JCTC Journal of Chemical Theory and Computation

Kinetic Model of Surface Segregation in Pt-Based Alloys

Yuguang Ma and Perla B. Balbuena*

Department of Chemical Engineering, Texas A&M University, College Station, Texas 77843

Received July 11, 2008

Abstract: A new vacancy model describes the kinetic process of surface segregation on clean surfaces via atomic movement through vacancies. The detailed segregation mechanism of two impurity metals (Au and Co) in Pt(111) is examined using periodic density functional theory, carefully evaluating energy barriers for each step in the segregation process. Au shows a strong surface segregation trend to the clean Pt(111) surface, while Co segregating to the surface is neither thermodynamically nor kinetically favorable.

Introduction

Platinum has become the most popular cathode electrocatalyst in polymer electrode membrane (PEM) fuel cells. However, its relatively sluggish kinetics for the oxygen reduction reaction (ORR) significantly limits the catalytic efficiency. In recent years, Pt-based alloys, composed of Pt and other transition metals, have been extensively investigated[1-7] owing to their enhanced ORR activity. The improvement of catalytic activities can be attributed to the modification of the geometric and electronic properties in the alloys. Surface segregation has been found to play a key role in the process. For example, Stamenkovic et al. compared the kinetics of the ORR on bulk Pt_3Co alloys and on Pt-skin structures produced by Pt segregating to the outmost layer of the annealed alloy surface. They noted that the ORR is uniquely active on Pt-skin surfaces and the reaction rate is four times that of the pure Pt catalyst.[6] On the other hand, Pt-based alloys, as cathode electrocatalysts, possess enhanced stability compared to pure Pt.[8,9] Some recent studies have revealed that the surface structure of these alloys differs considerably from that in the bulk: A pure Pt-skin is formed in the outmost layer of the alloys due to surface segregation of Pt atoms.[4,6,7,10-12] Consequently, the enhanced stability of Pt-based catalysts is also closely related to surface segregation.

Owing to the importance of surface segregation for Pt-based alloys, numerous studies have been carried out using modern experimental and computational techniques.[2,6,9-11,13-27] For many Pt alloy systems, segregation trends have been unambiguously identified, and the components of surface and subsurface regions have successfully been characterized. Nevertheless, some details of segregation are not completely understood. Particularly, the kinetics of surface segregation, which is critical to elucidate the driving forces and the process of this phenomenon, has barely been addressed at a molecular level.

In this work, we study the segregation mechanism using a vacancy model, on the basis of first principles calculations. The model includes a Pt slab with impurity metal atoms which allow the metal atoms moving to the surface through vacancies. A series of steps, including vacancy formation, metal atom migration, and vacancy regeneration, are proposed to describe the kinetics of surface segregation in this model. Two impurity metals, Au and Co, are employed due to their different segregation tendencies to clean Pt(111) surfaces:[28] Au shows strong surface segregation in Pt–Au alloys,[29,30] whereas Co is usually unable to enrich Pt–Co alloy surfaces.[16,17] We remark that in this model we do not include other effects such as the presence of oxygenated adsorbates on the surface, the acid medium, and the applied electric field; all of them may influence the studied phenomena. Thus, this work focuses on a simplified analysis which in a first approximation neglects all the external effects and focuses on the kinetics and thermodynamics driven and determined only by the nature of the alloy. This is the first step before introducing other complexities in the system.

Computational Models

For Pt-based alloys, surface segregation may take place by annealing the alloys at high temperatures.[6,31] Such an annealing process generally occurs by diffusion of atoms in solid state alloys, and vacancies are needed to facilitate the movement of metal atoms. The phenomenon can also be observed in acidic solution, often with dissolution of the alloyed metal and rearrangement of Pt atoms.[4,12,32] The dissolution of metal atoms results in some defects on the alloy surface, which are critical to the rearrangement process. Based on the experimental information, we propose a vacancy model to elucidate the plausible segregation mechanism, using the basic assumption that surface segregation can be achieved by the movement of metal atoms through the vacancies. As illustrated in Figure 1, defects are initially formed on the surface by metal atoms leaving the surface and then metal atoms located at the sublayers filling out those vacant sites. The process leads to new vacancies formed in the sublayers, which allow surface metal atoms to move back to the subsurface or bulk atoms to fill out the new vacancies.

* Corresponding author e-mail: balbuena@tamu.edu.
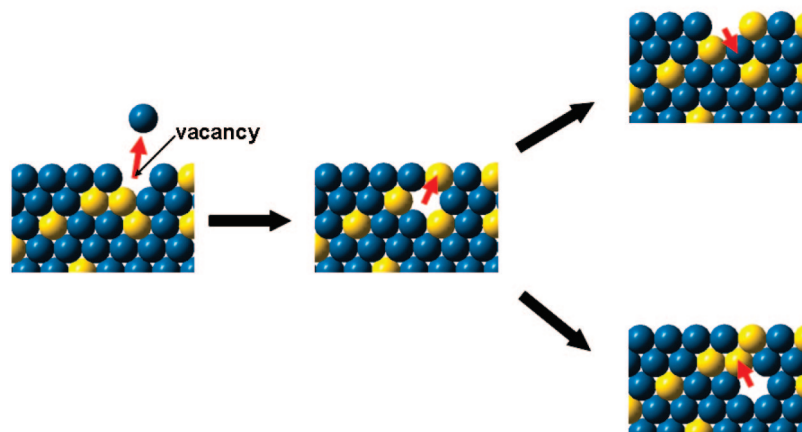
Figure 1. Plausible mechanism of surface segregation through vacancies.
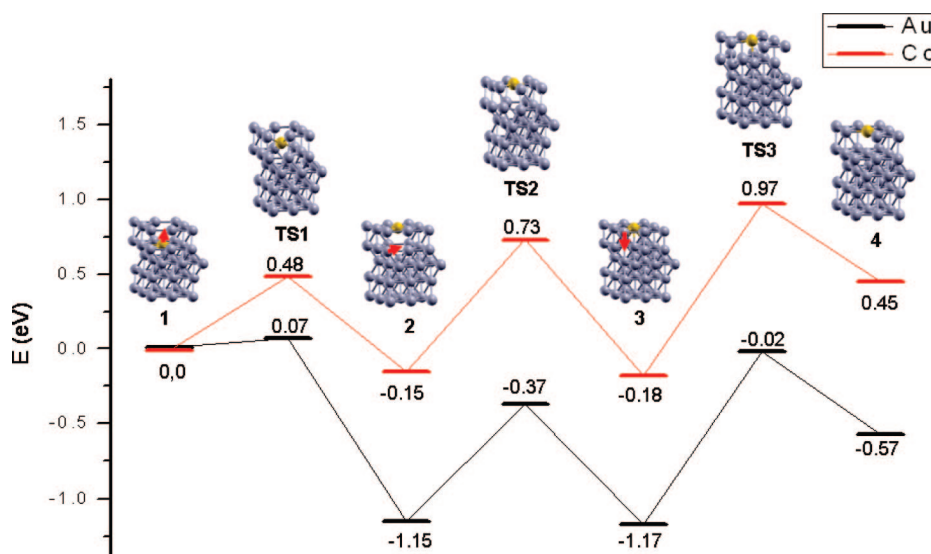
Figure 2. Calculated pathway and energetics for the impurity atoms segregating from the second layer to the Pt(111) surface. Grey and yellow spheres represent the Pt and impurity atoms, respectively.

Since Pt possesses a closed-packed fcc crystal structure and the most stable facet of the crystal is (111), our slab models were constructed using 3×3 (111) unit cells. Each slab consists of five-layers of metal atoms and six equivalent layers of vacuum. For each supercell, the initial position of the vacancy was set to the center of outmost layer, and the impurity atom was placed at the center of the second layer, as shown in Figure 2 (structure **1**). The atoms in the top three layers were allowed to relax to their lowest energy configuration, while the atoms of the bottom two layers were fixed to their bulk positions according to their optimized lattice constants.

The spin-polarized DFT calculations were performed using the Vienna *ab initio* simulation package (VASP),[33−37] in which the Kohn−Sham equations are solved by self-consistent algorithms. The valence electrons were described by plane wave basis sets with a cutoff energy of 350 eV, and the core electrons were replaced by the projector augmented wave (PAW) pseudopotentials[38,39] for improving the computation efficiency. The Brillouin zone was sampled with a $4 \times 4 \times 1$ Monkhorst-Pack k-point mesh for the periodic DFT calculations. The exchange-correlation functional was described within the generalized gradient approximation (GGA) proposed by Perdew,

Burke, and Ernzerhof (PBE).[40] The Methfessel-Paxon method was employed to determine electron occupancies with a smearing width of 0.2 eV. The geometries of the slabs were relaxed until the force was convergent to 10^{-4} eV/Å. The minimum energy paths, transition states, and energy barriers were determined by the nudged elastic band method.[41,42]

Results and Discussion

We study a plausible pathway for the impurity atoms, Au and Co, moving from the second layer to the topmost layer in the Pt(111) slab. The segregation process thus includes interlayer migration of the impurity atom, vacancy movement at the second layer, and vacancy regeneration at the surface layer, as depicted in Figure 2.

The average bond lengths of Pt−Au and Pt−Co are determined as 2.81 Å and 2.74 Å in structure **1**, respectively. In the first step, the impurity atom moves to the surface layer, leaving a vacancy in the second layer. Our calculations show that Au is both thermodynamically and kinetically favorable for this migration, with a total energy change of −1.15 eV and a barrier of 0.07 eV, indicating a strong surface segregation tendency. The corresponding values for Co, −0.15 eV and 0.48 eV, are

Letter

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **1993**
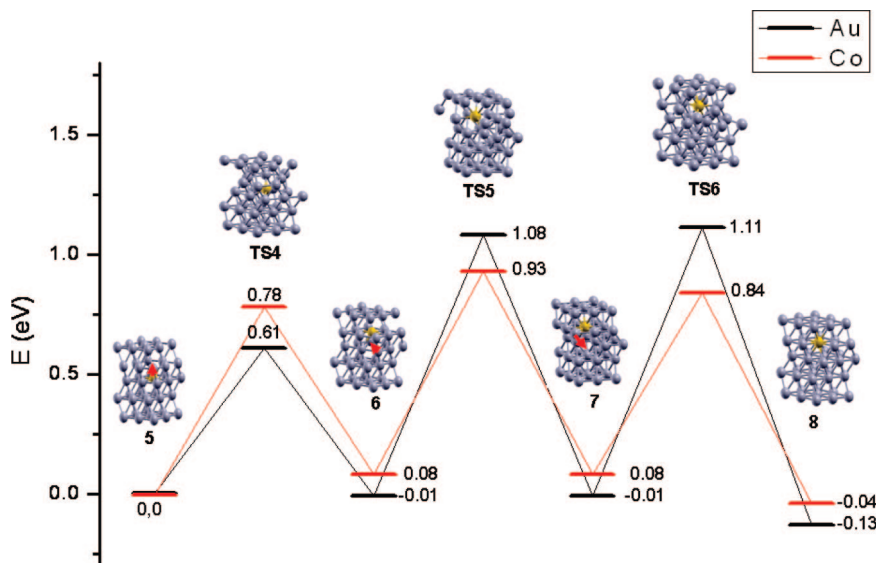
Figure 3. Calculated pathway and energetics for the impurity atoms segregating from the third layer to the second layer in Pt(111). Grey and yellow spheres represent the Pt and impurity atoms, respectively.

much higher than those of Au. Notably, the positions of the Au and Co atoms are considerably different in the transition state **TS1**, where the Au atom is closer to the surface layer (1.18 Å) than the Co atom (1.52 Å).

For continuing the surface segregation process, new vacancies must be formed at the surface. In acidic medium, this may be attained by dissolving surface metal atoms. The other approach, as we describe here, is via the migration of vacancies. This process is generally accomplished in two stages. The first stage (**2** to **3** in Figure 2) involves the intralayer movement of the Pt atoms neighboring the vacancies, and the second stage (**3** to **4** in Figure 2) includes the interlayer migration of the surface Pt atoms, leading to new vacancies in the surface. The detailed pathway can be found in Figure 2: From **2** to **3**, the total energy of the slab only changes very slightly, around -0.02 to -0.03 eV. The movement of a second layer Pt atom to the existent vacancy needs to surpass a barrier of 0.78 eV and 0.88 eV for the Pt−Au and Pt−Co alloys, respectively. The energy barrier arises primarily from the repulsion of neighboring Pt atoms. For the Pt−Au alloy, the Pt−Pt bond length decreases from 2.72−2.74 Å in **2** to 2.54−2.60 Å in **TS2**. A similar trend is observed for the Pt−Co alloy: the bond length varies from 2.73−2.74 Å in **2** to 2.55−2.61 Å in **TS2**. In the following step (**3** to **TS3**), the surface Pt atoms move down to occupy the vacancies in the second layer. The interlayer migration is determined as the rate-determining step (RDS) owing to its relatively high energy barrier (1.15 eV for both alloys). In the transition state (**TS3**), the migrated Pt atom is closer to the second layer than to the surface layer for both the Pt−Au and Pt−Co systems, suggesting an endothermic step in the conversion from **3** to **4**. Our calculations further verify that the step is not energetically favored, with endothermicity of 0.60 eV and 0.63 eV for Pt−Au and Pt−Co.

Previous experimental and theoretical studies have clearly revealed the segregation tendency for clean surfaces of Pt−Au and Pt−Co alloys. Au enriches strongly the surface of Pt-based alloys, whereas Co does not show segregation tendency. Our segregation model can successfully interpret these phenomena, particularly, from a kinetics viewpoint. The pathway from **1** to

4 describes a complete cycle in which an impurity atom in the second layer segregates to the Pt(111) surface. The process is exothermic (-0.57 eV) for the Pt−Au system, whereas it is endothermic (0.45 eV) for the Pt−Co system. This leads to the conclusion that Au surface segregation is thermodynamically favorable, consistent with previous studies.[26,28−30] Furthermore, our model reveals that Au segregation to the Pt(111) surface is kinetically feasible. Once vacancies are formed in the second layer, either Pt atoms or Au atoms at the surface can move down to the second layer. The steps correspond to conversions from **3** to **4** and from **2** to **1** in Figure 2. Comparing the two competitive steps, we notice that the latter one possesses a higher energy barrier (1.22 eV) than the former one (1.15 eV), indicating that Au antisegregation is unlikely to take place. For the Pt−Co alloy, the conversion from **2** to **1** has an energy barrier of 0.63 eV, much lower than that of the competitive step from **3** to **4** (1.15 eV), suggesting the surface Co atoms have a stronger tendency to migrate to the second layer compared to the Pt atoms. Accordingly, structure **4** most likely will not be formed, and thus the continuous segregation of Co cannot be reached. Therefore, Co segregating to the clean Pt(111) surface is both thermodynamically unfavorable and kinetically prohibitive according to our model.

We also examine a similar pathway involving segregation of impurity atoms from the third layer to the second layer. The complete process is shown in Figure 3. In the first step, the impurity atoms at the third layer move up to the vacancies at the second layer. The conversion from **5** to **6** only leads to a slight energy change, reported as -0.01 eV and 0.08 eV for the Pt−Au and Pt−Co systems. The energy barrier for the step (0.61 eV for Pt−Au and 0.78 eV for Pt−Co) is much higher than the similar conversion from **1** to **2**. This may be attributed to the steric effect: The half-open vacancies on the surface can provide more spacious room for the migration of metal atoms, while the vacancies at the second layer that are surrounded by atoms in all directions may result in space congestion in the process. For example, the distance of Au and the neighboring Pt atoms is 2.62−2.65 Å in **TS1** and is decreased to 2.60 Å in **TS4**. During interlayer movement of the Pt atom (**6** to **7**), energy

is approximately conserved. In this step, the barrier of the Pt−Au alloy is around 0.25 eV higher than that of the Pt−Co alloy. The step of **7** to **8** involving the downward movement of the Pt-atom, with an exothermicity of −0.12 eV for both alloys, differs considerably from the endothermic step of **3-4**. In addition, the energy barrier is also found to be lower than that in step of **3-4**, particularly, for the Pt−Co system. These results suggest that Pt diffusion from the second layer to the third layer is more energetically favorable than the similar diffusion from the surface to the second layer.

For the Pt−Co system, steps **5-6**, **6-7**, and **7-8** possess comparable barriers, which are 0.78 eV, 0.84 eV, and 0.75 eV, respectively. Since there is no RDS in the process, the segregation rate of Co is determined by all three steps. For the Pt−Au system, the segregation rate is mainly controlled by steps **5-6** and **6-7**, the intralayer and interlayer migration of Pt atoms. On the other hand, a small exothermicity in the conversion from **5** to **8** makes the segregation of Au/Co possibly feasible. However, as shown in Figure 3, each step in the process **5** to **8** is either mildly exothermic or mildly endothermic. Therefore, given enough energy to overcome the energy barriers, the segregation process is reversible, and the direction can be affected or determined by other related processes.

