both REMD and MREMD simulations). For 1E0G, convergence is slow in all simulations, both for REMD starting from the extended structure (Figure 5a) and starting from the native structure (Figure 5b) and for MREMD starting from the extended structure (Figure 5c) and starting from the native structure (Figure 5d). The final blue curves generated from all simulations are similar: there is a small peak around 1250 K and a high narrow peak around 1050 K. The small peak at higher temperature is associated with lowering of the average rmsd for this temperature to 16 Å as shown in Figure 6, and corresponds to formation of compact structures with local structures as α-helices. This peak converges very fast and is present in its final position even on red curves generated from the very short simulations. The high narrow peak at lower temperature is associated with a fast drop of the average rmsd at this temperature to 4 Å, as shown in Figure 6, and corresponds to formation of native structure with long-range antiparallel β structure. This peak converges

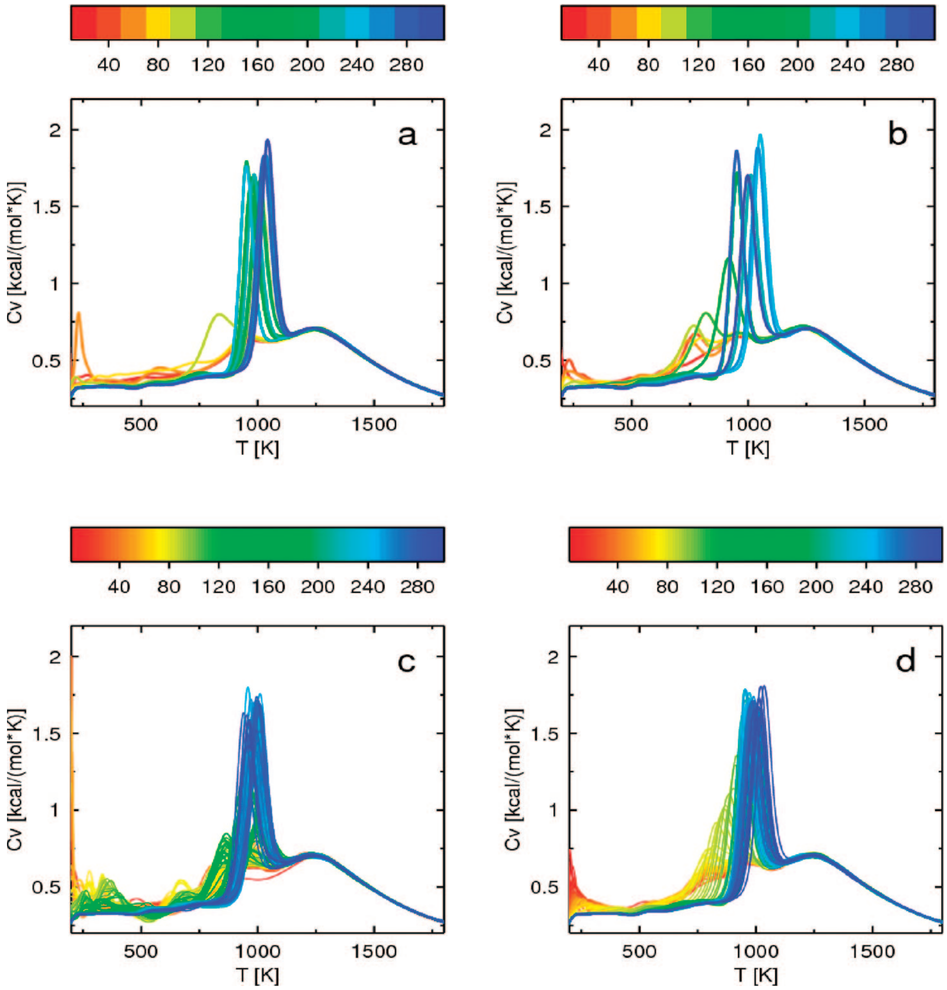


Figure 5. Plots of heat capacity as a function of temperature for 1E0G, calculated from windows of 20 mln consecutive steps corresponding to regular REMD starting from the extended structure (a) and from the native structure (b) and from windows of 2.5 mln steps corresponding to MREMD with multiplexing of 8 replicas per temperature starting from the extended structure (c) and from the native structure (d). The curves in all panels are colored according to the duration of simulation; the color scale (in million steps) is a rainbow from red (shortest) to blue (longest) shown above each panel.

slowly, first appearing at lower temperatures and slowly growing and shifting to higher temperatures. For simulations started from extended structures (Figure 6a and c), the average rmsd lowers to 4 Å only after long simulation. For simulations started from the native structure (Figure 6b and d), the average rmsd is around 4 Å for low temperatures from the very beginning, but convergence of the average rmsd is also slow.

To provide better insight into the convergence of ensemble averages, in Figure 7, we plot the standard deviation of rmsd(T) curves, $\sigma_{\mathrm{RMSDave}}$, from the rmsd curve averaged over the last 10 windows of the MREMD simulations started from the experimental structure, which was taken as reference. The quantity $\sigma_{\mathrm{RMSDave}}$ is defined by eq 6.

$$\sigma_{\mathrm{RMSDave}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}[\mathrm{rmsd}(T_i) - \mathrm{rmsd}^{\mathrm{ret}}(T_i)]^2} \qquad (6)$$

where $\mathrm{rmsd}^{\mathrm{ref}}$ denotes the reference rmsd and $N = 2001$ is the number of temperatures; $T_1 = 100$ K, $T_N = 2100$ K. We noted that the rmsd curves converge slower and are more sensitive to sampling scheme than the heat-capacity curves and, therefore, we used them to monitor convergence. Convergence was assumed to occur when $\sigma_{\mathrm{RMSDave}}$ dropped to 0.5 Å or below.

An analysis of the plots in Figure 7 shows that, for the 1E0G protein, MREMD simulations with multiplexing of 8 replicas per temperature do not converge 8 times faster than a single REMD simulation. The REMD simulation started from the extended structure converges in about 120 mln steps, whereas MREMD needs around 150 mln steps. The MREMD simulations started from the native structure converge faster, in about 110 mln steps, while the REMD simulation started from the native structure needs around 150 mln steps for convergence.

For 1E0G, the total computational expense necessary for convergence is smaller for REMD compared with that of MREMD simulations with multiplexing of 8 replicas per temperature. It can be noted, however, that the $\sigma_{\mathrm{RMSDave}}$ plots corresponding to MREMD simulations exhibit only small and high-frequency oscillations, as compared to those of REMD simulations which exhibit slower oscillations with larger amplitude. These oscillations are manifested in the rmsd curves (Figure 6a and b) as shifts of the region of the

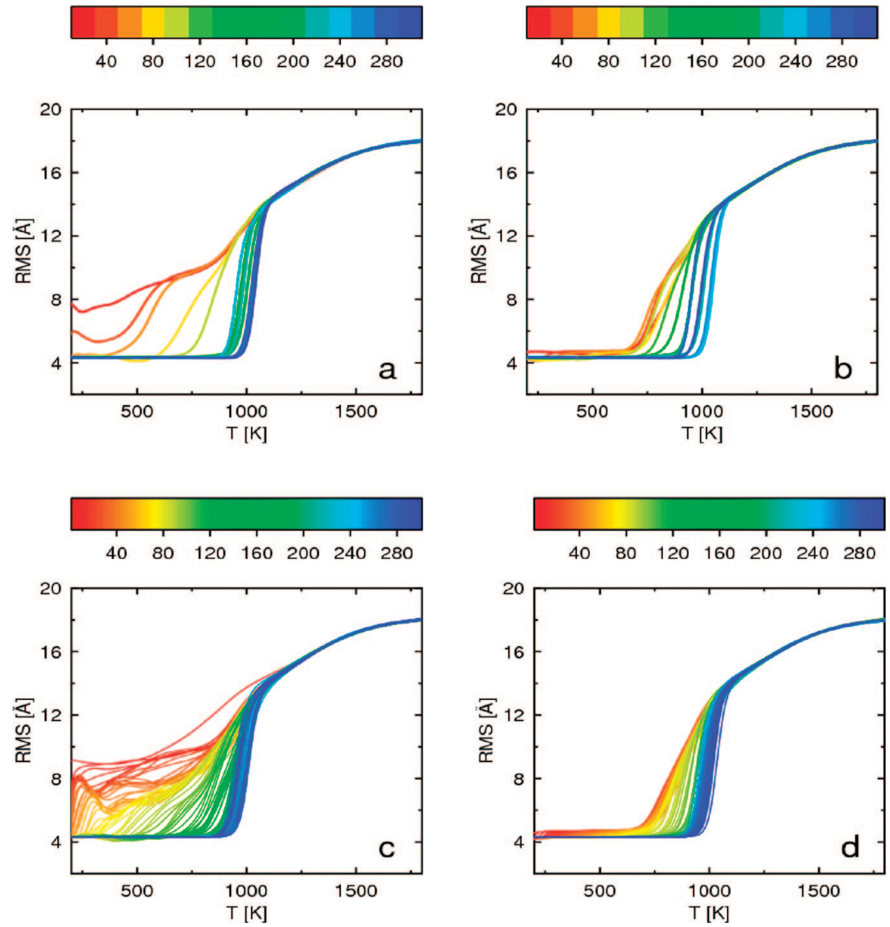


Figure 6. Convergence of the ensemble-averaged rmsd as a function of temperature for 1E0G, calculated from windows of 20 mln consecutive steps corresponding to regular REMD starting from the extended structure (a) and from the native structure (b) and from windows of 2.5 mln steps corresponding to MREMD with multiplexing of 8 replicas per temperature starting from the extended structure (c) and from the native structure (d). The curves in all panels are colored according to the duration of simulation; the color scale (in million steps) is a rainbow from red (shortest) to blue (longest) shown above each panel.

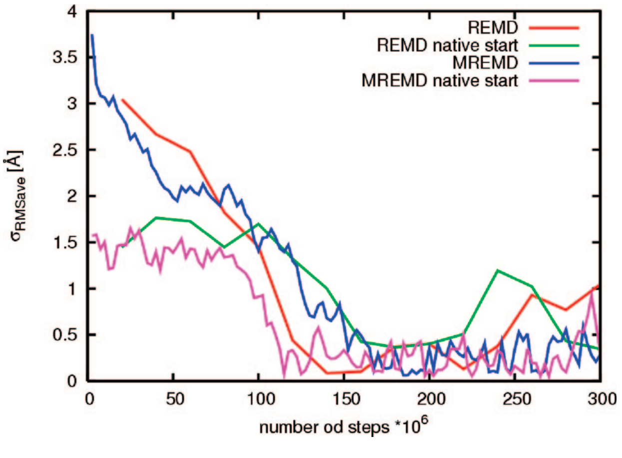


Figure 7. Plots of standard deviation of rmsd(*T*) curves, $\sigma_{RMSDave}$, from the rmsd curve averaged over the last 10 windows of the MREMD simulations started from the experimental structure, which was taken as reference, for 1E0G regular REMD simulation starting from the extended structure (red) and from the native structure (green); MREMD with multiplexing of 8 replicas per temperature starting from the extended structure (blue) and from the native structure (magenta)

inflection point (about $T = 1000$ K) with the progress of simulations. In addition, in REMD simulations, the average $\sigma_{RMSDave}$ after convergence is about 0.5 Å for the start from extended and 0.7 Å for the start from the native structutre. Conversely, in both MREMD simulations the average $\sigma_{RMSDave}$ is about 0.2 Å, and the $\sigma_{RMSDave}$ curves are virtually independent of each other after achieving convergence, exhibiting white noise behavior. This observation suggests that MREMD results in better averaging.

The convergence of the heat capacity curve with increased simulation length, calculated for 1LQ7 protein by WHAM, using consecutive windows of length 4 mln steps for REMD and 0.5 mln steps for MREMD, is shown in Figure 8. For 1LQ7, the convergence is fast for all simulations [i.e., (i) REMD started from the extended structure (Figure 8a) and (ii) started from the native structure (Figure 8b), (iii) MREMD started from the extended structure (Figure 8c) and (iv) started from the native structure (Figure 8d), (v) a series of 30 independent canonical MD simulations at temperatures of REMD simulations (Figure 8e), (vi) a series of 240 independent canonical MD simulations carried out at the temperatures of the MREMD simulation (8 trajectories per temperature) (Figure 8f), and (vii) a series of 8 independent REMD simulations (Figure 8g)]. The final blue curves generated from all simulations are similar with one wide peak around 1150 K. It can be noted that the heat-capacity curves

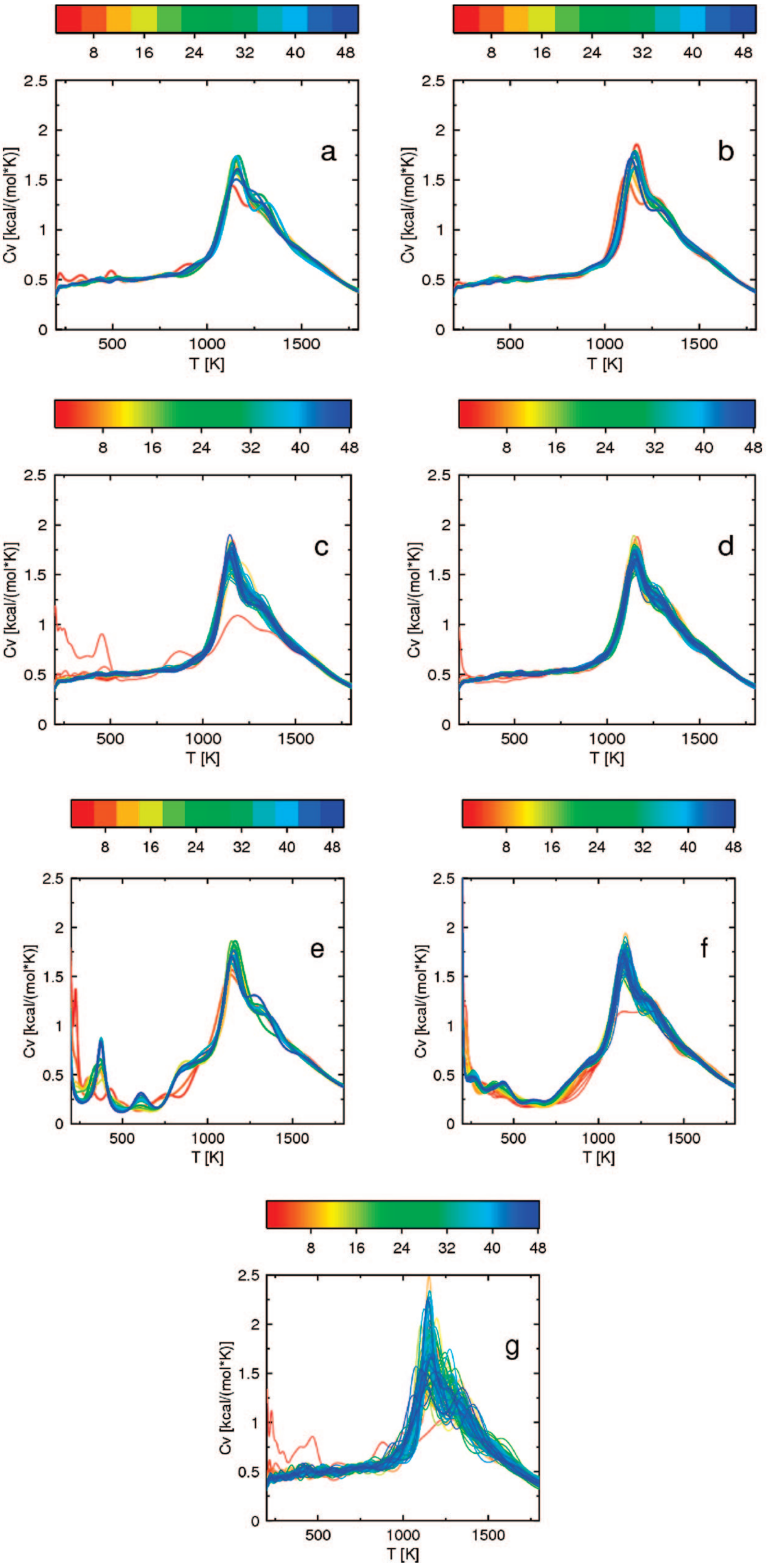


Figure 8. Plots of heat capacity as a function of temperature for 1LQ7, calculated from windows of 4 mln consecutive steps corresponding to regular REMD starting from the extended structure (a) and from the native structure (b), from windows of 0.5 mln steps corresponding to MREMD with multiplexing of 8 replicas per temperature starting from the extended structure (c) and from the native structure (d), from windows of 4mln steps from a series of 30 independent canonical MD simulations at temperatures of REMD simulations (e), from windows of 0.5 mln steps from a series of 240 independent canonical MD simulations carried out at the temperatures of the MREMD simulation (8 trajectories per temperature) (f), and a series of 8 independent REMD simulations (g). The curves in all panels are colored according to the duration of simulation; the color scale (in million steps) is a rainbow from red (shortest) to blue (longest) shown above each panel.

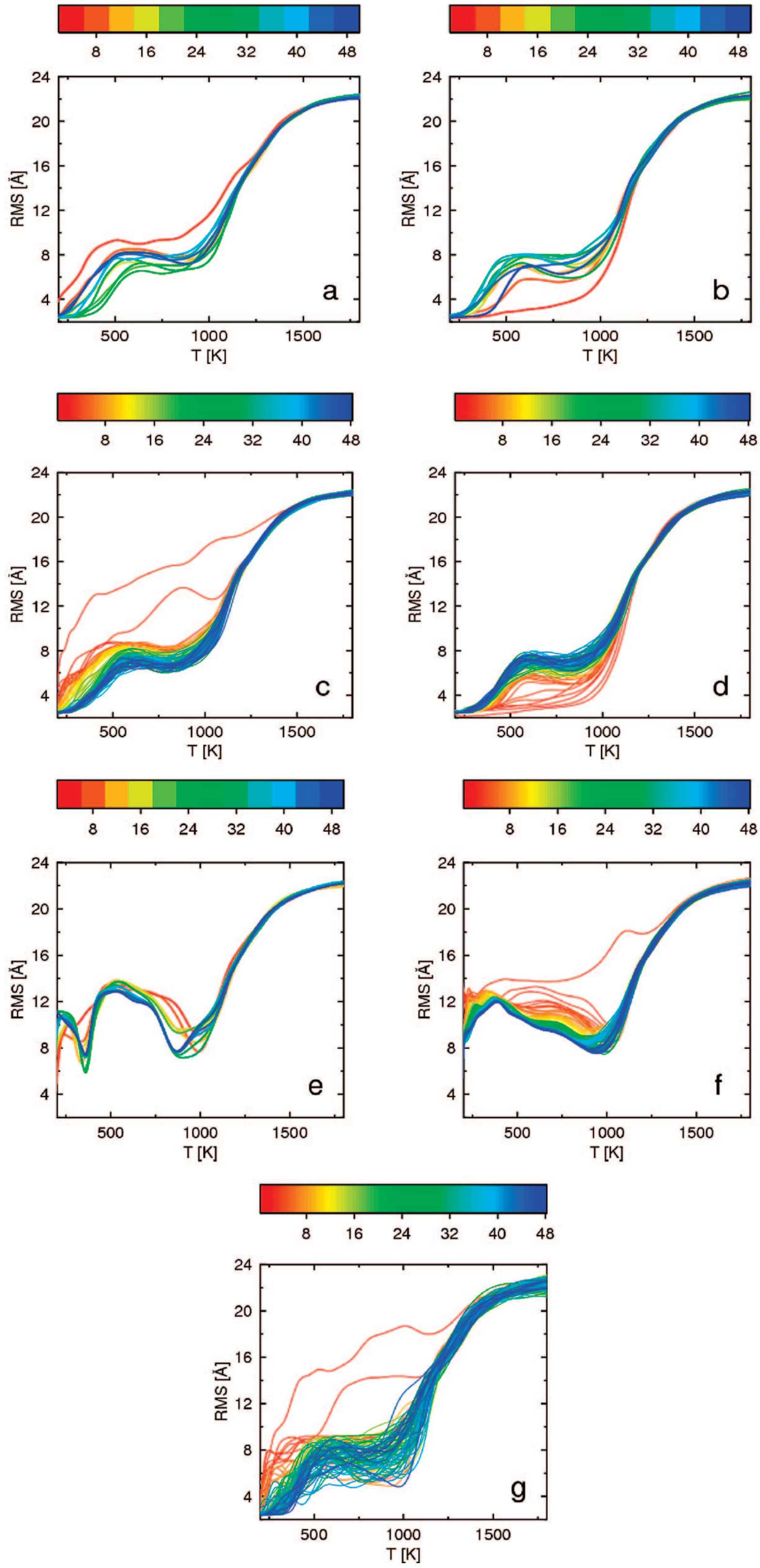


Figure 9. Convergence of the ensemble-averaged rmsd as a function of temperature for 1LQ7, calculated from windows of 4 mln consecutive steps corresponding to regular REMD starting from the extended structure (a) and from the native structure (b), from windows of 0.5 mln steps corresponding to MREMD with multiplexing of 8 replicas per temperature starting from the extended structure (c) and from the native structure (d), from windows of 4mln steps from a series of 30 independent canonical MD simulations at temperatures of REMD simulations (e), from windows of 0.5 mln steps from a series of 240 independent canonical MD simulations carried out at the temperatures of the MREMD simulation (8 trajectories per temperature) (f), and from a series of 8 independent REMD simulations (g). The curves in all panels are colored according to the duration of simulation; the color scale (in million steps) is a rainbow from red (shortest) to blue (longest) shown above each panel.

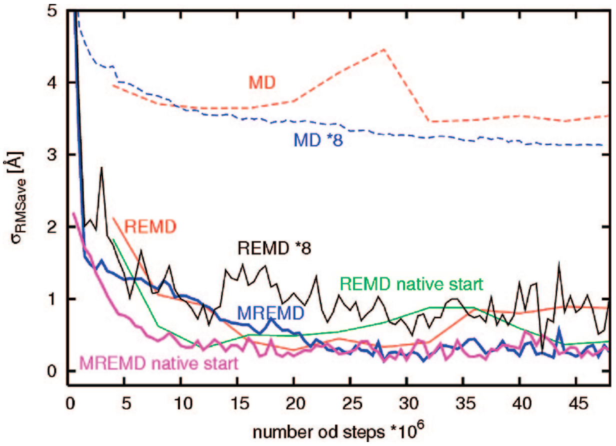


Figure 10. Plots of standard deviation of rmsd(T) curves, $\sigma_{\mathrm{RMSDave}}$, from the rmsd curve averaged over the last 10 windows of the MREMD simulations started from the experimental structure, which was taken as reference, for 1LQ7 regular REMD simulation starting from the extended structure (red line) and from the native structure (green line), MREMD with multiplexing of 8 replicas per temperature starting from the extended structure (blue line) and from the native structure (magenta line), and for a series of 30 independent canonical MD simulations at temperatures of REMD simulations (dashed red line), a series of 240 independent canonical MD simulations carried out at the temperatures of the MREMD simulation (8 trajectories per temperature) (dashed blue line), and a series of 8 independent REMD simulations (black line).

obtained from canonical MD simulations (Figure 8e and f) have many secondary peaks at low temperatures, which do not occur for REMD or MREMD simulations. This feature of canonical simulations results from nonergodicity of the system studied at low temperatures.

The ensemble-averaged rmsd curves for the 1LQ7 protein calculated for increasing simulation length, are shown in Figure 9 and the plots of $\sigma_{\mathrm{RMSDave}}$ are shown in Figure 10; as for 1E0G, the reference rmsd curve was calculated by averaging over 10 last windows of the MREMD simulation started from the experimental structure. For the 1LQ7 protein, UNRES generates two low-energy structures: a native-like structure with rmsd around 2 Å and a "mirror image" structure with rmsd around 9 Å. The temperature profile of the average rmsd depends on the balance of the free energies between these two minima as a function of temperature. For temperatures lower than 300 K, the native-like structures are more probable, while for temperatures between 300 and 1100 K, the "mirror image" structures with rmsd around 9 Å win. For the 1LQ7 protein, MREMD simulations with multiplexing of 8 replicas per temperature, the initial drop of $\sigma_{\mathrm{RMSDave}}$ is faster, which is best illustrated in Figure 10 by the relative shift of the REMD curve (red) with respect to the MREMD curve (blue) and of the REMD from the native start curve (green) with respect to the curve corresponding to the MREMD simulation started from the experimental structure (magenta) to the right. Both REMD and MREMD simulations, started from the extended structure, converge in about 15 mln steps. Simulations started from the native structure converge in about 10 mln steps. However, as for 1E0G, the $\sigma_{\mathrm{RMSDave}}$ curves corresponding to REMD simulations exhibit slow large-amplitude oscillations (with period of about 38 mln MD steps, which largely exceeds the time-window size which is 4 mln steps for REMD simulations) after convergence, while those corresponding to MREMD simulations exhibit white-noise behavior after convergence, which indicates better averaging in MREMD compared to REMD simulations.

It can also be noted that performing 8 independent REMD simulations is not equivalent to performing a single 8-plexed REMD simulation. The heat-capacity (Figure 8g) and ensemble-averaged rmsd (Figure 9g) curves corresponding to 8 independent REMD simulations are much more diffuse than those of the 8-plexed REMD simulation (Figure 8c and 9c). This is also manifested in the $\sigma_{\mathrm{RMSDave}}$ curve (the black line in Figure 10) corresponding to 8 independent REMD simulations which has a higher average $\sigma_{\mathrm{RMSDave}}$ value (about 1 Å) after the initial drop compared to that of both MREMD simulations (about 0.2 Å), and although, at the first glance, it exhibits a random-noise hehavior in the later part, the fast random oscillations are superposed on slow large-period oscillations. It can, therefore, be concluded that the added value of multiplexing is improved averaging even compared to the equivalent number of REMD simulations.

The rmsd curves obtained in canonical MD simulations indicate that the system does not contain a dominant amount of native structure even at low temperature (Figure 9e and f). This is also manifested in the corresponding $\sigma_{\mathrm{RMSDave}}$ curves (Figure 10), which are always above those corresponding to curves calculated from REMD or MREMD simulations. This observation clearly shows that canonical MD is not suitable to derive thermodynamic properties of systems which are not ergodic at all temperatures.

The improvement of ergodicity in REMD and MREMD simulations compared to canonical MD simulations carried out at the same temperatures can also be assessed by comparing the mean first passage times (MFPTs), where MFPT is defined as the average time (averaged over all trajectories) in which the native-like structure appears for the first time. For 1LQ7, we identified the appearance of a native-like structure with rmsd drop below 5 Å. For trajectories in which no native-like structures appear, the total trajectory time contributed to the average. The MFPTs and the number of trajectories which did not reach native-like (NONF) structures are as follows for the respective simulations: (i) MFPT = 175 ns and NONF = 20 for 30 independent MD simulations compared to (ii) MFPT = 42 ns and NONF = 1 for the corresponding REMD simulation; (iii) MFPT = 173 ns and NONF = 175 for 240 independent MD simulations at 30 temperatures with 8 trajectories per temperature compared to (iv) MFPT = 47 ns and NONF = 0 and MFPT = 49 ns and NONF = 3 for the 8 independent REMD simulations and the 8-plexed REMD simulation, respectively. The total simulation length was about 235 ns. These data clearly show that replica exchange leads to at least 4 times faster convergence to the native structure (which is the most probable one at low temperatures for 1LQ7 with the 4P force field used in this work). There is no appreciable difference in MFPT between REMD, 8-plexed REMD, and 8 independent REMD simulations. We must note at this point

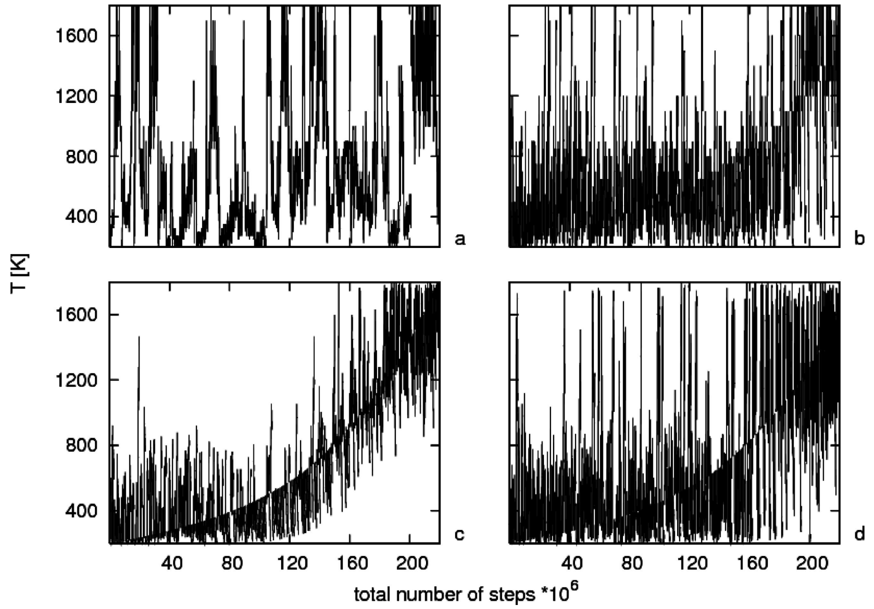


Figure 11. Diffusion in temperature replica space of REMD vs MREMD: combined plots of temperature as a function of time for all replicas for (a) REMD with 30 temperatures, (b) MREMD with 30 temperatures and multiplexing of 8 replicas per temperature, for a total of 240 replicas, (c) REMD with 240 temperatures, and (d) REMD with 240 temperatures and 10 times more frequent exchanges between replicas.