To make connection to the diffusion process of the impurity, we use the equation that gives the diffusion coefficient D due to atomic substitutional motion[43]

$$D = \frac{1}{6}\alpha^2 z v \chi_v \ \exp\left(\frac{-\Delta G_m}{kT}\right) \tag{1}$$

where

$$\alpha = a/\sqrt{2}$$

is the jump distance for an fcc metal with $a = 3.97$ Å the calculated lattice constant of the Pt lattice, z is the bulk coordination number $= 12$, and v is the vibration frequency of a metal atom, estimated as 10^{13} s^{-1}.[43] Taking ΔG_m as the calculated activation energy for migration of the Co impurity (0.48 eV), we can estimate the concentration of vacancies X_v required for a given value of the diffusion coefficient. Values of the diffusion coefficients of Co growing an ultrathin film on a Pt(111) surface have been reported[44] to follow an Arrhenius form as a function of temperature with a pre-exponential factor of 6.6×10^{-11} cm^2/s and an activation barrier of 0.90 eV; we used these experimental values to estimate the vacancy concentration needed for such diffusion using our calculated values and eq 1. For comparison, we also report the concentration of surface vacancies obtained by the Boltzmann factor using our calculated activation energy for surface vacancy formation in the case of Co impurities (−1.15 eV). The results based solely on our calculated vacancy energy formation shown in Table 1 are in closer agreement with the temperature dependence reported by a recent study of vacancy formation and atomic migration in fcc Al.[45]

Conclusions

The vacancy model presented the paper reasonably explains the well-known segregation trends of Au and Co to the Pt(111) surface. More importantly, the model provides an approach to evaluate the energy barriers in the process, which is crucial to study the kinetic properties of surface segregation. As mentioned

Table 1. Estimated Concentration of Vacancies Needed for Migration of Infinitely Diluted Co Impurities Using Eq *1* and Experimental Diffusion Coefficients from Ref 43 and Using the Activation Energies for Surface Vacancy Formation Obtained in This Work

temperature (K)	vacancy concentration according to experimental diffusion and eq 1	vacancy concentration based on calculated activation energy for vacancy formation
300	3.70691E-16	4.87946E-20
500	2.45578E-13	2.58835E-12
1000	3.2068E-11	1.60884E-06
1500	1.62692E-10	0.000137301
2000	3.66449E-10	0.001268399

in the Introduction, this is a systematic approach to evaluate the basic segregation steps of an infinitely diluted impurity determined only by the nature of the alloy. Future work using this approach may address other factors such as alloy composition, presence of adsorbates, and/or solvents, among others.

References

(1) Paulus, U. A.; Wokaun, A.; Scherer, G. G.; Schmidt, T. J.; Stamenkovic, V.; Radmilovic, V.; Markovic, N. M.; Ross, P. N. Oxygen Reduction on Carbon-Supported Pt-Ni and Pt-Co Alloy Catalysts. *J. Phys. Chem. B* **2002**, *106*, 4181–4191.

(2) Mun, B. S.; Watanabe, M.; Rossi, M.; Stamenkovic, V.; Markovic, N. M.; Ross, P. N., Jr. A study of electronic structures of Pt3M (M=Ti,V,Cr,Fe,Co,Ni) polycrystalline alloys with valence-band photoemission spectroscopy. *J. Chem. Phys.* **2005**, *123*, 204717/ 204711–204717/204714.

(3) Kitchin, J. R.; Norskov, J. K.; Barteau, M. A.; Chen, J. G. Modification of the surface electronic and chemical properties of Pt(111) by subsurface 3d transition metals. *J. Chem. Phys.* **2004**, *120*, 10240–10246.

(4) Toda, T.; Igarashi, H.; Uchida, H.; Watanabe, M. Enhancement of the electroreduction of oxygen on Pt alloys with Fe, Ni, and Co. *J. Electrochem. Soc.* **1999**, *146*, 3750–3756.

(5) Mukerjee, S.; Srinivasan, S.; Soriaga, M. P. Role of structural and electronic properties of Pt and Pt alloys on electrocatalysis of oxygen reduction. An in situ XANES and EXAFS investigation. *J. Electrochem. Soc.* **1995**, *142*, 1409–1422.

(6) Stamenkovic, V.; Schmidt, T. J.; Ross, P. N.; Markovic, N. M. Surface Composition Effects in Electrocatalysis: Kinetics of Oxygen Reduction on Well-Defined Pt3Ni and Pt3Co Alloy Surfaces. *J. Phys. Chem. B* **2002**, *106*, 11970–11979.

(7) Stamenkovic, V. R.; Fowler, B.; Mun, B. S.; Wang, G.; Ross, P. N.; Lucas, C. A.; Markovic, N. M. Improved Oxygen Reduction Activity on Pt3Ni(111) via Increased Surface Site Availability. *Science (Washington, DC, U. S.)* **2007**, *315*, 493–497.

(8) Yu, P.; Pemberton, M.; Plasse, P. PtCo/C cathode catalyst for improved durability in PEMFCs. *J. Power Sour.* **2005**, *144*, 11– 20.

(9) Menning, C. A.; Hwu, H. H.; Chen, J. G. Experimental and Theoretical Investigation of the Stability of Pt-3d-Pt(111) Bimetallic Surfaces under Oxygen Environment. *J. Phys. Chem. B* **2006**, *110*, 15471–15477.

Letter

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **1995**

(10) Stamenkovic, V.; Schmidt, T. J.; Ross, P. N.; Markovic, N. M. Surface segregation effects in electrocatalysis: kinetics of oxygen reduction reaction on polycrystalline Pt3Ni alloy surfaces. *J. Electroanal. Chem.* **2003**, *554−555*, 191–199.

(11) Stamenkovic, V. R.; Mun, B. S.; Mayrhofer, K. J. J.; Ross, P. N.; Markovic, N. M. Effect of Surface Composition on Electronic Structure, Stability, and Electrocatalytic Properties of Pt-Transition Metal Alloys: Pt-Skin versus Pt-Skeleton Surfaces. *J. Am. Chem. Soc.* **2006**, *128*, 8813–8819.

(12) Wan, L.-J.; Moriyama, T.; Ito, M.; Uchida, H.; Watanabe, M. In situ STM imaging of surface dissolution and rearrangement of a Pt-Fe alloy electrocatalyst in electrolyte solution. *Chem. Commun. (Cambridge, U. K.)* **2002**, *n/a*, 58–59.

(13) van den Oetelaar, L. C. A.; Nooij, O. W.; Oerlemans, S.; Denier van der Gon, A. W.; Brongersma, H. H.; Lefferts, L.; van Veen, J. A. R. Surface Segregation in Supported Pd-Pt Nanoclusters and Alloys. *J. Phys. Chem. B* **1998**, *102*, 3445–3455.

(14) Deckers, S.; Habraken, F. H. P. M.; Van der Weg, W. F.; Denier van der Gon, A. W.; Pluis, B.; Van der Veen, J. F.; Baudoing, R. Segregation at the platinum-nickel (Pt0.5Ni0.5)(111) surface studied by medium-energy ion scattering. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1990**, *42*, 3253–3259.

(15) Gauthier, Y.; Senhaji, A.; Legrand, B.; Treglia, G.; Becker, C.; Wandelt, K. An unusual composition profile. A LEED-TBIM study of Pt25Cu75(1 1 1). *Surf. Sci.* **2003**, *527*, 71–79.

(16) Gauthier, Y.; Baudoing-Savois, R.; Bugnard, J. M.; Hebenstreit, W.; Schmid, M.; Varga, P. Segregation and chemical ordering in the surface layers of Pt25Co75(111). A LEED/STM study. *Surf. Sci.* **2000**, *466*, 155–166.

(17) Gauthier, Y.; Baudoing-Savois, R.; Rosink, J. J. W. M.; Sotto, M. LEED study of Pt25Co75(111). *Surf. Sci.* **1993**, *297*, 193–201.

(18) Visser, R. F.; Roux, J. P. Segregation and oxidation studies on chromium, chromium-platinum (Cr5Pt) and Cr15Pt. *Appl. Surf. Sci.* **1991**, *51*, 115–124.

(19) Creemers, C.; Deurinck, P. Platinum segregation to the (111) surface of ordered Pt80Fe20: LEIS results and model simulations. *Surf. Interface Anal.* **1997**, *25*, 177–190.

(20) Hansen, P. L.; Molenbroek, A. M.; Ruban, A. V. Alloy Formation and Surface Segregation in Zeolite-Supported Pt-Pd Bimetallic Catalysts. *J. Phys. Chem. B* **1997**, *101*, 1861–1868.

(21) Grgur, B. N.; Markovic, N. M.; Ross, P. N., Jr. Electrooxidation of H2, CO, and H2/CO Mixtures on a Well-Characterized Pt70Mo30 Bulk Alloy Electrode. *J. Phys. Chem. B* **1998**, *102*, 2494–2501.

(22) Chen, W.; Severin, L.; Goethelid, M.; Hammer, M.; Cameron, S.; Paul, J. Electronic and geometric structure of clean Pt3Ti(111). *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 5620–5627.

(23) Brown, D.; Quinn, P. D.; Woodruff, D. P.; Noakes, T. C. Q.; Bailey, P. Surface and sub-surface segregation at the Pt25Rh75(1 1 1) surface. A medium energy ion scattering study. *Surf. Sci.* **2002**, *497*, 1–12.

(24) Attard, G. S.; Leclerc, S. A. A.; Maniguet, S.; Russell, A. E.; Nandhakumar, I.; Bartlett, P. N. Mesoporous Pt/Ru Alloy from the Hexagonal Lyotropic Liquid Crystalline Phase of a Nonionic Surfactant. *Chem. Mater.* **2001**, *13*, 1444–1446.

(25) Gasteiger, H. A.; Ross, P. N., Jr.; Cairns, E. J. LEIS and AES on sputtered and annealed polycrystalline Pt-Ru bulk alloys. *Surf. Sci.* **1993**, *293*, 67–80.

(26) Ruban, A. V.; Skriver, H. L.; Norskov, J. K. Surface segregation energies in transition-metal alloys. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 15990–16000.

(27) Menning, C. A.; Chen, J. G. Thermodynamics and kinetics of oxygen-induced segregation of 3d metals in Pt-3d-Pt(111) and Pt-3d-Pt(100) bimetallic structures. *J. Chem. Phys.* **2008**, *128*, 164703.

(28) Ma, Y.; Balbuena, P. B. Pt surface segregation in bimetallic Pt3M alloys: A density functional theory study. *Surf. Sci.* **2008**, *602*, 107–113.

(29) Tsong, T. T.; Ng, Y. S.; McLane, S. B., Jr. Surface segregation of a platinum-gold alloy: an atom probe field ion microscope investigation. *J. Chem. Phys.* **1980**, *73*, 1464–1468.

(30) Hoernstroem, S. E.; Johansson, L. I.; Flodstroem, A. Surface segregation in platinum-gold alloys studied using Auger electron spectroscopy. *Appl. Surf. Sci.* **1986**, *26*, 27–41.

(31) Beccat, P.; Gauthier, Y.; Baudoing-Savois, R.; Bertolini, J. C. Monotonous concentration profile and reconstruction at platinum-iron (Pt80Fe20)(111): LEED study of a catalyst. *Surf. Sci.* **1990**, *238*, 105–118.

(32) Watanabe, M.; Tsurumi, K.; Mizukami, T.; Nakamura, T.; Stonehart, P. Activity and stability of ordered and disordered Co-Pt alloys for phosphoric acid fuel cells. *J. Electrochem. Soc.* **1994**, *141*, 2659–2668.

(33) Kresse, G.; Furthmueller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169–11186.

(34) Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *49*, 14251–14269.

(35) Kresse, G.; Hafner, J. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B: Condens Matter Mater. Phys.* **1993**, *48*, 13115–13118.

(36) Kresse, G.; Hafner, J. Ab initio molecular dynamics of liquid metals. *Phys. Rev. B: Condens Matter Mater. Phys.* **1993**, *47*, 558–561.

(37) Kresse, G.; Furthmuller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.

(38) Bloechl, P. E. Projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953–17979.

(39) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758–1775.

(40) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(41) Mills, G.; Jonsson, H.; Schenter, G. K. Reversible work transition state theory: application to dissociative adsorption of hydrogen. *Surf. Sci.* **1995**, *324*, 305–337.

(42) Jonsson, H.; Mills, G.; Jacobsen, K. W. *Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions*; World Scientific: Singapore, 1998.

(43) Porter, D. A.; Easterling, K. E. *Phase transformations in metals and alloys*; Stanley Thornes: Cheltenham, 2000.

(44) Tsay, J. S.; Wu, Y. E.; Shern, C. S. Growth mode and inter-diffusion of Co on Pt(111). *Chin. J. Phys* **1997**, *35*, 610–616.

(45) Mantina, M.; Wang, Y.; Arroyave, R.; Chen, L. Q.; Liu, Z. K. First- principles calculation of self-diffusion coefficients. *Phys. Rev. Lett.* **2008**, *100*, 215901.

JCTC Journal of Chemical Theory and Computation

Assessment of the Performance of the M05−2X and M06−2X Exchange-Correlation Functionals for Noncovalent Interactions in Biomolecules

Edward G. Hohenstein,[†] Samuel T. Chill,[†] and C. David Sherrill*,[†,‡]

Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, and College of Computing, Georgia Institute of Technology, Atlanta, Georgia 30332-0280

Abstract: The highly parametrized, empirical exchange-correlation functionals, M05−2X and M06−2X, developed by Zhao and Truhlar have been shown to describe noncovalent interactions better than density functionals which are currently in common use. However, these methods have yet to be fully benchmarked for the types of interactions important in biomolecules. M05−2X and M06−2X are claimed to capture "medium-range" electron correlation; however, the "long-range" electron correlation neglected by these functionals can also be important in the binding of noncovalent complexes. Here we test M05−2X and M06−2X for the nucleic acid base pairs in the JSCH-2005 database. Using the CCSD(T) binding energies as a benchmark, the performance of these functionals is compared to that of a nonempirical density functional, PBE, and also to that of PBE plus Grimme's empirical dispersion correction, PBE-D. Due to the importance of "long-range" electron correlation in hydrogen-bonded and interstrand base pairs, PBE-D provides more accurate interaction energies on average for the JSCH-2005 database when compared to M05−2X or M06−2X. M06−2X does, however, perform somewhat better than PBE-D for interactions between stacked base pairs.

1. Introduction

Noncovalent interactions between biomolecules are important for the structure of DNA, RNA, and proteins.[1−5] The ability to describe these interactions computationally is essential for the *in silico* design of drug molecules.[6] In accurate quantum mechanical computations, a high degree of electron correlation must be included to reliably account for dispersion interactions. The current standard for accurately computing the interaction energy within a small, noncovalently bound complex is coupled-cluster with singles and doubles including perturbative triples [CCSD(T)].[7] Unfortunately, the applicability of CCSD(T) is hindered by the formal $\mathcal{O}(N^7)$ complexity of the method (more specifically, $\mathcal{O}(o^3v^4)$, where o and v are the number of occupied and virtual orbitals, respectively); to describe noncovalent interactions in large systems, less computationally expensive methods must be employed. The recently developed[8] spin-component scaled CCSD has been shown to produce results which closely match CCSD(T); however, the formal scaling of this approach remains high at $\mathcal{O}(N^6)$. Second-order Møller−Plesset perturbation theory (MP2) offers another approach for describing noncovalent interactions; with formal $\mathcal{O}(N^5)$ complexity, MP2 can be extended to much larger systems than are accessible with coupled-cluster methods. However, MP2 is substantially less accurate than the more rigorous coupled-cluster approaches.[9] While spin-component scaling[10−12] significantly improves the accuracy of MP2 on average, there are nevertheless cases where the accuracy is not as good as desired, or the system is too large for the computation to be feasible. More efficient methods that can effectively treat noncovalent interactions are necessary if complexes with 100 atoms or more are to be studied routinely. One approach is to reduce the computational scaling of coupled-cluster methods, and this is being actively pursued by several research groups.[13−16] Another approach is to attempt to improve the reliability of methods which are already applicable to larger systems.

Approximate density functional theory (DFT) is widely used today for examining a variety of chemical systems with dozens of atoms or more.[17] The application of DFT to noncovalently bound complexes has been limited due to the failure of most density functional approximations to describe dispersion interactions, which can be critical for noncovalent complexes. Dispersion interactions are inherently long-range electron correlation effects, which are not captured by the popular local or semilocal density functionals.[18−24] Several approaches exist for improving existing density functionals to handle dispersion effects. Among the more physically motivated approaches, Röthlisberger and co-workers have added effective atom-centered nonlocal potentials which have been fit to benchmark *ab initio* data;[25,26] Langreth, Lundqvist, and co-workers have introduced a van der Waals density functional (vdW-DF) which adds nonlocal terms to the correlation energy functional;[27] and Becke has proposed a novel approach that formulates the dispersion interaction in

  * Corresponding author e-mail: sherrill@gatech.edu.
  † School of Chemistry and Biochemistry.
  ‡ College of Computing.

Letter

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **1997**

terms of the dipole moment that would be created when considering an electron and its exchange hole.[28−30]

A more pragmatic and simple approach is to add empirical terms that model dispersion interactions.[31−35] These DFT-D approaches require only computation of interactions between atoms pairs; therefore, the additional computational expense is negligible. Alternatively, the highly parametrized empirical M05−2X and M06−2X functionals developed by Zhao and Truhlar have also shown promise for noncovalent interactions.[36−39] According to Zhao and Truhlar, the M05 and M06 series of functionals implicitly account for "medium-range" electron correlation because of the way they are parametrized, and this is sufficient to describe the dispersion interactions within many complexes.[39] These authors define "medium-range" correlation to be that found in complexes separated by about 5 Å or less.[40] While initial tests of these funtionals have been very promising,[41] they have yet to be fully benchmarked for biologically relevant noncovalent interactions.