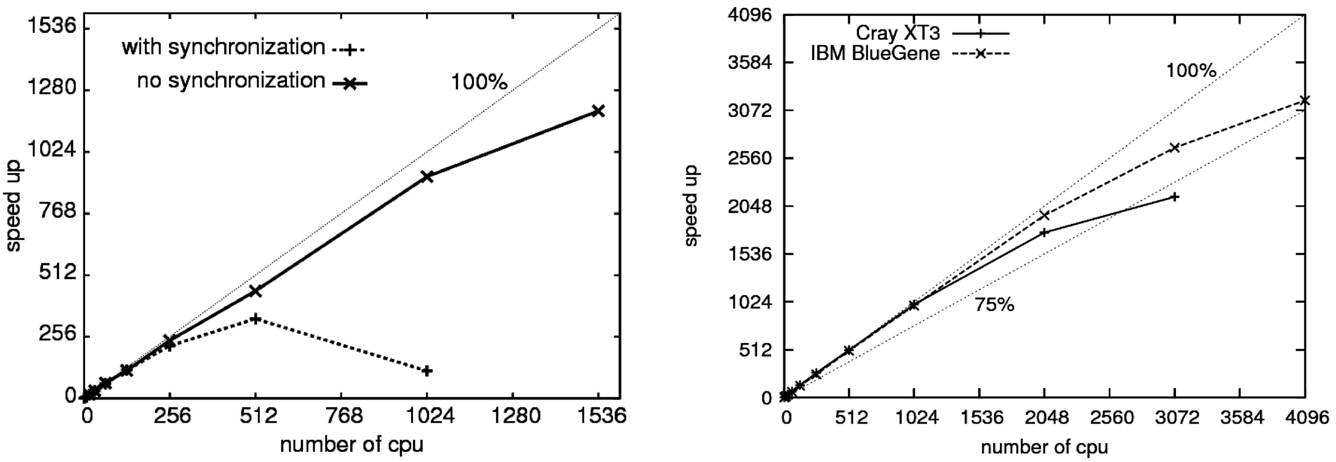


Figure 12. Scalability of the UNRES MREMD code. The left panel shows speed up curves calculated for MREMD using an AphaServer supercomputer with all replicas running for the same number of steps between exchanges (dashed line) and an improved code with no synchronization and exchange based only on the number of steps performed by the first replica (solid line). The dotted line shows perfect 100% speed up. The right panel shows speed up curves calculated for MREMD with no synchronization using a Cray XT3 computer (solid line) and IBM BlueGene (dashed line) supercomputers. Two dotted lines show 75% and 100% speedup lines for reference.

that the computed MFPTs do not have any values for folding kinetics because they correspond to all temperatures used in our simulations; they can be considered only as the ability of the system studied to reach the native structure given a simulation protocol.

Even if the total computational expense in the above simulations is the same for REMD and MREMD, MREMD finishes the simulations faster in real time by effectively using more processors for more replicas. This is especially attractive for UNRES simulations. For all-atom explicit solvent simulations, the large number of replicas that must be used to span a designated temperature range and the subsequent long time required for conformations sampled at high temperature to exchange for potential inclusion within the low temperature regime are seen as the main difficulties that are inherent in a REMD application. The large number of replicas is not a problem for united-residue (UNRES) simulations because the number of replicas scales as the number of degrees of freedom of the system and is several orders of magnitude smaller for UNRES when compared with all-atom simulations with explicit solvent.

MREMD simulations are more efficient than regular REMD simulations with the same number of replicas (each with a different temperature) because very closely spaced temperature exchanges must be made much more often for REMD. This difference between REMD and MREMD is illustrated in Figure 11, showing a comparison of diffusion in temperature replica space of REMD and MREMD in terms of temperature as a function of time for all replicas. REMD with 30 temperatures (Figure 11a) and MREMD with 30 temperatures and multiplexing of 8 replicas per temperature (Figure 11b) show good walks in temperature space, whereas for REMD with 240 temperatures (Figure 11c), each replica does not change its temperature very much from the starting one, and a clear monotonical pattern is visible even at the end of the simulations. By increasing the frequency of exchange 10 times for REMD with 240 temperatures (Figure 11d), we can obtain a proper random walk in temperature space, which is essential for REMD performance. The high frequency of exchange does not allow proper equilibration between exchanges and reduces parallel performance. Therefore, it is not practical to run REMD with so many replicas. MREMD does not involve these problems and can be used effectively even with thousands of replicas.

Parallel performance of the original REMD and MREMD code was limited by synchronization between all replicas on every exchange. As shown in Figure 12, this limitation restricts speedup for larger numbers of processors. When the restriction to perform exactly the same number of steps between exchanges for all replicas is lifted, and the exchange step is based only on the number of steps performed by the first replica, the improved algorithm scales almost linearly up to 4000 processors with over 75% speed up. In the improved algorithm, the exchanges of conformations between random replicas with neighboring temperatures are tried, not at the same number of MD steps for each replica but are forced by the replica with the lowest temperature, independent of the number of steps performed by the other replicas. With not too frequent exchanges, this algorithm allows for enough equilibration for each replica between exchanges. With processors of equal or closely spaced speed, the number of steps performed by each replica between exchanges does not vary too much, but even small variations remove waiting, which would reduce parallel performance very fast.

4. Conclusion

Replica exchange is the method of choice for studies of the thermodynamics of protein folding. Various thermodynamical properties are available as a function of temperature through histogram reweighting techniques (WHAM). Low free-energy minima are accessible through accelerated relaxation. Intrinsic parallelism of the algorithm can be extended effectively by multiplexing. Comparison of REMD versus MREMD shows that efficient sampling requires diffusion in temperature replica space; adding more temperature replicas means that the number of swaps grows quadratically and that either longer simulations are needed or exchanges must be attempted more frequently. The MREMD method takes advantage of both the multiple temperature aspect of REMD, as well as the large number of independent simulations to enhance sampling considerably. The simulation time should be long enough so that each trajectory can cover the entire conformational space, as well as the entire temperature space. Parallelization of the MREMD method has been enhanced by removing the synchronization step. Removing the restriction to perform exactly the same number of steps between exchanges has no consequence for the validity of the results in simulations performed using processors of equal or closely spaced speed because the number of steps performed by each replica between exchanges does not vary too much. These changes in the algorithm allow much better parallel performance. The improved algorithm scales almost linearly up to 4000 processors with over 75% of ideal speed up [i.e., (4000)(0.75) = 3000 time speed up].

Acknowledgment. This work was supported by grants from the NIH (GM-14312), the NSF (MCB05-41633), and Grant 0490/B/H03/2008/35 from the Polish Ministry of Science and Higher Education. This research was conducted by using the resources of (a) our Beowulf cluster at the Baker Laboratory, Cornell University, (b) the NSF Terascale Computing System at the Pittsburgh Supercomputer Center, (c) the John von Neumann Institute for Computing at the Central Institute for Applied Mathematics, Forschungszentrum Juelich, Germany, and (d) the Informatics Center of the Metropolitan Academic Network (ICMAN) in Gdańsk.

References

(1) Dobson, C. M. *Nature* **2003**, *426* (6968), 884–890.

(2) Cecconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S. *Science* **2005**, *309* (5743), 2057–2060.

(3) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57–83.

(4) Ueda, Y.; Taketomi, H.; Go, N. *Biopolymers* **1978**, *6*, 1531–1548.

(5) Cieplak, M.; Hoang, T. X.; Robbins, M. O. *Protein Struct. Funct. Genet.* **2002**, *1*, 104–113.

(6) Brown, S.; Fawzi, N. J.; Head-Gordon, T. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *19*, 10712–10717.

(7) Brown, S.; Head-Gordon, T. *Protein Sci.* **2004**, *4*, 958–970.

(8) Thirumalai, D.; Klimov, D. K. *Curr. Opin. Struct. Biol.* **1999**, *2*, 197–207.

(9) Ming, D. M.; Bruschweiler, R. *Biophys. J.* **2006**, *10*, 3382–3388.

(10) Liwo, A.; Oldziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *7*, 849–873.

(11) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Oldziej, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *7*, 874–887.

(12) Liwo, A.; Kazmierkiewicz, R.; Czaplewski, C.; Groth, M.; Oldziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. *J. Comput. Chem.* **1998**, *3*, 259–276.

(13) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *5*, 2323–2347.

(14) Liwo, A.; Oldziej, S.; Czaplewski, C.; Kozlowska, U.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *27*, 9421–9438.

(15) Oldziej, S.; Kozlowska, U.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. A* **2003**, *40*, 8035–8046.

(16) Kubo, R. *J. Phys. Soc. Jpn.* **1962**, *17*, 1100–1120.

(17) Liwo, A.; Arlukowicz, P.; Czaplewski, C.; Oldziej, S.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *4*, 1937–1942.

(18) Oldziej, S.; Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *43*, 16934–16949.

(19) Oldziej, S.; Lagiewka, J.; Liwo, A.; Czaplewski, C.; Chinchio, M.; Nanias, M.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *43*, 16950–16959.

(20) Oldziej, S.; Czaplewski, C.; Liwo, A.; Chinchio, M.; Nanias, M.; Vila, J. A.; Khalili, M.; Arnautova, Y. A.; Jagielska, A.; Makowski, M.; Schafroth, H. D.; Kazmierkiewicz, R.; Ripoll, D. R.; Pillardy, J.; Saunders, J. A.; Kang, Y. K.; Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *21*, 7547–7552.

(21) Liwo, A.; Khalili, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *7*, 2362–2367.

(22) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *28*, 13785–13797.

(23) Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *28*, 13798–13810.

(24) Khalili, M.; Liwo, A.; Scheraga, H. A. *J. Mol. Biol.* **2006**, *3*, 536–547.

(25) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *6*, 1604–1608.

(26) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *1−3*, 140–150.

(27) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *1−2*, 141–151.

(28) Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2003**, *2*, 775–786.

(29) Earl, D. J.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2005**, *23*, 3910–3916.

(30) Lei, H. X.; Duan, Y. *Curr. Opin. Struct. Biol.* **2007**, *2*, 187–191.

(31) Zuckerman, D. M.; Lyman, E. *J. Chem. Theor. Comput.* **2006**, *2*, 1200–1202.

(32) Zuckerman, D. M.; Lyman, E. *J. Chem. Theor. Comput.* **2006**, *2*, 1693–1693.

(33) Nanias, M.; Czaplewski, C.; Scheraga, H. A. *J. Chem. Theor. Comput.* **2006**, *3*, 513–528.

(34) Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **1992**, *12*, 9412–9426.

(35) Czaplewski, C.; Oldziej, S.; Liwo, A.; Scheraga, H. A. *Protein Eng., Des. Sel.* **2004**, *1*, 29–36.

(36) Chinchio, M.; Czaplewski, C.; Liwo, A.; Oldziej, S.; Scheraga, H. A. *J. Chem. Theor. Comput.* **2007**, *4*, 1236–1248.

(37) Lee, J.; Scheraga, H. A.; Rackovsky, S. *J. Comput. Chem.* **1997**, *9*, 1222–1232.

(38) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Oldziej, S.; Wachucik, K.; Scheraga, H. A. *J. Phys. Chem. B* **2007**, *1*, 260–285.

(39) Hagen, M.; Kim, B.; Liu, P.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* **2007**, *6*, 1416–1423.

(40) Shen, H.; Czaplewski, C.; Liwo, A.; Scheraga, H. A. *J. Chem. Theory Comput.* **2008**, *8*, 1386–1400.

(41) Green, D.; Meacham, K.; Hoesel, F.v. In *Parallelisation of the Molecular Dynamics Code GROMOS87 for Distributed Memory Parallel Architectures*; HPCN Europe 1995: Proceedings of the International Conference and Exhibition on High-Performance Computing and Networking; Springer-Verlag: London, 1995; pp 875−879.

(42) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *8*, 1011–1021.

(43) Bateman, A.; Bycroft, M. *J. Mol. Biol.* **2000**, *4*, 1113–1119.

(44) Dai, Q. H.; Tommos, C.; Fuentes, E. J.; Blomberg, M. R. A.; Dutton, P. L.; Wand, A. J. *J. Am. Chem. Soc.* **2002**, *37*, 10952–10953.

(45) Daggett, V. *Curr. Opin. Struct. Biol.* **2000**, *2*, 160–164.

(46) Day, R.; Daggett, V. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *38*, 13445–13450.

(47) Fersht, A. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *22*, 14122–14125.

CT800397Z

JCTC Journal of Chemical Theory and Computation

Graph Measures Reveal Fine Structure of Complexes Forming in Multiparticle Simulations

Florian Lauck,[‡] Volkhard Helms,* and Tihamér Geyer*

Zentrum für Bioinformatik, Universität des Saarlandes,
D-66041 Saarbrücken, Germany

Received September 23, 2008

Abstract: Modern simulation techniques are beginning to be used to study the dynamic assembly and disassembly of multiprotein systems. In these many-particle simulations it can be very tedious to monitor the formation of specific structures such as fully assembled protein complexes or virus capsids above a background of monomers and partial complexes. However, such analyses can be performed conveniently when the spatial configuration is mapped onto a dynamically updated interaction graph. On the example of Monte Carlo simulations of spherical particles with either isotropic or directed mutual attractions, we demonstrate that this combined strategy allows for an efficient and also detailed analysis of complex formation in many-particle systems.

Introduction

Proteins are the main building blocks and functional units of biological cells and are involved in signal transduction, energy metabolism, immune response, and the processing of DNA and RNA. Interestingly, many individual proteins do not work independent of each other. About half of them form protein complexes that consist of two or more proteins, which can be either identical or different in size, shape, and function.[1,2] Thus, a chain of metabolic reactions catalyzed by the components of a multienzyme complex can be executed in a shorter time, because intermediate substrates need not to diffuse through the cell. They will simply be handed over from one enzyme to the next one.[3,4] Another functional advantage of organizing single proteins into complexes is their modularity. A certain subunit can be used within more than one complex to accomplish the same task.[2,5]

In the past, the formation of protein complexes was first studied on a microscopic level, where the details of the binding interfaces of the proteins were investigated experimentally by X-ray crystallography, chemical shift mapping, or site-directed mutagenesis or theoretically with docking algorithms at atomic resolution.[6,7] For the modeling of larger protein assemblies, combinatorial methods have been developed, which can efficiently find the best configuration built from a number of building blocks.[8] Parallel to these static docking approaches, explicitly time-dependent simulations were used to investigate complex formation. These range from atomistic Brownian dynamics simulations of the association of two small proteins [9−12] up to coarse-grained simulations of protein pairs with reaction patches[13] or of virus capsids.[14,15] The largest system assembled sofar by simulation techniques is the nuclear pore complex consisting of 456 protein units in yeast.[16] While it is not a problem to interpret the results of simulations with only two or three particles, it can become a tedious and computationally demanding task to monitor, whether and when complexes with tens to hundreds of constituents are formed correctly. This problem becomes even more pronounced when more realistic simulations of large numbers of different proteins are considered, where more than one complete complex can be formed or where the complete complex is in a dynamic equilibrium with its components. In such a simulation, partial complexes of various sizes may be found together with complete ones and even with complexes which are assembled incorrectly.[17]

Recently, large databases have become available that provide convenient access to experimental and computationally generated data sets on protein−protein interactions

---

* Corresponding author e-mail: volkhard.helms@bioinformatik.uni-saarland.de; tihamer.geyer@bioinformatik.uni-saarland.de.

[‡] Current address: Department of Biopharmaceutical Sciences, University of California, San Francisco, CA 94143.

10.1021/ct800396v CCC: $40.75

(PPI).[18,19] By neglecting the atomistic details of the components, protein interaction networks are being generated from this data which then can be analyzed with graph algorithms and statistical methods developed for networks.[20,21] In these networks, each protein type typically denotes one node (or vertex), and a known interaction between two proteins is a link. Thus, it is relatively easy to identify a specific pattern corresponding to a certain multiprotein complex. For situations such as the one described above, where one wants to identify a few target complexes in a sea of monomers and partially assembled intermediates, we will show that this task can be conveniently performed by mapping the spatial simulation onto a protein interaction graph and subsequent analysis by well-known efficient tools. This protein interaction network built from the simulation differs from the usual PPI networks, however, because each of its nodes now denotes an individual copy of a protein type. Additionally, the graph is generated dynamically. This means that links appear and disappear over time, as the proteins bind and unbind from each other during the spatial simulation.

Similar considerations arise when, for example, the formation of van der Waals clusters is studied. Here, the objective may be to identify specific “magic” configurations or deviations thereof.[22,23] These magic configurations are characterized by their high degree of symmetry, which in turn leads to easily identifiable patterns in the interaction network.

Here we aim to demonstrate the advantages of combining these two views on particle interactions: the spatial view from the simulation and the network view via the associated graph. For this, a toy system of spherical particles with either isotropic or directed interactions was simulated with a standard Monte-Carlo algorithm and concurrently mapped onto a graph which reflects the current interactions between the particles. To avoid that complexes are essentially fixed in space, the graph was also used to identify the next neighbors of a displaced particle, which are then also displaced by a fraction of this move. This idea can even be taken further so that each complex is treated as one rigid pseudoparticle. This in turn would allow reduction of the number of particles in the simulation and thus speed it up significantly.

Methods

Spatial Simulation. The spatial simulation uses a standard Metropolis Monte-Carlo (MC) approach to model the diffusion of the particles in a stochastic manner. In the canonical *NVT* ensemble the probability p_j for a specific state j of the system is

$$p_j = \frac{e^{-\beta E_j}}{Z} \tag{1}$$

with the energy E_j of this state, the thermal energy $\beta^{-1} = k_B T$ with the Boltzmann constant k_B and the temperature T, and the partition function $Z = \sum e^{-\beta E_j}$. We used a rejection sampling algorithm which creates a new state $j+1$ of the system with the new energy E_{j+1} by displacing one of the particles at a time. In the Metropolis MC scheme, the transition probability into $j+1$ is defined as $a = p_{j+1}/p_j$ which can also be written as

$$a = e^{-\beta(E_{j+1} - E_j)} \tag{2}$$

by inserting eq 1. In the following sections, energies are given in units of the thermal energy $k_B T$, and lengths are given in dimensionless arbitrary units (a.u.).

Interactions and Bonds. Without external forces, the total energy of the simulated system is the sum of all particle−particle interactions. The interaction energy V_W (W for well) between two isotropic spheres of radius σ and center−center distance r was modeled by a repulsive core and an attractive well potential of width $r_C - \sigma$ and depth ε:

$$V_W(r_{ij}) = \begin{cases} \left(100 + 500\dfrac{\sigma}{r_{ij}}\right) k_B T & \text{if } r_{ij} \le \sigma \\ -\varepsilon & \text{if } \sigma < r_{ij} \le r_C \\ 0 & \text{if } r_{ij} > r_C \end{cases} \tag{3}$$

In the examples presented here, the particles had a diameter of $\sigma = 6$ a.u. and $r_C = 7$ a.u. In simulations performed with this interaction, a link between particles i and j was added when $r_{ij} < r_C$.

To model attractive ring patches, the well potential V_W was multiplied with an angle-dependent term as in Wilber et al.:[24]

$$G_{ij}(r_{ij}, \theta_{ij}) = \exp\left[\frac{(\theta_{ij} - \nu)}{2\sigma_{PW}^2}\right] \tag{4}$$

Here, θ_{ij} denotes the angle between the interparticle vector r_{ij} and the vector Ω_i describing the orientation of particle i. ν is the opening angle of the ring, i.e., the angle between Ω_i and the ring, while σ_{PW} describes the angular patch width. In this case, the pairwise potential $V_R(r_{ij})$ between two particles i and j was defined as

$$V_R(r_{ij}) = V_W(r_{ij}) \times G_{ij}(r_{ij}, \theta_{ij}) \times G_{ji}(r_{ji}, \theta_{ji}) \tag{5}$$

In simulations using this ring potential, a link was added in the graph for $V_R(r_{ij}) < 0.4 k_B T$.

Network Measures. The number of individual interactions shorter than r_C equals the number of links in the graph and can be derived from both the spatial and the graph domain. For point particles, the theoretical maximal number of interactions between N particles is $N(N - 1)/2$, because interactions are bidirectional and self-interactions are excluded. For particles with nonzero radius that may not overlap, as were used here, the maximal number of interactions is smaller.

A connected component of an undirected graph is a set of vertices that are all reachable from each other. This means that two vertices are in the same connected component if and only if there exists a path between them. The connected components distribution or complex size distribution shows how many complexes of a certain size exist in a graph. The degree distribution $P(k)$ gives the probability that a selected vertex has exactly k links. It is defined as:

$$P(k) = \frac{N(k)}{N} \tag{6}$$

where $N(k)$ is the number of vertices having degree k, and N is the total number of vertices in the graph. The clustering coefficient C_{k_i} of a node i is a measure for the connectivity in the neighborhood of i. It is calculated from the number of links that interconnect the neighbors of i, n_i, and the degree of i, k_i:

$$C_{k_i} = \frac{2n_i}{k_i(k_i - 1)} \tag{7}$$

The function $C(k)$ is defined as the average clustering coefficient of all nodes with degree k. The distance distribution $D(l)$ gives the number of shortest paths[25] of a certain length l.

An overview over commonly used network measures can be found, for example, in a review by Costa et al.[26] The network analysis of the spatial simulations was implemented using the boost graph library.[27]

Simulation Setup. The simulation runs modeled the particle diffusion in a cubic box with periodic boundary conditions. During initialization, each particle was assigned a random position and a random orientation. During one MC iteration, a trial move was generated for one particle after the other, and it was checked whether the new state of the system should be accepted. If the new move was accepted and interactions changed relative to the previous configuration, the interaction graph was updated correspondingly. After all particles had the chance to move once, the interaction graph was analyzed with respect to the properties of the complete graph such as the total number of links or its average connectivity. These measures were used to monitor the convergence of the simulation. Moreover, the connected components were determined, which were individually analyzed for their size and connectivity in a second step. At that time, we also searched for specific patterns which allow identification of, for example, completely assembled icosahedra. Additionally, conventional properties of the simulation were determined as its total energy or the radius of gyration.

Results and Discussion

As introduced before, this study aims at introducing a new way of studying particle interactions in a spatial simulation. The particle simulations themselves are standard MC simulations, and the graph analysis is done by well-established mathematical graph algorithms. The novel aspect introduced in this study is the dynamic updating and subsequent analysis of the interaction graph.

Spherical Particles. A first impression about the monitoring of association processes by graph measures will be given on the example of simulations with 200 particles that used a spherically symmetric attractive well potential of different depths $\varepsilon = 0...-4k_BT$. The simulation box had a volume of (70 a.u.)^3 with periodic boundary conditions.

Commonly, the convergence of simulations is monitored by the time course of the total energy of the system. Here the energy was found to converge exponentially, as complexes formed from the initially randomly distributed particles. Figure 1a shows the equilibrium value of this real space measure vs the well depth ε (filled triangles). As expected, for $\varepsilon < 1k_BT$, the total energy was only slightly smaller than zero, meaning that only a few particles were bound by the weak attractive potential. Around $\varepsilon = 1k_BT$, the total energy strongly decreased, indicating a transition from independent particles to a large cluster. This transition into a single large cluster is also clearly visible in the behavior of the radius of gyration. For small ε, its value corresponds to independent particles distributed over the volume of the simulation box. For $\varepsilon > 2k_BT$ its value of $\approx 5\sigma$ indicates a densely packed cluster which essentially does not get any denser once ε is large enough.

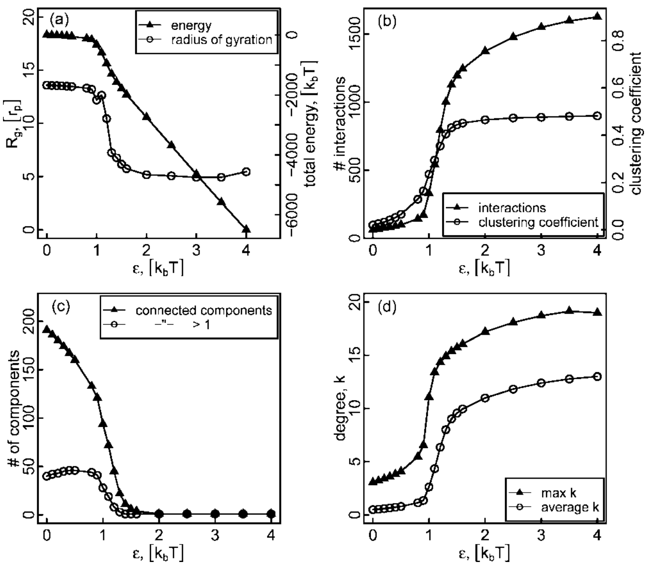


Figure 1. Complex formation monitored via real-space and graph measures. Results from simulations of 200 particles with a spherical attractive well potential for different well depths ε. All graphs show a transition around $\varepsilon = 1k_BT$, above which the particles form a large complex. The data points were connected by lines as guides to the eye. (a) Real space measures total energy (filled triangles) and radius of gyration (circles). (b) Number of interactions, i.e., links in the network derived from the cluster (triangles) and average clustering coefficient (circles) of the complete graph. (c) Total number of connected components (triangles) and number of connected components containing more than one particle (circles). (d) Maximum (triangles) and average (circles) degrees of the nodes of the cluster's graph.

Of course, the graph measures derived from the interaction graph describe the same behavior; it is the very same set of simulations. The measures shown in Figure 1b, the number of interactions and the average clustering coefficient, are the analogues to the total energy and the radius of gyration. The number of interactions again is small for small ε, increases strongly around $\varepsilon = 1k_BT$, and then saturates because the cluster cannot be compacted any further once the most dense packing is reached. The increase of the number of interactions for $\varepsilon > 1.5k_BT$ can be attributed to the behavior of the particles on the surface of the cluster, which try to penetrate deeper into the cluster, thus making the surface smoother. The clustering coefficient also increases around $\varepsilon = 1k_BT$ before it saturates, indicating the dense packing inside the cluster.

Whereas the measures of panels a and b put the emphasis on the central large cluster at $\varepsilon > 1k_BT$, Figures 1c and 1d highlight the behavior of the particles for weak attractions. According to panel c, the total number of connected components decreases monotonically with increasing ε. For $\varepsilon \leq 1k_BT$, the number of connected components of size larger than one stays roughly constant. In this regime the simulation box contains many small complexes with average sizes from two to four particles. At increasing ε they start to collapse into large clusters until for $\varepsilon \geq 2k_BT$, there is only one large component (complex) left including all 200 particles.

This behavior is also reflected by the average and maximal degrees of the particles shown in Figure 1d: a maximal degree of three for small ε indicates a complex of size four. In a densely packed hexagonal three-dimensional cluster, a particle has 12 direct neighbors. With the width of the attractive potential chosen for these simulations, also the second neighbors may be counted as bonded if the cluster is packed densely enough. This leads to the maximal degree of 18 for large ε. Again, the clustering transition can be seen clearly.

These simple graph measures can be computed from the simulations with little additional overhead. When compared to an analysis in coordinate space alone, they readily provide a wealth of insight into the behavior of the particles at weak interactions, about the clustering transition, and about the internal structure of the single complex for strong attraction. Certainly, particles with a spherically isotropic attraction are an extremely simple toy system, which can also be fully understood without the graph analysis. When more structured interactions lead to the formation of specific spatial structures as is the case in the following two examples, the advantages of including the graph analysis will become more obvious.

Particles with Equatorial Ring Patches. When the attraction between the particles was restricted to their equatorial region, membrane-like two-dimensional structures developed as expected. In the following example, an additional attractive well potential similar to eq 3 with $r_C = 4$ a.u. and a depth of $\varepsilon_W = 4k_BT$ was applied between the particles and the bottom wall of the simulation box. The simulation box had a volume of (100 a.u.)^3. Periodic boundary conditions were applied in the other two dimensions, and the top wall of the box opposite of the attractive surface simply reflected the particles. This setup mimics the formation of ordered layers of, for example, proteins on solid surfaces.[28]

Figure 2 shows four snapshots from a spatial simulation with 200 particles. Initially, the particles formed small clusters or adsorbed to the bottom wall (top panels). However, because of construction, the clusters in solution were less stable than those adsorbed to the wall. Consequently, these small clusters eventually fell apart because of thermal fluctuations, and all particles were finally attached to the bottom wall. There, their mutual interaction led to the formation of a hexagonally structured layer. The apparent gap between the two seed clusters seen in the third snapshot at 10^5 iterations was finally filled by the remaining free particles. Although the last snapshot still shows a clear grain boundary between the two domains originating from the seed clusters, this boundary finally disappeared because of reorganization of the particles, and a single, hexagonally ordered layer of adsorbed particles was formed (snapshot not shown).

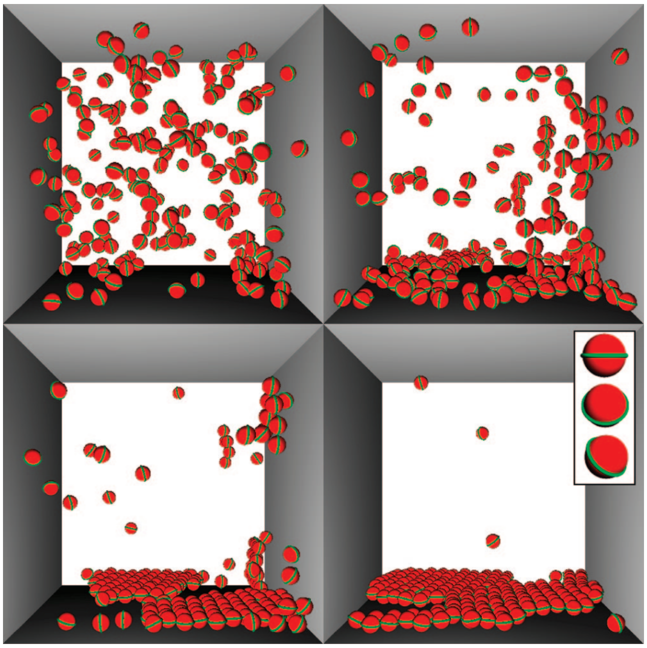


Figure 2. Formation of an ordered layer, visualized in the spatial domain: Snapshots after 10^3, 10^4, 10^5, and 2×10^5 iterations (top left to bottom right) from a simulation of 200 particles with attractive equatorial patches and an adhesive bottom wall of the simulation box. The inset in the last snapshot shows a close-up view of one of the particles in three different orientations. The ring indicates the region where two particles attract each other. According to the given interactions, the particles adsorb to the surface as an ordered layer with a hexagonal structure.

A much richer understanding of the dynamic processes during the formation of the surface layer is obtained when the measures derived from the interaction graph are considered, too. Figure 3a shows the sizes of the connected components that occur during the equilibration phase of the simulation. A dot denotes that at least one cluster of the corresponding size was present at a given iteration. This statistical representation illustrates several trends: (i) from the many small clusters that form in the beginning, a number of clusters of sizes around 30 particles emerged. (ii) Many of these intermediate-size clusters slowly shrank again. (iii) Several clusters constantly grew in size, finally forming one large adsorbate. The stepwise increase of these clusters during the first 10^4 iterations suggests that several seeds quickly associated into a single cluster that then grew by attracting individual particles from the bulk. (iii) The fraction of the small few-particle clusters in the bulk constantly decreased both in cluster size and number.

The hexagonal structure shown in Figure 2d results in a special signature in the degree distribution $P(k)$, as illustrated in Figure 3. Here, $P(k)$ was averaged over the simulation after the equilibration phase, i.e., between iterations 2×10^5 and 10^6. Most of the graph nodes had $k = 6$ as in a two-dimensional hexagonal lattice. The nodes with $k < 6$ are the ones with missing neighbors found on the rims of the adsorbed layer. The few nodes with $k = 7$ are particles in

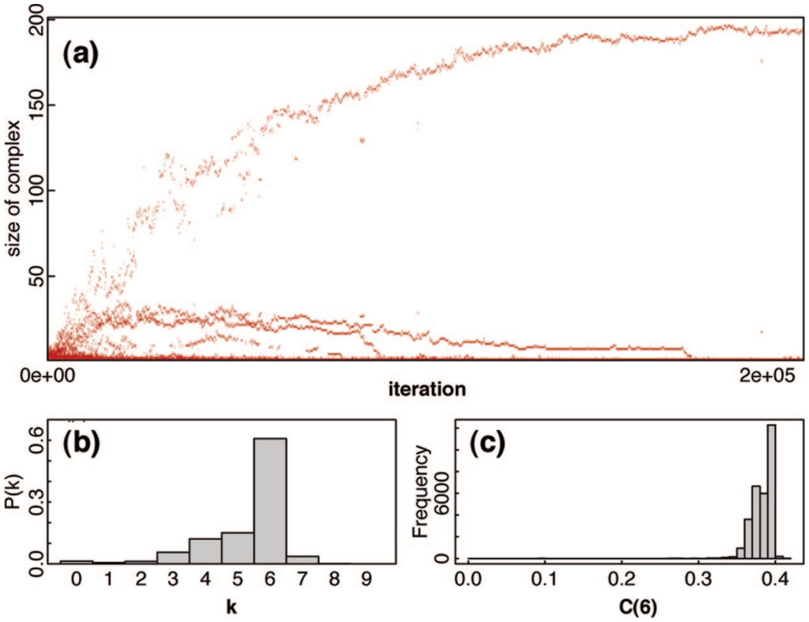


Figure 3. Formation of an ordered layer, visualized in the network domain. (a) Sizes of the connected components that formed during the equilibration phase of the simulation shown in Figure 2. (b) Averaged degree distribution and (c) average clustering coefficient distribution of the particles with $k = 6$ during the steady state after 2×10^5 iterations.

the region of the grain boundary where the hexagonal lattice was disturbed.

The graph analysis also allows focusing on a certain group of particles, as shown for the clustering coefficient of the nodes with $k = 6$ in Figure 3c. In a perfect hexagonal lattice, all nodes have $C(6) = 0.4$. This value also occurred most frequently in the equilibrated particle layer. However, because of the rearrangement dynamics around the grain boundary, some of the six neighbors of a certain particle sometimes separated too far from each other so that no bond was established and the connectivity in this region was reduced. One could now use this information to identify these less interconnected and thus less stable regions of the cluster and examine their dynamics in more detail.

Compared to the first example, the graph analysis in this slightly more complex scenario provides additional insights into the dynamics of the particle adsorption. These would be hard to obtain and even harder to visualize from the spatial view alone. Together with the few snapshots of Figure 2, the cluster size evolution of Figure 3 provides an intuitive visualization of the dynamics of the growth and shrinking stages of the clusters. For example, around iteration 8×10^4, the large cluster, consisting of about 160 particles at that time, temporarily fell apart into a larger and a smaller part of 130 and 30 particles, respectively. Such events had occurred already earlier, too, around iteration 5×10^4. However, here the smaller fragment is not so well visible in Figure 3a among the other smaller clusters.

Icosahedral Complexes. The previous example showed how graph measures can be used to collect and visualize detailed information about the dynamics during structure formation and how regular structures (and irregularities therein) can be easily identified and interpreted. The last application presented here is the formation of icosahedral complexes. In this case, the objective is to identify these highly regular structures among all the particles in the simulation, to monitor their formation, and to assess their quality.

With the opening angle of the attractive ring patch set to 58°, one particle can bind to five other particles, which leads to the formation of icosahedral complexes of twelve particles.[24] A spatial representation of a completely assembled icosahedron and its graph representation are shown in the second row of Figure 4. Each of the particles has exactly five neighbors. Consequently, the degree distribution is zero everywhere except for $P(k = 5) = 1$ (see third column of Figure 4). Among the five neighbors, five of the ten possible links are established, leading to a clustering coefficient of $C(5) = 1/2$ for each of the particles (fourth column). A third signature of an icosahedron is its distribution of path lengths, which consists of 30 paths of lengths one and two, respectively, and six paths of length three that connect the pairs of the diametrically opposite particles (last column).

Any perturbation in the assembly of an icosahedron either changes the number of particles or their connectivity. This in turn is reflected in perturbations of $P(k)$, $C(k)$, and the distribution of path lengths. A correct icosahedron can thus be identified by the correct size and the correct network measures. In contrast to the correctly assembled icosahedron from the second row, the third row of 4 illustrates the signatures of a slightly misfolded icosahedral complex. Here, $P(k)$ has an additional nonzero entry for $k = 4$ and $P(5) < 1$. The $C(k)$ and $D(l)$ distributions are perturbed, too. Because of the partly opened structure, some shortest paths have the length four. These different behaviors allow easy and unambiguous assess to the quality of an icosahedron, even if the spatial representation seems correct. Figure 4 also shows two examples of incorrectly assembled complexes of $N = 8$ and $N = 16$ particles, respectively.

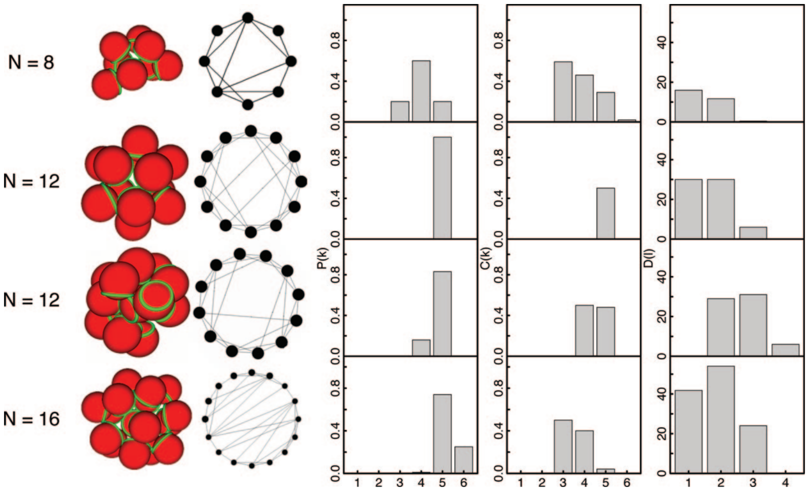


Figure 4. Correctly assembled clusters can be identified via their network signatures. Examples of complexes found in the simulations of icosahedron formation. The spatial representation of the chosen complex is shown in the first column, the corresponding network graph in the second column. The following columns give the degree distribution *P*(*k*), the degree-averaged clustering coefficient distribution *C*(*k*), and the distribution of path lengths *D*(*l*), respectively. The second row belongs to a correctly assembled icosahedron with *N* = 12 particles, while the third row shows a slightly misfolded icosahedron. The first and last rows belong to complexes of *N* = 8 and *N* = 16 particles, respectively.

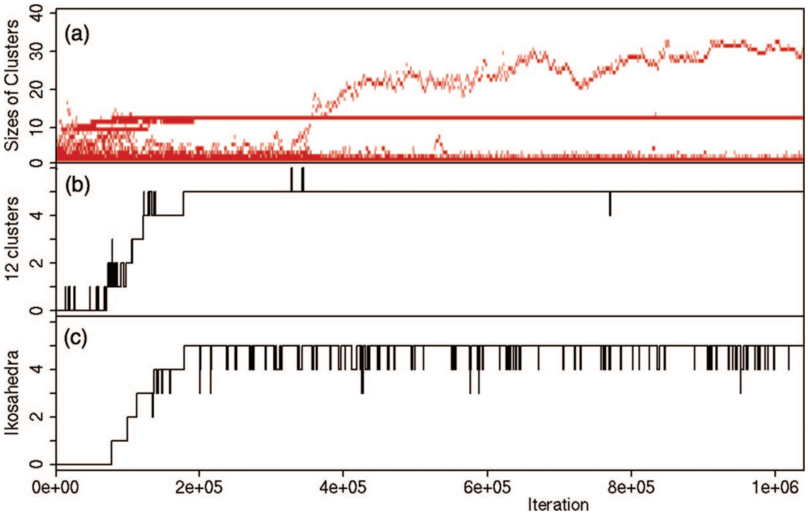


Figure 5. Formation of clusters of size 12 and of icosahedra during a simulation of 100 particles with attractive ring patches with a 58° opening angle. (a) Sizes of the formed clusters, (b) number of clusters of size 12, and (c) number of correctly assembled icosahedra.

With respect to the dynamics of complex formation, the first signature of icosahedron formation during a simulation is the occurrence of connected components of size 12, see Figure 5a. This simulation was performed with 100 particles in a box of size (70 a.u.)^3 with periodic boundary conditions. Obviously, this number of particles would allow for the formation of up to eight complete icosahedra. However, it was unclear whether this maximum number was kinetically accessible on a typical simulation time scale. Here, five to six complexes of size 12 assembled during the first 10^5 iterations, see Figure 5b. From the remaining fragments, a large unstructured complex of about 30 particles formed at a later stage. For each of the complexes of size 12, we computed *P*(*k*), *C*(*k*), and *D*(*l*) and compared them to the distributions expected for a correct icosahedron (cf. Figure 4). This filtering gave four to five correct icosahedra (see Figure 5c). Comparing panels b and c shows that there were essentially five icosahedra present in the simulation run. Because of the random displacements, one of the monomers sometimes left its correct position in the complex by breaking some of the bonds. However, these deviations were not strong enough to break the complex apart completely. Note that the nearly correct icosahedron from Figure 4 (third row) was taken from this simulation.

Using Information from the Graph for Cooperative Moves in the Simulation. A shortcoming of the Monte-Carlo approach used so far is that the particles are moved independently. With the interaction potential used, moving a nonbonded particle does not change the total energy, and this new state is always accepted. Thus, the particles diffusively move through the simulation volume as long as they are not bonded to any other particle. But as soon as they are bonded to one or more of the other particles, any

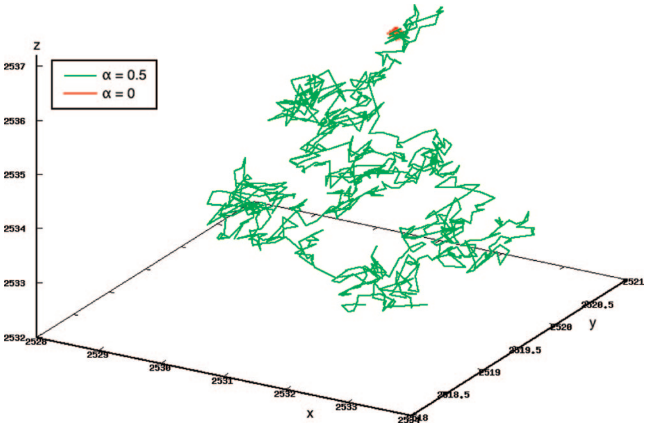


Figure 6. Diffusion of complexes is restored when the Monte-Carlo trial moves are propagated through the network. Trajectory of an icosahedron with (green, $\alpha = 0.5$) and without (red, $\alpha = 0$) coupling of the Monte-Carlo moves to the next neighbors. When only one particle is moved at a time, the cluster itself remains fixed in space, whereas a diffusive movement is observed when a fraction α of the random displacements of the Monte-Carlo algorithm are propagated through the interaction network and the connected next neighbors are displaced, too. The trajectories shown here are representative parts of simulations of twelve particles, initialized to form a correctly assembled icosahedral complex.

displacement of the chosen particle is strongly suppressed, because this would involve the breaking of bonds and the corresponding increase of the energy. Hence, complexes become essentially fixed in space with such a one-particle-at-a-time updating scheme.

This artificial behavior can be elegantly overcome by making use of the information from the graph about the connected components. Our idea was to treat each formed complex as one pseudoparticle that is propagated as a whole. Keeping reality in mind, we suggest that when one particle in a complex is displaced, its neighbors should also be dragged along via the bonds. With the information from the graph, it is now straightforward to not only displace a single particle in the spatial Monte-Carlo algorithm, but to also propagate this displacement along the bonds. Here we implemented a simple variant, where the next graph neighbors of the chosen particle are also displaced by a fraction $\alpha < 1$ of the test move.

Figure 6 shows how such a propagation of the trial moves to the next neighbors with $\alpha = 0.5$ makes a complete icosahedral complex diffuse through the simulation volume (green trajectory). With $\alpha = 0$, i.e., when only a single particle is displaced at a time, the center of mass of the complex only exhibits small fluctuations around its otherwise fixed position (red trajectory). This implementation could be extended by considering not only nearest neighbors but also second-nearest neighbors up to an arbitrary distance.

When cluster formation is considered, the coupling effectively reduces the relative displacements of neighboring particles. This has the same effect as slightly extending the range of the interaction potential. Consequently, with increasing α, clusters form already at smaller potential depths ε (data not shown).

Summary

In summary, we have shown how a dynamically updated graph can be used conveniently to analyze and observe complex formation during a spatial simulation. Since the graph analysis is independent of the simulation, the Monte Carlo approach chosen here for simplicity can be replaced by a different model, as for instance Brownian dynamics or force field simulations. The criterion for placing an interaction surely needs to be adjusted appropriately. Then, for each pair of particles, a separate meaningful definition of an interaction has to be chosen based for example on a maximal relative distance and possibly some angles measuring their relative orientation and/or on a threshold to the interaction energy.

On the graph side, it appears worthwhile to explore the usefulness of more advanced graph measures in the future. For simulations at higher concentrations, for example, the complete complexes start to interfere with each other. These interacting complexes could be distinguished by means of a biconnected or k-connected components analysis as well as a cluster search when a simple analysis of connected components would fail.[26,29] Particles and interactions that connect two clusters can then be identified via their betweenness.[29] For simulations of more advanced particles that form clusters with a specific shape, a motif search as used by Pržulj et al.,[30] Middendorf et al.,[31] or Baskerville et al.[32] could be applied.

The results presented here focused on problems with a known geometry of the correctly formed complex. Moreover, the first example showed that the combination of spatial simulation and network analysis may also be advantageous in situations where the final structure is not known a priori or when, such as in the rather amorphous focal adhesions, a single "correct" configuration does not even exist. Then, the network analysis can be used to identify and to separately monitor the convergence of the "clustered part" of the simulation. Comparing subsequent network snapshots additionally allows us to distinguish the more stable regions of such dynamic assemblies from the more dynamic growth regions.

References

(1) Aloy, P.; Böttcher, B.; Ceulemans, H.; Leutwein, C.; Mellwig, C.; Fischer, S.; Gavin, A.; Bork, P.; Superti-Furga, G.; Serrano, L.; Russell, R. B. *Science* **2004**, *303*, 2026–2029.

(2) Gavin, A. C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L. J.; Bastuck, S.; Dumpelfeld, B.; Edelmann, A.; Heurtier, M. A.; Hoffman, V.; Hoefert, C.; Klein, K.; Hudak, M.; Michon, A. M.; Schelder, M.; Schirle, M.; Remor, M.; Rudi, T.; Hooper, S.; Bauer, A.; Bouwmeester, T.; Casari, G.; Drewes, G.; Neubauer, G.; Rick, J. M.; Kuster, B.; Bork, P.; Russell, R. B.; Superti-Furga, G. *Nature* **2006**, *440*, 631–636.

(3) Milne, J. L. S.; Shi, D.; Rosenthal, P. B.; Sunshine, J. S.; Domingo, G. J.; Wu, X.; Brooks, B. R.; Perham, R. N.;

Henderson, R.; Subramaniam, S. *EMBO J.* **2002**, *21*, 5587–5598.

(4) Elcock, A. H.; McCammon, J. A. *Biochemistry* **1996**, *35*, 12652–12658.

(5) Krause, R.; von Mering, C.; Bork, P.; Dandekar, T. *BioEssays* **2004**, *26*, 1333–1343.

(6) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195–2199.

(7) Halperin, I.; Ma, B. Y.; Wolfson, H.; Nussinov, R. *Proteins* **2002**, *47*, 409–443.

(8) Inbar, Y.; Benyamini, H.; Nussinov, R.; Wolfson, H. J. *J. Mol. Biol.* **2005**, *349*, 435–447.

(9) Northrup, S. H.; Boles, J. O.; Reynolds, J. C. L. *Science* **1988**, *241*, 67–70.

(10) Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **1997**, *72*, 1917–1929.

(11) Elcock, A. H.; Gabdoulline, R. R.; Wade, R. C.; McCammon, J. A. *J. Mol. Biol.* **1999**, *291*, 149–162.

(12) Spaar, A.; Dammer, C.; Gabdoulline, R. R.; Wade, R. C.; Helms, V. *Biophys. J.* **2006**, *90*, 1913–1924.

(13) Schluttig, V.; Alamanova, D.; Helms, V.; Schwarz, U. S. *J. Chem. Phys.* **2008**, *129*, 155106.

(14) Berger, B.; Shor, P. W.; Tucker-Kellog, L.; King, J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7732–7736.

(15) Hagan, M. F.; Chandler, D. *Biophys. J.* **2006**, *91*, 42–54.

(16) Alber, F.; Dokudovskaya, S.; Veenhoff, L. V.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, A.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Rout, M. R.; Sali, A. *Nature* **2007**, *450*, 683–694.

(17) Zlotnick, A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5549–15550.

(18) BOND Web Portal. http://bond.unleashedinformatics.com (accessed Dec. 15, 2008).

(19) Xenarios, I.; Salwinski, L.; Duan, X. J.; Higney, P.; Kim, S.; Eisenberg, D. *Nucleic Acids Res.* **2002**, *30*, 303–305.

(20) Albert, R.; Barabási, A. L. *Rev. Mod. Phys.* **2002**, *74*, 47–97.

(21) Barabási, A.-L.; Oltvai, Z. N. *Nature Rev. Gen.* **2004**, *5*, 101–113.

(22) Grigoryan, V. G.; Alamanova, D.; Springborg, M. *Phys. Rev. B* **2006**, *73*, 115415.

(23) Alamanova, D.; Grigoryan, V. G.; Springborg, M. *J. Phys. Chem.* **2007**, *111*, 12577–12587.

(24) Wilber, A. W.; Doye, J. P. K.; Louis, A. A.; Noya, E. G.; Miller, M. A.; Wong, P. *J. Chem. Phys.* **2007**, *127*, 085106.

(25) Dijkstra, E. W. *Numer. Math.* **1959**, *1*, 269–271.

(26) Costa, L. d. F.; Rodrigues, F. A.; Travieso, G.; Boas, P. R. V. *Adv. Phys.* **2007**, *56*, 167–242.

(27) Boost C++ libraries. http://www.boost.org (accessed Dec. 15, 2008).

(28) Quinn, A.; Mantz, H.; Jacobs, K.; Bellion, M.; Santen, L. *Eur. Phys. Lett.* **2008**, *81*, 56003.

(29) Girvan, M.; Newman, M. E. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7821–7826.

(30) Pržulj, N.; Corneil, D. G.; Jurisica, I. *Bioinformatics* **2004**, *20*, 3508–3515.

(31) Middendorf, M.; Ziv, E.; Wiggins, C. H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3192–3197.

(32) Baskerville, K.; Grassberger, P.; Paczuski, M. *Phys. Rev. E* **2007**, *76*, 036107.

CT800396V

J. Chem. Theory Comput. **2009,** *5,* 649–658

JCTC Journal of Chemical Theory and Computation

Protein Influence on Electronic Spectra Modeled by Multipoles and Polarizabilities

Pär Söderhjelm,[†] Charlotte Husberg,[†] Angela Strambi,[‡] Massimo Olivucci,[‡] and
Ulf Ryde*,[†]

*Department of Theoretical Chemistry, Lund University, Chemical Centre,
P.O. Box 124, SE-221 00 Lund, Sweden, and Dipartimento di Chimica, Università di
Siena, via Aldo Moro 2, I-53100 Siena, Italy*

Received October 30, 2008

Abstract: We have developed automatic methods to calculate multipoles and anisotropic polarizabilities for all atoms and bond centers in a protein and to include such a model in the calculation of electronic properties at any level of quantum mechanical theory. This approach is applied for the calculation of the electronic spectra of retinal in rhodopsin at the CASPT2//CASSCF level (second-order multiconfigurational perturbation theory) for the wild-type protein, as well as two mutants and isorhodopsin in QM/MM structures based on two crystal structures. We also perform a detailed investigation of the importance and distance dependence of the multipoles and the polarizabilities for both the absolute and the relative absorption energies. It is shown that the model of the surrounding protein strongly influences the spectrum and that different models give widely different results. For example, the Amber 1994 and 2003 force fields give excitation energies that differ by up to 16 kJ/mol. For accurate excitation energies, multipoles up to quadrupoles and anisotropic polarizabilities are needed. However, interactions with residues more than 10 Å from the chromophore can be treated with a standard polarizable force field without any dipoles or quadrupoles.

Introduction

Electronic spectroscopy is one of the most important sources of information about biological systems. Therefore, one of the major goals in theoretical chemistry has been to obtain methods that accurately predict electronic spectra of chromophores in proteins. This is a formidable task, because it is well-known that the surrounding medium strongly affects the spectrum of a chromophore. Therefore, many attempts have been made to model such effects also in theoretical calculations. For pure solvents, detailed methods exist, which explicitly include the closest solvent molecules in the calculations using a combination of accurate quantum mechanical (QM) methods, detailed electrostatic models, and statistical simulation methods.[1] Thereby, the most important effects are included: electrostatics, polarization, dispersion, exchange repulsion, and dynamic effects.

For proteins, calculations have not reached this level yet, but many attempts have been made to include at least some of these effects, for example, by point-charge models[2−11] and multipole models[12] for the electrostatics, atomic isotropic polarizabilities for polarization,[13−17] extended QM systems (typically at the semiempirical level) to include some dispersion and exchange repulsion,[18] self-consistent reaction-field methods for solvation effects,[19] and sampling of many conformations.[3,14,20] However, few investigations have included several of those effects at the same time.[21] The reason for this is mainly the heterogeneous nature of a protein, which requires a very large number of parameters that vary with geometry. In this Article, we take a first step in this direction, by implementing QM methods to calculate a multicenter-multipole + anisotropic polarizabilities model for a full protein, as well as methods to include such a detailed model of the electrostatics and polarization into accurate ab initio

---

* Corresponding author phone: +46-46 2224502; fax: +46-46 2228648; e-mail: ulf.ryde@teokem.lu.se.

[†] Lund University.

[‡] Università di Siena.

10.1021/ct800459t CCC: $40.75

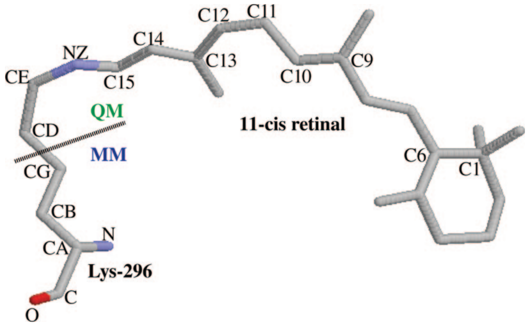


Figure 1. Structure of the protonated Schiff-base 11-cis retinal chromophore in rhodopsin with the quantum (QM) and molecular mechanics (MM) systems, as well as atom names marked out.

QM calculations. Moreover, we study the influence and distance dependence of the various terms in this model on both absolute and relative absorption energies. Such information can point out how much of the model needs to be recalculated if dynamics effects are included by repeating the calculations on several conformations.