In this letter, we analyze the performance of M05−2X and M06−2X for the nucleic acid base pairs from the JSCH-2005 database.[42] This database contains stacked, hydrogen-bonded, and interstrand base pairs, totaling 124 in all. Using the CCSD(T) interaction energies as a benchmark, M05−2X and M06−2X are compared to a density functional that fails to describe dispersion interactions, the generalized gradient approximation (GGA) of Perdew, Burke, and Wang (PBE).[43] In contrast to M05−2X and M06−2X, which are parametrized to fit a large number of experimental and higher-level theoretical data, all parameters in PBE (other than those in the underlying local spin density approximation) are physical constants. Grimme's empirical dispersion correction[34,35] is then applied to the PBE binding energies to assess the performance of the increasingly popular DFT-D approach. The JSCH-2005 database provides a rigorous test of M05−2X and M06−2X for the description of some of the types of noncovalent interactions that are important in biomolecules.

2. Theoretical Methods

The DFT-D approach of Grimme is a widely applicable method for correcting the performance of standard density functionals.[34,35,44] This method utilizes a damped R^{-6} term to model the dispersion interactions (eqs 1, 2, and 3).

$$E_{DFT-D} = E_{DFT} - s_6 \sum_{i=1}^{N-1} \sum_{j>i}^{N} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}) \qquad (1)$$

$$C_6^{ij} = \sqrt{C_6^i C_6^j} \qquad (2)$$

$$f_{dmp}(R_{ij}, R_r, d) = \frac{1}{1 + e^{-d(R_{ij}/R_r - 1)}} \qquad (3)$$

The geometries of the base pairs in the JSCH-2005 set were taken from Jurečka et al.[42] All density functional computations were performed using Q-Chem 3.1.[45] The empirical dispersion correction was computed using a locally modified version of PSI 3.3[46] (although these terms have since been added directly to Q-Chem by one of the authors, CDS). For use with the PBE functional, Grimme recommends a global scaling factor, s_6, of 0.7 for his dispersion correction term.[35] All energy computations were performed using Dunning's aug-cc-pVDZ basis set, which

**Table 1.** Performance of Various Methods for 124 Nucleic Acid Dimers from the JSCH-2005 Set[a]

method	MSE[b]	MUE[c]	STD[d]	MAX[e]
38 H-Bonded Complexes[g]				
PBE/aug-cc-pVDZ	2.59	2.59	0.81	4.85
PBE-D/aug-cc-pVDZ[f]	−0.53	0.70	0.79	3.17
M05−2X/aug-cc-pVDZ	1.98	1.98	0.62	3.22
M06−2X/aug-cc-pVDZ	1.62	1.62	0.58	2.76
32 Interstrand Complexes[h]				
PBE/aug-cc-pVDZ	1.83	1.84	2.06	9.27
PBE-D/aug-cc-pVDZ[f]	0.16	0.58	1.36	5.03
M05−2X/aug-cc-pVDZ	1.02	1.23	1.41	5.66
M06−2X/aug-cc-pVDZ	0.74	1.08	1.41	5.53
54 Stacked Complexes[i]				
PBE/aug-cc-pVDZ	7.48	7.57	2.46	14.28
PBE-D/aug-cc-pVDZ[f]	1.41	1.53	0.79	3.05
M05−2X/aug-cc-pVDZ	2.48	2.59	0.96	4.56
M06−2X/aug-cc-pVDZ	0.95	1.08	0.80	2.78
JSCH-2005 Nucleic Acids				
PBE/aug-cc-pVDZ	4.52	4.56	3.28	14.28
PBE-D/aug-cc-pVDZ[f]	0.49	1.03	1.28	5.03
M05−2X/aug-cc-pVDZ	1.95	2.05	1.17	5.66
M06−2X/aug-cc-pVDZ	1.10	1.25	1.00	5.53

[a] Empirical terms added to counterpoise corrected PBE/aug-cc-pVDZ interaction energies. All energies are in kcal mol⁻¹. Errors are computed relative to the benchmark interaction energies of Hobza and co-workers.[42] [b] Mean signed error. [c] Mean unsigned error. [d] Standard deviation. [e] Maximum absolute deviation. [f] PBE/aug-cc-pVDZ with Grimme's dispersion correction.[34] [g] Average benchmark interaction energy of −20.79 kcal mol⁻¹. [h] Average benchmark interaction energy of −1.00 kcal mol⁻¹. [i] Average benchmark interaction energy of −7.84 kcal mol⁻¹.

would be the most tractable of Dunning's augmented basis sets in studies of large systems.[47] It is possible that larger basis sets would improve the performance of M05−2X and M06−2X; however, in order for these methods to be applicable to larger systems, they must be reliable with modest basis sets. We comment on the performance of PBE and PBE-D with larger basis sets below. All binding energies were counterpoise corrected using the scheme of Boys and Bernardi.[48] All density functional computations were performed using a Lebedev grid with 302 angular points for each of the 100 radial points included. This grid is larger than those used by default by most electronic structure program packages, but it appears necessary to avoid artifacts due to numerical integration for noncovalent interactions, particularly when using meta-GGA functionals.[24]

3. Results and Discussion

The JSCH-2005 set of nucleobase dimers has been used previously as a benchmark for dispersion corrected DFT methods,[49−51] including the DFT-D method of Grimme.[34,35] Counterpoise corrected PBE/aug-cc-pVDZ interaction energies are reported here to allow comparison to a standard DFT approximation. Results for the methods tested are summarized in Table 1. Examples of complexes included in the JSCH-2005 database are shown in Figure 1. For the entire set of molecules, PBE/aug-cc-pVDZ underestimates the binding energy (overestimates the interaction energy) between the base pairs by 4.52 kcal mol⁻¹ on average. PBE only overestimates the binding energy for two complexes out of the 124 base pairs in the set. This is due to the failure of PBE to capture the dispersion interactions within the complexes. Only the electrostatic,
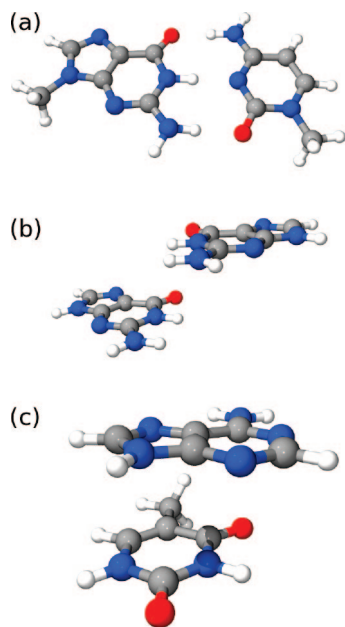
Figure 1. Representative geometries for (a) hydrogen-bonded base pairs (mG···mC WC), (b) interstrand base pairs (G···G IS), and (c) stacked base pairs (TA08/3.16 A//Ts), using the nomenclature of ref 42.

exchange, and induction interactions are captured by this functional. Since this functional almost always underestimates the binding energy, the addition of empirical dispersion terms substantially improves the performance. The mean signed error in interaction energies is reduced to 0.49 kcal mol^{-1}, and the mean unsigned error is reduced from 4.56 to 1.03 kcal mol^{-1}. This simple correction greatly increases the accuracy of PBE/aug-cc-pVDZ without increasing the computational cost in any significant way. In another study of this test set, Antony and Grimme[51] found that PBE-D with a polarized triple-ζ basis set yields a mean unsigned error of 0.82 kcal mol^{-1}, demonstrating that only minor improvements are obtained for DFT-D by increasing the basis set size. On the other hand, larger basis sets reduce the size of the basis set superposition error, and Antony and Grimme argue that polarized triple-ζ basis sets are large enough that one can dispense with the counterpoise correction.

M05−2X/aug-cc-pVDZ describes the interactions between the complexes in the JSCH-2005 database more accurately than PBE/aug-cc-pVDZ. M05−2X tends to underestimate the binding energies by about 2 kcal mol^{-1} relative to the CCSD(T) benchmark values. M06−2X also tends to underbind these complexes, but it reduces the mean unsigned error to 1.25 kcal mol^{-1}. While this represents a tremendous improvement over traditional density functionals for noncovalent interactions, nevertheless, the M05−2X and M06−2X functionals do not perform better than Grimme's simple DFT-D approach. The PBE-D/aug-cc-pVDZ approximation has a lower mean signed error, mean unsigned error, and maximum error. The M05−2X and M06−2X functionals are only improved with respect to the standard deviation. M05−2X performs markedly worse than PBE-D with respect to the mean signed error and mean unsigned error. M06−2X is comparable to, although slightly worse than, PBE-D.

Dispersion interactions are especially important in the description of stacked base pairs. The failure of PBE to describe dispersion interactions is especially evident for the stacked subset of the JSCH-2005 set. For these complexes, PBE/aug-cc-pVDZ grossly underestimates the attractive interactions. The maximum error of PBE for the stacked base pairs is 14.28 kcal mol^{-1} for a complex which CCSD(T) predicts to be bound by 14.57 kcal mol^{-1}. Obviously, PBE cannot be used to treat systems that include stacking interactions. The inclusion of empirical dispersion terms improves the PBE binding energies immensely; the mean unsigned error is reduced from 7.57 to 1.53 kcal mol^{-1}. The "medium-range" electron correlation apparently captured by M05−2X is sufficient to reduce the unsigned error for stacked complexes to 2.59 kcal mol^{-1}. While this improves greatly over PBE, it does not improve over PBE-D. Of the four methods tested, M06−2X is the best able to describe stacking interactions between base pairs. The mean unsigned error for M06−2X/aug-cc-pVDZ is 1.08 kcal mol^{-1}. This suggests that for stacked base pairs, when the electron clouds of each monomer are in relatively close proximity to one another, the electron correlation included in the M06−2X functional is sufficient to describe the relevant dispersion interactions, as also indicated by previous work.[41]

Hydrogen bonded complexes are dominated by electrostatic interactions, and thus traditional density functionals are capable of computing acceptable binding energies. PBE/aug-cc-pVDZ has a mean unsigned error of 2.95 kcal mol^{-1} for the hydrogen-bonded base pairs, and one should note that the average binding energy of the hydrogen-bonded complexes is much larger than the binding energies of stacked complexes. The inclusion of an empirical dispersion term to the PBE/aug-cc-pVDZ energies further improves the binding energies. PBE-D/aug-cc-pVDZ has a mean unsigned error of 0.70 kcal mol^{-1}. Neither M05−2X or M06−2X match the quality of the PBE-D energies for the hydrogen-bonded systems. A notable result is that PBE, M05−2X, and M06−2X all uniformly underestimate the attraction between hydrogen-bonded base pairs. While dominated by electrostatic interactions, dispersion interactions also play an important role in the binding of these complexes. The performance of PBE-D versus M05−2X and M06−2X for hydrogen-bonded complexes shows that it is not sufficient to include only the "medium-range" electron correlation which M05−2X and M06−2X are said to capture.

The third subset of complexes in the JSCH-2005 database is comprised of interstrand base pairs. These complexes are bound primarily by a combination of electrostatic interactions and dispersion interactions. The intermonomer separation in these complexes is larger than it is between the stacked and hydrogen-bonded bases; as a consequence, these complexes are not bound as strongly. PBE/aug-cc-pVDZ has a mean unsigned error for this set of molecules of 1.84 kcal mol^{-1}; however, the strongest binding energy in this subset is only 5.20 kcal mol^{-1}, and hence PBE fails severely for this subset. The addition of Grimme's dispersion terms reduces the unsigned error to 0.58 kcal mol^{-1}. Due to the importance of "long-range" electron correlation in these complexes, neither M05−2X or M06−2X produce binding energies as accurate as PBE-D. M05−2X/aug-cc-pVDZ and M06−2X/aug-cc-pVDZ have mean unsigned errors of 1.23 and 1.08 kcal mol^{-1}, respectively. Thus interstrand base pairs are

Letter

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **1999**

an example of a chemically interesting system where "medium-range" electron correlation is not sufficient to compute accurate binding energies. For both interstrand and hydrogen-bonded base pairs, Grimme's simple DFT-D approach appears to provide more reliable results than meta-GGA's which hope to effectively capture "medium-range" electron correlation effects through the addition of terms depending on the electronic kinetic energy density and through significant parametrization. It might be noted that, to our knowledge, no theoretical connection between electronic kinetic energy density and dispersion interactions has been demonstrated.

4. Conclusions

We have shown that M05−2X and M06−2X provide significant improvements over traditional density functionals for the noncovalent interactions exemplified by the JSCH-2005 test set. The apparent ability of these functionals to capture "medium-range" (≤ 5 Å) electron correlation allows them to describe stacking interactions between base pairs accurately, especially in the case of M06−2X. However, in hydrogen-bonded and interstrand base pairs, much of the dispersion interaction is "long-range" (> 5 Å) and is not captured by M05−2X and M06−2X. The addition of Grimme's empirical dispersion terms to the PBE functional provided a better description of the interactions between base pairs. Dispersion interactions can be modeled accurately with an R^{-6} term. Although M05−2X and M06−2X represent a significant step forward in density functionals able to describe noncovalent interactions, more improvement will be required before a density functional can outperform DFT-D methods which add dispersion empirically. Approaches which include terms explicitly tailored to capture the physics of long-range dispersion interactions[25−28,30] appear to be the most promising.

Supporting Information Available: Binding energies for hydrogen bonded, interstrand, and stacked base pairs. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(2) Shieh, H.-S.; Berman, H. M.; Dabrow, M.; Neidle, S. *Nucleic Acids Res.* **1980**, *8*, 85–97.

(3) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149–5155.

(4) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 1210–1250.

(5) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23–28.

(6) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118–1134.

(7) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(8) Takatani, T.; Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2008**, *128*, 124111.

(9) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656–10668.

(10) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095–9102.

(11) Hill, J. G.; Platts, J. A. *J. Chem. Theory Comput.* **2007**, *3*, 80–85.

(12) Distasio, R. A.; Head-Gordon, M. *Mol. Phys.* **2007**, *105*, 1073–1083.

(13) Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330–8343.

(14) Schutz, M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3941–3947.

(15) Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *121*, 10935.

(16) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 074116.

(17) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules, Vol. 16 of International Series of Monographs on Chemistry*; Oxford: New York, 1989.

(18) Tsuzuki, S.; Lüthi, H. P. *J. Chem. Phys.* **2001**, *114*, 3949–3957.

(19) Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624–1626.

(20) Allen, M. J.; Tozer, D. J. *J. Chem. Phys.* **2002**, *117*, 11113–11120.

(21) Hobza, P.; Šponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315–1325.

(22) Kristyán, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175–180.

(23) Kurita, N.; Sekino, H *Int. J. Quantum Chem.* **2003**, *91*, 355–362.

(24) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334–338.

(25) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, 153004.

(26) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. B* **2005**, *71*, 195119.

(27) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(28) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.

(29) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 024101.

(30) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2006**, *124*, 014104.

(31) Zimmerli, U.; Parrinello, M.; Koumoutsakos, P. *J. Chem. Phys.* **2004**, *120*, 2693–2699.

(32) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555–569.

(33) Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *116*, 515–524.

(34) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(35) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

(36) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103.

(37) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.

(38) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

(39) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(40) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289–300.

(41) Gu, J. D.; Wang, J.; Leszczynski, J.; Xie, Y. M.; Schaefer, H. F. *Chem. Phys. Lett.* **2008**, *458*, 164–166.

(42) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(43) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(44) Grimme, S.; Antony, J.; Schwabe, T.; Mück-Lichtenfeld, C. *Org. Biomol. Chem.* **2007**, *5*, 741–758.

(45) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P., Jr.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Voorhis, T. V.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

(46) Crawford, T. D.; Sherrill, C. D.; Valeev, E. F.; Fermann, J. T.; King, R. A.; Leininger, M. L.; Brown, S. T.; Janssen, C. L.; Seidl, E. T.; Kenny, J. P.; Allen, W. D. *J. Comput. Chem.* **2007**, *28*, 1610–1616.

(47) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(48) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(49) Lin, I.; Rothlisberger, U. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2730–2734.

(50) Morgado, C.; Vincent, M. A.; Hillier, I. H.; Shan, X. *Phys. Chem. Chem. Phys.* **2007**, *9*, 448–451.

(51) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287–5293.

JCTC Journal of Chemical Theory and Computation

United-Atom Discrete Molecular Dynamics of Proteins Using Physics-Based Potentials

Agustí Emperador,[†,‡] Tim Meyer,[†,‡] and Modesto Orozco*[,†,‡,§,ll]

Joint IRB-BSC research program in Computational Biology, Institute for Research in Biomedicine (IRB), Josep Samitier 1-5, Barcelona 08028, Spain, Barcelona Supercomputing Centre (BSC), Jordi Girona 29, Barcelona 08034, Spain, Departament de Bioquímica i Biología Molecular, Facultat de Biología, Universitat de Barcelona, Avgda Diagonal 645, Barcelona 08028, Spain, and National Institute of Bioinformatics, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain

Abstract: We present a method for the efficient simulation of the equilibrium dynamics of proteins based on the well established discrete molecular dynamics algorithm, which avoids integration of Newton equations of motion at short time steps, allowing then the derivation of very large trajectories for proteins with a reduced computational cost. In the presented implementation we used an all heavy-atoms description of proteins, with simple potentials describing the conformational region around the experimental structure based on local physical interactions (covalent structure, hydrogen bonds, hydrophobic contacts, solvation, steric hindrance, and bulk dispersion interactions). The method shows a good ability to describe the flexibility of 33 diverse proteins in water as determined by atomistic molecular dynamics simulation and can be useful for massive simulation of proteins in crowded environments or for refinement of protein structure in large complexes.

Introduction

Under native conditions proteins exist not as unique structures but as dynamic ensembles of conformers, some of them which can be quite distant from the most probable structure of the protein. The spontaneous sampling of different conformations helps the proteins to perform their biological action, since most protein actions imply relevant conformational changes. Thus, many studies[1−10] have reported that the biological relevant transition (i.e. that required for biological action) corresponds in many cases to the easiest deformation movements of the relaxed protein, suggesting that evolution has driven proteins to optimize not only their structure but also their intrinsic deformability patterns.[11−16]

A better understanding of the dynamical behavior of proteins is then crucial to bridge structure and function in proteins.

Despite recent advances[10] the experimental study of protein flexibility is still very difficult and impossible to carry out at the full proteome-level. Theoretical methods are then the logical alternative, and many examples of the power of these methods have been published in the last years.[17,18] One of the most useful tools to describe protein dynamics is atomistic molecular dynamics (MD), where solvated proteins are represented at full atomistic detail by means of physical potentials[19−25] derived from high level quantum mechanical calculations and experimental data on condensed phase and model systems. The impact of MD in the study of protein flexibility has been impressive.[26,27] However, despite recent improvements in software and hardware, atomistic MD simulations of proteins are still too expensive to allow a full-proteome description of flexibility or to simulate simultaneously the dynamics of hundreds of thousands of proteins in crowded cellular-like environments. Therefore, less rigorous, but more efficient simulation, tools, such as normal

---

* Corresponding author e-mail: modesto@mmb.pcb.ub.es.
† Institute for Research in Biomedicine.
‡ Barcelona Supercomputing Centre.
§ Universitat de Barcelona.
ll National Institute of Bioinformatics, Parc Científic de Barcelona.

mode analysis,[18] Brownian MD,[28] or discrete molecular dynamics[29] (DMD), are necessary.

DMD is an ultrafast but still physically-based technique where interactions are represented with discontinuous potentials with a steplike profile. In the simple case of coarse grained models with strictly structure-based potentials at residue-level resolution, the interactions between protein residues are represented by means of simple infinite square wells.[30] Within this model, the particle moves in ballistic regime (constant velocity) in a flat potential region (within the well), interrupted by collisions at the walls of the potential wells. Upon each collision momentum and energy conservation laws are imposed. The simplicity of the potential functional avoids the need to integrate Newton's equations of motion every femtosecond, since trajectory becomes fully deterministic and is followed from collision to collision. This allows the obtaining of very long trajectories in a simple PC, the performance being increased in diffusive or slow transitional movements. Despite its simplicity, DMD has been successful in describing different aspects of protein dynamics[31−35] and nucleic acids,[36,37] including macromolecular aggregation[38−44] and transitions.[45] The reliability of the DMD discontinuous potentials to describe protein interactions has been shown, for example, in *ab initio* protein folding[46] and protein oligomerization studies.[47−49] Very recently, our group has shown that DMD coupled to a simple Gō-like Hamiltonian[50] (similar to those used in normal mode analysis) is able to reproduce reasonably well the C_α dynamics of proteins,[30] with a very small computational cost, suggesting that DMD can be an alternative to Gō-based Brownian dynamics[28] or normal mode analysis.[18] Unfortunately, as formulated in our previous work the technique presents two intrinsic shortcomings. First, since it is based exclusively on Gō-like potentials it can only reproduce pseudoharmonic deformability around a reference structure, without any possibility of describing large (nonharmonic) conformational transitions. Second, no atomic-detailed information is provided, limiting the range of applicability of the technique in cases where resolution higher than the backbone is needed.

In this paper we present and validate a new atomistic DMD model based on a simple pseudophysical force-field, which represents a hybrid between pure physical potentials[51−57] and empirical Gō-like models linked to strictly harmonic/pseudoharmonic potentials.[50] The method defines the Hamiltonian based on a reference structure, but instead of using all atomic or residue contacts as Gō-restraints, only physically meaningful interactions are considered. The method provides atomic-detailed information on the protein, while guaranteeing that the sampling will not produce protein conformations unrealistically different from the experimentally determined structure. The technique is fast and seems to be a promising tool for the structure refinement, for the representation of moderate conformational transitions, and for the analysis of protein-protein docking, and multiprotein systems.

Methods

Discrete Molecular Dynamics Algorithm. DMD assumes that particles defining a molecule follow fully deterministic
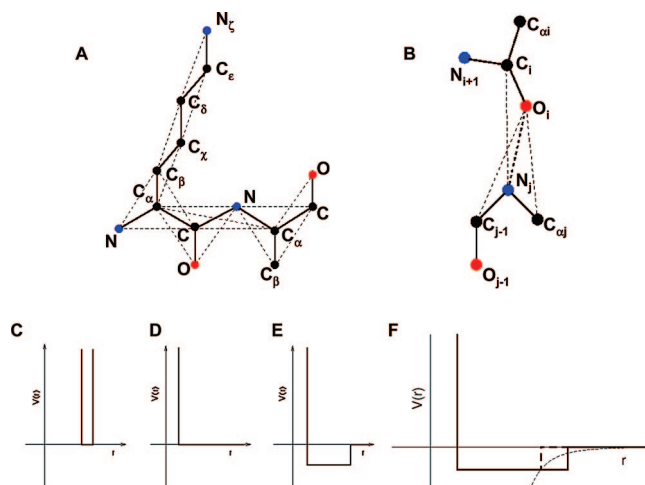

Figure 1. (A) Schematic diagram for the all-atom protein model. A Lys-Ala sequence is shown. Solid thick lines represent covalent bonds, and thin dashed lines show pseudobonds (see text). (B) Schematic diagram for the hydrogen bond. The pseudobonds (dashed lines) keep the directionality of the bond. (C) DMD square well potential for covalently or noncovalently bound particles. In the case of covalent bonding the well depth is infinite. (D) DMD hardcore repulsive potential for unbound particles. (E) Square well for hydrophobic interactions and salt bridges, composed by the hardcore repulsion at short distances and a potential step at the cutoff distance for these interactions. (F) Dispersion potential established between C_αs. The well depth depends on the distance in the native conformation (see text).

Table 1. Different Metrics Indicating the Similarities of DMD Ensembles to Reference Structures (MD-Averaged: MD) or Experimental Conformation (EXP) and with Ensembles Obtained by Atomistic MD Simulations in Water

parameter	DMD vs MD	DMD vs EXP[d]
RMSd[a]	2.6 ± 0.8	3.2 ± 0.7
Tm-score[a]	0.4 ± 0.2	1.5 ± 0.3
R_{gyr}[a,b]	1.1 ± 0.6	0.8 ± 0.6
SAS(tot)[b,c]	12% ± 7%	11% ± 7%
SAS(apolar)[b,c]	10% ± 7%	9% ± 7%
Conserv helix	96% ± 8%	93% ± 12%
Conserv β-sheet	85% ± 12%	85% ± 12%
Conserv turn	96% ± 3%	96% ± 3%
no. of native contacts	80% ± 11%	87% ± 10%

[a] Average RMSd, R_{gyr}, and Tm-scores are in Å. [b] Computed as $dif = ((1/n)\sum_n(\alpha_{ref} - \alpha_{DMD})^2)^{0.5}$, where α is the descriptor, n is the number of proteins in the test set, α_{ref} is the value of the descriptor in the reference conformation (MD-averaged or experimental conformation), and α_{DMD} is the average deviation obtained along the 5-10 ns portion of the DMD trajectory. [c] Solvent accessible surfaces are represented as relative values for every protein. [d] Values in the last column always refer to average relative deviations. All values were obtained considering an all-heavy atom representation of proteins.

trajectories defined by their starting positions and velocities and simple (mono-or multiple) square-well potentials for particle-particle interactions. Particles move with constant velocities which change only after collision with the walls of the wells (see Figure 1 for explanation). As a consequence, trajectory can be analytically determined, integration of equations of motions at fixed time steps (as in MD) is not needed, and the calculation progresses directly from collision
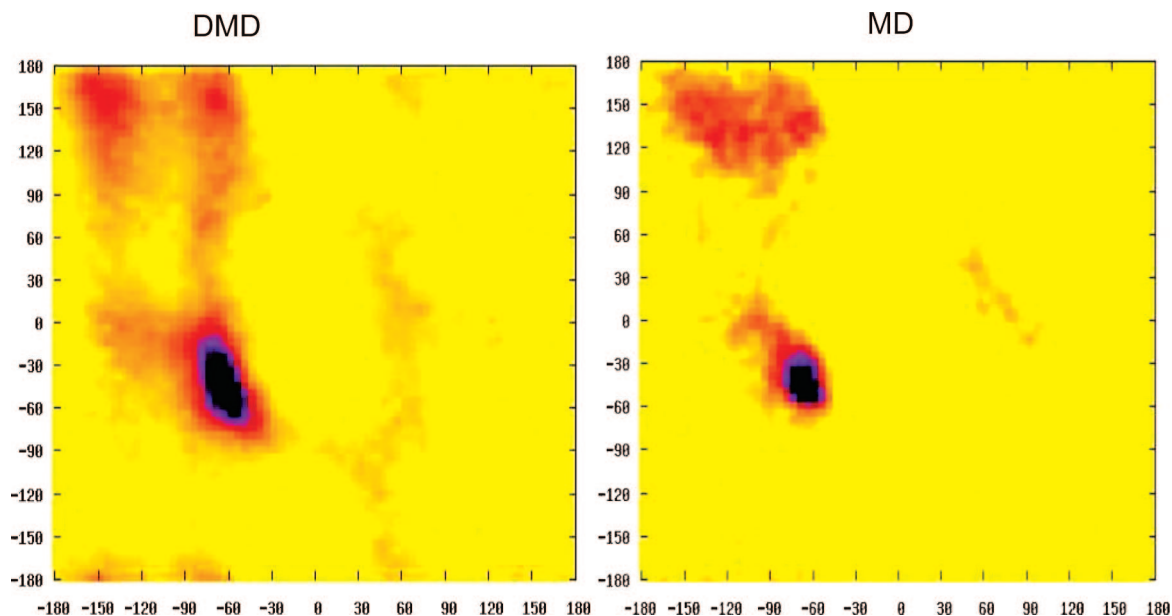
Figure 2. Ramachandran map for the proteins calculated with the current DMD sampling technique and all-atoms MD simulations with AMBER. In both cases 10 ns trajectories were used.
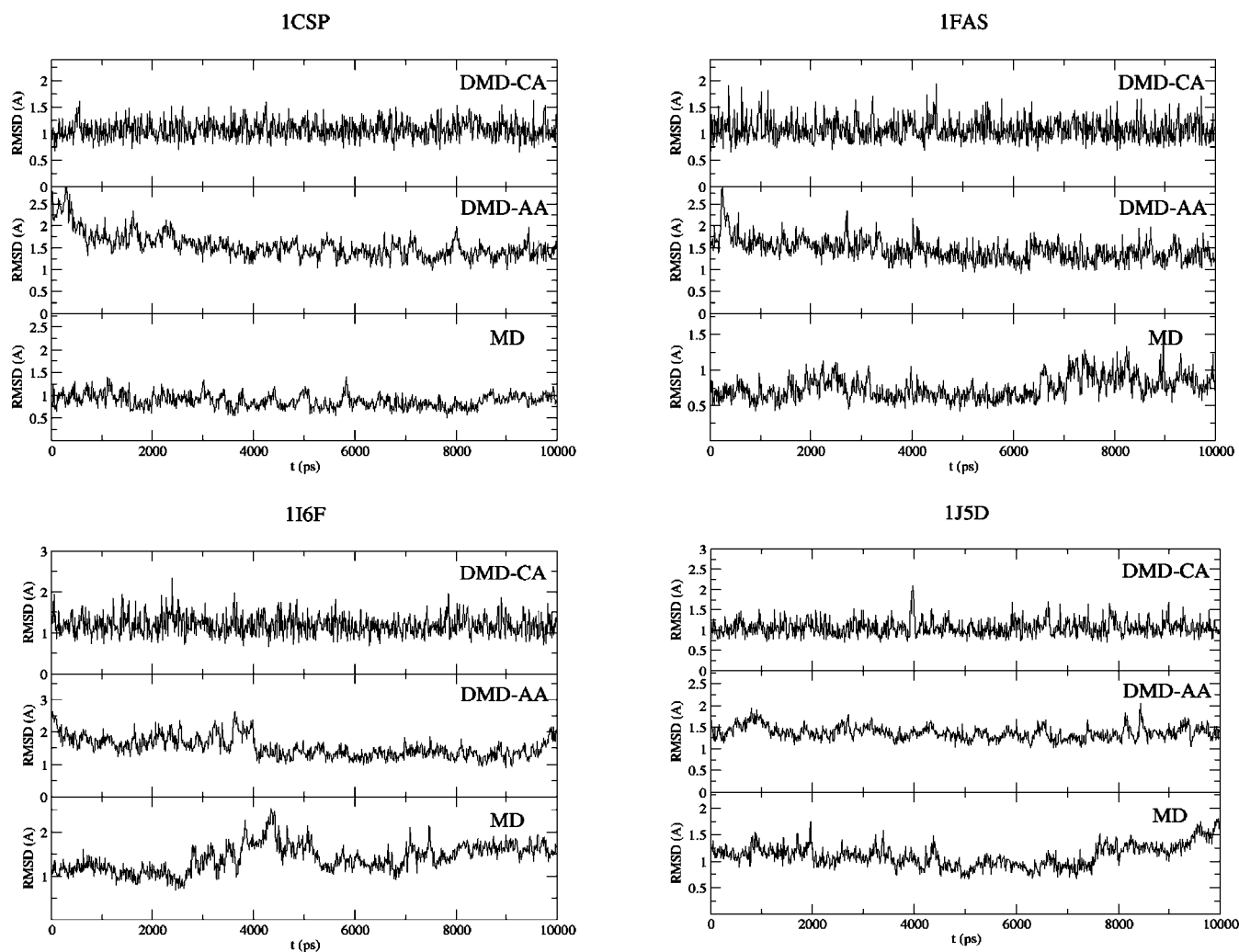
Figure 3. RMSd (in Å^2) with respect to reference structure (MD-averaged one) of selected proteins in DMD using a Gō-like potential and C$_\alpha$ representation of proteins (DMD-CA), present "all atoms" DMD implementation (DMD-AA) and atomistic MD simulation with AMBER-force-field (MD). In all cases only the C$_\alpha$ coordinates are considered to compute the RMSd.

2004 *J. Chem. Theory Comput., Vol. 4, No. 12, 2008*

Emperador et al.

(event) to collision. If an efficient algorithm for predicting collisions is used[58] the method can be extremely efficient allowing simulation of very long time periods. In the present implementation the CPU time needed to determine collisions scales as $\sim N^{1.1}$ with the number of particles (N) considered.