As a test case, we use the retinal chromophore in the protein rhodopsin, which already has been the subject for many theoretical studies of the protein influence on the spectrum.[22] This G protein-coupled receptor is responsible for the vision in mammals, and the protein is responsible for the tuning of the same chromophore in the various color pigments in the cones of our eyes.[23] The 11-cis retinal chromophore is covalently bound to the Lys-296 residue (numbering according to the bovine enzyme) via a protonated Schiff-base linkage (Figure 1). After the absorption of a photon, the chromophore is converted to all-trans retinal, which triggers the neuronal signal.

In this Article, we study the wild-type protein, as well as two mutants: In the first, the counterion to the protonated Schiff-base, Glu-113 is converted to Asp, with one CH_2 group less, leading to a movement of a charged group close to the chromophore.[24] In the second, the small Gly-121 group, which forms van der Waals contact with the β-ionone ring of the chromophore, is converted to a much more bulky Leu residue.[25] Finally, we also study isorhodopsin, in which the 11-cis retinal chromophore is replaced with 9-cis retinal.[26] All calculations are based on previous QM/MM structures of the chromophore[10,11] in two different crystal structures[27,28] and a well-tested CASPT2 protocol for the spectra calculations.[29]

Methods

Structures. All calculations are based on previously determined QM/MM structures,[10,11] in which the chromophore, Lys-296, and two nearby water molecules were optimized by the CASSCF (complete active space self-consistent field) method and 6-31G* basis set, keeping the rest of the protein fixed at the crystal structure (and treated by the Amber 1994 force field[30]). Two different crystal structures were used, one 2.8 Å-resolution structure of bovine rhodopsin (1HZX)[27] and a newer structure of the same protein at 2.2 Å resolution (1U19).[28] Following the QM/MM calculations, these residues were charged: Met-1, Ala-348 (amino and carboxy terminals), all seven Arg residues, Lys-245 and 248, Asp-190 and 300, Glu-25, 113, 134, 150, 197, 232, 247, and 249, as well as His-195. All of these residues, except the counterion Glu-113, are solvent-exposed and ensure that the full protein is neutral. All of the other amino acids (including three Asp, nine Glu, and nine Lys residues) were kept neutral, in agreement with experiments for this membrane-bound protein.[31] For isorhodopsin, the chromophore, Lys-296, and the two water molecules were optimized, whereas in the Glu113Asp and Gly121Leu mutants, the mutated groups were also optimized.[10] The full retinal molecule, as well as a $CH_3NH=$ model of the Lys-296 protonated Schiff-linkage (from the NZ, HZ, CE, HE, and CD atoms; cf., Figure 1), were treated at the QM level, whereas the rest of Lys-296, the two water molecules, as well as Asp-113 or Leu-121 were treated at the MM level. Calculations with the 1U19 crystal structure were restricted to the wild-type protein and isorhodopsin.

Spectra Calculations. The electronic spectra were calculated at the CASPT2//CASSCF level (complete active space second-order perturbation theory), as has been described before.[10,29] The same QM system as in the QM/MM optimizations was used. The wave function was optimized at the CASSCF level (complete active space self-consistent field)[32] with the 6-31G* basis set and an active space consisting of 12 electrons in 12 active orbitals. An average over the three lowest states was optimized, but the polarization field for the lowest state was used. Next, the energy of the ground state was calculated by the CASPT2 method,[33,34] with the 22 core orbitals frozen and with a level shift of 0.2 to remove intruder states. To be consistent with the old calculations, the new zeroth order Hamiltonian of CASPT2[35] was not employed (i.e., the keyword IPEA was set to 0). Finally, the CASSCF energy was optimized again for the three-level average, but using the polarization field for the second state, and the energy for the second state was optimized with another CASPT2 calculation. All calculations were run with a developmental version of the Molcas 7.1 software.[36]

Multipoles and Polarizabilities. The influence of the surrounding protein was simulated by a detailed multicenter-multipole + polarizabilities model. This classical model consisted typically of charges, dipoles, and quadrupoles, as well as anisotropic polarizabilities for all atoms and bond centers in the protein (in total 10 824 and 11 012 centers for the HZX and U19 structures, respectively).

The quantum-mechanical and classical subproblems were solved in a self-consistent manner at the CASSCF level. In each CASSCF iteration, the electric field from the charge density of either the ground state or an excited state was computed at each center and added to the field from multipoles and induced dipoles. The induced dipoles were updated to this field until self-consistency within the classical system was obtained. The resulting electric field from the classical system (polarization field) was used to perturb the Hamiltonian operator, thus contributing to the change in charge density in the next CASSCF iteration. Similar approaches have been used before.[15,37−39] The final perturbation obtained by this procedure was applied in the CASPT2 calculation. In the computed energy, both the internal polarization energy of the classical system and the coupling between the quantum-mechanical and classical systems were included.

The multipoles and polarizabilities were obtained with the LoProp approach,[40] using the Molcas software. The calculations were performed either at the Hartree−Fock level with the 6-31G* basis set[41] or at the density functional B3LYP[42] level with the aug-cc-pVDZ basis set.[43] Each basis set was turned into the atomic natural orbital form (as required by the LoProp procedure) by a linear transformation that does not affect the orbital optimization. The properties were calculated for the whole protein by dividing it into the individual amino-acid residues, which all were capped with CH_3CO- and $-NHCH_3$ groups (dipeptides). The effect of the capping groups was removed by calculating the properties also of the overlapping $CH_3CONHCH_3$ fragments and subtracting them from the properties of the corresponding dipeptides, the molecular fractionation with conjugate caps approach, which has been shown to give excellent results.[44,45] A separate calculation was performed on every residue in the structure, with the correct geometry from the QM/MM structure. Lys-296 was truncated at the CG atom, which was converted to a hydrogen atom. Test calculations were also performed, in which Lys-296 was instead truncated at the CD atom (cf., Figure 1). However, this led to very unstable results, because the hydrogen atoms from the truncated CD atom on the QM and MM sides are only ∼0.5 Å apart.

To accomplish a well-defined classical treatment of intramolecular polarization, the centers were divided into groups. In the calculation of the electric field at a certain center, multipoles and induced dipoles from a list of groups were excluded, as described in the Molcas manual.[32] For the LoProp model, these exclusion lists were constructed so that polarization between any two centers that belong to the same fragment (dipeptide) in the LoProp calculations is omitted, because this polarization has already been treated quantum-mechanically (a rigorous motivation can be found in ref 45). For the polarizable Amber force field (ff02), we followed the Amber practice to ignore polarization between atoms separated by one, two, or three bonds. When testing the distance-dependent transition between these two models, a “generous” exclusion protocol was used to ensure that no unphysical polarization occurs. Thus, polarization between two centers treated by different models was omitted if it would have been omitted in either of the two models. A simpler method, using the LoProp exclusion rule for the Amber part as well, was also tested and gave similar results.

The splitting and splicing of the protein were performed with local software, as described in the Supporting Information. The calculations for a full protein took ∼400 CPU hours at the HF/6-31G* level and ∼6000 CPU hours at the B3LYP/aug-cc-pVDZ level. For isorhodopsin and the two mutants, only groups with new geometries (Lys-296, the two water molecules, and the mutated residue, if any) were recalculated (in fact, in the original calculations,[10,11] hydrogen atoms were reoptimized in U19-isorhodopsin and the HZX-G121L mutant, but these changes were not considered in this investigation, because the other structures were not changed).

In the 1HZX structure, residues 236−240 and 331−333 are missing. In the LoProp model, residues 235 and 330 were truncated by $-NHCH_3$ groups, and residues 241 and 334 were truncated by CH_3CO- groups (this is necessary to make the QM calculations possible). However, with the Amber force fields, no such truncating groups were used, following the original calculations[10,11] and because these capping groups are not available with all force fields.

Following the previous calculations, no effects of the surrounding solvent were included. Test calculations with a spherical cavity surrounding the protein (i.e., with a radius of 44−50 Å) indicated that the solvent effect is only ∼2 kJ/mol. However, its inclusion is complicated by the fact that the solvent reaction field should not (in variance to the polarization field) relax for the excited state.

Result and Discussion

We have calculated the electronic spectra of rhodopsin, isorhodopsin, and two mutants with the CASPT2//CASSCF approach, performing a detailed study of the influence of the surrounding protein. The use of two different crystal structures gives us a first indication of the reproducibility of the results and of their conformational dependence. Although the two lowest excitation energies are included in the CASSCF state average, we concentrate the discussion on the lowest (and main) excitation, which experimentally is found around 498 nm (240 kJ/mol).[46] The goal of this Article is not to reproduce this value (which would require a better basis set in the CASPT2//CASSCF step and an extensive sampling of conformations of the protein and the chromophore), but to take a first step in this direction by investigating what is needed for an accurate modeling of the surrounding protein for a single structure. This is done by using a very detailed model of the electrostatic and polarizing effect of the surroundings, consisting of a multicenter-multipole model up to quadrupoles and anisotropic polarizabilities for all atoms and bond centers of the protein. We then examine what parts of this model are actually needed, both for absolute and for relative excitation energies, by removing the various terms stepwise. We also study the distance dependence of the various terms, that is, how far from the chromophore they have a significant effect.

Spectra. Calculated excitation energies for the various structures are collected in Table 1. The treatment of the surroundings is described by a vector of four entries in square brackets. The four entries are the charges, which may be LoProp (+), various Amber force fields (ff94, ff03, or ff02[30,47,49]), or integer charges only for the charged residues, Asp, Glu, Lys, and Arg (i); the dipoles, which may be LoProp (+) or none (−); the quadrupoles, which may be LoProp (+) or none (−); and the polarizabilities, which may be LoProp anisotropic polarizabilities (a), scalar isotropic polarizabilities, taken as the average of the diagonal elements of anisotropic polarizabilities (s), polarizabilities from Amber (ff02[49]), or none (−). Thus, our standard and most accurate model is [+,+,+,a], whereas the original investigations[10,11] used a [ff94,−,−,−] model.

Table 1. Calculated Excitation Energies for Rhodopsin, Isorhodopsin, and the Glu113Asp (E113D) and Gly121Leu (G121L) Rhodopsin Mutants, Using Two Different Crystal Structures, U19[28] and HZX[27a]

model				U19				HZX			
				Rh		Iso-Rh		Rh	Iso-Rh	E113D	G121L
q	dipole	q-pole	pol	6-31G*	VDZ	6-31G*	VDZ	VDZ	VDZ	VDZ	VDZ
+	+	+	a	257.5	260.7	269.0	277.0	268.4	283.3	265.5	285.3
+	+	+	s	257.6	259.3	269.0	274.2	268.0	281.6	265.7	282.2
+	+	−	a	253.9	257.8	262.1	266.6	267.8	278.5	266.0	272.4
+	−	−	a	263.2	259.4	278.6	276.0	271.7	291.5	268.6	294.8
i	−	−	a		326.6		332.0	306.0	313.1	291.4	298.2
−	−	−	a	230.6	237.4	234.9	241.4	245.3	253.9	246.0	247.9
+	+	+	−	259.9	263.4	271.0	274.3	270.1	283.2	256.5	279.5
+	+	−	−	248.7	251.9	255.8	259.1	263.5	270.8	252.7	264.1
+	−	−	−	267.4	262.7	279.6	274.2	274.0	289.0	260.3	284.3
i	−	−	−		322.1		326.4	311.3	315.1	294.6	305.3
−	−	−	−		214.9		217.9	219.6	228.7	217.9	224.2
ff94	−	−	−		258.1		269.8	275.9	292.4	264.1	289.7
ff03	−	−	−		272.0		285.4	287.0	303.9	267.0	300.7
ff02	−	−	−		264.1		279.5	277.1	293.1	264.6	290.5
ff02	−	−	ff02		254.4		264.7	269.7	285.1	265.6	284.5
ff94CD[b]	−	−	−		230.2		243.7	253.6	266.8	243.0	261.1
ff03CD[b]	−	−	−		283.2		305.7	283.7	318.4	280.3	319.6
−[c]	−	−	−		215.5			219.7	227.2	218.6	223.8
ff94CD[c]	−	−	−		233.2			250.3	262.4	240.2	269.7
exp[d]					240.2		247.7	240.2	247.7	234.7	251.9

[a] Two different methods were used to calculate the multipoles and polarizabilities: HF/6-31G* and B3LYP/aug-cc-pVDZ (VDZ). The treatment of the surroundings is described by the four entries in model: the charges (q), which may be calculated by the LoProp procedure[40] (+), taken from the Amber force fields (ff94,[30] ff03,[47] or ff02[49]), taken as simple integer charges for the charged residues, Asp, Glu, Lys, and Arg, (i), or be ignored (−); the dipoles and quadrupoles (q-pole), which both may be calculated by LoProp (+) or be ignored (−); and the polarizabilities (pol), which may be fully anisotropic and calculated by LoProp (a), be scalar and isotropic, taken as the trace of the anisotropic ones (s), be taken from the Amber 2002 force field[49] (ff02), or be ignored (−). [b] In these calculations, the CD atom was converted to a hydrogen junction atom and charges were present on this atom, as well as on the CG and HG atoms (in all of the other calculations, instead the CG atom was converted to a hydrogen junction atom, and no charge was present on the CD, HD, and HG atoms; cf., Figure 1). [c] Data from refs 10 and 11. In the ff94CD calculations, charges were present on the CG, HG, and HD atoms, but not on the CD atom. [d] Data from refs 24, 25, 26, and 46.

It can be seen that the isolated chromophore (retinal + CH$_3$N$^+$H− from the protonated Lys-296 Schiff base) in vacuum [−,−,−,−] gives an excitation energy of 215 kJ/mol in U19 and 220 kJ/mol in the HZX crystal. This is within 1 kJ/mol of previously published values,[10,11] and ∼25 kJ/mol lower than the experimentally observed excitation energy in rhodopsin. The other structures gave slightly different excitation energies: 3 (U19) or 9 (HZX) kJ/mol higher excitation energy for Iso-Rh, 2 kJ/mol lower energy for the E113D mutant, and 5 kJ/mol higher energy for the G121L mutant. These values reflect the influence of the geometry on the excitation energies.

If the chromophores are inserted into the rhodopsin protein with the full LoProp model [+,+,+,a], the spectra shift significantly: The excitation energy shifts by 46−49 kJ/mol for Rh and by 48−61 kJ/mol for Iso-Rh and the mutants. The spectral shift is in the right direction as compared to the vacuum calculation (i.e., the excitation energy increases). However, the correction is too large, giving rise to too high excitation energies, and errors compared to experiments similar to those of the vacuum calculations (18−36 kJ/mol). This is an effect of the small basis sets used in the calculations and the fact that a single minimized structure is used. However, that is no problem in this investigation, because it is a constant factor, and we will only discuss the relative effect of the various terms in the multipole + polarizability model.

Multipole Model. We will first study the effect of the multipoles on the calculated spectra. From Table 1, it can be seen that if the quadrupoles are ignored, the spectra shift by 4−15 kJ/mol without polarizabilities (i.e., the difference between [+,+,+,−] and [+,+,−,−]) and by 1−13 kJ/mol with the polarizabilities ([+,+,+,a]−[+,+,−,a]). The excitation energies are reduced (by on average 11 and 5 kJ/mol), with one exception. Relative energies (i.e., between isorhodopsin and rhodopsin or between rhodopsin and its mutants) change by up to 12 kJ/mol (Table 2). Thus, the effect of the quadrupoles is quite small, but it cannot be ignored if you aim at accurate results.

If the dipoles are also removed, the spectra shift by 8−24 kJ/mol without and by 2−22 kJ/mol with polarizabilities (relative to the calculations without quadrupoles). However, this time, the excitation energies increase (by 16 and 10 kJ/mol on average), so that the energies go back to close to the original [+,+,+,a/−] energies again (average difference 4−5 kJ/mol). Relative energies change by up to 19 kJ/mol. Thus, the dipoles have a slightly larger effect on the excitation energies than do the quadrupoles.

The charges have an even larger effect on the excitation energies: The spectra shift by 42−62 kJ/mol without and by 14−36 kJ/mol with polarizabilities if the charges are removed (as compared to the [+,−,−,a/−] calculations). In all cases, the excitation energies are reduced (by 54 and 29 kJ/mol on average). Relative energies change by a much smaller amount, up to 14 kJ/mol. Thus, the charges have a large effect, as could be expected.

We have also examined the effect of various sets of point charges on the excitation energies. Calculations with CD

Table 2. Shifts in Excitation Energies (in kJ/mol) for Isorhodopsin Relative to Rhodopsin in the Two Crystal Structures and for the Glu113Asp and Gly121Leu Mutants Relative to the Wild-Type Protein Calculated with the Various Methods[a]

model				isorhodopsin		mutants			
q	dipole	q-pole	pol	U19	Hzx	E113D	G121L	MAD	max
+	+	+	a	16.3	14.9	−2.9	16.9	6.0	8.8
+	+	+	s	14.9	13.6	−2.3	14.2	4.8	7.4
+	+	−	a	8.8	10.7	−1.8	4.6	3.8	7.1
+	−	−	a	16.6	19.8	−3.2	23.1	8.8	12.3
i	−	−	a	5.4	7.1	−14.6	−7.8	7.8	19.5
-	−	−	a	4.0	8.6	0.7	2.6	5.0	9.1
+	+	+	−	10.9	13.1	−13.6	9.4	4.9	8.1
+	+	−	−	7.2	7.3	−10.8	0.6	4.2	11.1
+	−	−	−	11.5	15.0	−13.7	10.3	5.3	8.2
i	−	−	−	4.3	3.8	−16.7	−6.0	9.0	17.7
−	−	−	−	3.0	9.1	−1.7	4.6	4.3	7.1
ff94	−	−	−	11.7	16.5	−11.8	13.8	5.4	9.0
ff03	−	−	−	13.4	16.9	−20.0	13.7	7.9	14.5
ff02	−	−	−	15.4	16.0	−12.5	13.4	6.3	8.5
ff02	−	−	ff02	10.3	15.4	−4.1	14.8	3.8	7.9
ff94CD[b]				13.5	13.2	−10.6	7.5	5.3	6.0
ff03CD[b]				22.5	34.7	−3.4	35.9	17.0	27.2
experimental data				7.5	7.5	−5.5	11.7	0.0	0.0

[a] In addition, the mean average deviation (MAD) and the maximum deviation (max) as compared to the experimental data[24−26,46] are given. [b] In these calculations, the CD atom was converted to a hydrogen junction atom and charges were present on this atom, as well as on the CG and HG atoms (in all of the other calculations, instead the CG atom was converted to a hydrogen junction atom, and no charge was present on the CD, HD, and HG atoms; cf., Figure 1).

junctions in the MM system and employing the Amber 1994 force field[30] (ff94CD in Table 1) give results that differ from those of our full multipole model [+,+,+,−] by 15−33 kJ/mol (always more negative; the original calculations[10,11] used a similar approach, but they included charges on the HD atoms, but not on the CD junction atom; these results differ from the ff94CD results by 3−9 kJ/mol). However, this means that there are two hydrogen capping atoms, one in the QM system and one among the point charges, both representing the CD atom of Lys-296 and only 0.5 Å apart. Such a procedure makes the calculations very unstable. For example, it can be seen from Table 1 that if the charges are shifted to Amber 2003 charges[47] instead (ff03CD), the excitation energies change by 30−62 kJ/mol, showing that the results are completely unreliable. The same happens if we use a LoProp model with two capping atoms from CD; the results become unstable and counterintuitive.

Therefore, all of our calculations are performed with a hydrogen atom representing CD in the QM system and a charge representing the CG atom of Lys-296, whereas no charges are included for the CD, HD, and HG atoms. The results of such calculations with the Amber 1994 and 2003 force fields are also included in Table 1 (ff94 and ff03). It can be seen that the deletion of these three charges changes the excitation energies by 21−29 kJ/mol for ff94 and by 3−20 kJ/mol for ff03. Now, the results with the two force fields are much closer to the LoProp model, with differences of 1−9 kJ/mol (with a varying sign) for ff94 and 2−21 kJ/mol (always smaller excitation energies) for ff03. The two force fields differ by 3−16 kJ/mol, which shows that there still is some ambiguity in the use of a point-charge model of the surrounding protein. However, relative energies are more stable and change by less than 8 kJ/mol.

In both Amber force fields, the charges were obtained from QM calculations, using the RESP approach.[48] However, in the 1994 force field,[30] the electrostatic potential was taken from vacuum HF/6-31G* calculations, which are supposed to overestimate the dipole moments in a way similar to the average polarizing effect in solution. In the 2003 force field,[47] the electrostatic potential is instead calculated using the B3LYP/cc-pVTZ method (which gives almost correct dipole moments), explicitly polarized with a (protein-like) continuum solvent with a dielectric constant of 4. Thus, two different methods are used to obtain charges that are polarized in an average way similar to a protein-like surrounding. However, as we see from Table 1, the methods give quite different spectral shifts of excitation energies in a protein.

We have also used the charges of a third Amber force field, the polarizable 2002 force field (ff02).[49] These charges were also obtained with the B3LYP/cc-pVTZ method, but without the continuum solvent (i.e., in vacuum). Of course, the charges are not intended to be used without the polarizabilities, but this is similar to the comparison with the LoProp multipoles without any polarization. The results of the Amber 2002 charges (without the polarizabilities) differ by 1−10 kJ/mol from our full multipole model, that is, by an amount similar to that of the other Amber charges and are always intermediate between those of the ff94 and the ff03 charges.

An even simpler model of the surroundings is obtained with only integer charges for the residues with a net charge [i,−,−,a/−] models. From Table 1, it can be seen that such a model changes the results as compared to the [+,−,−,a/−] models by 22−67 kJ/mol (but only 3 kJ/mol in one case). However, it can also be seen that the results go in the opposite direction as compared to the calculations without any multipole model of the protein with differences of 45−109 kJ/mol and that the results are far (13−66 kJ/mol) from the full LoProp model. This shows that it is a poor model.

Polarizabilities. We next study the effect of the polarizabilities. If we replace all of the anisotropic polarizabilities (symmetric 3 × 3 tensors) with scalar isotropic polarizabilities (a single number, the average of the diagonal elements in the anisotropic tensors), there is only a minimal change in the excitation energies (less than 3 kJ/mol).

However, if the polarizabilities are completely removed, the spectra change by 0−13 kJ/mol if there is a multipole model of the protein and by 16−28 kJ/mol without any multipoles. Relative energies change by a similar amount. In general, the polarizabilities increase the excitation energies, but with only charges or with the full multipole model, the excitation energies decrease in many cases. This is a significant effect, so for accurate results, polarizabilities are important.

The polarizable Amber 2002 force field[49] performs rather well, with differences of 0−12 kJ/mol as compared to the full LoProp model. Relative energies differ by up to 6 kJ/mol. This is similar to the LoProp charges and polarizabilities [+,−,−,a], which give errors of up to 10 kJ/mol.

Distance Dependence. Next, we studied the distance dependence of the multipoles and polarizabilities; that is, we removed (zeroed) multipoles or polarizabilities for all residues outside a certain distance from the chromophore (minimum distance between any atom in the residue and the chromophore). This gives further information about the influence of the multipoles and polarizabilities on the spectra (it lets us identify accidental coincidences). Moreover, it gives information about the range of each type of interactions, which is important especially if we intend to study several conformations of the protein (it tells us how far out the multipoles or polarizabilities need to be recalculated for each new conformation). We will consider both absolute and relative excitation energies, but we concentrate on the U19 crystal and the Rh and Iso-Rh states. To ensure that the same residues are used for the two states, distances were calculated only for the Rh state and were used also for the Iso-Rh state.

The distance dependence of the conversion of anisotropic polarizabilities to isotropic polarizabilities is shown in Figure 2a. It can be seen that the total effect (i.e., the difference between the energies at the distance 35 Å = only anisotropic polarizabilities and 0 Å = only isotropic polarizabilities) is 1−3 kJ/mol, as was also reported in Table 1. However, the curves show that this small difference is coincidental: The true distance variation is actually up to 11 kJ/mol, showing that the anisotropy needs to be considered for accurate results. However, the results are accurate to within 4 kJ/mol from 6 Å. Likewise, the two curves for Rh and Iso-Rh are almost parallel, meaning that the relative excitation energy is almost constant, 12−17 kJ/mol, and insensitive to whether anisotropic or isotropic polarizabilities are used. Thus, the effect of the anisotropic polarizabilities is quite short-ranged.

Next, we considered the effect of removing the polarizabilities completely. As we saw in Table 1, the polarizabilities had a rather small effect on the spectrum, up to 9 kJ/mol, but only 3 kJ/mol for the two U19 structures. Again, the distance dependence in Figure 2b shows that this is coincidental: The total effect of the polarizabilities is up to 46 kJ/mol, and the curves do not level out (within 4 kJ/mol) until at 10 Å. The curves for Rh and Iso-Rh again run reasonably in parallel, but the variation is larger, ranging from 9 to 17 kJ/mol, but converging at 10 Å to 16 ± 1 kJ/mol. Thus, the polarizabilities are important both for the absolute and for the relative excitation energies.

Next, we looked at the effect of the quadrupoles. Figure 2c shows that the effect is up to 12 kJ/mol for rhodopsin and up to 26 kJ/mol for isorhodopsin. The curves level out around 10 Å. The difference between the two curves is 9−22 kJ/mol, but 16 ± 2 kJ/mol from 6 Å. Thus, we can conclude that the quadrupole interactions are quite short-range, as expected.

Interestingly, the effect of the dipoles is similar, but somewhat smaller: up to 10 kJ/mol for rhodopsin and 17 kJ/mol for isorhodopsin. The curves level out at 6 Å. The effect on the difference in excitation energy between Rh and Iso-Rh is even smaller, and the difference is 9 ± 1 kJ/mol, outside 2 Å.

Finally, we also tested the effect of going from a point-charge model to a model with all charges zeroed, except for residues with a net charge, for which a unit integer charge was used at the center of the charge. From Figure 2e, it can be seen that the effect is extensive, up to 97 kJ/mol, and it does not level out until 15−25 Å. For the relative excitation energy, the convergence is better, and the results are fully converged at 15 Å. However, there are variations of up to 22 kJ/mol at shorter distances.

From this, we can conclude that the polarizabilities are necessary, at least at distances up to 10 Å, although the effect of going to scalar polarizabilities is smaller. The effect of quadrupoles and dipoles is also extensive, but also converges within ∼10 Å. Therefore, a proper approach to use to save time if an extensive conformation sampling is employed (as is most like necessary to obtain accurate results) is to use a polarizable force field with scalar polarizabilities and only point charges for all residues outside a distance of 10 Å from the chromophore. There is no reason to use a nonpolarizable force field, because polarizabilities are needed to get the close interactions right. On the other hand, the use of scalar polarizabilities (and omission of dipoles and quadrupoles) makes the treatment of general parameters for the long-range interaction much easier (there is no need to rotate any vectors and matrices from a standard orientation to the actual orientation in the protein). Thus, we recommend the use of the polarizable Amber 2002 force field, or something similar, for the long-range interactions, outside ∼10 Å.

Such an approach is illustrated in Figure 2f, where we go from the full LoProp model to the Amber 2002 force field. It can be seen that the results of ff02 differ from that of the LoProp model by 6−12 kJ/mol, as was also seen in Table 1. Moreover, the difference increases to ∼20 kJ/mol around 4 Å. However, then it rapidly decreases, and at 10 Å and outward, the difference is less than 2 kJ/mol. The relative excitation energy between Iso and Rh converges even more rapidly and is 16 ± 3 kJ/mol already at 8 Å. This shows that it is an excellent approach to replace the LoProp model with the simpler ff02 outside a radius of 10 Å from the chromophore.

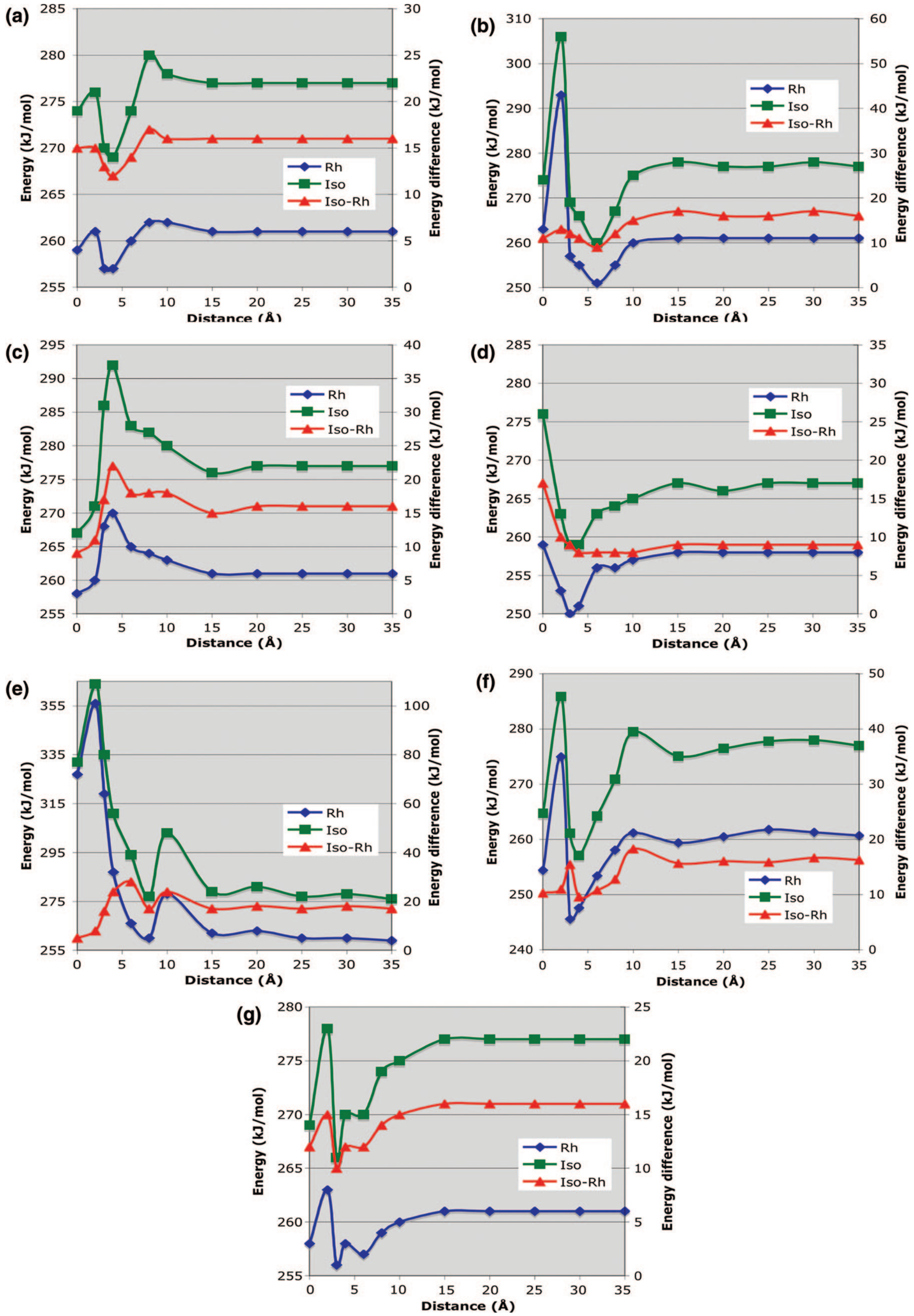


Figure 2. Distance dependence of the effect of the various multipoles and polarizabilities for the first excitation energy of Rh (◆) and Iso-Rh (■) in the U19 crystal. In addition, the difference between the two curves is given (▲, and right axis). All multipoles or polarizabilities for all residues within the given distance from the chromophore are changed. (a) Conversion from anisotropic (35 Å) to isotropic (0 Å) polarizabilities; (b) deletion of the anisotropic polarizabilities (35 Å; no polarizabilities at 0 Å); (c) deletion of the quadrupoles; (d) deletion of the dipoles; (e) conversion of LoProp charges (35 Å) to integer charges (0 Å); (f) conversion from the full LoProp model to the polarizable Amber 2002 force field; and (g) conversion of B3LYP/aug-cc-pVDZ properties to HF/6-31G* properties.

Method and Basis-Set Dependence. Another way to speed up the calculations is to employ a cheaper approach to calculate the LoProp properties. To test this, we have for the U19 structure calculated the multipoles and polarizabilities at two levels of theory, HF/6-31G* and B3LYP/aug-cc-pVDZ. From Table 1, it can be seen that the two methods give quite similar results: With the full multipole and polarizability model [+,+,+,a], the calculated excitation energies differ by only 3 (Rh) or 8 (Iso-Rh) kJ/mol (note that different methods are used only for the calculations of the multipoles and polarizabilities; the excitation energies are always calculated with the same method and basis set, CASPT2/6-31G*). In fact, the cheaper HF/6-31G* is closer to experiments (i.e., lower excitation energies), but this is only fortuitous. Similar differences are obtained for the other calculations, but with only charges [+,−,−,a/−], the HF/6-31G* properties give higher excitation energies (by 3−5 kJ/mol).

This small difference is quite unexpected, because polarizabilities normally depend strongly on the basis sets (and method): diffuse functions are mandatory to obtain converged and accurate polarizabilities, and the double-ζ basis is actually probably still somewhat small to obtain converged results.[50] Likewise, the HF/6-31G* method should overestimate calculated dipole moments, and this is actually employed in the Amber 1994 force field to get results that should simulate the polarizing effect in water solution.[30] In our calculations, we see no such tendency.

The distance dependence of the transition from B3LYP/aug-cc-pVDZ to HF/6-31G* properties is shown in Figure 2g. It can be seen that the actual effect is somewhat larger, up to 7 kJ/mol for Rh and 12 kJ/mol for Iso-Rh. The results are converged at ~8 Å.

For relative excitation energies (the difference between Rh and Iso-Rh), the difference between the two methods is even smaller, typically 0−1 kJ/mol, but 4−5 kJ/mol for the [+,+,+,a] and [+,+,+,s] calculations. The distance dependence in Figure 2g shows that the difference is converged around 8 Å. Thus, if errors of up to 8 kJ/mol are acceptable, much computer time can be saved by calculating all of the properties at the HF/6-31G* level. If this method is used only for residues outside 8 Å, the error is less than 3 kJ/mol.

Conclusions

In this Article, we have performed a detailed investigation of how the surrounding protein is best modeled in theoretical calculations of excitation energies. As a test case, we have used the retinal chromophore in rhodopsin and calculated the excitation energies at the CASPT2//CASSCF level. We have studied both absolute and relative excitation energies, in the latter case for variations in the chromophore (retinal and 9-cis retinal) and in the protein (two rhodopsin mutants, close to the chromophore). Theoretical excitation energies are typically rather poor (as compared to the experimental uncertainty), so theoretical investigations often concentrate on relative excitation energies. Moreover, we have tested the stability of the results by studying two different crystal structures, giving rise to slightly different QM/MM structures. We have used a very detailed model of the surroundings, a multicenter-multipole model up to quadrupoles and with anisotropic polarizabilities in all atoms and bond centers, as the reference point, and then studied if any parts of this model may be removed without compromising the results. The investigation has led to a number of important and interesting observations and conclusions.

Polarizabilities have a strong influence on the spectrum, up to 46 kJ/mol, and can therefore not be ignored in any detailed study of the effect of the protein on the spectrum of a bound chromophore. The use of anisotropic polarizabilities is important at distances shorter than 6−10 Å. Therefore, we strongly recommend the use of anisotropic polarizabilities in calculations of excitation energies.

The polarizable Amber 2002 force field gives errors of up to 10 kJ/mol, as compared to the full LoProp model (which is calculated for the right conformation of the protein). Thus, a polarizable force field does not solve the problem, but it can be used to speed up the calculations for residues more than 10 Å from the chromophore.

The effect of quadrupoles is up to 26 kJ/mol, that is, quite significant for absolute excitation energies. For relative energies, the effect is up to 13 kJ/mol, but it levels out already around 6 Å.

The effect of dipoles is somewhat smaller, up to 11 kJ/mol for absolute energies and 9 kJ/mol for relative energies.

The effect of point charges is very large, up to 60 kJ/mol for absolute energies and up to 14 kJ/mol for relative energies. The effect is long-ranged.

Different sets of simple point-charge models give quite different results. For example, the Amber 1994 and 2003 charges give an effect that differs by up to 16 kJ/mol for absolute and 8 kJ/mol for relative energies. This shows that calculations with a fixed (standard) point-charge model (as in most previous calculations[2−11]) give quite uncertain results. The results are even more unreliable if there are point charges very close to the chromophore. However, it is possible that the effect is reduced if the geometry is optimized with the respective force field.

Interestingly, multipoles and polarizabilities calculated at the HF/6-31G* level of theory give an effect on the excitation energies quite similar to those calculated at the much more expensive B3LYP/cc-pVDZ level of theory, with errors of less than 8 kJ/mol (6 kJ/mol for relative energies). The results converge within 8 Å from the chromophore. Thus, if errors of this size are acceptable, such an approach may save much computer time.

A natural question is why such seemingly accurate results have been obtained with theoretical calculations if the results are so sensitive to the model of the surroundings. The answer is that only relative excitation energies are normally considered. This is illustrated by Table 2, which shows the difference in excitation energies of rhodopsin and isorhodopsin, as well as the calculated shift in the excitation energies of the E113D and G121L mutants (i.e., the relative energies studied in this Article, for which experimental data are available). It can be seen that all calculations except one (ff03 with CG junctions) give mean absolute deviations of less than 10 kJ/mol and maximum errors of less than 20 kJ/mol. In fact, even the vacuum model [−,−,−,−] gives excellent results with a maximum error of 7 kJ/mol. Thus, accurate relative energies are easily obtained, whereas correct absolute energies are much harder to calculate.

Our results give us a firm indication of how accurate absolute excitation energies of chromophores in proteins should be calculated. Clearly, both point charges and polarizabilities need to be considered for accurate results, and they must be calculated for the actual conformation of the protein, using an approach similar to the one used in this Article. If the program used allows it, we see no reason not to include a full multipole model for the electrostatics. In fact, previous results indicate that the multipole model is not converged until quadrupoles are included,[21,51] as was done in this Article. There are no computational reasons not to include dipoles and quadrupoles in the calculations; they do not make the calculations significantly slower, neither the calculations of the multipoles, nor the spectra calculations. Instead, it is the polarizabilities that are expensive; they increase the calculation time for the properties by a factor of 6 and that for the spectra calculation by a factor of 2−3, mainly due to slower convergence of the CASSCF iterations. However, all calculations in this Article were run in less than 3 days (two CASSCF and CASPT2 calculations), which is not prohibitively much. In fact, this is much less than the time taken for the calculation of the multipoles and polarizabilities for the surrounding protein at the B3LYP/aug-cc-pVDZ level (∼250 CPU days, but it is trivially parallelizable). However, for accurate results, the basis sets in the CASSCF//CASPT2 calculations should be increased.

Finally, it is most likely important to sample many different conformations of the surrounding protein (and the chromophore). This can be easily done with molecular dynamics simulations. However, the present results indicate that the properties should be recalculated for each new conformation of the protein, at least for residues within 10 Å of the chromophore. However, much time can be saved by using the polarizable Amber 2002 force field for residues outside this distance or by calculating their properties with the HF/6-31G* method (which takes ∼15 CPU days for the whole protein). In future publications, we will test such an approach.

Acknowledgment. This investigation has been supported by grants from the Swedish research council, the Crafoord foundation, and AstraZeneca, as well as by computer resources of Lunarc at Lund University.

Supporting Information Available: Instructions for calculating multipoles for a protein, and a template of the restop file. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Öhrn, A.; Karlström, G. *Mol. Phys.* **2006**, *104*, 3087–3099.

(2) Pierloot, K.; De Kerpel, J. O. A.; Ryde, U.; Roos, B. O. *J. Am. Chem. Soc.* **1997**, *199*, 218–226.

(3) Rajamani, R.; Gao, J. *J. Comput. Chem.* **2002**, *23*, 96–105.

(4) Yamada, A.; Kakitani, T.; Yamamoto, S.; Yamato, T. *Chem. Phys. Lett.* **2002**, *366*, 670–675.

(5) Hayashi, S.; Ohime, I. *J. Phys. Chem. B* **2000**, *104*, 10678–10691.

(6) Hayashi, S.; Tajkhorshid, E.; Pebay-Peyroula, E.; Royant, A.; Landau, E. M.; Navarro, J.; Schulten, K. *J. Phys. Chem. B* **2001**, *105*, 10124–10131.

(7) Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2002**, *83*, 1281–1297.

(8) Fujimoto, K.; Jun-ya, H.; Hayashi, S.; Shigeki, K.; Nakatsuji, H. *Chem. Phys. Lett.* **2005**, *414*, 239–242.

(9) Hoffmann, M.; Wanko, M.; Strodel, P.; König, P. H.; Frauenhiem, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. *J. Am. Chem. Soc.* **2006**, *128*, 10808–10818.

(10) Coto, P. B.; Strambi, A.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17154–17159.

(11) Strambi, A.; Coto, P. B.; Ferré, N.; Olivucci, M. *Theor. Chem. Acc.* **2007**, *118*, 185–191.

(12) Bravaya, K.; Bochenkova, A.; Granovsky, A.; Nemukhin, A. *J. Am. Chem. Soc.* **2007**, *129*, 13035–13042.

(13) Warshel, A.; Chu, Z. T.; Hwang, J.-K. *Chem. Phys.* **1991**, *158*, 303–314.

(14) Warshel, A.; Chu, Z. T. *J. Phys. Chem. B* **2001**, *105*, 9857–9871.

(15) Thompson, M. A.; Schenter, G. K. *J. Phys. Chem.* **1995**, *99*, 6374–6386.

(16) Houjou, H.; Inoue, Y.; Sakurai, M. *J. Phys. Chem. B* **2001**, *105*, 867–879.

(17) Matsuura, A.; Sato, H.; Houjou, H.; Saito, S.; Hyashi, T.; Sakurai, M. *J. Comput. Chem.* **2006**, *27*, 1623–1630.

(18) Ren, L.; Martin, C. H.; Wise, K. J.; Gillespie, N. B.; Luecke, H.; Lanyi, J.; Spudich, J. L.; Birge, R. R. *Biochemistry* **2001**, *40*, 13906–13914.

(19) Houjou, H.; Inoue, Y.; Sakurai, M. *J. Am. Chem. Soc.* **1998**, *120*, 4459–4470.

(20) Luzhkov, V.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4491–4499.

(21) Krauss, M. *Comput. Chem.* **1995**, *19*, 199–204.

(22) Wanko, M.; Hoffmann, M.; Strodel, P.; Kozlowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

(23) Kochendoerfer, G. G.; Lin, S. W.; Sakmar, T. T.; Mathies, R. A. *Trends Biochem. Sci.* **1999**, *24*, 300–305.

(24) Sakmar, T. P.; Franke, R. R.; Khorana, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8309–8313.

(25) Han, W.; Lin, S. W.; Smith, S. O.; Sakmar, T. P. *J. Biol. Chem.* **1996**, *271*, 32330–32336.

(26) Hurley, J. B.; Ebrey, T. G.; Honig, B.; Ottolenghi, M. *Nature* **1977**, *270*, 540–542.

(27) Teller, D. C.; Okada, T.; Behnke, C. A.; Palczewski, K.; Stenkamp, R. E. *Biochemistry* **2001**, *40*, 7761–7772.

(28) Okada, T.; Sugihara, M.; Bondar, A. N.; Elstner, M.; Entel, P.; Buss, V. *J. Mol. Biol.* **2004**, *342*, 571–583.

(29) Andruniów, T.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *52*, 17908–17913.

(30) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.

(31) Fahmy, K.; Jager, F.; Beck, M.; Zvyaga, T. A.; Sakmar, T. P.; Siebert, F. *Proc. Natnl. Acad. Sci. U.S.A.* **1993**, *90*, 10206–10210.

(32) Roos, B. O.; Taylor, P. R. *Chem. Phys.* **1980**, *48*, 157–173.

(33) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(34) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218–1226.

(35) Ghigo, G.; Roos, B. O.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **2004**, *396*, 142–149.

(36) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222–239.

(37) Luzhkov, V.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4491–4499.

(38) Gao, J.; Byun, K. *Theor. Chem. Acc.* **1997**, *96*, 151–156.

(39) Poulsen, T. D.; Kongsted, J.; Osted, A.; Ogilby, P. R.; Mikkelsen, K. V. *J. Chem. Phys.* **2001**, *115*, 2393–2400.

(40) Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **2004**, *121*, 4494–4450.

(41) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.

(42) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(43) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(44) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599–3605.

(45) Söderhjelm, P.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*, 617–627.

(46) Kandori, H.; Schichida, Y.; Yoshisawa, T. *Biochemistry (Moscow)* **2001**, *66*, 1197–1209.

(47) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(48) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(49) Cieplak, P.; Caldwell, J.; Kollman, P. A. *J. Comput. Chem.* **2001**, *22*, 1048–1057.

(50) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2004**, *120*, 9903–9906.

(51) Söderhjelm, P.; Krogh, J. W.; Karlström, G.; Ryde, U.; Lindh, R. *J. Comput. Chem.* **2007**, *28*, 1083–1090.

CT800459T

JCTC Journal of Chemical Theory and Computation

Bound Ligand Conformer Revealed by Flexible Structure Alignment in Absence of Crystal Structures: Indirect Drug Design Probed for HIV-1 Protease Inhibitors

Alok Juneja,[†] Henning Riedesel,[†] Milan Hodoscek,[‡] and E. W. Knapp*,[†]

Institute of Chemistry & Biochemistry, Freie Universität Berlin, Fabeckstr. 36a, D-14195 Berlin, Germany, and National Institute of Chemistry, Hajdrihova 19, SI-1001 Ljubljana, Slovenia

Received November 11, 2008

Abstract: In the absence of structural knowledge on the target protein, the bound ligand conformer (BLC) can be constructed approximately by an indirect drug-design approach that uses a set of ligands binding to the same target. Once the bound ligand conformer (BLC) is known, different strategies of drug design can be pursued. The indirect drug-design approach of the present study is based on the assumption that a set of ligands with chemically different architecture binding to the same target protein carry hidden information of their corresponding true BLCs. It is shown how this information can be extracted by pairwise flexible structure alignment (FSA) using molecular dynamics (MD) simulations with attractive intermolecular interactions that derive from the molecular similarity of the ligands and allow the ligands to adopt the same space. The FSA approach is performed with a newly designed module overlap in the experimental CHARMM-29a1, which soon will become publicly available. Combining the conformations obtained from FSA of different ligand pairs yields consensus ligand conformers (CLCs) that should be similar to the BLCs. This procedure was validated on HIV-1 protease (HIV-P), where at present 44 crystal structures with bound ligands of sufficiently diverse chemical composition are available. The FSA approach identifies four different clusters of HIV-P BLCs. These clusters are consistent with the H-bond patterns of the ligands bound to HIV-P in the crystal structures exhibiting four different binding modes. The cluster-specific CLCs are indeed very similar (rmsd $\approx$ 2 Å) to the corresponding BLCs from the crystal structure, demonstrating the feasibility of the present approach.

Introduction

Fighting diseases more effectively requires an increased effort on the design of new drugs. Rational drug design, a major tool in the development of new drugs, is a challenging task for structural, chemical, and computational biology. Most of the known theoretical approaches to drug design are structure-based direct methods, such as docking,[1−6] which employ knowledge of the target structure and the binding pocket. When the target structure is unknown, purely ligand-based indirect drug design is still possible in case several ligands of chemically different architecture are known to bind to the same target.[6−11] One indirect drug design approach is the quantitative structure−activity relationship (QSAR).[12−17] QSAR relates variations in drug activity to changes in chemical composition, characterizing the drugs by suitable molecular descriptors.[18−20] Specific descriptors characteristic for active molecules, the so-called pharmacophores,[21−24] are used in virtual screening studies to classify drugability of molecules from a given database.[25−31] A variant is 3D-QSAR, which employs information of the spatial arrangement of the ligand components[32−35] and may also use

---

* Corresponding author phone: 0049-30-838-54387; fax: 0049-30-838-56921; e-mail: knapp@chemie.fu-berlin.de.

[†] Freie Universität Berlin.

[‡] National Institute of Chemistry.

10.1021/ct8004886 CCC: $40.75

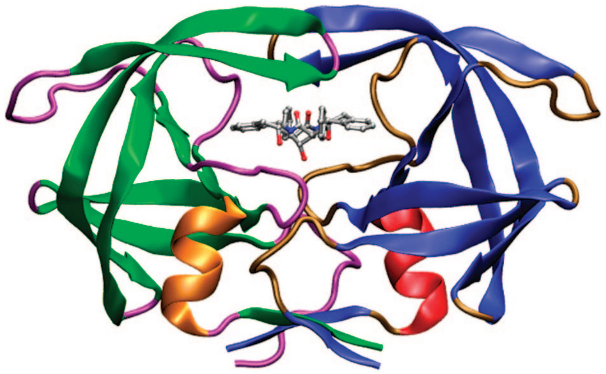


Figure 1. Crystal structure of HIV-1 protease[55] with the HIV-P ligand L10 (see Table 1) as a ball and stick model. The homodimer is represented as rubber band model possessing a *C*2 rotational symmetry with the rotation axis oriented vertically in the drawing plane. The binding pocket extends over both monomers.

information of the binding pocket.[33,36] However, in the latter case the QSAR approach is no longer indirect. Another 3D-QSAR approach is comparative molecular field analysis (CoMFA),[33,34,37,38] which in addition to steric aspects is also using electrostatic information of the ligands. More recently, in the comparative molecular similarity indices analysis (CoMSIA) steric and electrostatic aspects of ligands were complemented by H-bond and hydrophobic patterns.[38−41] However, in contrast to the present approach, these methods generally do not yield coordinates of the target bound ligand conformer (BLC).

Here, we report an indirect drug design approach, which yields approximate BLC coordinates. We strictly use information of the chemical architecture and composition of ligands that bind to the same target (protein binding pocket). Hence, we employ an approach of indirect drug design. Structures cocrystallized with ligands are only used after the study is completed to compare with the true BLC. In this study, we exploit the conformational variability of a set of ligands that bind to the same target protein to gain information on the BLC. Flexible structure alignment (FSA) of these ligands leads to consensus ligand conformers (CLC), which need to be compatible with the diverse chemical architectures of the ligands. Since these ligands bind in the same pocket of the target protein, it is expected that they assume a common geometry in the binding pocket (the true BLC) which should be reflected by the CLCs obtained with FSA. For a sufficiently large set of ligands with different chemical architectures, the CLCs obtained by FSA should be similar to the BLCs. Consequently, one can employ an ensemble of these CLCs as an integral part of the pharmacophore to design new drugs or to perform virtual screening on molecular libraries.[25−31] In earlier work we used a similar FSA approach to align two proteins with equivalent function.[42]

The purpose of the present study is to test this FSA approach by considering a protein−ligand model system for which a large number of BLCs are known. Presently, one of the best studied protein−ligand systems is HIV-1 protease (HIV-P) for which crystal structures with 44 chemically diverse bound ligands are available (see Table 1). The human immunodeficiency virus (HIV) is derived from immature polyproteins that contain HIV-P.[43] During viral replication, HIV-P cleaves the peptide bonds of the Gag and Pol polyproteins yielding the appropriately sized active protein components of the mature virus. HIV-P is proteolytically active as a homodimer possessing *C*2 rotational symmetry (Figure 1)[44,45] with the binding cavity extending symmetrically over both subunits of the homodimer.

The FSA approach in this study is performed with a newly designed module overlap in the modified CHARMM-29a1, which is not yet publicly available. It employs MD simulations of ligand pairs using a modified energy function. This energy function employs conventional intramolecular interactions, but instead of conventional intermolecular interactions it uses energy terms based on molecular similarity allowing molecular groups from different ligands to attract each other in proportion to their similarities and to adopt the same space. The measure of similarity is based on three features: (i) volume/shape, (ii) charge, and (iii) electrostatic potential. Although the latter two features both refer to electrostatics, they differ in a practical sense, since charges focus on short-range while the electrostatic potential describes also long-range aspects of electrostatic interactions. With this similarity-based energy function, MD simulations of a ligand pair with subsequent energy minimization yield superimposed structures of ligand pairs in low energy CLCs, which can be compared with the BLCs of the crystal structures.

In short, our working hypothesis is that in a pairwise FSA ligands that bind to the same target protein but have different intrinsic flexibility adopt CLCs close to the BLCs of the crystal structures. Hence, where the true BLC is not available, we can use the ensemble of these CLCs as integral part of pharmacophores to find new drugs for specific targets by virtual screening.[25−31]

Methods

Abbreviations. FSA, flexible structure alignment; RSA, rigid structure alignment; HIV-P, HIV-1 protease; MD, molecular dynamics; rmsd, root-mean-square deviation; BLC, bound ligand conformer; CLC, consensus ligand conformer.

Overview. In parts A1 and A2 we describe how molecules of completely different chemical composition can be characterized and their similarity measured. In part B we introduce the set of test data for HIV-P, for which the structures of 44 cocrystallized ligands are available. In part C we describe the general methodology and the modified energy function which is the basis of the FSA approach. In part D the procedure for structure alignment is described. It comprises preparation of ligand structures for MD simulation (part D1), performance of the FSA (part D2), and generation of the CLCs (part D3). Technical details are given respectively in parts S2, S5, and S9 of the Supporting Information.

A1: Estimating Molecular Similarity. The concept of molecular similarity, originally introduced by Carbo to compare charge densities $\rho_a(\vec{r})$ and $\rho_b(\vec{r})$ of molecules a and b, respectively, and known as the Carbo Index (CI),[46,47] is defined by

$$\mathrm{CI(a,b)} = \frac{\int \rho_{\mathrm{a}}(\vec{r})\rho_{\mathrm{b}}(\vec{r})\mathrm{d}\vec{r}}{\left(\int \rho_{\mathrm{a}}^{2}(\vec{r})\mathrm{d}\vec{r}\right)^{1/2}\left(\int \rho_{\mathrm{b}}^{2}(\vec{r})\mathrm{d}\vec{r}\right)^{1/2}} \qquad (1)$$

If charge densities of molecules a and b are identical in shape but differ in magnitude, then $\rho_{\mathrm{a}}(\vec{r}) = c\rho_{\mathrm{b}}(\vec{r})$, where c being an arbitrary positive constant, yields CI(a,b) = 1. Hence, CI quantifies similarities in shape but not in magnitude. An alternative definition of molecular similarity, the Hodgkin Index (HI)

$$\mathrm{HI(a,b)} = \frac{2\int \rho_{\mathrm{a}}(\vec{r})\rho_{\mathrm{b}}(\vec{r})\mathrm{d}\vec{r}}{\int \rho_{\mathrm{a}}^{2}(\vec{r})\mathrm{d}\vec{r} + \int \rho_{\mathrm{b}}^{2}(\vec{r})\mathrm{d}\vec{r}} \qquad (2)$$

was proposed by Hodgkin and Richard.[47−49] In contrast to the CI, the HI is sensitive to both shape and magnitude of the density distributions ρ_{a} and ρ_{b}. Both similarity indices vary between +1 and −1, extremes which correspond to identity and complementarity, respectively. The values of the similarity indices CI(a,b) and HI(a,b) are identical if $\int\rho_{\mathrm{a}}^{2}\mathrm{d}\vec{r} = \int\rho_{\mathrm{b}}^{2}\mathrm{d}\vec{r}$ but generally $|\mathrm{CI(a,b)}| > |\mathrm{HI(a,b)}|$.

A2: Features Characterizing Molecules. We use a combination of three different features to characterize the structure of molecules. These are shape and volume, distribution of atomic partial charges, and electrostatic potential. The latter two features are related, since they both are based on electrostatics, but charges focus on the short-range aspect of electrostatics only, while the electrostatic potential also considers long-range contributions.