The basic DMD formalism assumes that the position of a particle after some period of time (the minimum collision time) is determined analytically using

$$\vec{r}_i(t+t_c) = \vec{r}_i(t) + \vec{v}_i(t)t_c \qquad (1)$$

where $\vec{r}_i$ and $\vec{v}_i$ stand for positions and velocities, and t_c is the minimum amongst the collision times t_{ij} between each pair of particles i and j, given by

$$t_{ij} = \frac{-b_{ij} \pm \sqrt{b_{ij}^2 - v_{ij}^2(r_{ij}^2 - d^2)}}{v_{ij}^2} \qquad (2)$$
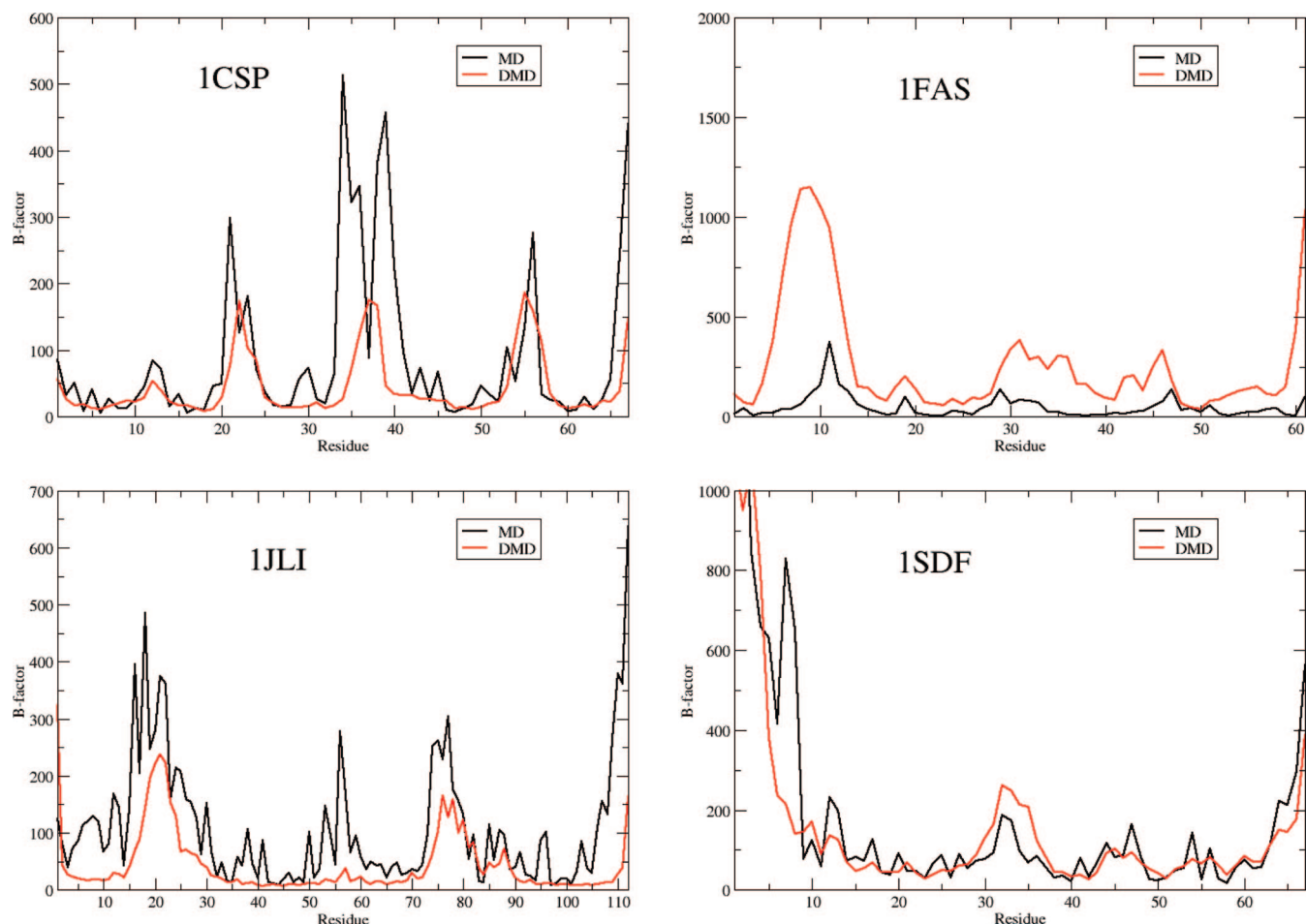
Table 2. Indexes Representative of the Dynamic Behavior Obtained in DMD and MD Samplings[a]

	DMD	MD	MD vs MD
variance (heavy atoms) [b]	25.6 ± 11.4	13.4 ± 6.4	6.1-53.3
reduced variance[b]	6.1 ± 4.5	6.2 ± 4.8	1.8-30.3
entropy (Schlitter, heavy atoms)[b]	56.9 ± 15.2	47.1 ± 2.7	39.9-64.3
complexity[c]	$9\% \pm 2\%$	$4\% \pm 2\%$	1 %-8%
Z-score	71.4 ± 24.6	----	62-255
Γ index	0.61 ± 0.13	----	0.8-1.0
correlation B-factors (to exp)	0.43 ± 0.12	0.66 ± 0.10	0.4-0.8
correlation B-factors (to MD)	0.54 ± 0.15	----	0.4-0.9
Δ_L (all)	0.40 ± 0.08	0.32 ± 0.07	0.20-0.57
Δ_L (exposed) - Δ_L (buried)	0.13 ± 0.07	0.18 ± 0.06	0.08-0.30
Δ_L (α-helix) - Δ_L (β-sheet)	-0.19 ± 0.15	-0.12 ± 0.13	−0.33-0.16

[a] In all cases an all-heavy atoms representation of the protein is used, and represented values correspond to averages over 33 proteins. See Methods for description of the different metrics. Values in the column labeled by MD correspond to the meta-trajectory obtained by adding individual CHARMM, OPLS, and AMBER trajectories for individual proteins. The values in the last column (labeled as MD vs MD) correspond to those obtained when instead of the original meta-trajectory the original CHARMM, OPLS, and AMBER trajectories are compared among each other. The last column gives the range of uncertainty expected in atomistic MD simulation as reported by comparing the 4 different trajectories available for each protein. [b] Values used in the average were previously referred to as the number of residues in the protein. [c] Values used in the average were previously referred to as the total number of eigenvectors in the protein.
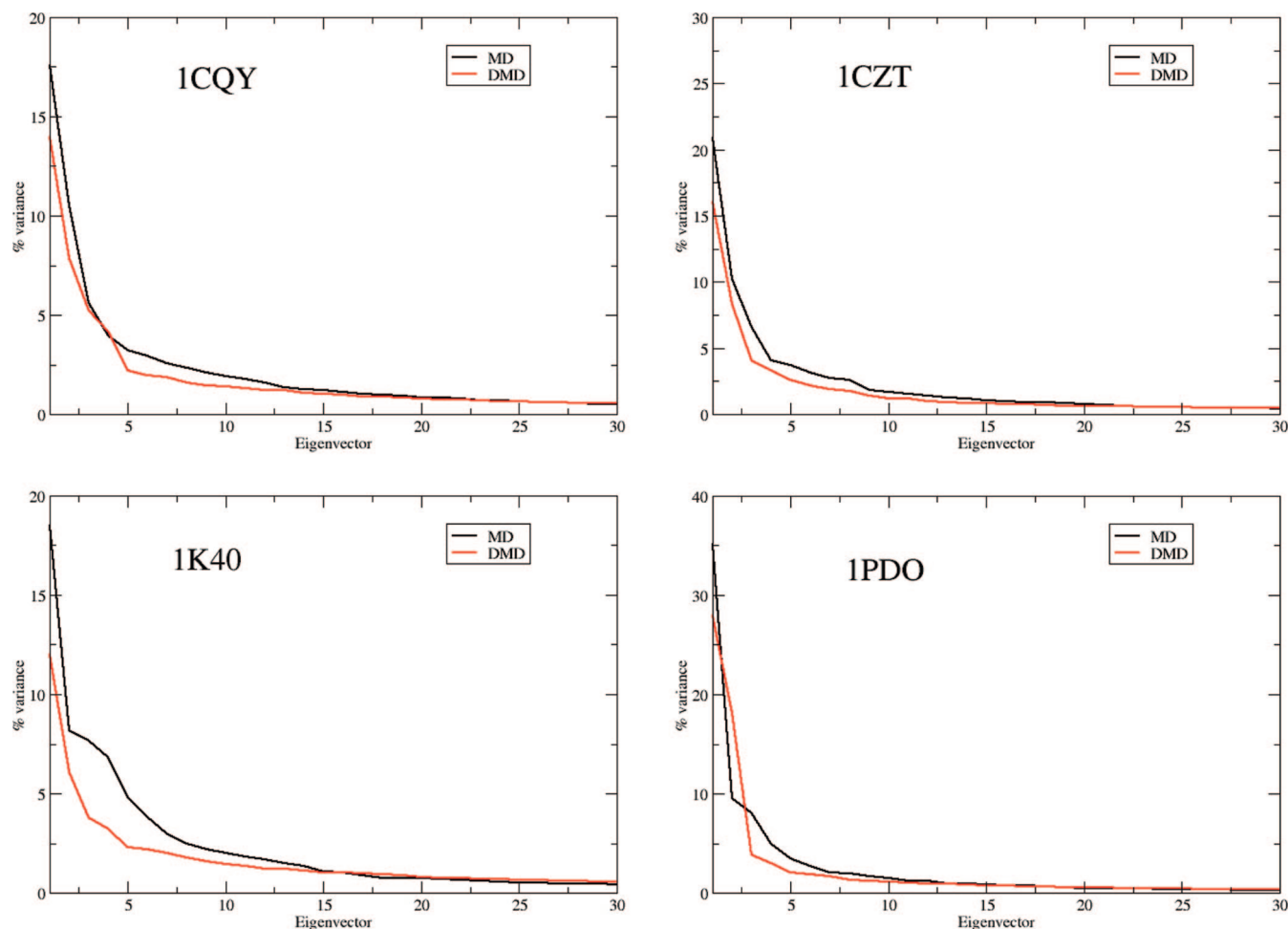
Figure 4. Percentage of the total variance explained by individual eigenvectors of different rank for selected proteins as determined from MD and DMD[5] ensembles.

Figure 5. B-factors (in Å^2) for selected proteins as determined from MD and DMD ensembles.

where r_{ij} is the square modulus of $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$, v_{ij} is the square modulus of $\vec{v}_{ij} = \vec{v}_j - \vec{v}_i$, $b_{ij} = \vec{r}_{ij} \cdot \vec{v}_{ij}$, and d is the distance corresponding to a discontinuity (the wall) in the potential (the signs + and − before the square root are used for particles approaching one another and moving apart, respectively).

When two particles collide, there is a transfer of linear momentum into the direction of the vector $\vec{r}_{ij}$

$$m_i \vec{v}_i = m_i \vec{v}_i' + \Delta \vec{p} \tag{3}$$

$$m_j \vec{v}_j + \Delta \vec{p} = m_j \vec{v}_j' \tag{4}$$

where the prime indices denote the variables after the event (collision).

In order to calculate the change in velocities upon collision the velocity of each particle is projected in the direction of the vector $\vec{r}_{ij}$ so that the conservation equations become one-dimensional along the interatomic coordinate, which implies that

$$m_i v_i + m_j v_j = m_i v_i' \, m_j v_j' \tag{5}$$

$$\frac{1}{2} m_i v_i^2 + \frac{1}{2} m_i v_i^2 = \frac{1}{2} m_i v_i'^2 + \frac{1}{2} m_i v_i'^2 + \Delta V \tag{6}$$

where the change in potential energy (ΔV) is the depth of the square well that defines the interatomic potential.

The transferred momentum can be easily determined from

$$\Delta p = \frac{m_i m_j}{m_i + m_j} \left\{ \sqrt{(v_j - v_i)^2 - 2\frac{m_i + m_j}{m_i m_j} \Delta V} - (v_j - v_i) \right\} \tag{7}$$

Note that the two particles can go out of the well as long as

$$\Delta V < \frac{m_1 m_2}{2(m_1 + m_2)} (v_j - v_i)^2 \tag{8}$$

Otherwise, the particles do not cross the potential energy discontinuity at the wall of the well and remain inside the well ($\Delta V = 0$). In this case eq 7 reduces to

$$\Delta p = \frac{m_i m_j}{m_i + m_j} \left\{ \sqrt{(v_j - v_i)^2} - (v_j - v_i) \right\} \tag{9}$$

which taking the negative solution of the root leads to the definition of the transferred momentum:

$$\Delta p = \frac{2 m_i m_j}{m_i + m_j} (v_i - v_j) \tag{10}$$

Since the proteins we want to simulate are considered to be in a thermal bath, but kinetic and potential energy are continuously exchanged during the simulation, we maintain the temperature through a Berendsen coupling.[25]
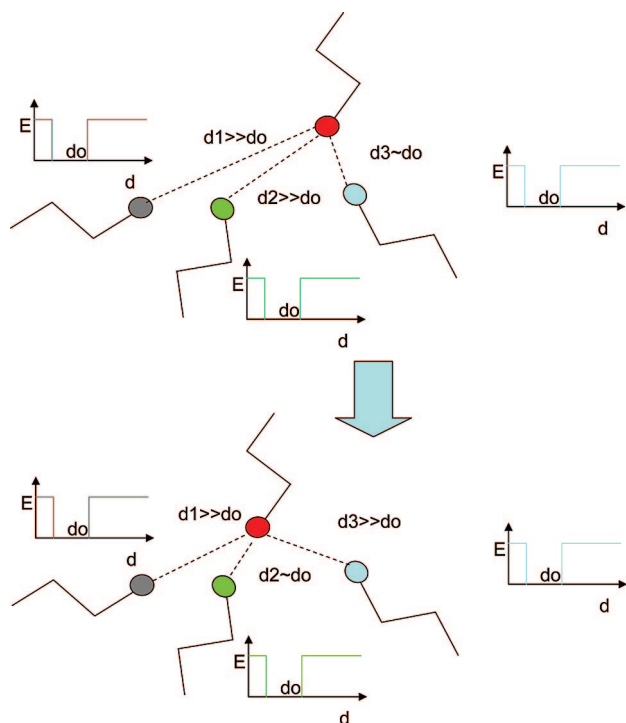
Figure 6. Schematic representation of how DMD can be used with competing interactions for side-chain (or backbone) refinement. In the example a bead in red is at an interaction distance for another bead (in blue), but the system Hamiltonian is defined by three competing finite square wells, all centered at the same equilibrium distance d_o. During the simulation the system is then allowed to exit the original well (in blue) entering into any of the two other alternating wells (in gray and green). This opens the possibility of changing from a native (red-blue) to a non-native contact (red-gray or red-green).

Force-Field Definition for DMD. The force-field was aimed at performing atomistic simulations of proteins around equilibrium structures but enabling moderate nonharmonic conformational transitions (not allowed in a pure Gō-Cα model) and introducing physical considerations (missing in current normal mode or Gaussian elastic model systems). To this purpose, we defined a simple united atom force-field consisting of "bonded" and "nonbonded" terms. As usual "bonded" terms account for stretchings (a-b), bendings (a-b-c), and torsions (a-b-c-d). Stretching, bendings, and torsions involving double or conjugated bonds were represented by means of infinite square wells with a width corresponding to 5% of the length of the bond/pseudobond distance[41] (a-b bond for stretching, a-c pseudobond for bending, and a-d pseudobond for torsions; see Figure 1). Equilibrium values for bonds and pseudobonds were taken from the reference structure (in our case the MD-averaged one). In this version of the force-field no explicit terms were used to represent torsions around single bonds (the only fixed torsions were that of the peptide bonds and sidechain rings), but forbidden regions in the Ramachandran plot were not accessible due to the steric interactions between backbone O and C_β atoms (see Results).

Up to six nonbonded terms were considered: i) a hardcore infinite repulsive term to avoid nuclei overlap, ii) a backbone

hydrogen bond potential which is designed to enforce the maintenance of the secondary structure of the reference conformation, iii) a salt bridge term to preserve native ionic contacts, iv) a potential accounting for residue-residue hydrophobic interactions, v) a weak C_α-C_α based dispersion term, and finally vi) a pseudosolvation term for exposed residues to avoid the tendency of exposed polar side chains to collapse onto the surface of the protein.

Hardcore repulsions between all nonbonded particles at $d_{min} = R_{hc}{}^i + R_{hc}{}^j$ were established (see Table S1 in the Supporting Information for hardcore radii of each atom type) using an infinite repulsive wall (see Figure 1). Hydrogen bonds were defined with pseudobonds between backbone O_i and N_j that were located at a distance $R_c < 4$ Å in the reference conformation as long as i) the N_jO_i and N_jC_i axes are almost parallel (maximum angle allowed of 45°) and ii) the $C_{j-1}O_i$, $C_{\alpha j}O_i$, and N_jO_i axes are almost coplanar (see schematic diagram in Figure 1). To keep the geometry of the hydrogen bonds, pseudobonds between C_i and N_j, O_i and C_{j-1}, and O_i and $C_{\alpha j}$ were added. These pseudobonds are defined as finite potential wells centered at the interatomic distance d_{ab} in the native conformation, whose width is 5% of d_{ab}. Salt bridges were defined between suitable atom types (see Table S2 in the Supporting Information) that were located at a distance lower than $R_c = 6$ Å in the native conformation. Hydrophobic interactions were defined between the central atoms of hydrophobic side chains (see Table S3 in the Supporting Information) that were located at $R_c \leq 6$ Å. Salt bridges and hydrophobic interactions were defined as wells between d_{min} and R_c. Hydrogen bonds and hydrophobic and saline interactions are defined by deep (15 kcal/mol) finite wells in order to preserve the contact topology of the native conformation. Therefore the method is not well suited for *ab initio* folding prediction. The pseudosolvation term is based on a volume exclusion term for exposed polar residues as a hardcore potential of 4 Å for the atoms specified in Table S4 in the Supporting Information. Finally, dispersion terms at the residue level were introduced by using a C_α-based finite square well. The potential associated to this well is defined by an infinite wall at the hardcore repulsion distance $R_{hc}{}^i + R_{hc}{}^j$ and a step of depth $V(r_0)$ at the interparticle distance r_0 plus 1 Å in the native conformation. The distance-dependent $V(r_0)$ term is computed as

$$V(r_0) = -\varepsilon \left(\frac{d}{r_0} \right)^6 \tag{11}$$

where $d = R_{hc}{}^i + R_{hc}{}^j$, and the factor hardness constant ε was fitted to 2 after training with a selected set of ultrarepresentative proteins in our database (see below).

Overall, the force-field tries to maintain the structure close to the reference conformation but enables large movements if they do not disrupt favorable physical interactions. The force-field is then expected to keep sampling within reasonable limits around the reference structure but providing a more realistic exploration of conformational space.