To describe shape and volume of a molecule, we represent its atoms (m) by three-dimensional Gaussian distribution functions $g_{\mathrm{m}}(\vec{r})$[50,51]

$$g_{\mathrm{m}}(\vec{r}) = \frac{\sqrt{V_{\mathrm{m}}}}{\sigma_{\mathrm{m}}^{3/2}\pi^{3/4}}\exp\left(\frac{-(\vec{r}-\vec{r}_{\mathrm{m}})^{2}}{2\sigma_{\mathrm{m}}^{2}}\right) \qquad (3)$$

where V_{m} is the van der Waals (vdW) volume of atom m, $\vec{r}_{\mathrm{m}}$ is the position of the center of atom m, σ_{m} is the width of the Gaussian accounting approximately for the atomic radius $R_{\mathrm{m}} \approx \sigma_{\mathrm{m}}$. We normalized $g_{\mathrm{m}}(\vec{r})$ to yield for the self-overlap the volume of atom m, i.e., $\int g_{\mathrm{m}}^{2}(\vec{r})\mathrm{d}\vec{r} = V_{\mathrm{m}}$. The volume common to atoms m and n is given by the overlap integral

$$S_{\mathrm{mn}} = \int (g_{\mathrm{m}}(\vec{r})g_{\mathrm{n}}(\vec{r}))\mathrm{d}\vec{r} = \sqrt{V_{\mathrm{m}}V_{\mathrm{n}}}\left(\frac{2\sigma_{\mathrm{m}}\sigma_{\mathrm{n}}}{\sigma_{\mathrm{m}}^{2}+\sigma_{\mathrm{n}}^{2}}\right)^{3/2}\exp\left(\frac{-(\vec{r}_{\mathrm{m}}-\vec{r}_{\mathrm{n}})^{2}}{2(\sigma_{\mathrm{m}}^{2}+\sigma_{\mathrm{n}}^{2})}\right) \qquad (4)$$

A molecule is represented as the sum of $g_{\mathrm{m}}(\vec{r})$ corresponding to its individual atoms. The similarity of a pair of molecules can then be calculated as the overlap integral of the molecular distribution functions. The approach to estimate molecular similarity based on atom-centered Gaussians is not restricted to volume and shape similarity[52] but can be applied to any atom based molecular property. Accordingly, a molecular distribution function referring to property P of molecule can be defined as the sum of $g_{\mathrm{m}}^{(P)}(\vec{r})$ referring to the corresponding atomic property P. Hence, for property P (vol, charge, epot for volume/shape, charge, electrostatic potential, respectively) a molecule a, which consists of n_{a} atoms, is described by the distribution function

$$\rho_{\mathrm{a}}^{(P)}(\vec{r}) = \sum_{\mathrm{m}=1}^{n_{\mathrm{a}}} g_{\mathrm{m}}^{(P)}(\vec{r}) \qquad (5)$$

Estimation of the similarly between two molecules a and b with respect to property P requires computation of the overlap integral of the two corresponding molecular distribution functions $\rho_{\mathrm{a}}^{(P)}(\vec{r})$ and $\rho_{\mathrm{b}}^{(P)}(\vec{r})$

$$S^{(P)}(\mathrm{a,b}) = \int \rho_{\mathrm{a}}^{(P)}(\vec{r})\rho_{\mathrm{b}}^{(P)}(\vec{r})\mathrm{d}\vec{r} = \sum_{\mathrm{m}}^{n_{\mathrm{a}}}\sum_{\mathrm{n}}^{n_{\mathrm{b}}} S_{\mathrm{mn}}^{(P)} \qquad (6)$$

Hence, the atom-based form of HI that measures the similarity with respect to property P for a pair of molecules (a, b) is given by

$$\mathrm{HI}^{(P)}(\mathrm{a,b}) = \frac{2S^{(P)}(\mathrm{a,b})}{S^{(P)}(\mathrm{a,a}) + S^{(P)}(\mathrm{b,b})} \qquad (7)$$

where $S^{(P)}(a,b)$ is the overlap integral for the molecule pair (a, b) and $S^{(P)}(\mathrm{a,a})$ and $S^{(P)}(\mathrm{b,b})$ are the self-overlaps of molecules a and b, respectively.

The result of volume overlap of two atoms n and m yields, in analogy to eq 4

$$S_{\mathrm{mn}}^{(\mathrm{vol})} = w_{\mathrm{m}}^{(\mathrm{vol})}w_{\mathrm{n}}^{(\mathrm{vol})}\left(\frac{2R_{\mathrm{m}}^{(\mathrm{vol})}R_{\mathrm{n}}^{(\mathrm{vol})}}{(R_{\mathrm{m}}^{(\mathrm{vol})})^{2}+(R_{\mathrm{n}}^{(\mathrm{vol})})^{2}}\right)^{3/2}\exp\left(\frac{-(\vec{r}_{\mathrm{m}}-\vec{r}_{\mathrm{n}})^{2}}{2(kR_{\mathrm{m}}^{(\mathrm{vol})})^{2}+2(kR_{\mathrm{n}}^{(\mathrm{vol})})^{2}}\right) \qquad (8)$$

where the weighting factors $w_{\mathrm{n}}^{(\mathrm{vol})}$ replace the volume factors in eq. (4), $R_{\mathrm{n}}^{(\mathrm{vol})}$ is the effective atomic radius for the volume of atom n and k is a factor that scales all atomic radii (i.e., k determines the degree of localization of the Gaussians). The default values of k and $w_{\mathrm{n}}^{(\mathrm{vol})}$ are unity. Likewise, for charge overlap between atoms m and n with atomic partial charges q_{m} and q_{n} (given in units of the elementary charge) and $R_{\mathrm{n}}^{(\mathrm{charge})}$, the effective atomic radii for the charges, we have

$$S_{\mathrm{mn}}^{(\mathrm{charge})} = q_{\mathrm{m}}q_{\mathrm{n}}\left(\frac{2R_{\mathrm{m}}^{(\mathrm{charge})}R_{\mathrm{n}}^{(\mathrm{charge})}}{(R_{\mathrm{m}}^{(\mathrm{charge})})^{2}+(R_{\mathrm{n}}^{(\mathrm{charge})})^{2}}\right)^{3/2}\exp\left(\frac{-(\vec{r}_{\mathrm{m}}-\vec{r}_{\mathrm{n}})^{2}}{2(kR_{\mathrm{m}}^{(\mathrm{charge})})^{2}+2(kR_{\mathrm{n}}^{(\mathrm{charge})})^{2}}\right) \qquad (9)$$

where the charges of the atom pairs enter as first power instead of square root to enhance the similarity measure. In the present application we use $R_{\mathrm{n}}^{(\mathrm{charge})} = R_{\mathrm{n}}^{(\mathrm{vol})}$.

As similarity measure for the electrostatic potentials of two charges q_{m} and q_{n} at positions $\vec{r}_{\mathrm{m}}$and $\vec{r}_{\mathrm{n}}$, respectively, we use

$$S_{\mathrm{mn}}^{(\mathrm{epot})} = q_{\mathrm{m}}q_{\mathrm{n}}\exp(-\gamma^{(\mathrm{epot})}|\vec{r}_{\mathrm{m}}-\vec{r}_{\mathrm{n}}|) \qquad (10)$$

where $\gamma^{(\mathrm{epot})}$ is a range parameter governing the length scale over which the electrostatic potentials of atoms m and n are compared (a large $\gamma^{(\mathrm{epot})}$ corresponds to short distances, the default value is unity). A derivation of the expression for the electrostatic potential overlap, eq 10, is given in the part

Table 1. HIV-1 Protease (HIV-P) with the 44 Ligands Considered in This Study

no.	ligand (inhibitor/substrate)[a]	PDB id	resol (Å)
L1	TMC114 or UIC-94017 (017)	1T3R	1.20
L2	AQ148 (ARQ)	3AID	2.50
L3	ABT-378 (AB1)	1MUI	2.80
L4	SB203386 (IM1)	1SBG	2.30
L5	BOC-PHM-TYR-ILE-GLY	1MTR	1.75
L6	BMS-182193	1ODW	2.10
L7	L-739,622 (3IN)	2BPZ	2.50
L8	L-738,317 (1IN)	2BPW	2.80
L9	L-735,524 (MK1)	2BPX	2.80
L10	L-700,417 (VAC)	4PHV	2.10
L11	KNI-272 (KNI)	1HPX	2.00
L12	SB203238 (GAN)	1HBV	2.30
L13	AHA455 (A1A)	2BQV	2.10
L14	UCSF8 (THK)	2AID	1.90
L15	AHA006 (NMB)	1AJV	2.00
L16	U100313 (U02)	2UPJ	3.00
L17	TPV	1D4S	2.50
L18	AKC (AKC4P_133A)	2BB9	1.35
L19	HOE/BAY 793 (BAY)	1VIJ	2.40
L20	TS-126 (IPF)	2A1E	1.30
L21	A-74704	9HVP	2.80
L22	A-84538 (RIT)	1HXW	1.80
L23	A79285 (A85)	1DIF	1.70
L24	CGP 53820 (C20)	1HIH	2.20
L25	GR137615 (G37)	1HTG	2.00
L26	ROC	1FB7	2.60
L27	KI2-PHE-GLU-GLU- NH2	1NH0	1.03
L28	SD146 (146)	1QBT	2.10
L29	MSA367 (MS3)	1EC3	1.80
L30	p1-p6 substrate	1KJF	2.00
L31	MA-CA substrate	1KJ4	2.90
L32	CA-p2 substrate	1F7A	2.00
L33	p2-NC substrate	1KJ7	2.00
L34	RT-RH substrate	1KJG	2.00
L35	U89360E (U0E)	1AXA	2.00
L36	MVT-101	4HVP	2.30
L37	JG-365	7HVP	2.40
L38	U-85548E	8HVP	2.50
L39	LP-130 (LP1)	1ODY	2.00
L40	acetyl-pepstatin	5HVP	2.00
L41	CH_2-CBG-Asn-Tyr-CH_2-Pro-ILE-Val-NH_2	1CPI	2.05
L42	SKF108738 (HEF)	1HEF	2.20
L43	U75875	1HIV	2.00
L44	SDZ283-910	1A8G	2.50

[a] Five ligands (L30−L34) are natural substrate oligopeptides. All other ligands listed in the table are inhibitors.

S1 of Supporting Information. Although charge distribution and electrostatic potential are derived from the same set of atomic partial charges, they focus on different aspects of these charges. The charge distribution considers the similarity at short distances accounting for direct interactions of the ligands with atoms in the binding pocket like hydrogen bonds and salt bridges, while the electrostatic potential considers also the long-range aspect of the charge distribution.

A similarity measure considering contributions from all three properties (volume/shape, charge, and electrostatic potential) is the weighted sum of the individual HI yielding the combined HI

$$\mathrm{HI}_{\mathrm{combined}} = \frac{w_{\mathrm{vol}}}{w_{\mathrm{sum}}}\mathrm{HI}_{\mathrm{vol}} + \frac{w_{\mathrm{charge}}}{w_{\mathrm{sum}}}\mathrm{HI}_{\mathrm{charge}} + \frac{w_{\mathrm{epot}}}{w_{\mathrm{sum}}}\mathrm{HI}_{\mathrm{epot}} \quad (11)$$

where $w_{\mathrm{sum}} = w_{\mathrm{vol}} + w_{\mathrm{charge}} + w_{\mathrm{epot}}$; $\mathrm{HI}_{\mathrm{vol}}$, $\mathrm{HI}_{\mathrm{charge}}$, and $\mathrm{HI}_{\mathrm{epot}}$ are the individual Hodgkin indices and w_{vol}, w_{charge}, w_{epot} (default values are unity) are the relative weights for volume/shape, charge, and electrostatic potential overlap, respectively. The resulting similarity index $\mathrm{HI}_{\mathrm{combined}}$ can vary formally between −1 and +1, but in practice it will not come close to −1. As the value of $\mathrm{HI}_{\mathrm{combined}}$ approaches unity the similarity of the two molecules increases.

B: HIV-1 Protease Ligand Test Set. To test whether FSA of ligands with the same function can provide information on the ligand conformers in the binding pocket of the target protein, we consider HIV-1 protease (HIV-P), for which X-ray structures of 44 different HIV-P ligand complexes are available (listed in Table 1). Of these 39 are HIV-P inhibitors (less flexible than peptides), and the remainder (L30−L34) are natural substrate oligopeptides[53] that are cleaved by HIV-P during viral replication. Although the HIV-P binding cavity possesses *C*2 rotational symmetry (Figure 1),[44,45] only one inhibitor (L19) exhibits the same symmetry while five of the inhibitors (L6, L10, L15, L21, L28) possess mirror symmetry. All the remaining HIV-P ligands, including the natural substrate oligopeptides, possess no symmetry. These inhibitor substrates, cocrystallized with HIV-P, are available in the protein data bank (PDB).[54]

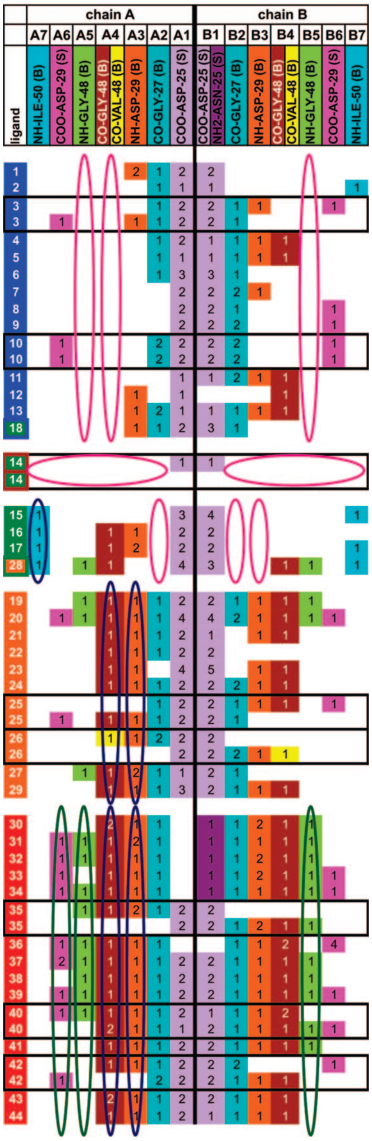
Because of the *C*2 symmetry of the HIV-P binding pocket, several ligands (L3, L10, L14, L25, L26, L35, L40, L42) appear in the crystal structures in two opposite orientations with slight variations in the conformation. For the analysis of the H-bond pattern of the ligand with the protein, we used both crystal structures (see Table 2, discussed in part F of Results and Discussion). For analyses of HIV-P ligand coordinates, we used only the first of the two alternative ligand orientations listed in the PDB data file. This is justified, since the rmsd between alternate conformers of the same HIV-P ligand overlaid with the Kabsch algorithm[56,57] are generally small (see Table SIV).

C: Maximizing the Similarity of Molecules. In the search for CLCs of pairs of chemically disparate molecules, we monitored the similarity measured by $\mathrm{HI}_{\mathrm{combined}}$, eq 11. As the conformers of the pair of molecules are being adjusted to maximize their similarity, each of the two molecules must at the same time assume a reasonable low-energy structure. To achieve a compromise between these two criteria, we constructed an effective potential energy function

$$E(\mathrm{a,b}) = E_{\mathrm{intra}}(\mathrm{a}) + E_{\mathrm{intra}}(\mathrm{b}) - e_{\mathrm{S}}\mathrm{HI}_{\mathrm{combined}}(\mathrm{a,b}) \quad (12)$$

that simultaneously accounts for both. The intramolecular contributions to E [$E_{\mathrm{intra}}(\mathrm{a})$, $E_{\mathrm{intra}}(\mathrm{b})$] are taken to be the true physical ones (i.e., bonded interactions depending on bond length, bond angle, torsions; and nonbonded interactions depending on van der Waals and Coulomb terms within the same molecule). The true intermolecular contribution is replaced by a nonphysical term proportional to the combined Hodgkin index, where the parameter e_{S} is set so that all three terms in eq 12 are of comparable magnitude. Different values of e_{s} are used during the FSA approach as given in part S2, S5, and S9 of Supporting Information. The gradient of this effective potential energy function can be used for MD simulations, which will yield CLCs of the molecule pair under consideration. In these combined CLCs both molecules are structurally aligned in a low energy conformation such that similar parts of the two molecules are superimposed.

Table 2. HIV-P H-Bond Pattern



	chain A								chain B								
	A7	A6	A5	A4		A3	A2	A1	B1		B2	B3	B4		B5	B6	B7
ligand	NH-ILE-50 (B)	COO-ASP-29 (S)	NH-GLY-48 (B)	CO-GLY-48 (B)	CO-VAL-48 (B)	NH-ASP-29 (B)	CO-GLY-27 (B)	COO-ASP-25 (S)	COO-ASP-25 (S)	NH2-ASN-25 (S)	CO-GLY-27 (B)	NH-ASP-29 (B)	CO-GLY-48 (B)	CO-VAL-48 (B)	NH-GLY-48 (B)	COO-ASP-29 (S)	NH-ILE-50 (B)
1						2	1	2	2								
2							1	1	1								1
3							1	2	2		1	1				1	
3		1				1	1	2	2		1						
4							1	2	1		1	1	1				
5							1	1	1		1	1	1				
6							1	3	3		1						
7								2	2		2	1					
8								2	2		1					1	
9								2	2		2					1	
10		1					2	2	2		2					1	
10		1					2	2	2		2					1	
11								1	1		2	1	1				
12						1		1					1				
13						1	2	1	1		1	1	1				
18						1	1	2	3		1						
14								1	1								
14																	
15	1							3	4								1
16	1			1		1		2	2								
17	1			1		2		2	2								1
28	1		1	1				4	3				1		1		1
19			1	1		1	1	2	2		1	1	1		1		
20		1	1	1		1	1	4	4		2	1	1		1	1	
21				1		1	1	2	1			1	1				
22				1		1	1	2	2								
23				1		1		4	5			1	1				
24				1		1	1	2	2		2	1	1				
25							1	2	2		1	1	1			1	
25		1		1		1	1	2	2		1						
26					1	1	2	2	2								
26								2	2		2	1		1			
27			1	1		2	1	1	2		1						
29				1		1	1	3	2		1	1	1				
30				2		1	1			1	1	2	1		1		
31		1	1	1		2	1			1	1	1	1		1		
32		1	1	1		1	1			1	1	2	1		1		
33		1		1		1	1			1	1	2	1		1	1	
34		1	1	1		1	1			1	1	1	1		1	1	
35			1	1		2	1	2	2								
35								2	2		1	2	1		1		
36		1	1	1		1	1				1	1	2			4	
37		2	1	1		1	1	2	2		1	1	1		1		
38			1	1		1	1	2	2		1	1	1		1		
39		1	1	1		1	1	2	2		1	1	1		1	1	
40		1	1	1		1	1	2	1		1	1	2				
40				2		1	1	1	2		1	1	1		1	1	
41				1		1	1	2	2		1	1	1		1		
42				1		1	1	2	2		2					1	
42		1					2	2	2		1	1	1				
43				2		1	1	2	2		1	1	1				
44				1		1	1	2	1		1	1	1				

Column 1: ligand number; Row 1: polypeptide chain A or B of HIV-P; Row 2: formal numbering of the H-bond partner in HIV-P (an H-bond is formed if the participating non-hydrogen atoms are closer than 3.5 Å); Row 3: specification of H-bond partners in format (atom_name)-(amino_acid_type)-(amino_acid_number)-(backbone(B)/side_chain(S)). For mutant HIV-Ps, the residues 1 (Asn replaces Asp) and 4 (Val replaces Gly) possess different side chains. From these, B1, A4, and B4 are H-bonds with some HIV-P ligands as indicated. Each column has a color code referring to the H-bond partner in HIV-P. Each line characterizes the H-bond pattern for a specific HIV-P ligand labeled in column 1. For the ligands (L3, L10, L14, L25, L26, L35, L40, L42), alternate conformations were found in the HIV-P crystal structures. For these ligands, both H-bond patterns are given in subsequent lines. These cases are highlighted by black boxes as for instance for ligand L3. A filled cell denotes an H-bond of the ligand with the particular group of HIV-P; the digits (1 or 2) count the number of H-bonds formed between the ligand and this group. Empty cells denote absence of possible H-bonds. H-bond patterns persistent for a group of ligands are highlighted by ellipses. Red (blue) ellipses denote strict absence (presence) of specific H-bonds. Green ellipses indicate dominant occurrences of a particular H-bond. The five substrate HIV-P ligands (L30−L34) belong to the red cluster. The clustering of the HIV-P ligands is based on the similarity of the H-bond pattern and indicated by the color code in column 1. A detailed description of H-bond pattern is given in Table SV.

D: Computational Procedures.

D1: Model Preparation of Ligands and Generation of Initial Conformers. To avoid the bias of crystal structure information of the ligand conformers used in this study rigorously, we were choosing the initial conformers for the MD simulations to differ as much as possible from the conformers they adopt in the HIV-P binding pocket. The average rmsd of the initial conformers relative to the BLCs is 4.36 Å. More details are given in part S2 and Figure S1 and S2 of the Supporting Information.

D2: Pairwise Flexible Structure Alignment (FSA). The alignment of a pair of flexible molecules is performed by MD simulation using the effective potential energy function, eq 12. Seven hundred twenty different starting arrangements of the ligand pairs were used. The most similar ligand pair conformer is extracted from the final 80 ps of 105 ps trajectory runs at room temperature and used for further analysis, as detailed in part S5 of the Supporting Information. The CPU time for single 105 ps trajectory is 85 min on AMD Athlon 2.2 GHz.

D3: Generating Consensus Ligand Conformers (CLCs). Since the pairwise FSA of the 44 HIV-P ligands yields for each ligand 43 different conformers, it is necessary to determine consensus ligand conformers (CLCs) from them. Alternatively we could have used multiple FSA considering all HIV-P ligands in a single alignment attempt and thus obtained CLCs directly. However, in such an approach one would assume that all 44 HIV-P ligands obey the same binding mode and give equivalent conformers. But, as a result of this study we can show that the 44 HIV-P ligands actually exhibit four different binding modes. Hence, the direct use of a multiple FSA approach would obscure this information and yield less than optimum results. Pairwise FSA keeps information on different binding mode and allows one to cluster the ligands according to mutual similarity of their conformers. After this analysis, CLCs are computed for each of the similarity clusters, as explained in part S9 of the Supporting Information.

Results and Discussions

Overview. Before using the concept of FSA to generate conformers of HIV-P ligands that can be related to the corresponding bound ligand conformers (BLCs) fitting in the HIV-P binding pocket, we carried out a number of studies for each of the 44 HIV-P ligands independently. In part A below, we discuss (i) how much ligand conformers can deviate from the corresponding BLCs, (ii) how close ligand conformers can come by chance to their corresponding BLCs through equilibrium fluctuations during a short MD simulation at room temperature, and (iii) how much the ligand conformers of such an MD simulation can deviate on average from the corresponding BLCs. In part B, we analyze the similarity of ligand conformers adopted in the HIV-P binding pocket of the crystal structure. As a typical result we compared reference ligand L1 with all other ligands L*x* in the main text. The comparison of the other ligand pairs is found in Figure S4 of the Supporting Information. In part C, we examine in detail the conformers of reference ligand

L1 resulting from FSA with all other HIV-P ligands and compared the computed ligand conformers with the conformer adopted by ligand L1 in the crystal structure. Part D provides an overview of the pairwise FSA results considering all combinations of ligand pairs, and part E contains an analysis of the degrees of similarities between the different pairs of HIV-P ligands computed by the FSA approach. These results are clustered to assign the HIV-P ligands to different classes (part E) using solely the FSA data without consideration of the crystal-structure information. The binding modes of the HIV-P ligand conformers in the crystal structures based on the H-bond pattern were analyzed in part F and compared with the results of the cluster analysis based on the FSA results. In part G we use the results of ligand classifications based on the FSA to construct consensus ligand conformers (CLC) for each ligand cluster separately, as described in part D3 of Methods. Finally, in part H we visualize the CLCs of the four obtained HIV-P ligand clusters.

A: Comparing HIV-P Ligand Conformers Obtained by MD Simulations with BLCs from Crystal Structures. First, we examined how close and how distant HIV-P ligand conformers may come by chance to the corresponding crystal structure conformers if conventional MD simulations are performed for each ligand separately in vacuo using the CHARMM energy function. For this purpose we performed MD simulations of 225 ps (including 5 ps for heating) at constant energy ($T \approx 300$ K), employing only the last 100 ps of the trajectory to allow for sufficient equilibration. In Figure 2a the RMSDs of these ligand conformers relative to the crystal structures are displayed. The RMSDs of ligands 1 to 18 (displayed in the left half of Figure 2a) are smaller than that of 19 to 44. This correlates with the size of the ligands (see Table SII of the Supporting Information), since larger ligands generally lead to larger RMSDs. For the energy-minimized starting structures of the MD simulation, the conformers with the largest deviation from the crystal structures were selected from 100 randomly generated structures (see part D1 in Methods). The RMSDs of these selected conformers are in the range 4−5.5 Å, larger than the time-averaged RMSDs (about 4 Å) and the minimum RMSDs (about 3 Å) occurring during the MD simulation runs. This range of rmsd values clearly shows that the HIV-P ligands spontaneously adopt conformers close to the conformer in the HIV-P binding pocket only rarely. Hence, MD simulations of HIV-P ligands performed independently from information of the HIV-P binding pocket generally do not yield conformers close to the BLCs. The actual variation in ligand conformers between crystal structure, average structure of an MD simulation of the individual ligand, and the CLC from the FSA approach as described later in detail in connection with Figure 7 is displayed for ligand L1 in Figure 2b. Analogue figures for all other 43 HIV-P ligands are shown in Figure S9 of the Supporting Information.

B: Analysis of the Similarity of HIV-P Ligand Conformers in the Crystal Structure. For the FSA method to be successful, there must be significant similarities among the BLCs from the 44 different HIV-P ligands. To meet this condition, we first performed pairwise rigid structure alignments (RSA) of the BLCs. This was done in two different ways: (i) Protein structure-based RSA, in which the ligands were ignored and HIV-Ps from the 44 crystal structures were aligned pairwise by the Kabsch algorithm.[56,57] The resulting translational and rotational transforms were then used to superimpose the ligands. For these RSA, the combined

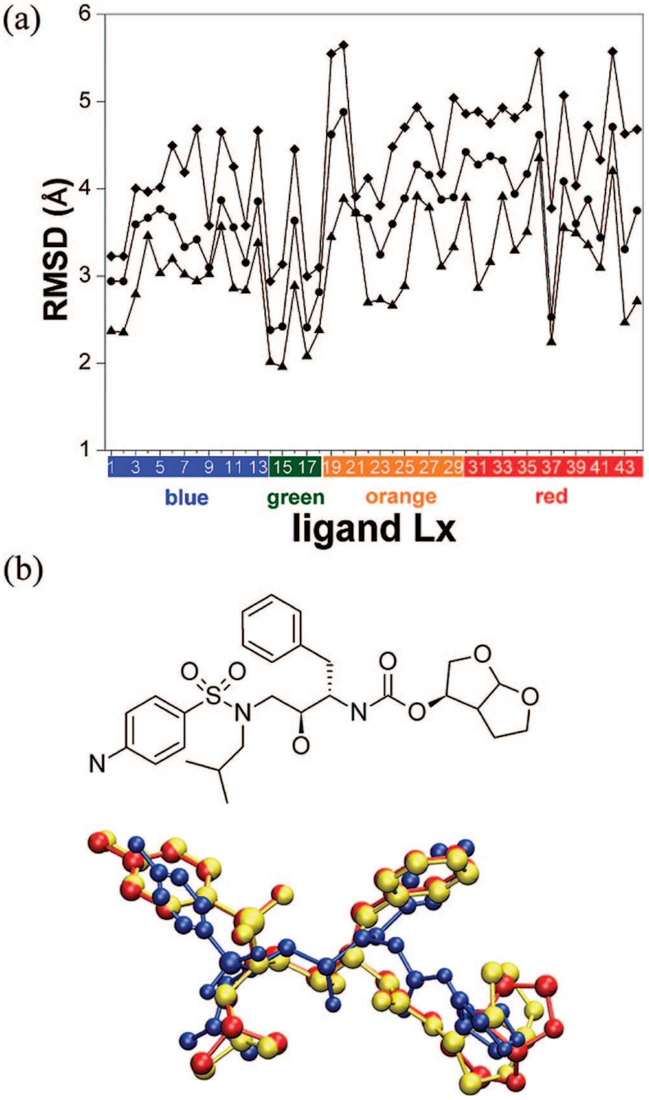


Figure 2. Part a: HIV-P ligand conformers obtained by MD simulations for all 44 HIV-P ligands considered using the last 100 ps of a 225 ps trajectories in vacuo at constant energy ($T \approx 300$ K). The RMSDs of the ligand conformers are given relative to the crystal structures. They refer to the starting conformers of the MD simulations (top line, which coincides with middle line of Figure S2), the conformers with minimum deviation to the crystal structures obtained by MD simulation (bottom line), and the average conformers (middle line). The ligand numbers L*x* at the abscissa are defined in Table 1; their background colors refer to the four clusters of ligands classified according to structure similarity as discussed later in connection with Figure 6. Part b: Chemical architecture of HIV-P ligand L1 (top) and three-dimensional structures (bottom) of the bound ligand conformer from the crystal structure (yellow), average structure from MD simulation of ligand L1 alone (blue) with 2.95 Å rmsd from crystal structure, and consensus ligand conformer (CLC) after FSA (red) as described in connection with Figure 7 with 1.65 Å rmsd.

Hodgkin indices $\mathrm{HI}_{\mathrm{combined}}$ were generally between 0.4 and 0.5 (red circles in Figure 3) for all HIV-P ligand pairs involving the reference ligand L1 and ligands L2−L44. (ii) Ligand conformer-based RSA, in which the torsion angles of the ligands were fixed by use of stiff torsion potentials (as described in the CHARMM script $\mathrm{Script}_{\mathrm{FSA}}$ in part S8 of the Supporting Information) to obtain approximately rigid ligand models. With these rigid ligand models, the FSA procedure was applied to all HIV-P ligand pairs involving the reference ligand L1 and all other ligands L2−L44. The resulting $\mathrm{HI}_{\mathrm{combined}}$ values (green diamonds in Figure 3) (essentially between 0.5 and 0.6) are larger than the values obtained by the protein structure-based RSA. Results of RSA for all other HIV-P ligand pairs, which are qualitatively similar, are given in Figure S4 of the Supporting Information.

C: Pairwise Flexible Structure Alignment of Reference Ligand L1 with all HIV-P Ligands. FSA of HIV-P ligands were performed for all 946 = (44 × 43)/2 possible ligand pairs. The results for 43 ligand pairs involving the reference ligand L1 and all other ligands are displayed in Figure 3. These results are typical of those for all the other combinations of ligand pairs, which are shown in 43 additional graphs in Figure S4 of the Supporting Information. The $\mathrm{HI}_{\mathrm{combined}}$ for the initial conformers of the ligand pairs used in the MD simulations to perform the FSA are displayed as gold stars in Figure 3. These low $\mathrm{HI}_{\mathrm{combined}}$ values, generally between 0.3 and 0.4, correspond to ligand conformers that are significantly dissimilar from the BLCs. During the MD simulation the similarity increases, as indicated by $\mathrm{HI}_{\mathrm{combined}}$ values that vary between 0.5 and 0.65 (blue triangles in Figure 3) and are often considerably larger than those of the starting conformers and also slightly larger than those obtained by ligand conformer-based RSA of the HIV-P ligand conformers of the crystal structures (green diamonds in Figure 3). It is encouraging that the dependence of the $\mathrm{HI}_{\mathrm{combined}}$ on the ligand Lx is similar for the conformer pairs (L1, Lx) obtained by FSA (blue triangles) and from the crystal structures (green diamonds), although for the latter the $\mathrm{HI}_{\mathrm{combined}}$ are slightly smaller.

The open symbols in Figure 3 refer to comparisons of the L1 conformer from the crystal structure with L1 conformers obtained by FSA of L1 with the 43 other HIV-P ligands. Since different conformers of the same L1 ligand were considered in this comparison, both rmsd (open brown triangles, right scale in Figure 3) and $\mathrm{HI}_{\mathrm{combined}}$ (open cyan diamonds, left scale in Figure 3) were evaluated. It is noteworthy that the RMSDs relative to the L1 crystal-structure conformer are all close to 2 Å (open brown triangles, right scale), which is about as small as the minimum rmsd (= 2.3 Å) obtained from the last 100 ps of a vacuum MD simulation of L1 at room temperature (see Figure 2a) and is considerably smaller than the rmsd (= 3.9 Å) from the initial conformer used for the FSA approach (black solid line in Figure 3). Parallel with small RMSDs we observed $\mathrm{HI}_{\mathrm{combined}}$ values generally greater than 0.5 when comparing the L1 conformer from the crystal structure with the L1 conformers obtained by FSA with the 43 other HIV-P ligands (open cyan diamonds). These results, which are typical of all ligand pairs (Lx, Ly) (see Figure S4 in the Supporting Information), indicate the success of the FSA approach in generating ligand conformers that are close to the corresponding BLCs.

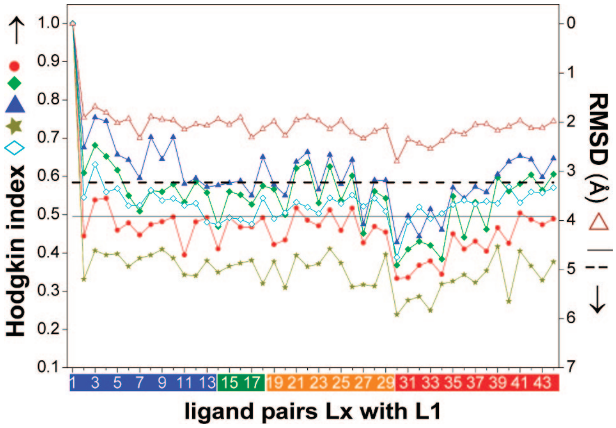


Figure 3. Results of pairwise structure alignment of the HIV-P reference ligand L1 with 43 other HIV-P ligands Lx, $x = 2$, 3, ... 44, listed on the abscissa. Ligands are defined in Table 1. The left scale refers to $\mathrm{HI}_{\mathrm{combined}}$, eq 11, the right scale refers to RMSDs (in reverse direction) given relative to the crystal structures. For the sake of completeness, we formally assigned $\mathrm{HI}_{\mathrm{combined}} = 1$ and rmsd = 0.0 aligning ligand L1 with itself. $\mathrm{HI}_{\mathrm{combined}}$ (left scale) is displayed for ligand conformer-based RSA using ligand conformers from the crystal structures (green diamonds), for protein structure-based RSA in which the HIV-P crystal structures were aligned by the Kabsch algorithm[56,57] (red circles), for initial conformers of the HIV-P ligands used for the MD simulation of FSA (gold stars), for FSA results of ligand pairs (L1, Lx) consisting of the reference L1 and all other ligands Lx of HIV-P (blue triangles), for L1 ligand conformer in the crystal structure with the L1 conformers obtained by pairwise FSA with all other 43 HIV-P ligands Lx (open cyan diamonds), and the corresponding rmsd (right scale) (open brown triangles). The rmsd relative to the crystal structure of the L1 conformer used as starting conformer in pairwise FSA (solid black line); rmsd of the same L1 conformer after energy minimization (dashed black line). N numbers on the abscissa refer to ligands designated in Table 1; their background colors refer to the four clusters of ligands classified according to structural similarity as discussed later in connection with Figure 6.

Analyzing the dependencies of $\mathrm{HI}_{\mathrm{combined}}$ in Figure 3 with respect of the different types of clusters labeled by the background color code of the ligand numbers at the abscissa, we observe significant lower similarities of L1 with the red labeled ligands. This is typical for all blue labeled ligands like L1, which are members of the same ligand cluster. As will be seen subsequently (in part E), this deviation can be explained by the large dissimilarity of the ligands belonging to the blue cluster relative to the ligands of the red cluster. The cluster analysis shows that these two ligand clusters are the most dissimilar and adopt different binding modes in the HIV-P binding cavity.

D: Pairwise Alignment Overview of Data from all HIV-P Ligand Pairs. The data for all 946 = 44 × 43/2 pairs of HIV-P ligands yield 44 diagrams like the one of Figure 3, which displays the data of all 43 ligand pairs involving Ly = L1 as reference ligand. Since these diagrams provide too many details, they are deferred to the Supporting

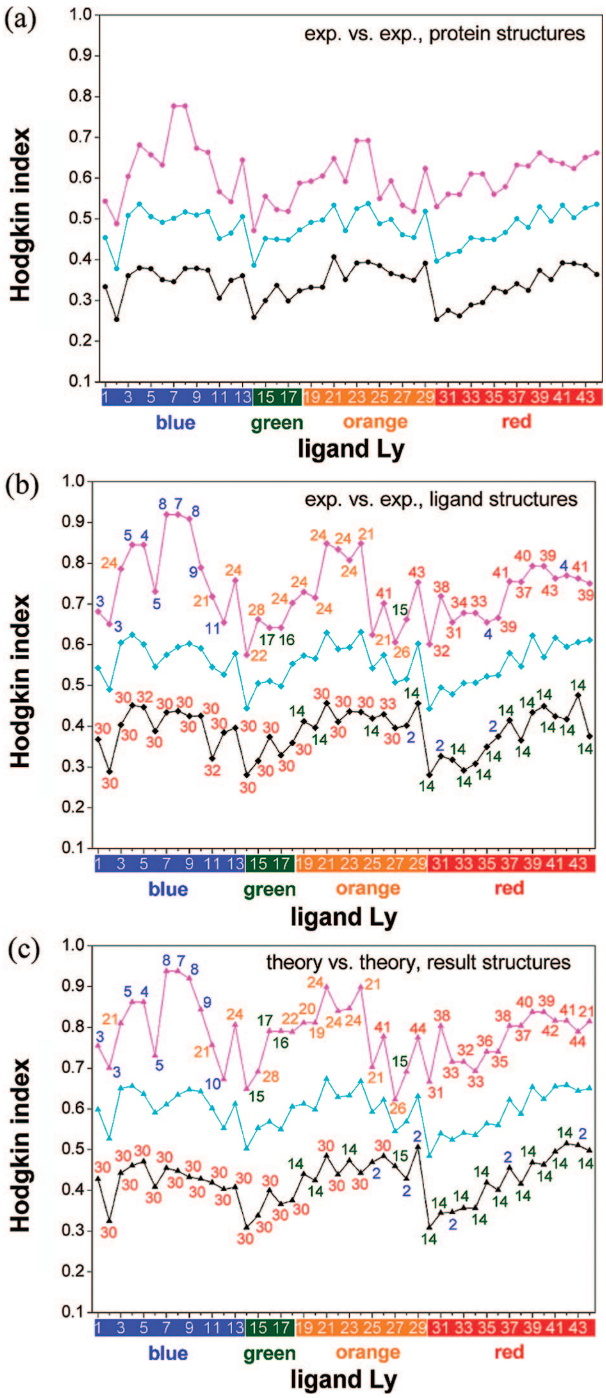


Figure 4. Maxima (magenta), minima (black), and averages (cyan) of $\mathrm{HI}_{\mathrm{combined}}$, eq 11, from HIV-P ligand pairs formed between specific reference ligands L*y* displayed on the abscissa and other 43 HIV-P ligands. *N* numbers on the abscissa refer to ligands listed in Table 1. (a) $\mathrm{HI}_{\mathrm{combined}}$ obtained by protein structure-based RSA of HIV-P ligand conformers form the crystal structures (corresponding to red circles in Figure 3 and Figure S4); (b) $\mathrm{HI}_{\mathrm{combined}}$ obtained by ligand conformer-based RSA of HIV-P ligand conformers from the crystal structures (corresponding to green diamonds in Figure 3 and Figure S4); (c) $\mathrm{HI}_{\mathrm{combined}}$ obtained by FSA of HIV-P ligand pairs (corresponding to blue triangles in Figure 3 and Figure S4). The numbers at the maxima (magenta) and minima (black) of $\mathrm{HI}_{\mathrm{combined}}$ in parts b and c refer to the ligand numbers L*x* which, combined with the reference ligand L*y*, yield the maximum and minimum of $\mathrm{HI}_{\mathrm{combined}}$, respectively.

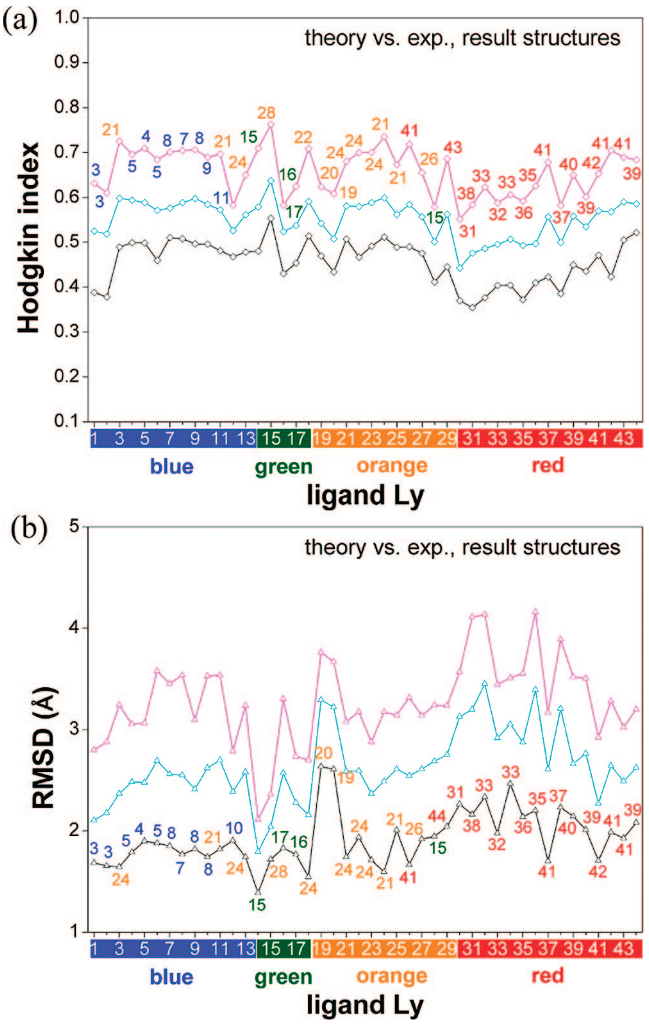


Figure 5. Comparison of ligand conformers obtained by pairwise FSA with conformers from crystal structures. Maxima (magenta), minima (black), and averages (cyan) of $\mathrm{HI}_{\mathrm{combined}}$ (a) and rmsd (b) for ligand conformers formed between specific reference ligand conformers L*y* (the BLCs) displayed on the abscissa and other 43 HIV-P ligands L*x* (obtained by FSA). *N* numbers at maxima (magenta) of $\mathrm{HI}_{\mathrm{combined}}$ in part a and minima (black) of rmsd in part b refer to ligand number L*x* which by pairwise FSA yields the conformer of ligand L*y* (marked on the abscissa) most similar to the crystal structure. Note that the data points at L*y* = L1 use the information from the curve (open cyan diamonds) in Figure 3. *N* numbers on the abscissa identify ligands listed in Table 1.

Information (Figure S4). Figures 4 and 5 contain an overview and summary of these results displaying maxima, minima, and averages of $\mathrm{HI}_{\mathrm{combined}}$ and rmsd for HIV-P ligand pairs formed between a particular reference ligand L*y* (denoted at the abscissa) and the 43 other ligands (L*x*). Parts a and b of Figure 4 show maxima, minima, and averages [referring to ligand pairs (L*y*, L*x*) with the average running over L*x*] of $\mathrm{HI}_{\mathrm{combined}}$ obtained from the BLC of the HIV-P crystal structures using protein structure-based (Figure 4a) and ligand conformer-based (Figure 4b) RSA (see part B of Results and Discussion).

The maxima of $\mathrm{HI}_{\mathrm{combined}}$ obtained by RSA, which are around 0.6 and 0.8 for the protein structure-based and ligand conformer-based alignment of the BLC, respectively, indicate a large degree of similarity. On the other hand, the minima of $HI_{combined}$ around 0.3 and 0.4 for protein structure-based and ligand conformer-based alignment, respectively, indicate low similarity. The average of $HI_{combined}$ is typically the mean of the maxima and minima, which suggests that for a given HIV-P reference ligand there are generally ligands with high and with low degrees of similarity. As we will see later, this large variability in HIV-P ligand conformers is a trait of HIV-P ligands that possess different binding modes.

In parts a and b of Figure 4 only experimental data on the similarity of the HIV-P ligand conformers of the crystal structures are analyzed, while in part c only computational data on ligand conformers obtained by FSA are compared. Comparing the dependencies of $HI_{combined}$ on the reference ligand L*y* in parts b and c, one observes a nearly quantitative agreement in particular for the maximal and the average $HI_{combined}$. This suggests a close relationship between the ligand conformers obtained by FSA and the corresponding ligand conformers in the HIV-P crystal structures, although the former did not use any information from the crystal structures. The numbers at the maximal (minimal) $HI_{combined}$ in parts b and c of Figure 4 label the HIV-P ligand L*x* that is most similar (distant) to the reference ligand L*y* on the abscissa. That these numbers for the experimental (Figure 4b) and computational data (Figure 4c) often agree suggests again close similarity between experimental and computational ligand conformers. Furthermore, with few exceptions these numbers possess the same color as the background color of the reference ligand L*y* from the abscissa which shows that the ligand pairs with maximal (minimal) similarity belong to the same (most distant) cluster and therefore possess the same (very different) binding mode in the HIV-P binding pocket.

The ligand pairs with minimum similarity in Figures 4b and 4c are the ligands L30, L2, L14. This agrees with the results of the cluster analysis (Figure 6, see part E), which demonstrates that L30 is most distant from ligands of the blue and green cluster, L2 is most distant from ligands of the orange cluster, and L14 is most distant from ligands of the red cluster. This is a clear demonstration that the FSA approach can indeed be used to generate ligand conformers which are close to the BLCs.

A direct comparison between theory and experiment for the HIV-P ligand conformers from crystal structures and from pairwise FSA is given in Figure 5. Here, $HI_{combined}$ (part a) and rmsd (part b) are displayed as functions of the reference ligand L*y* for which ligand pairs with all other HIV-P ligands were considered to evaluate averages, maxima, and minima. The maxima of $HI_{combined}$ vary roughly between 0.6 and 0.75, and the minima of rmsd are generally below 2.0 Å, which indicates high similarity within subsets of HIV-P ligands. Simultaneously, the minima of $HI_{combined}$ generally vary between 0.4 and 0.5, and the maxima of rmsd between 3.0 Å and 4.0 Å, which demonstrates a low similarity between subsets of HIV-P ligands possessing different binding modes. The $HI_{combined}$ (rmsd) are slightly higher (lower) for the ligands of the first two clusters (on the left side of Figure 5, marked by blue and green background colors) compared to ligands of the other two clusters (orange and red background colors). This behavior is likely due to the generally smaller size of the HIV-P ligands in the first two clusters compared with the ligands of the other two clusters. Table SII in the Supporting Information, which exhibits the HIV-P ligand structures, indicates the size differences. The numbers at the

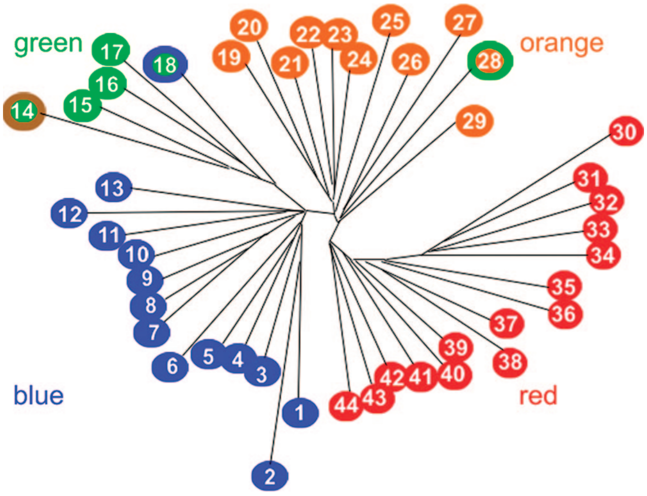


Figure 6. Phylogenetic tree of HIV-P ligands generated by T-REX[59] based on the similarity measure $HI_{combined}$, eq 11, obtained by FSA. Besides the tree structures the closeness of the ligand symbols in the diagram is also a qualitative measure of structure similarity. On the basis of similarity, four clusters can be clearly discriminated. The background color code reflects the results of an H-bond pattern analysis of the crystal structures (Table 2). Small differences in the FSA based cluster analysis appeared for the ligands L14 and L18 that appear to belong to the green cluster and for L28 that belongs to the orange cluster. The five natural substrate oligopeptides (L30−L34) belong to the red cluster. Note that the color code denoting the ligands at the abscissa of Figure 2−5, 7 uses the results of the classification based on FSA as displayed above.

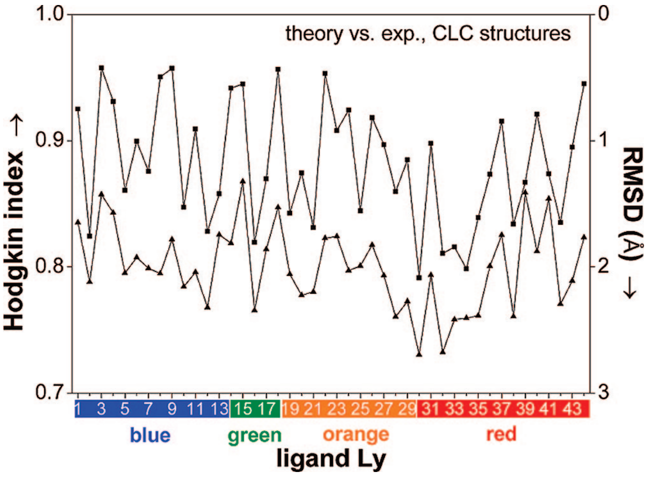


Figure 7. Cluster specific HIV-P ligand CLCs obtained by FSA using the cluster result from Figure 8. Combined Hodgkin indices, $HI_{combined}$, eq 11, (left scale, squares) and RMSDs (right scale, in reverse direction, triangles) of the CLCs displayed relative to the ligand conformers in the crystal structures. CLCs were generated without crystal structure information, as explained in the Method part D3.

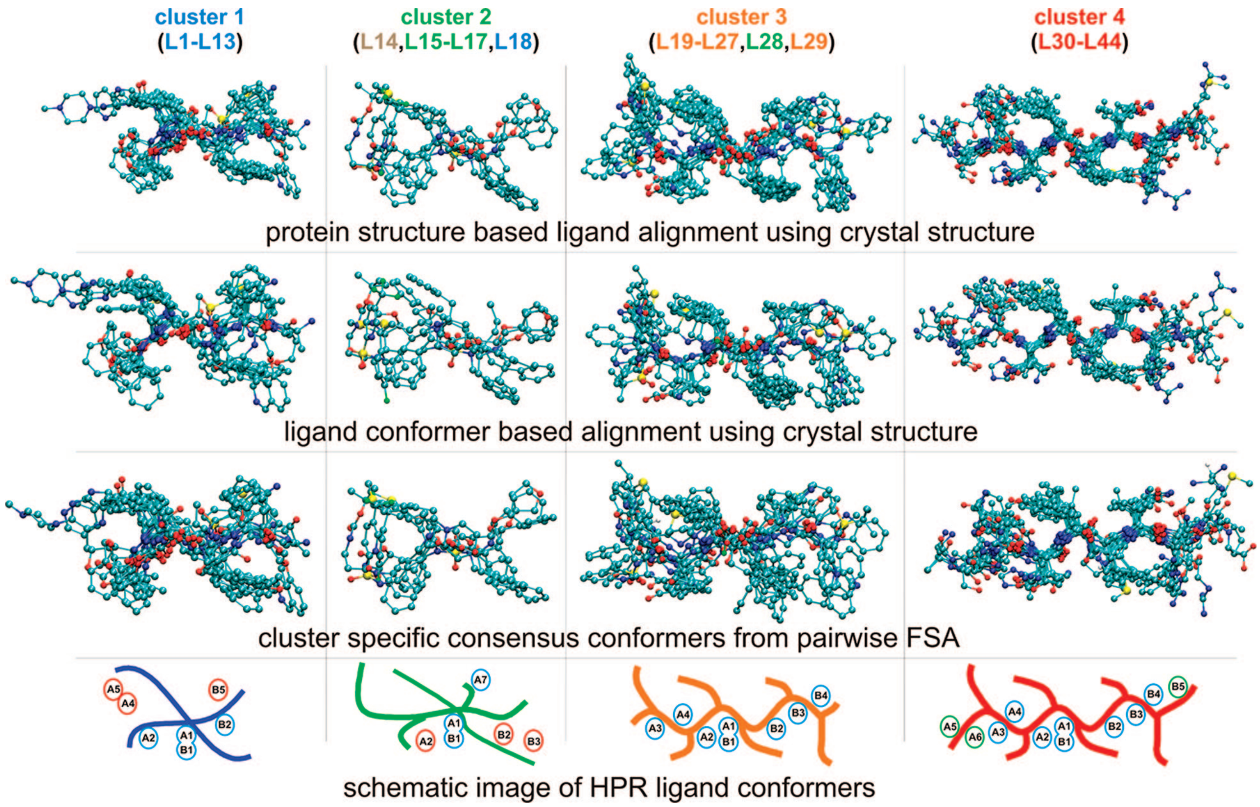


Figure 8. Structure alignments of HIV-P ligands for four clusters found from pairwise similarity based on FSA (see Figure 6). First row: protein structure-based RSA of ligand conformers in crystal structures. Second row: ligand conformer-based RSA using ligand conformers from crystal structures. Third row: structure alignment of consensus conformers from pairwise FSA. Last row: schematic representation of ligand conformers from protein structure-based RSA of first row. Circles indicate dominant H-bond patterns found from analysis of crystal structures. Labels in circles refer to residue identified in Table 2. Red circles indicate strict absence; blue circles strict presence of H-bonds; green circles dominant occurrences of H-bonds (circles obey same color code as ellipses in Table 2).

maximal $\mathrm{HI}_{\mathrm{combined}}$ in Figure 5a label the HIV-P ligand Lx, which by FSA yields the Ly conformer (on the abscissa) most similar to the crystal structure. Evidently FSA with ligand pairs belonging to the same cluster yield the highest similarity with the BLCs (see color code of ligand numbers in Figure 5). The same applies to the minimal RMSDs in Figure 5b.

E: Clustering HIV-P Ligands Based on Similarity Obtained by FSA. The preceding analysis of HIV-P ligand similarity suggests a heterogeneous set of ligands that may involve different binding modes. To purse this idea, we applied a cluster algorithm to explore the molecular similarities which consider functional groups and the scaffold of the ligands. The combined Hodgkin indices $\mathrm{HI}_{\mathrm{combined}}$, eq 11, of all $43 \times 44/2 = 946$ ligand pairs were evaluated by the FSA approach. We used the inverse of the $\mathrm{HI}_{\mathrm{combined}}$ as a measure of the distance between ligands (see Table SVI of the Supporting Information) to classify the ligands in different clusters by the neighbor-joining cluster algorithm.[58] We then used the results of this cluster algorithm to construct a phylogenetic tree by the program T-REX.[59] The resulting tree, displayed in Figure 6, shows four distinct ligand clusters. The smallest cluster (green, L14−L18) consists of five members. Together with the second largest cluster with 13 members (blue, L1−L13), it harbors the HIV-P ligands of smaller size. The third cluster contains 11 members (orange, L19−L29), while the largest cluster, containing the natural substrate oligopeptides, has 15 members (red, L30−L44). The colored circles around L14, L18, and L28 highlight differences in ligand classification derived from the H-bond patterns of the HIV-P ligands in the crystal structures (see part F, Table 2).

F: Classification of HIV-P Ligands Based on H-Bond Patterns from Crystal Structures. We can use the information from the crystal structures to classify HIV-P ligands. For this purpose, we analyzed the H-bond patterns of the HIV-P ligands, which have been discussed in numerous publications.[53,55,60−98] The binding affinities of these HIV-P ligands are also available[53,55,60−102] (see Table SV, Supporting Information) and their correlations with the H-bond patterns have been studied (Table SV), but no clear relations were found.

In the present study H-bonds are assumed to be present if the distances between the participating non-hydrogen atoms are smaller than 3.5 Å. The HIV-P ligands are listed in Table 2 in four groups (blue, green, orange, red, except for the outlier L14, which possesses a single H-bond and very low binding affinity, Table SV) corresponding to the H-bond patterns to which they belong. Using the same color code as in Figure 6, which displays the clustering results based on the FSA approach, Table 2 provides an abbreviated list of H-bonds that are significant for the clustering of the ligands. A complete list of the H-bonds formed between the ligands and the HIV-P homodimer is given in Table SV of the Supporting Information. The H-bonds, which are most relevant for discrimination between different groups of ligands, are highlighted by elongated ellipses (red ellipses for absence of possible H-bonds, blue ellipses for presence of H-bonds, and green ellipses for dominant, but not exclusive, occurrences of specific H-bonds).