Molecular Dynamics Simulations. MD trajectories were taken when available from our µMODEL database (http://mmb.pcb.ub.es/MODEL) or computed directly for this work. In all cases protein structures were titrated, neutralized by ions, minimized, hydrated, heated, and equilibrated (for at

United-Atom Discrete Molecular Dynamics of Proteins

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2007**

least 0.5 ns) using an established protocol,[17] followed by at least 10 ns of production. Trajectories in the micromodel database (see below) were collected using three all-atom force fields (AMBER,[51] CHARMM,[52,53] OPLS/AA[54−57]). The μMODEL proteins considered in this work are 1AGI, 1BFG, 1BJ7, 1BSN, 1CHN, 1CQY, 1CSP, 1CZT, 1EMR, 1FAS, 1FVQ, 1I6F, 1IL6, 1J5D, 1JLI, 1K40, 1KTE, 1KXA, 1LIT, 1LKI, 1OOI, 1OPC, 1PDO, 1PHT, 1SDF. Additional proteins which are not present in the μmodel databases and whose MD simulations are presented here for the first time were obtained using only the AMBER[51] force-field: 1ARK, 1BPI, 1CEI, 1SRO, 1UBQ, 2GB1, 3CI2, 4ICB.

The Particle Mesh Ewald approach in conjunction with periodic boundary conditions was used to address long-range nonbonded interactions.[59] Integration of the equations of motion proceeded with a time step of 1 fs; vibrations of bonds involving hydrogen atoms were removed by the SHAKE/RATTLE algorithm.[60,61] Jorgensen's TIP3P model[62,63] was used to represent aqueous solvent. Calculations were performed with AMBER8[64] and NAMD2.6[65,66] computer programs. Results presented in the manuscript are always referred to as the AMBER force-field results (those for which more proteins trajectories are available), but comparison between trajectories obtained with the three different atomistic force-fields are used as reference.

Training and Testing Databases. A reduced training set of proteins (1I6F, 1SDF, 1JLI) simulated with very long (0.1 μs) trajectories was used to fit the adjustable parameters in our model. Testing was then performed with proteins representing all metafolds as described in the micromodel database,[17] and several extra proteins were selected as examples of mobile structures in the protein databank (see above). Proteins in the database are structurally and functionally diverse and represent a quite complete sampling of single domain proteins, which are the most challenging for representation with simplified models.

Metrics for Trajectory Comparison. The samplings obtained from MD and DMD simulations were compared by using a variety of metrics, which are briefly discussed (the reader is addressed to suitable references[18] for more details). Most of the comparison relies on a preprocessing of the trajectories by the essential dynamics procedure,[67] where the covariance matrix built from the ensemble of configuration is treated by principal component analysis (PCA) to derive a set of eigenvectors (v_i) determining the nature of the essential deformation movements and a set of eigenvalues (λ_i) defining the amount of variance explained by each deformation movement. Note that eigenvalues can be transformed into stiffness constants using the harmonic expression

$$k_l = \frac{k_b T}{\lambda_l} \tag{12}$$

where T is the absolute temperature, and k_b is the Boltzman's constant.

Size of the accessible configuration space was analyzed by inspection of the corresponding variance and the entropy computed by diagonalization of the mass-weighted covariance matrix using Schlitter's method[68]

$$S = \frac{k_B}{2} \ln\left(1 + \frac{e^2}{\alpha^2}\right) \tag{13}$$

where $\alpha_i = \hbar\omega_i/k_bT$, with ω being the eigenvalues (in frequency units) obtained by diagonalization of the mass-weighted covariance matrix, and the sum extends to all the nontrivial vibrations of the system.

The reduced variance (Var_{red}) measures the size of the deformation space when only the most prevalent deformation modes are considered (here only the first 3 eigenvectors were considered)

$$Var_{red} = \sum_{i=1}^{3} \lambda_i \tag{14}$$

where i stands for the rank order of the eigenvalues.

The complexity of the deformability space is measured as the minimum number of eigenvectors needed to explain 90% of variance. Note that this parameter does not necessarily correlate with the size of the conformational space, since samplings with large variance might be explained by a small number of very soft modes, indicating a wide but simple deformation space.

The overlap of deformation space indicates the similarity between the essential spaces of two trajectories and was computed by analyzing the overlap of the respective important spaces,[69−71] defined as those necessary to complete 90% of the total variance in the trajectories generated with MD

$$\gamma_{XY} = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{m}(v_i^X \cdot v_j^Y)^2 \tag{15}$$

where X and Y index the two methods to be compared, i and j index the eigenvectors (ranked on the basis of their contribution to structural variance), and m is the number of eigenvectors in the "important space". For a finite time trajectory structural variation generates noise in the similarity computation that needs to be corrected by including the self-similarity terms, deriving then a relative similarity index[69−71]

$$\Gamma = \frac{2\sum_{i=1}^{m}\sum_{j=1}^{m}(v_i^X \cdot v_j^Y)^2}{\sum_{i=1}^{m}\sum_{j=1}^{m}(v_i^{Y\prime} \cdot v_j^{Y\prime})^2 + \sum_{i=1}^{m}\sum_{j=1}^{m}(v_i^{X\prime} \cdot v_j^{X\prime})^2} \tag{16}$$

where the self-similarity products ($v_i^Y \cdot v_j^Y$) were obtained by comparing first and second halves of trajectories.

Note that similarity indexes are dependent on the number of eigenvalues considered converging to 1 for the entire eigenvector/eigenvalue (complete basis set). Statistical significance of a given similarity index should be then quantified by Z-score

$$Z_{score} = \frac{(\gamma_{XY}(observed)) - (\gamma_{XY}(random))}{std(\gamma_{XY}(random))} \tag{17}$$

where random models were obtained by diagonalization of a pseudocovariance matrix obtained from a simple coarse grained DMD simulation, where only chemical bond and hard sphere potentials were considered. As described elsewhere,[30] this pseudorandom model is more restrictive (yield-

ing to lower but more realistic Z-scores) than a pure background model obtained by assuming gas phase behavior in the proteins. The standard deviation (*std* in the equation) was obtained by considering 500 different pseudotrajectories.

Secondary structure was determined for DMD_{AA} and MD trajectories using standard H-bond annotation criteria. Additionally, Ramachandran's maps were created to investigate the population of forbidden regions along DMD simulations. In all cases calculations were performed using 5000 equally-spaced snapshots collected from the last 5 ns portion of the trajectories.

Local residue mobility was analyzed by means of the isotropic B-factors

$$\text{B-factor} = \frac{8}{3}\pi^2 \langle \Delta r^2 \rangle \qquad (18)$$

where $\langle \Delta r^2 \rangle$ stands for the oscillations of residues around equilibrium positions. B-factors distributions were compared by means of the Spearman correlation coefficient to obtain a dimensionless measure of the distribution of residue fluctuations.

Regional mobility was analyzed by means of Lindemann's index[17,72] determined as

$$\Delta_L = \frac{\left(\sum_i \langle \Delta r_i^2 \rangle / N \right)^{1/2}}{a'} \qquad (19)$$

where a' is the most probable nonbonded near-neighbor distance, N is the number of atoms, and $\langle \Delta r^2 \rangle$ is the mean-square displacement of an atom from its equilibrium position. Lindemann's index provides a well accepted description of the macroscopic behavior of a molecular system or a part of it: if Δ_L is lower than 1.5, it is considered a solid; if higher, it is considered a molten solid; and for ever higher values, it is considered a liquid.[73]

Results and Discussion

We generated 10 ns trajectories for the proteins mentioned above and compared them with the trajectories obtained from state-of-the-art all-atom explicit solvent molecular dynamics simulations. It is worth pointing out that within 10 ns DMD allows sampling a bigger conformational space than within 10 ns of MD, because the implicit treatment of the solvent and hydrogen atoms increases the conformational freedom of the protein and reduces the complexity of the energy landscape.

Despite the lack of a complex network of residue-residue harmonic (or infinite) restraints, as happened in Gō-based models, the structures sampled in current DMD simulations remain quite close to the reference conformation (see Table 1). Thermal noise explains around 1-1.5 Å of the deviation between DMD and the MD-averaged conformation, the rest being due mostly to flexible loops as noted in the TM-scores shown in Table 1.

We were concerned that the use of physical interactions and the removal of many residue-residue restriction potentials from our current force-field might expand and perhaps distort the protein in those regions with less physical contacts. Fortunately, such an effect is negligible as seen in measures

like the solvent accessible surface and radii of gyration, which are not larger than those expected from a normal MD simulation.[11] Another potential concern was the structural stability of secondary elements, since no specific torsional terms are applied to Φ and Ψ bonds. However, the analysis demonstrates that the secondary structure is well preserved (from both MD and experimental values; see Table 1), with just a moderate decrease expected from thermal vibration. Furthermore, Ramachandran's maps obtained from DMD simulations are close to those that can be derived from the same proteins from atomistic MD simulations (see Figure 2), indicating that besides its simplicity DMD is not sampling artefactual local structures. Finally, the analysis shows that DMD simulations maintain very well the pattern of native residue-residue contacts (Table 1), even in cases where there are not obvious physical interactions that can be traced between the contact partners.

At this point we should remark that we have used averaged MD structures (obtained by averaging and partial minimization of atomistic MD ensembles) as reference structures to define the DMD force-field, and, accordingly, structural analysis should always be performed with respect to these structures. However, for the sake of completeness, we extend the analysis to consider experimental structures (X-ray or NMR) as reference. Results in Table 1 demonstrate DMD ensembles are always close to experimental structure, remarking the suitability of the technique to represent real structural properties of proteins.

The force-field definition used in this DMD implementation pursues not only to maintain the samplings close to the reference structure but also to reduce the full harmonicity intrinsic to Gō-like methods. Inspection of RMSd fluctuation profiles (see randomly selected examples in Figure 3) demonstrates that the pseudophysical potential is able to maintain the samplings close to the reference conformations but at the same time allow local transitions, temporal oscillations in the trajectories, and in summary a more realistic deformation pattern than the pure harmonic behavior observed in sampling generated by NMA-predicted essential movements, Brownian MD based on C_α-C_α harmonic restraints, or our Gō-like C_α implementation of DMD (see Figure 3).

The total size of the deformation space sampled by DMD and MD simulations is quite similar, as noted in the total variances and molecular entropies (both given as values per residue) shown in Table 2. Furthermore, the distribution of variance along modes in DMD simulations is quite realistic as shown in the variance vs eigenvector profiles shown in Figure 4, the complexity and the reduced variance metrics displayed in Table 2, giving a clear improvement with respect to Gō-like C_α methods.[24] In summary, present heavy atom DMD simulations reproduce well the extension of the deformation pattern determined by atomistic MD simulation and at the same time balance properly the importance of the different deformation modes. It seems then that the Hamiltonian definition used here is able to reduce some of the artefacts arising from the use of residue-residue harmonic (or Gō-like) potentials but keeping at the same time structures close to those used as reference.

We also compared DMD and MD essential deformation spaces to determine to what extent both methodologies detect the same type of essential deformations. Results in Table 2 demonstrate that there is a good overlap between MD and DMD trajectories, since Γ values between DMD samplings and AMBER MD trajectories are not far from those obtained when the individual MD trajectories (CHARMM, AMBER, and OPLS) are compared. The statistical significance of the computed similarity becomes evident when looking at the associated Z-scores which are in the order of 10^2 (within the range of Z-scores obtained when atomistic MD trajectories with different force-fields are compared), thus ruling out the possibility of a fortuitous similarity. In other words, the DMD algorithm reported here is able to capture not only the global pattern of flexibility of proteins but also the intrinsic nature of the deformation modes.

In order to determine whether or not DMD is able to reproduce also well the residue flexibility we computed residue B-factors from the DMD ensembles comparing the values with those obtained by MD simulations and when available X-ray data. Results summarized in Table 2 (for randomly selected examples see Figure 5) demonstrate that MD and DMD values correlate well with Spearman's coefficients within the range of those obtained when the MD trajectories with different force-fields are compared (Table 2). Moreover, DMD computed B-factors correlate also well with available X-ray values (see Table 2); even such experimental data were never considered to refine the method. Finally, the ability of the DMD to distribute properly flexibility among different protein regions is also clear by inspecting Lindemann's indexes (see Table 2), which demonstrate that the balance between solid (interior of protein)/liquid (exterior of the protein) which is found in atomistic MD simulation is well reproduced by our DMD calculations.

Conclusions

All the preceding analysis, performed on a very large set of representative proteins and using state of the art methods as reference, demonstrates that the pseudophysical DMD method can reasonably reproduce the flexibility of proteins as determined by atomistic MD simulations in explicit solvent, avoiding the need to integrate the equations of motions every femtosecond. Without important modifications the method can be used to refine portions of the protein just adding competing wells that will allow a given residue to change partners to optimize the overall energy (see Figure 6), and just adding a long range attractive potential the technique can be used to study protein/ligand diffusion and protein-protein interactions in the context of flexible macromolecules. The performance of the method in these scenarios will be the subject of future investigation.

Supporting Information Available: Hardcore radii of each atom type, atom types defining salt bridges and hydrophobic interactions, atom types used to simulate solvation effects in our model, and a figure comparing the sampling obtained with DMD to that obtained with MD for the protein 1FVQ. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Ma, J.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8502–8507.

(2) Daniel, R. M.; Dumm, R. V.; Finney, J. L.; Smith, C. J. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 69–92.

(3) Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. *Science* **2002**, *295*, 1520–1523.

(4) Luo, J.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13152–13156.

(5) Hinsen, K.; Thomas, A.; Field, M. J. *Proteins* **1999**, *34*, 369–382.

(6) Waldron, T. T.; Murphy, K. P. *Biochemistry* **2003**, *42*, 5058–5064.

(7) Yang, L.-W.; Bahar, I. *Structure* **2005**, *13*, 893–904.

(8) Sacquin-Mora, S.; Lavery, R. *Biophys. J.* **2006**, *90*, 2706–2717.

(9) Remy, I.; Wilson, I. A.; Michnick, S. W. *Science* **1999**, *283*, 990–993.

(10) Henzler-Wildman, K. A.; Ming, L.; Vu, T.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450*, 913–916.

(11) Teague, S. J. *Nat. Rev. Drug Discovery* **2008**, *2*, 527–541.

(12) Marvin, J. S.; Hellinga, H. W. *Nature Struct. Biol.* **2001**, *8*, 795–798.

(13) Falke, J. J. *Science* **2002**, *295*, 1480–1481.

(14) Kenakin, T. *Trends Pharmacol. Sci.* **1995**, *16*, 188–192.

(15) Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. *Protein Sci.* **2002**, *11*, 184–197.

(16) Shoichet, B. K.; Baase, W. A.; Kuroki, R.; Matthews, B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 452–456.

(17) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 796–801.

(18) Rueda, M.; Chacon, P.; Orozco, M. *Structure* **2007**, *15*, 565–575, 2007.

(19) Karplus, M.; McCammon, J. A. *Sci. Am.* **1986**, *254*, 42–51.

(20) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590.

(21) Allen, M. P.; Tildesley, D. J. Computer Simulation of Liquids; Clarendon Press: Oxford, 1989.

(22) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*; Cambridge University Press: Cambridge, 1987.

(23) Warshel, A. *Nature* **1976**, *260*, 679–683.

(24) Van Gunsteren, W. F.; Karplus, M. *Biochemistry* **1982**, *21*, 2259–74.

(25) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(26) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

(27) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679–6685.

(28) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.

(29) Alder, B. J.; Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys* **1959**, *31*, 459–466.

(30) Emperador, A.; Carrillo, O.; Rueda, M.; Orozco, M. *Biophys. J.* **2008**, *95*, 2127–2138.

(31) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. *Biophys. J* **2005**, *88*, 147–155, 2005.

(32) (a) Zhou, Y. Q.; Karplus, M. *Nature* **1999**, *401*, 400–403. (b) Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Folding Des.* **1998**, *3*, 577–587.

(33) Linhananta, A.; Zhou, Y. *J. Chem. Phys.* **2002**, *117*, 8983–8995.

(34) Luo, Z.; Ding, J.; Zhou, Y. *Biophys. J.* **2007**, *93*, 2152–2161.

(35) Zhou, Y.; Linhananta, A. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 154–162.

(36) Ding, F.; Sharma, S.; Chalasani, P.; Demidov, V. V.; Broude, N. E.; Dokholyan, N. V. *RNA* **2008**, *14*, 1164–73.

(37) Sharma, S. F.; Ding, F.; Dokholyan, N. V. *Biophys. J.* **2007**, *92*, 1457–1470.

(38) Ding, F.; Dokholyan, N. V.; Buldyrev, S. V.; Stanly, E. H.; Shakhnovich, E. I. *J. Mol. Biol.* **2002**, *324*, 851–857.

(39) Chen, Y.; Dokholyan, N. V. *J. Mol. Biol.* **2005**, *354*, 473–482.

(40) Ding, F.; LaRocque, J. J.; Dokholyan, N. V. *J. Biol. Chem.* **2005**, *280*, 40235–40240.

(41) Khare, S. D.; Ding, F.; Gwanmesia, K. M.; Dokholyan, N. V. *PLoS Comput. Biol.* **2005**, *1*, 230–235.

(42) Marchut, A. J.; Hall, C. K. *Biophys. J.* **2006**, *90*, 4574–4584.

(43) Nguyen, H. D.; Hall, C. K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16180–16185.

(44) Peng, S.; Ding, F.; Urbanc, B.; Buldyrev, S. V.; Cruz, L.; Stanley, H. E.; Dokholyan, N. V. *Phys. Rev. E* **2004**, *69*, 041908.

(45) Ding, F.; Borreguero, J. M.; Buldyrey, S. V.; Stanley, H. E.; Dokholyan, N. V. *Proteins: Struct., Funct., Genet.* **2003**, *53*, 220–228.

(46) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010–1018.

(47) Urbanc, B.; Cruz, L.; Buldyrev, S. V.; Bitan, G.; Teplow, D. B.; Stanley, H. E. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17345–17350.

(48) Yun, S.; Urbanc, B.; Cruz, L.; Bitan, G.; Teplow, D. B.; Stanley, H. E. *Biophys. J.* **2007**, *92*, 4064–4077.

(49) Urbanc, B.; Cruz, L.; Buldyrev, S.V.; Bitan, G.; Teplow, D. B.; Stanley, H. E. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6015–6020.

(50) Taketomi, H.; Ueda, Y.; Gô, N. *Int. J. Pept. Protein Res.* **1975**, *7*, 45–459.

(51) Cornell, W.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc* **1995**, *117*, 5179–5197.

(52) 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. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(53) MacKerell, A. D., Jr.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946–11975.

(54) Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1997**, *18*, 1955–1970.

(55) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(56) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. *J. Phys. Chem.* **1994**, *98*, 13077–13082.

(57) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(58) Smith, W. S.; Hall, C. K.; Freeman, B. D. *J. Comput. Phys.* **1997**, *134*, 16–30.

(59) Darden, T. L.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(60) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1997**, *23*, 327–341.

(61) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24–34.

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

(63) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910–8922.

(64) Case, D. A., Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. L.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. University of California, San Francisco, 2004.

(65) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283–312.

(66) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(67) Amadei, A.; Linssen, A. B.; Berendsen, H. J. *Proteins* **1993**, *17*, 412–425.

(68) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617–621.

(69) Hess, B. *Phys. Rev. E* **2000**, *62*, 8438–8448.

(70) Noy, A.; Meyer, T.; Rueda, M.; Ferrer, C.; Valencia, A.; Perez, A.; Orozco, M.; de la Cruz, X.; Luque, F. J. *J. Biomol. Struct. Dyn.* **2006**, *23*, 357–484.

(71) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350–364.

(72) Zhou, Y.; Vitkup, D.; Karplus, M. *J. Mol. Biol.* **1999**, *285*, 1371–1375.

(73) Zhou, Y.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14429–14432.

JCTC Journal of Chemical Theory and Computation

Al_5O_4: A Superatom with Potential for New Materials Design

Ujjal Das and Krishnan Raghavachari*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405

Received June 19, 2008

Abstract: The $Al_5O_4^-$ cluster displays a highly symmetric (D_{4h}) planar ring structure and magic cluster stability. The enhanced stability of this nonstoichiometric cluster is the result of an unconventional electronic distribution within the cluster, which is different from that found in stoichiometric Al_2O_3. The corresponding neutral Al_5O_4 cluster exhibits a strong electron affinity (3.5 eV) that is very close to that of a chlorine atom (3.6 eV). When interacting with the electropositive metals (M = Li, Na, K, etc.), the neutral cluster captures one electron and forms a "binary salt" composed of $Al_5O_4^-$ anion and M^+ cation. Interestingly, the geometric and electronic structure of bare $Al_5O_4^-$ is completely retained in the salt structure. This suggests that Al_5O_4 behaves as a superatom and Al_5O_4M is reminiscent of a diatomic ionic molecule such as NaCl or KCl. We have also demonstrated that Al_5O_4M can be used as a building block to construct new solid state materials. A detailed structural analysis of the monomer, dimer, and trimer of Al_5O_4M reveals that while M tends to coordinate with the cluster oxygen in monomeric Al_5O_4M, the binding preference is significantly changed in the presence of multiple metal and cluster ions. In this case, M favors coordinating to the terminal Al atoms in the cluster where the four highest occupied molecular orbitals are distributed. Based on these observations, we have designed new 1-D, 2-D, and 3-D extended networks using Al_5O_4M as the building block. The 3-D periodic lattice displays a structure similar to that in zeolites and, therefore, may exhibit behavior useful for applications as molecular sieves.

Introduction

Clusters are ensembles of bound atoms intermediate in size between a molecule and the bulk solid. There are certain clusters with exceptionally high stability compared to their immediate neighbors. These are known as magic clusters. A few well-known examples of magic clusters are C_{60},[1] Ti_8C_{12},[2] Au_{20},[3] Al_{13}^-,[4] the endohedral $M@Au_{12}$ cages,[5−7] the aromatic B_{12} cluster,[8] and the polyhedral $(V_2O_5)_n$ cages.[9,10] Their high stability is typically characterized by high symmetry in the structure, saturated electronic shell, chemical inertness, and a large energy separation between the highest occupied and the lowest unoccupied molecular orbitals. In recent years, the quest to assemble these stable individual cluster units into new solid state materials has gathered a lot of momentum.[11−19] One big advantage in this approach is the opportunity to tune the electronic and magnetic

properties of the new materials based on the user choice.[18] However, there are many practical difficulties associated with this process. First of all, the clusters must retain their structural integrity in the assembled unit. Failure to do so often leads the cluster units to form oligomers. This is nicely documented in the all-metal aromatic cluster, $M_n(Al_4^{2-})$, where M represents an alkali or alkaline earth metal.[20] When brought together, two such units initiate interatomic bonding within the clusters rather than being stabilized into two separate units held together by electrostatic interactions. There may be other significant factors to be considered in such cluster-assembled materials related to the size of the clusters, heat of the lattice formation, etc. For example, Castleman and co-workers have shown that the size of the Al_{13} superatom[4] is too big to fit with counterions such as the alkali metals. They have recently proposed new super-alkali motifs (Na_3O and K_3O) to overcome this size mismatch.[16] Nevertheless, the cluster-based approach to design

---

* Corresponding author e-mail: kraghava@indiana.edu.
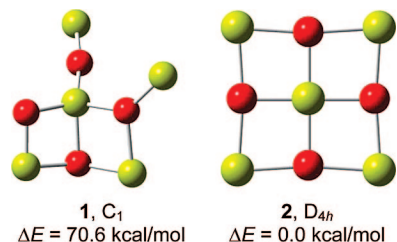
Figure 1. Lowest energy structure of Al_5O_4 (**1**) and $Al_5O_4^-$ (**2**). Also shown are their symmetry and relative stability.

new nanomaterials with desired properties continues to be very fascinating. Herein, we report an aluminum oxide cluster anion, $Al_5O_4^-$, which shows magic cluster stability, and the corresponding neutral behaves like a superatom. Using first principles electronic structure calculations, we have shown that this metal oxide cluster can be assembled into different extended systems having diverse geometric and electronic properties. Given that the predicted 3D systems show zeolite-shaped structures, such materials, if synthesized, might find useful applications as molecular sieves.

Computational Methods

The geometry optimizations of the title species and other discrete molecular systems reported in this paper are performed using density functional theory (DFT). The B3LYP hybrid exchange-correlation functional, in conjunction with the multiply polarized $6-311+G(3df)$ basis set (valence triple-ζ $6-311G$ + three d-type and one f-type polarization functions + diffuse sp-functions) is used in the calculations.[21-23] Geometry optimizations of extended systems, in 1D and 2D, are performed imposing periodic boundary conditions (PBC).[24,25] Since PBC calculations using the hybrid DFT functionals such as B3LYP are prohibitively expensive, we use the PW91 functional with a moderate sized basis set, $6-311G(d)$. All these calculations are performed using the development version of the Gaussian quantum chemistry software package.[26]

Geometry optimizations on the 3D extended systems have been performed using the DFT-based Vienna *ab initio* simulation package (VASP).[27-30] The Perdew–Wang (PW91) exchange-correlation functional within the generalized gradient approximation (GGA) has been applied here.[31-33] The calculations use ultrasoft pseudopotentials[34] and a plane wave basis set with an energy cutoff of 400 eV. The Brillouin zone is sampled using a $11 \times 11 \times 11$ k-point grid under the Monkhorst-Pack scheme.[35] The energy convergence criteria during geometry relaxation is set to 10^{-7} eV per atom. The geometries are well converged with respect to the plane-wave energy cutoff and k-point sampling. At the optimized geometry, the forces are converged to better than 0.6 meV/Å.

Results and Discussion

The $Al_5O_4^-$ cluster was easily identified as a high intensity peak at 199 amu in the experimental mass spectra.[36] The lowest energy isomers of neutral (**1**) and anionic (**2**) forms of Al_5O_4 are shown in Figure 1. The neutral cluster possesses
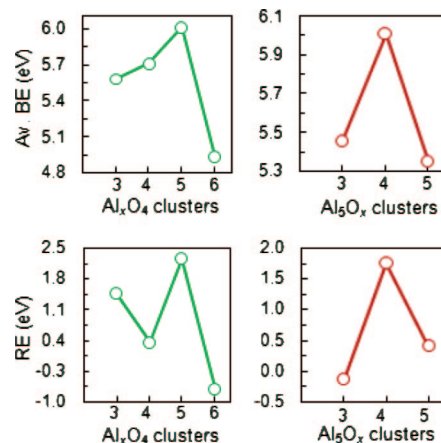
Figure 2. The top two boxes show the average atomic binding energy (see text) in $Al_xO_4^-$ ($x = 3-6$) and $Al_5O_x^-$ ($x = 3-5$) cluster series, respectively. The bottom two boxes show the local stability of each cluster with respect to the disproportionation schemes shown in eqs 3 and 4, respectively.

39 valence electrons. Interestingly, addition of one more electron to the neutral species produces a highly symmetric structure with alternating Al, O atoms along the periphery and an Al atom at the center of the ring (**2**). The central Al_c-O bonds (1.78 Å) in **2** are in general shorter than the terminal Al_t-O bonds (1.90 Å). It should be pointed out that though a B3LYP/$6-31+G(d)$ calculation suggests a perfect square planar (D_{4h}) structure for the cluster anion, use of a larger basis set such as $6-311+G(3df)$ makes it slightly nonplanar. In this case, the five Al atoms are still in a plane, but the position of pairs of trans-oxygen atoms above and below this plane introduces a slight puckering and reduces the overall symmetry to D_{2d}. However, the deviation from planarity does not bring any notable changes in the structural parameters. Moreover, the energy difference between the planar D_{4h} structure and the slightly puckered D_{2d} structure is less than 1 kcal/mol. When zero-point energy corrections and finite temperature effects are considered, the difference in energy becomes even less. Hence we conclude that the vibrationally averaged structure of $Al_5O_4^-$ is planar.[37]

The relative stability of $Al_5O_4^-$ with respect to its neighboring clusters can be estimated from the binding energies (BE) of their constituent atoms. The binding energy (BE) of aluminum in $Al_xO_y^-$ is defined as the difference between the electronic energy of $Al_xO_y^-$ and that of $Al_{x-1}O_y^-$ + Al (eq 1). The binding energy of oxygen is also defined in a similar way (eq 2).

$$BE(Al) = E(Al_xO_y^-) - E(Al_{x-1}O_y^-) - E(Al) \qquad (1)$$

$$BE(O) = E(Al_xO_y^-) - E(Al_xO_{y-1}^-) - E(O) \qquad (2)$$

The average binding energy per atom (i.e., the average of the Al and O binding energies as defined above) in two different cluster series ($Al_xO_4^-$ and $Al_5O_x^-$) is shown in the top two boxes in Figure 2. In the first case (top-left), the number of metal atoms has been altered, while in the second case (top-right), the number of oxygen atoms has been changed. In both cases, a sharp peak at $Al_5O_4^-$ suggests that

Al$_5$O$_4$: Potential for New Materials Design

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2013**

the fragmentation of the heavier species may end up enhancing the population of Al$_5$O$_4^-$. Also note that the average binding energy in Al$_5$O$_4^-$ is approximately 6 eV per atom. This is comparable to the binding energy of atoms in other stable clusters such as fullerenes (e.g., C$_{60}$) and met-cars (e.g., Ti$_8$C$_{12}$).[38]

As an alternative way to measure the stability of the clusters, we can consider two disproportionation schemes shown in eqs 3 and 4.

$$2Al_xO_4^- = Al_{x-1}O_4^- + Al_{x+1}O_4^- \tag{3}$$

$$2Al_5O_x^- = Al_5O_{x-1}^- + Al_5O_{x+1}^- \tag{4}$$

The relative stability of the species on the left-hand side of the equation compared to the species on the right-hand side gives a good estimation of the local stability of the individual cluster. They are presented in the lower panels of Figure 2. Once again a sharp peak at Al$_5$O$_4^-$ indicates a strong resistance toward changes in the current composition. All these observations justify the magic cluster stability of Al$_5$O$_4^-$.

The chemical bond analysis in Al$_5$O$_4^-$ reveals why this cluster is so much more stable. While in stoichiometric Al$_2$O$_3$, both Al and O are in their formal oxidation states (Al^{3+} and O^{2-}), the electronic distributions in the nonsto-ichiometric aluminum oxide clusters are different. The central Al atom in Al$_5$O$_4^-$ contributes three electrons toward bonding, while the remaining four metal atoms donate one electron each. This leaves a total of eight electrons (3+4+1) in the anionic cluster which are then equally shared by four oxygen atoms. All the atoms in the cluster now have saturated electronic shells (3s^2 for the four terminal aluminum atoms and 2p^6 for the central aluminum and the four oxide ions), and the resulting strong ionic bonding contributes toward the enhanced stability of the cluster. Such a bonding scenario also explains the large structural change in going from the neutral to the anionic cluster and suggests that neutral Al$_5$O$_4$ should have a strong electron affinity.

The adiabatic electron binding energy in Al$_5$O$_4^-$, which is equivalent to the adiabatic electron affinity (EA) of neutral Al$_5$O$_4$, is computed to be 3.46 eV. The corresponding vertical electron detachment energy is computed to be 3.75 eV. These two energies reconcile well with the experimental negative ion photoelectron spectra of Al$_5$O$_4^-$ where the first peak appears at around 3.8 eV with a tail at 3.5 eV.[36,39] The difference of 0.3 eV between these two energies is significant and, in fact, supports our previous observation that there is a considerable geometric change between the neutral and the anion cluster. Note that the EA of Al$_5$O$_4$ is very close to that of the chlorine atom (3.6 eV), the element with the highest EA in the periodic table. The very strong electron binding energy, high symmetry, and high stability of the anionic cluster motivate us to verify if Al$_5$O$_4$ behaves as a superatom.

To verify this, we have considered the neutral "salt" Al$_5$O$_4$K. The geometries of its two lowest energy isomers are shown in Figure 3.[40] The only difference between **3** and **4** is the relative position of K. While K occupies a position facing an O atom in **3**, it occupies a position opposite a
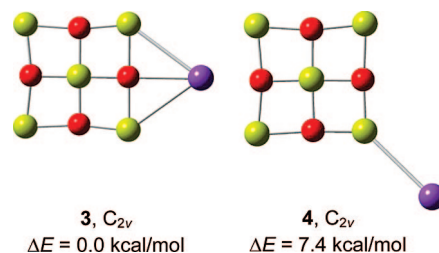
Figure 3. The two lowest energy isomers of Al$_5$O$_4$K and their relative stability.

Table 1. Selected Bond Lengths, Bond Angles, and NBO Atomic Charges of Al$_5$O$_4^-$ and Two Isomers of Al$_5$O$_4$K

parameters	Al$_5$O$_4^-$	Al$_5$O$_4$K (**3**)	Al$_5$O$_4$K (**4**)
		Bond Lengths (Å)	
Al$_c$−O	1.785	1.813/1.787	1.808/1.767
Al$_t$−O	1.902	1.817/1.905	1.813/1.939
		Bond Angles (°)	
O−Al$_t$−O	83.1	84.2/81.1	87.6/81.7
		NBO Charges (e)	
O	−1.57	−1.59	−1.57
Al$_c$	2.08	2.11	2.09
Al$_t$	0.80	0.87	0.84
K	NA	0.77	0.82

terminal Al atom in **4**. **3** is energetically more stable than **4**. The binding energies of K in **3** and **4** are −48 and −41 kcal/mol, respectively. The distance between K and O in **3** is 2.66 Å, which is nearly 0.7 Å shorter than the distance between K and Al in **4**. However, we will see later that this binding preference is actually reversed when the cluster is surrounded by multiple metal atoms. As expected, the presence of K atom in Al$_5$O$_4$K reduces the high symmetry observed in bare Al$_5$O$_4^-$. This is reflected in the corresponding structural parameters shown in Table 1. However, it is surprising that the net changes in the bond lengths and bond angles between Al$_5$O$_4^-$ and Al$_5$O$_4$K are not very significant. For example, the maximum deviations in Al$_c$−O and Al$_t$−O distances and in the O−Al$_t$−O angle between these two species are 0.03 Å, 0.09 Å, and 4.5°, respectively. The maximum deviations are observed in the vicinity of K. As the distance of the cluster atoms from K increases, these differences become even smaller. Moreover, NBO population analysis shows that the charge on K atom in Al$_5$O$_4$K is +0.8e, clearly suggesting that the alkali atom loses one electron on bond formation. A careful analysis reveals that there are almost no changes in the individual atomic charges when comparisons are made between bare Al$_5$O$_4^-$ and the Al$_5$O$_4$ segment in Al$_5$O$_4$K. This indicates that the electron lost by K is mostly captured by the cluster moiety and that the compound Al$_5$O$_4$K is actually a "binary salt" consisting of K$^+$ and Al$_5$O$_4^-$ ions. All these observations suggest that the electronic and structural integrity of Al$_5$O$_4^-$ are retained even when it interacts with the metal atoms.