For eight ligands (L3, L10, L14, L25, L26, L35, L40, L42) crystal structures with two different ligand orientations are available. In these cases, both H-bond patterns are given in Tables 2 and SV. Except for L14 the H-bond pattern of these ligands possesses mirror symmetry, which is due to the *C*2 symmetry of the HIV-P homodimer. L14 binds poorly, exhibiting one or two H-bonds in the two orientations.

Comparing the four classes of HIV-P ligands obtained by an analysis of the H-bond patterns (Table 2) with the clusters resulting from the similarities of pairwise FSA, one finds that the results of the two classifications are nearly identical (Figure 6). If we exclude HIV-P ligand L14, which is a poor binder, only two other ligands fall into similarity clusters that differ from the results of H-bond classification. Ligand L18 (L28) belongs to the blue (green) cluster according to the H-bond pattern but belongs to the green (orange) cluster based on FSA (Figure 6). Note that the ligands L15−L17, L28 of the green cluster have a polar ring structure in the center of the ligand scaffold in common, which induces a specific H-bond pattern. In L18 the corresponding polar ring in the ligand scaffold is off-center, giving rise to in a different H-bond pattern, which accordingly places the ligand in the blue cluster. The similarity of L18 to ligands of the blue or green cluster is nearly equal (see Figure S6 of the Supporting Information) with a slight preference for the green cluster. This contrasts with L28, which, based on similarity, is too large to fit into the green cluster of small ligands. Therefore, according to FSA it is placed in the orange cluster, while according to the H-bond pattern it should be placed in the green cluster.

G: Cluster-Specific Consensus Ligand Conformers Based on Pairwise FSA. For each of the 44 HIV-P ligands considered, the pairwise FSA method yields 43 different conformers. From these conformers one can derive CLCs without using crystal structure information. Since four different clusters of ligands were found based on pairwise similarity (see Figure 6), a CLC can be derived for each ligand L*y* using only the L*y* conformers obtained by pairwise FSA with the other ligands belonging to the same cluster. The procedure to generate such CLCs is described in part D3 of Methods. The resulting values of $\mathrm{HI}_{\mathrm{combined}}$ and RMSDs of the CLCs with the BLC of the crystal structures are given in Figure 7. The $\mathrm{HI}_{\mathrm{combined}}$ based on the CLCs vary now between 0.8 and 0.95. They are considerably larger than the corresponding maximal $\mathrm{HI}_{\mathrm{combined}}$ varying between 0.6 and 0.75 based on conventional FSA that disregards the results of ligand classification (Figure 5a). The RMSDs of the CLCs are around 2.0 Å. This value is as small as the minimal rmsd taken from the 43 ligand conformers obtained by conventional FSA (Figure 5b). However, in the latter case we do not know which one of the 43 ligand conformers yields the minimal rmsd. This demonstrates clearly how important the information of the ligand-cluster analysis is for the construction of cluster-specific CLCs.

H: Visualizing the HIV-P Ligand Clusters. Structure alignments of HIV-P ligands were performed for the BLCs and for conformers obtained by pairwise FSA. Two different pairwise rigid-structure alignment (RSA) methods were used for the ligand conformers from the crystal structures: (i) ligand conformer-based RSA using the pairwise FSA machinery with rigid torsion angles to align the BLCs; (ii) protein structure-based RSA using the Kabsch algorithm[56,57] to align the HIV-P crystal structures (without considering the ligands) by minimizing the rmsd, applying the resulting translation and rotation transforms to superimpose the corresponding ligands on their bound conformers. The results of the protein structure-based RSA are depicted in the first row of Figure 8 which shows the overlaid ligand structures for the four different ligand clusters found by analyzing the FSA results. The blue and green clusters contain 13 and 5 small ligands, while the orange and red clusters contain 11 and 15 large ligands, respectively.

For the ligand conformer-based RSA, the transforms (translations and rotations) necessary to perform the structural overlay of ligands are obtained by the same method used to generate CLCs for a set of ligand conformers as described in part S9 of the Supporting Information. The results of this structure alignment are shown in the second row of Figure 8.

The first two rows in Figure 8 display structure overlays of the HIV-P ligand conformers as they appear in the crystal structures obtained through the two different overlay techniques discussed above. The third row of Figure 8 shows the results of structural overlay of HIV-P ligands, where the ligand conformers were generated by the FSA method without any information from the HIV-P crystal structures. Here, we consider the cluster-specific CLCs obtained by the FSA method as described in preceding part G. The results of these structure overlays are strikingly similar to the RSA of the ligand conformers of the crystal structures of the first two rows of Figure 8. This finding demonstrates again the feasibility of the FSA approach determining the ligand conformers in the protein binding pocket without using knowledge on the protein structure.

The last row of Figure 8 summarizes schematically the similarity pattern of the HIV-P ligand conformers in the four different clusters derived by the FSA procedure. The most relevant H-bonds, which characterize the four different binding modes of the HIV-P ligands, are also indicated. It is interesting that the natural substrate peptides, which are cleaved by HIV-P, all belong to the red ligand cluster (right column in Figure 8), which possesses the most complex binding mode. The other HIV-P ligands belonging to the red cluster have the same binding mode found by an analysis of the H-bond pattern of the crystal structures (Table 2) and thus possess the same conformation. Moreover, ligands from the same cluster often possess chemically similar scaffolds, as evident from the chemical composition of the HIV-P ligands displayed in Table SII of the Supporting Information. Hence, for HIV-P ligands, differences in ligand binding modes also correlate with differences in ligand scaffolds. If this is a general feature of ligands, it will be more difficult to find new drugs possessing different scaffolds.[103]

Conclusions

In the present study we used MD simulations on ligand pairs from a set of ligands binding to the same target protein to perform flexible structure alignment (FSA). To make this procedure feasible, we used a specialized energy function. While it includes conventional intramolecular interactions, the true intermolecular interactions are replaced by attractive interactions derived from similarity measures accounting for volume/shape, atomic partial charges, and electrostatic potential of the ligands. These attractive terms allow equivalent molecular groups from a pair of flexible ligands to be superimposed, yielding consensus ligand conformers (CLC) that are most similar to each other. The assumption of the FSA approach is that the CLCs contain information on the true bound ligand conformer (BLC). Thus, if a sufficient number of ligands of different chemical architecture binding to the same target are available, it should be possible to obtain approximate BLCs without using any experimental structural information of the target protein.

We explored the FSA approach using a test set of 44 HIV-P crystal structures. For HIV-P, we have nowadays one of the largest databases of protein−ligand structures comprising many chemically diverse ligands. Since this database was 10 years ago only half as large, this study could not have been performed much earlier (see Figure S7).

Analyzing the similarity of the HIV-P ligand conformers obtained from the pairwise FSA, we found four different ligand clusters. There are two clusters of small and two clusters of large ligands. Ligands belonging to the same cluster exhibit similar backbone architecture. The five natural substrate oligopeptides all belong to the same cluster of large ligands. The four clusters agree well with a classification based on H-bond pattern of the ligands in the HIV-P binding pocket.

RMSDs of HIV-P ligands averaged over MD trajectories of independent ligands are 4 Å, if measured relative to the bound ligand conformers (BLCs) in the crystal structure. Even for closest approach to the BLCs in these MD trajectories the RMDSs are still 3 Å. In contrast, the consensus ligand conformers (CLCs) agree well with the BLC of the crystal structures with an rmsd of about 2 Å. Hence, the present approach of indirect drug design offers a possibility to generate coordinates of the BLCs without using any structural information from the target of the ligand. The application to 44 HIV-P ligands demonstrated that the procedure works well although the ligands possess four different binding modes. The ensemble of coordinates of CLCs belonging to the same binding mode that are generated by the present FSA approach approximate the true BLCs. In order that the FSA approach of indirect drug design presented in this study can work, one needs several (say five or more) ligands of different chemical architecture binding to the same target that possess the same binding mode as for instance the case for the considered 44 HIV-P ligands. In combination with the chemical composition (i.e., electrostatics, H-bond pattern, and hydrophobicity) of the ligands, the ensemble of CLCs can be used to define pharmacophores, thus opening alternative ways for drug design.

Acknowledgment. We thank Dr. Martin Karplus for providing the program CHARMM. We also acknowledge preliminary work from Srdjan Pokorni. We are grateful for useful suggestions and proofreading by Dr. Dennis Diestler. This work was supported by the Deutsche Forschungsgemeinschaft (Sfb 765 Project C1).

Supporting Information Available: Derivation of the electrostatic potential overlap. Details of the preparation of initial ligand conformers and generating consensus ligand conformers. Procedure for pairwise FSA by MD simulation with schematic demonstration. The RMSDs of 100 random structures generated for each of the 44 ligands. CHARMM scripts to perform structural overlap and generate consensus ligand conformers. 2D structures of 44 HPR ligands. $HI_{combined}$ and rmsd between alternate ligand conformers. Detailed hydrogen bonding pattern of bound HPR ligands. Pairwise structure alignment of the HPR reference ligands L*y* (L1, L2, ... L44) with all other 43 HPR ligands L*x* (x=1, 2, ... 44). $HI_{combined}$ from the initial conformers of the HPR ligands. Visualization of HPR ligand clusters on the basis of H-bond pattern. Distance matrix to generate the HPR ligand clusters. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kuntz, I. D. *Science* **1992**, *257*, 1078–1082.

(2) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. *J. Comput. Chem.* **2004**, *13*, 505–524.

(3) Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. *J. Med. Chem.* **1994**, *37*, 1035–1054.

(4) DesJarlais, R. L.; Seibel, G. L.; Kuntz, I. D.; Furth, P. S.; Alvarez, J. C.; Montellano, P. R. O. d.; DeCamp, D. L.; Babe, L. M.; Craik, C. S. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6644–6648.

(5) Hol, W. G. J. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 767–778.

(6) Richards, W. G. *Pure Appl. Chem.* **1994**, *66*, 1589–1596.

(7) Loew, G. H.; Villar, H. O.; Alkorta, I. *Pharm. Res.* **1993**, *10*, 475–486.

(8) Sheridan, R. P.; Rusinko, A.; Nilakantan, R.; Venkataraghavan, R. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8165–8169.

(9) Marshall, G. R.; Cramer, R. D. *Trends Pharmacol. Sci.* **1988**, *9*, 285–289.

(10) Bender, A.; Glen, R. C. *Org. Biomol. Chem.* **2004**, *2*, 3204–3218.

(11) Willett, P.; Barnard, J. M.; Downs, G. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996.

(12) Lewis, R. A. *J. Med. Chem.* **2005**, *48*, 1638–1648.

(13) Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*; VCH: Weinheim, 1993; Vol. 1.

(14) Yang, G.-F.; Huang, X. *Curr. Pharm. Des.* **2006**, *12*, 4601–4611.

(15) Tropsha, A.; Golbraikh, A. *Curr. Pharm. Des.* **2007**, *13*, 3494–3504.

(16) Leach, A. R., *Molecular Modelling: Principles and applications*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, 2001.

(17) Waterbeemd, H. v. d.; Rose, S. In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C.-G., Ed.; Academic Press: New York, 2008; Chapter 23, pp 491−513.

(18) Labute, P. *J. Mol. Graph. Model.* **2000**, *18*, 464–477.

(19) Labute, P. *Methods Mol. Biol.* **2004**, *275*, 261–278.

(20) Gozalbes, R.; Doucet, J. P.; Derouin, F. *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 93–102.

(21) Mason, J. S.; Good, A. C.; Martin, E. J. *Curr. Pharm. Des.* **2001**, *7*, 567–597.

(22) Güner, O. F. *Curr. Top. Med. Chem.* **2002**, *2*, 1321–1332.

(23) Sippl, W. In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C.-G., Ed.; Academic Press: New York, 2008; Chapter 28, pp 572−586.

(24) Langer, T.; Hoffmann, R. D. *Expert Opin. Drug Discovery* **2006**, *1*, 261–267 .

(25) Kontoyianni, M.; Madhav, P.; Suchanek, E.; Seibel, W. *Curr. Med. Chem.* **2008**, *15*, 107–116.

(26) Mclnnes, C. *Curr. Opin. Chem. Biol.* **2007**, *11*, 494–502.

(27) Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H. N.; Sastry, G. N. *Curr. Protein Pept. Sci.* **2007**, *8*, 329–351.

(28) Villoutreix, B. O.; Renault, N.; Lagorce, D.; Sperandio, O.; Montes, M.; Miteva, M. A. *Curr. Protein. Pept. Sci.* **2007**, *8*, 381–411.

(29) Cavasotto, C. N.; Orry, A. J. W. *Curr. Top. Med. Chem.* **2007**, *7*, 1006–1014.

(30) Jain, A. N. *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 396–403.

(31) Klebe, G. *Drug Discovery Today* **2006**, *11*, 580–594.

(32) Martin, Y. C. In *3D QSAR in Drug Design: Recent Advances*; Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds. Kluwer Academic: The Netherlands, 1998; Vol. 3, pp 3−23.

(33) Langer, T.; Bryant, S. D. In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C.-G., Ed. Academic Press: New York2008; Chapter 29, pp 587−604.

(34) Akamatsu, M. *Curr. Top. Med. Chem.* **2002**, *2*, 1381–94.

(35) Mor, M.; Rivara, S.; Lodola, A.; Lorenzi, S.; Bordi, F.; Plazzi, P. V.; Spadoni, G.; Bedini, A.; Duranti, A.; Tontini, A.; Tarzia, G. *Chem. Biodivers.* **2005**, *2*, 1438–1451.

(36) Sippl, W. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 825–830.

(37) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110* (18), 5959–5967.

(38) Böhm, M.; Stürzebecher, J.; Klebe, G. *J. Med. Chem.* **1999**, *42*, 458–477.

(39) Klebe, G. In *3D QSAR in Drug Design: Recent Advances*; Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds. Kluwer Academic: The Netherlands, 1998; Vol. 3, pp 87−104.

(40) Klebe, G.; Abraham, U. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 1–10.

(41) Suh, M.-E.; Park, S.-Y.; Lee, H.-J. *Bull. Korean Chem. Soc.* **2002**, *23*, 417–422.

(42) Ullmann, G. M.; Hauswald, M.; Jensen, A.; Kostic, N. M.; Knapp, E. W. *Biochemistry* **1997**, *36*, 16187–16196.

(43) Gandhi, R.; Bartlett, J. G.; Linkinhoker, M. Johns Hopkins University Division of Infectious Diseases and AIDS Service, 1999.

(44) Toh, H.; Ono, M.; Saigo, K.; Miyata, T. *Nature* **1985**, *315*, 691–692.

(45) Pearl, L. H.; Taylor, W. R. *Nature* **1987**, *329*, 351–354.

(46) Carbo, R.; Leyda, L.; Arnau, M. *Int. J. Quantum Chem.* **1980**, *17*, 1185–1189.

(47) Good, A. C. *J. Mol. Graphics* **1992**, *10*, 144–151.

(48) Hodgkin, E. E.; Richards, W. G. *Int. J. Quantum Chem.* **1987**, 105–110.

(49) Good, A. C.; Richards, W. G. In *3D QSAR in Drug Design: Ligand-Protein Interactions and Molecular Similarity*, Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds. Kluwer Academic: The Netherlands, 1998; Vol. 2, pp 321−338.

(50) Good, A. C.; Hodgkin, E. E.; Richards, W. G. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188–191.

(51) Good, A. C.; Richards, W. G. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 112–116.

(52) Meyer, A. Y.; Richards, W. G. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 427–439.

(53) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. *Structure* **2002**, *10*, 369–381.

(54) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(55) Bone, R.; Vacca, J. P.; Anderson, P. S.; Holloway, M. K. *J. Am. Chem. Soc.* **1991**, *113*, 9382–9384.

(56) Kabsch, W. *Acta Crystallogr.* **1976**, *A32*, 922–923.

(57) Kabsch, W. *Acta Crystallogr.* **1978**, *A34*, 827–828.

(58) Saitou, N.; Nei, M. *Mol. Biol. Evol.* **1987**, *4*, 406–425.

(59) Makarenkov, V. *Bioinformatics* **2001**, *17*, 664–668.

(60) Fitzgerald, P.; McKeever, B.; VanMiddlesworth, J.; Springer, J.; Heimbach, J.; Leu, C.; Herber, W.; Dixon, R.; Darke, P. *J. Biol. Chem.* **1990**, *265*, 14209–14219.

(61) Erickson, J.; Neidhart, D. J.; VanDrie, J.; Kempf, D. J.; Wang, X. C.; Norbeck, D. W.; Plattner, J. J.; Rittenhouse, J. W.; Turon, M.; Wideburg, N.; Kohlbrenner, W. E.; Simmer, R.; Helfrich, R.; Paul, D. A.; Knigge, M. *Science* **1990**, *249*, 527–533.

(62) Swain, A.; Miller, M.; Green, J.; Rich, D.; Schneider, J.; Kent, S.; Wlodawer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805–8809.

(63) Miller, M.; Schneider, J.; Sathyanarayana, B.; Toth, M.; Marshall, G.; Clawson, L.; Selk, L.; Kent, S.; Wlodawer, A. *Science* **1989**, *246*, 1149–1152.

(64) Rutenber, E.; Fauman, E.; Keenan, R.; Fong, S.; Furth, P.; Montellano, P. O. d.; Meng, E.; Kuntz, I.; DeCamp, D.; Salto, R.; Rose, J. R.; Craik, C. S.; Stroud, R. M. *J. Biol. Chem.* **1993**, *268*, 15343–15346.

(65) Priestle, J.; Fässler, A.; Rösel, J.; Tintelnot-Blomley, M.; Strop, P.; Grütter, M. *Structure* **1995**, *3*, 381–389.

(66) Munshi, S.; Chen, Z.; Li, Y.; Olsen, D. B.; Fraley, M. E.; Hungate, R. W.; Kuo, L. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *D54*, 1053–1060.

(67) Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. *Biochemistry* **1991**, *30*, 1600–1609.

(68) Ringhofer, S.; Kallen, J.; Dutzler, R.; Billich, A.; Visser, A. J. W. G.; Scholz, D.; Steinhauser, O.; Schreiber, H.; Auer, M.; Kungl, A. J. *J. Mol. Biol.* **1999**, *286*, 1147–1159.

(69) Brynda, J.; Rezacova, P.; Fabry, M.; Horejsi, M.; Stouracova, R.; Sedlacek, J.; Soucek, M.; Hradilek, M.; Lepsik, M.; Konvalinka, J. *J. Med. Chem.* **2004**, *47*, 2030–2036.

(70) Bäckbro, K.; Löwgren, S.; Österlund, K.; Atepo, J.; Unge, T.; Hultén, J.; Bonham, N. M.; Schaal, W.; Karlén, A.; Hallberg, A. *J. Med. Chem.* **1997**, *40*, 898–902.

(71) Hong, L.; Hartsuck, J. A.; Foundling, S.; Ermolieff, J.; Tang, J. *Protein Sci.* **1998**, *7*, 300–305.

(72) Hong, L.; Treharne, A.; Hartsuck, J. A.; Foundling, S.; Tang, J. *Biochemistry* **1996**, *35*, 10627–10633.

(73) Abbenante, G.; March, D. R.; Bergman, D. A.; Hunt, P. A.; Garnham, B.; Dancer, R. J.; Martin, J. L.; Fairlie, D. P. *J. Am. Chem. Soc.* **1995**, *117*, 10220–10226.

(74) Geremia, S.; Demitri, N.; Wuerges, J.; Benedetti, F.; Berti, F.; Tell, G.; Randaccio, L. *ChemMedChem* **2006**, *1*, 186–188.

(75) Thaisrivongs, S.; Skulnick, H. I.; Turner, S. R.; Strohbach, J. W.; Tommasi, R. A.; Johnson, P. D.; Aristoff, P. A.; Judge, T. M.; Gammill, R. B.; Morris, J. K.; Romines, K. R.; Chrusciel, R. A.; Hinshaw, R. R.; Chong, K.-T.; Tarpley, W. G.; Poppe, S. M.; Slade, D. E.; Lynn, J. C.; Horng, M.-M.; Tomich, P. K.; Seest, E. P.; Dolak, L. A.; Howe, W. J.; Howard, G. M.; Schwende, F. J.; Toth, L. N.; Padbury, G. E.; Wilson, G. J.; Shiou, L.; Zipp, G. L.; Wilkinson, K. F.; Rush, B. D.; Ruwart, M. J.; Koeplinger, K. A.; Zhao, Z.; Cole, S.; Zaya, R. M.; Kakuk, T. J.; Janakiraman, M. N.; Watenpaugh, K. D. *J. Med. Chem.* **1996**, *39*, 4349–4353.

(76) Silva, A. M.; Cachau, R. E.; Sham, H. L.; Erickson, J. W. *J. Mol. Biol.* **1996**, *255*, 321–340.

(77) Andersson, H. O.; Fridborg, K.; Löwgr; en, S.; Alterman, M.; Mühlman, A.; Björsne, M.; Garg, N.; Kvarnström, I.; Schaal, W.; Classon, B.; Karlén, A.; Danielsson, U. H.; Ahlsén, G.; Nillroth, U.; Vrang, L.; Öberg, B.; Samuelsson, B.; Hallberg, A.; Unge, T. *Eur. J. Biochem.* **2003**, *270*, 1746–1758.

(78) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. *J. Mol. Biol.* **2000**, *301*, 1207–1220.

(79) Hong, L.; Zhang, X.; Hartsuck, J.; Tang, J. *Protein Sci.* **2000**, *9*, 1898–1904.

(80) Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. *J. Med. Chem.* **1991**, *34*, 3340–3342.

(81) Hoog, S. S.; Zhao, B.; Winborne, E.; Fisher, S.; Green, D. W.; DesJarlais, R. L.; Newlander, K. A.; Callahan, J. F.; Abdel-Meguid, S. S.; Moore, M. L.; Huffman, W. F. *J. Med. Chem.* **1995**, *38*, 3246–3252.

(82) Murthy, K.; Winborne, E.; Minnich, M.; Culp, J.; Debouck, C. *J. Biol. Chem.* **1992**, *267*, 22770–22778.

(83) Thanki, N.; Rao, J.; Foundling, S. I.; Howe, W. J.; Moon, J. B.; Hui, J. O.; Tomasselli, A. G.; Heinrikson, R. L.; Thaisrivongs, S.; Wlodawer, A. *Protein Sci.* **1992**, *1*, 1061–1072.

(84) Smith, A. B.; Charnley, A. K.; Harada, H.; Beiger, J. J.; Cantin, L.-D.; Kenesky, C. S.; Hirschmann, R.; Munshi, S.; Olsen, D. B.; Stahlhut, M. W.; Schleif, W. A.; Kuo, L. C. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 859–863.

(85) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. *Structure* **1995**, *3*, 581–590.

(86) Jhoti, H.; Singh, O. M. P.; Weir, M. P.; Cooke, R.; Murray-Rust, P.; Wonacott, A. *Biochemistry* **1994**, *33*, 8417–8427.

(87) Ekegren, J. K.; Unge, T.; Safa, M. Z.; Wallberg, H.; Samuelsson, B.; Hallberg, A. *J. Med. Chem.* **2005**, *48*, 8098–8102.

(88) Thaisrivongs, S.; Watenpaugh, K. D.; Howe, W. J.; Tomich, P. K.; Dolak, L. A.; Chong, K.-T.; Tomich, C.-S. C.; Tomasselli, A. G.; Turner, S. R.; Strohback, J. W.; Mulichak, A. M.; Janakiraman, M. N.; Moon, J. B.; Lynn, J. C.; Horng, M.-M.; Hinshaw, R. R.; Curry, K. A.; Rothrock, D. J. *J. Med. Chem.* **1995**, *38*, 3624–3637.

(89) Kempf, D.; Marsh, K.; Denissen, J.; McDonald, E.; Vasavanonda, S.; Flentge, C.; Green, B.; Fino, L.; Park, C.; Kong, X.; Wideburg, N.; Saldivar, A.; Ruiz, L.; Kati, W.; Sham, H.; Robins, T.; Stewart, K.; Hsu, A.; Plattner, J.; Leonard, J.; Norbeck, D. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2484–2488.

(90) Lange-Savage, G.; Berchtold, H.; Liesum, A.; Budt, K.-H.; Peyman, A.; Knolle, J.; Sedlacek, J.; Fabry, M.; Hilgenfeld, R. *Eur. J. Biochem.* **1997**, *248*, 313–322.

(91) March, D. R.; Abbenante, G.; Bergman, D. A.; Brinkworth, R. I.; Wickramasinghe, W.; Begun, J.; Martin, J. L.; Fairlie, D. P. *J. Am. Chem. Soc.* **1996**, *118*, 3375–3379.

(92) Stoll, V.; Qin, W.; Stewart, K. D.; Jakob, C.; Park, C.; Walter, K.; Simmer, R. L.; Helfrich, R.; Bussiere, D.; Kao, J.; Kempf, D.; Sham, H. L.; Norbeck, D. W. *Bioorg. Med. Chem.* **2002**, *10*, 2803–2806.

(93) Kervinen, J.; Thanki, N.; Zdanov, A.; Tino, J.; Barrish, J.; Lin, P. F.; Colonno, R.; Riccardi, K.; Samanta, H.; Wlodawer, A. *Protein Pept. Lett.* **1996**, *3*, 399–406.

(94) Kervinen, J.; Lubkowski, J.; Zdanov, A.; Bhatt, D.; Dunn, B. M.; Hui, K. Y.; Powell, D. J.; Kay, J.; Wlodawer, A.; Gustchina, A. *Protein Sci.* **1998**, *7*, 2314–2323.

(95) Rutenber, E. E.; McPhee, F.; Kaplan, A. P.; Gallion, S. L.; Hogan, J. C., Jr.; Craik, C. S.; Stroud, R. M. *Bioorg. Med. Chem.* **1996**, *4*, 1545–1558.

(96) Jadhav, P. K.; Ala, P.; Woerner, F. J.; Chang, C.-H.; Garber, S. S.; Anton, E. D.; Bacheler, L. T. *J. Med. Chem.* **1997**, *40*, 181–191.

(97) Abdel-Meguid, S. S.; Metcalf, B. W.; Carr, T. J.; Demarsh, P.; DesJarlais, R. L.; Fisher, S.; Green, D. W.; Ivanoff, L.; Lambert, D. M.; Murthy, K. H. M.; Petteway, S. R., Jr.; Pitts, W. J.; Tomaszek, T. A., Jr.; Winborne, E.; Zhao, B.; Dreyer, G. B.; Meek, T. D. *Biochemistry* **1994**, *33*, 11671–11677.

(98) Surleraux, D. L. N. G.; Tahri, A.; Verschueren, W. G.; Pille, G. M. E.; Kock, H. A. d.; Jonckers, T. H. M.; Peeters, A.; Meyer, S. D.; Azijn, H.; Pauwels, R.; Bethune, M.-P. d.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B. T. P. *J. Med. Chem.* **2005**, *48*, 1813–1822.

(99) Rutenber, E. E.; Voss, J. J. D.; Hoffman, L.; Stroud, R. M.; Lee, K. H.; Alvarez, J.; McPhee, F.; Craik, C.; Montellano, P. R. O. d. *Bioorg. Med. Chem.* **1997**, *5*, 1311–1320.

(100) Ermolieff, J.; Lin, X.; Tang, J. *Biochemistry* **1997**, *36*, 12364–12370.

(101) Barrish, J. C.; Gordon, E.; Alam, M.; Lin, P.-F.; Bisacchi, G. S.; Chen, P.; Cheng, P. T. W.; Fritz, A. W.; Greytok, J. A.; Hermsmeier, M. A.; Humphreys, W. G.; Lis, K. A.; Marella, M. A.; Merchant, Z.; Mitt, T.; Morrison, R. A.; Obermeier, M. T.; Pluscec, J.; Skoog, M.; Slusarchyk, W. A.; Spergel, S. H.; Stevenson, J. M.; Sun, C.-q.; Sundeen, J. E.; Taunk, P.; Tino, J. A.; Warrack, B. M.; Colonno, R. J.; Zahler, R. *J. Med. Chem.* **1994**, *37*, 1758–1768.

(102) Ghosh, A. K.; Sridhar, P. R.; Leshchenko, S.; Hussain, A. K.; Li, J.; Kovalevsky, A. Y.; Walters, D. E.; Wedekind, J. E.; Grum-Tokars, V.; Das, D.; Koh, Y.; Maeda, K.; Gatanaga, H.; Weber, I. T.; Mitsuya, H. *J. Med. Chem.* **2006**, *49*, 5252–5261.

(103) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894–2896.

CT8004886