A third isomer of Al$_5$O$_4$K where K interacts with the central Al atom of the cluster from the top has also been considered. The optimized structure is shaped like an "umbrella" (C$_{4v}$ symmetry) and is considerably less stable than the two other isomers (**3** and **4**). The lower stability in

this isomer is perhaps due to the unfavorable interaction between the electron deficient central aluminum atom and the K^+ ion.

The metal-halide salts are generally characterized by very large dipole moments. For example, the dipole moments of NaCl and KCl in the gas phase are 9 and 10 D, respectively.[41,42] The dipole moment of the potassium salt of Al_{13}^- anion is 11 D.[43,44] Similarly, we have computed a very large dipole moment in Al_5O_4K (13 D in **3** and 19 D in **4**). The orientations of the dipole vectors in these molecules, from the center of the cluster anion toward the cation, resemble the directions of dipole vectors in model compounds such as NaCl or KCl. This is again strongly indicative that Al_5O_4K indeed behaves as a diatomic ionic molecule.

The large energy difference between the highest occupied (HOMO) and the lowest unoccupied (LUMO) molecular orbital is a good measure of the chemical inertness of a species. For example, the HOMO–LUMO energy gap in $Al_{13}K$, which also behaves like a stable diatomic ionic molecule as shown by Bowen and co-workers, is observed to be 1.3 eV.[38] As_7K_3 is another stable cluster where this gap is measured to be 2.2 eV by Castleman et al.[15] Using TDDFT calculations, we have computed the energy required to excite an electron from the doubly filled HOMO of Al_5O_4K to the LUMO without allowing any geometric changes. This gives an estimation of the HOMO–LUMO gap in the system, which is found to be 1.6 eV. The relatively large energy gap indicates that Al_5O_4K is also chemically inert and especially suitable to serve as a building block for designing new materials. However, note that the structure and the HOMO–LUMO energy gap of Al_5O_4K described here are purely computational. This and the remaining gas phase molecules reported in this study are subject to the experimental verifications.

The integrity of Al_5O_4K structure has also been verified by exchanging the positions of terminal Al and K atoms in **3** and **4**. The K atom is now a part of the cluster unit. Since we already know from population analysis that formal charges on both of these atoms are close to +1, this alternation does not perturb the electronic distribution of the system. During geometry optimization one of these structures rearranges back into **3**, while the other optimized structure appears to be far less stable than either **3** or **4**. This confirms yet again that **3** is most stable among all these isomers.

Next we try to incorporate the individual cluster unit into real 1-D, 2-D, and 3-D networks. For this, we need to understand the orientation of the counterions around $Al_5O_4^-$ and vice versa. First we consider the species, $Al_5O_4K_4$, and set the overall charge to +3 to make sure that it corresponds to the $Al_5O_4^-$ ion surrounded by four K^+ ions. **5** and **6** in Figure 4 show the positions of the K^+ ions around an $Al_5O_4^-$ unit. **6** is about 36 kcal/mol more stable than **5**. In fact, **5** is not a minimum and has an imaginary vibrational mode that leads to **6**. This clearly suggests that the corner positions are preferred over the center of each side of the cluster when multiple potassium ions are present. A similar tendency was previously observed by Hoffmann and co-workers where they found that Li^+ ions bind more strongly at the four corners of the planar tetracoordinate carbon unit, C_5^{2-}.[14] This trend

5, D_{4h}
$\Delta E = 35.5$ kcal/mol

6, D_{4h}
$\Delta E = 0.0$ kcal/mol
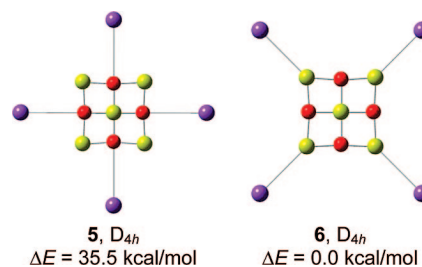
Figure 4. Two possible isomers of $Al_5O_4K_4^{3+}$ showing the preferred orientation of the cations around the cluster anion.
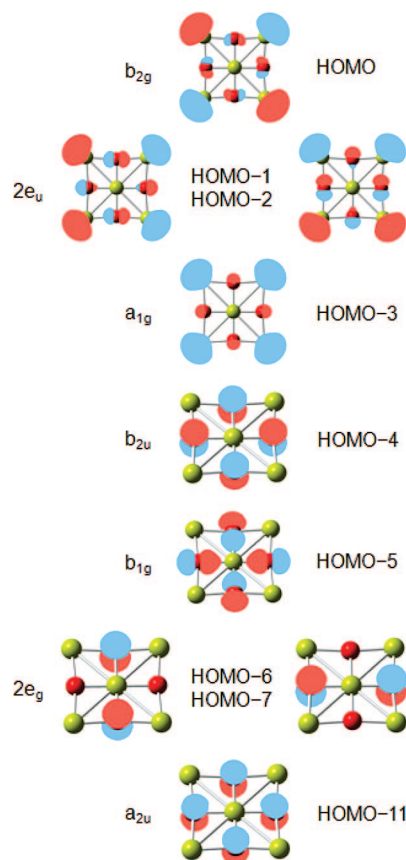
Figure 5. Molecular orbital (MO) energy diagram of $Al_5O_4^-$ showing selected orbitals. Note that the four highest occupied orbitals represent lone pairs on terminal Al atoms.

appears to be different from our earlier observation that isomer **3** is more stable than **4**. However, the higher stability of **6** relative to **5** can be understood from a careful analysis of the interactions between the species. Figure 5 displays the high-lying molecular orbitals of $Al_5O_4^-$ in their respective energy order. Note that the lone pair electrons on the terminal Al atoms in $Al_5O_4^-$ are distributed among the four highest occupied molecular orbitals. In contrast, the orbitals associated with the electron pairs on the oxygen atoms appear lower down the energy scale. As a result, the strong electrophilic interactions between K^+ ions and the lone pairs on the terminal Al atoms explain the higher stability of **6**.

The preference for metal-Al coordination over metal-O coordination can also be understood from the position of the Dyson orbitals in $Al_5O_4^-$. As shown by Guevara-Garcia et al., the Dyson orbitals of three final states after detaching

Al$_5$O$_4$: Potential for New Materials Design

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2015**

7, D$_{4h}$
ΔE = 33.5 kcal/mol

8, D$_{4h}$
ΔE = 0.0 kcal/mol
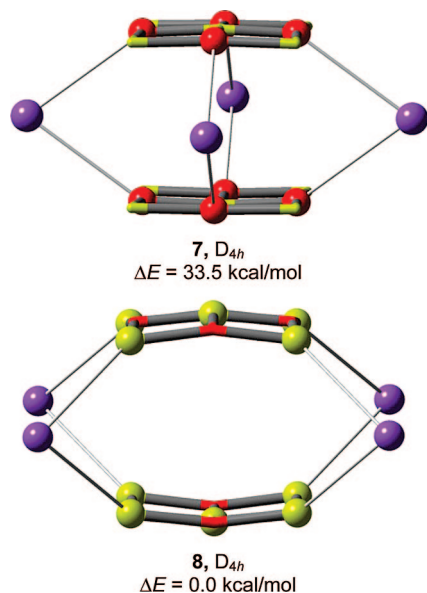
Figure 6. Two possible isomers of (Al$_5$O$_4$)$_2$K$_4{}^{2+}$ showing the preferred nonplanar orientation of the clusters.
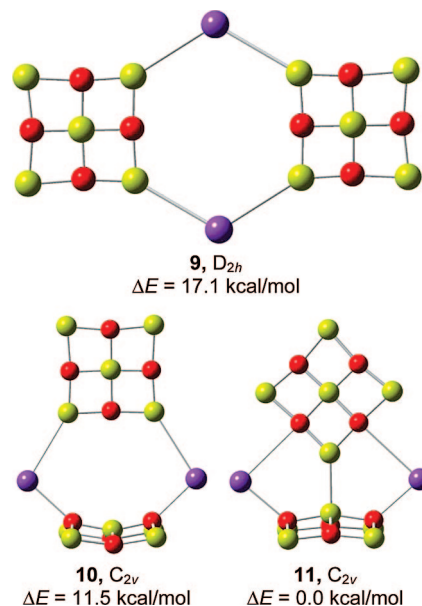

9, D$_{2h}$
ΔE = 17.1 kcal/mol

10, C$_{2v}$
ΔE = 11.5 kcal/mol

11, C$_{2v}$
ΔE = 0.0 kcal/mol

Figure 7. Three different isomers of (Al$_5$O$_4$K)$_2$ dimer and their relative stability.

an electron from this cluster consist chiefly of Al (terminal) 3s functions with minor antibonding oxygen contributions.[45] This suggests that in the interaction of neutral Al$_5$O$_4$ with K, the electron released by the metal will be mostly captured by the terminal aluminums. Also note that the absence of Dyson orbital amplitude at the central Al makes it the most preferred position for a nucleophilic attack. This is in fact observed in the reaction of Al$_5$O$_4{}^-$ with the nucleophiles such as water, ammonia, and methanol.[36,45,46]

Next, we consider different structural isomers of (Al$_5$O$_4$)$_2$K$_4{}^{2+}$ in order to understand the possible nonplanar arrangements of the cluster assemblies. We begin with **7** where the two cluster units are stacked on top of each other and the four K atoms are positioned in a separate layer in between the planes of the clusters. This is somewhat analogous to the structure of sandwich complexes such as ferrocene. As shown in Figure 6, the metal atoms in this configuration are actually facing the oxygen atoms in the cluster. However, we have found that **7** is a first order saddle point on the potential energy surface. When the displacement vectors of the corresponding imaginary frequency are followed, the optimization leads to structure **8**. **8** is 34 kcal/mol more stable than **7**. In **8**, the metal ions are facing the corner Al atoms, consistent with the previously observed trend.

As we have mentioned briefly in the Introduction, it is possible sometimes that the structural integrity of the individual cluster may collapse when two such units interact with each other. This is usually due to the formation of intercluster atomic bonds, which initiates the collapse. If so, then predicting an extended system using such clusters as building blocks has to be regarded with some caution. We want to verify this point for the Al$_5$O$_4$K unit. In particular, we want to see if Al–Al bond formation leading to a thermodynamically more stable product is feasible. Since each terminal aluminum atom contains one nonbonding electron pair, it is unlikely that the interaction between two terminal atoms will stabilize the system to any significant

extent. However, since the central aluminum is electron deficient, it can induce an ion-dipole type interaction with a polar end of another cluster, which may result in some bonding interaction between the two clusters.

To gain insight into how two Al$_5$O$_4$K units interact with each other, we have considered many different structures and found dimers **9**–**11** as the three lowest energy isomers (Figure 7). Note the relative positions of the cations in these isomers. **9** is planar and the K atoms are only pointed toward the terminal aluminums. **10** shows a nonplanar arrangement of the clusters where the alkali metals are facing both Al$_t$ and O atoms. **11** shows a different nonplanar arrangement of the clusters where the K atoms are only facing the cluster oxygen atoms. Additional ion-dipolar interaction between the central (in +3 state) and terminal (contains lone pair) aluminums is also found in **11**, which makes it the most stable among all three dimers. As a result of this interaction, the central metal atom slightly projects out of the cluster surface. The Al$_c$–Al$_t$ distance is 2.61 Å. Nonetheless, the structural integrity of the two clusters in **11** is mostly retained. We will see next which of these three structural patterns is actually translated in an extended system.

Alternative isomers considering the interactions between the oxygens of one cluster and the aluminums of another cluster also have been generated. Their optimized geometries are comparatively less stable than the ones described above. This suggests that the dimerization via formation of Al–Al bonds is energetically more preferable than Al–O bond formation.

We have already observed that in the presence of multiple cations, the Al$_t$–K interactions are energetically favored over O–K interactions for both planar and nonplanar arrangement of the counterions (isomers **6** and **8**). This gives a first order approximation that in an extended system the structural pattern found in **9** should get preference over the other dimers. This has been further verified here. Shown in Figure 8 are the optimized structures of three trimers designed based

12, D_{2h}
ΔE = 0.0 kcal/mol

13, D_{2d}
ΔE = 10.1 kcal/mol

14, D_{2d}
ΔE = 0.7 kcal/mol
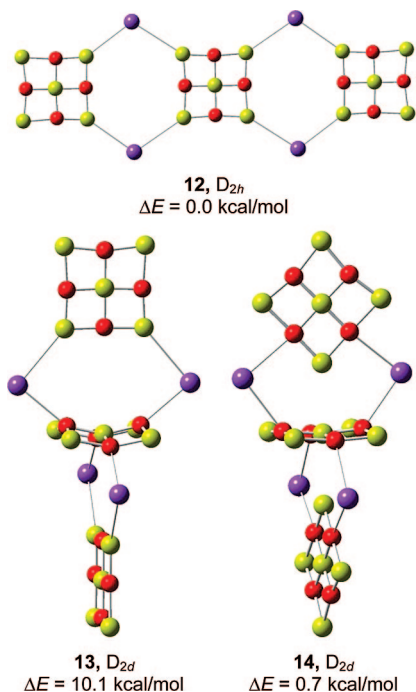
Figure 8. Three different isomers of $(Al_5O_4K)_3$ trimer and their relative stability. One additional K^+ ion in the trimer structure has been added to maintain the structural building blocks.

on the corresponding dimer structures. One additional K is placed to maintain the structural building blocks (leading to a high symmetry), and the overall charge is set to +1 to retain the proper electronic distribution. Note the stability order of these three isomers. The large energy difference observed in case of the dimers favoring the nonplanar configurations does no longer exist. Instead the planar configuration **12** is now energetically competitive with **14** and, in fact, slightly more stable. This indicates that the ion-dipolar interactions are not as dominant as the number of clusters in the oligomer increases. These results strongly suggest that the structural pattern found in **4**, **9**, and **12** will lead to a stable extended system.

Optimized structures of the extended 1-D and 2-D networks are displayed in Figure 9. **15** is obtained by expanding structure **4** in one dimension. On the other hand, **16** repeats the trend observed in **9** and **12** in two dimensions. **16** belongs to the *P4mm* space group. The length of the unit cell vectors in these two systems is given in Table 2. In each unit cell, there is exactly one cluster anion and one K^+ ion, thus maintaining the overall charge neutrality. The Al_t-K distances are slightly elongated than the values observed in the discrete molecular systems. However, the Al−O distances do not show any significant changes. To examine how the HOMO−LUMO energy gaps of the cluster motifs evolve as they are brought together to form the solid, a comparison is made in Figure 10. Note that the HOMO−LUMO gap of 1.6 eV in Al_5O_4K increases significantly to 3.2 eV in the corresponding dimer. This is because when two Al_5O_4K units interact with each other, the corresponding HOMOs are stabilized while the LUMOs are destabilized, increasing the overall energy gap. However, the gap in the trimer unit is reduced to 2.3 eV due to orbital interactions of the additional cluster. Interestingly, a similar band gap is computed for the
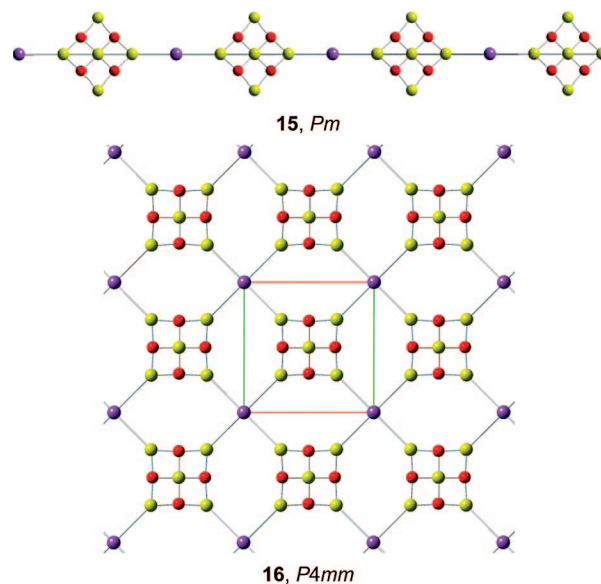
15, *Pm*

16, *P4mm*

Figure 9. Optimized 1-D and 2-D networks using Al_5O_4K as building blocks.

Table 2. Space Group, Unit Cell Parameters (Å), z Value, and Volume (Å³) of Different Extended Networks

system	space group	z	a	b	c	V
15	*Pm*	1	12.23			
16	*P4mm*	1	8.90	8.90		
17	*P4̄m2*	1	7.15	7.15	13.26	677.35
18	*P4/mmm*	1	8.17	8.17	3.11	207.95

extended networks **15** and **16**. This shows that the electronic properties of the cluster motifs are more or less retained in the extended structures.

Based on the above observations, we have designed a 3-D network displayed in Figure 11. **17** belongs to the $P\bar{4}m2$ space group, and the crystal has a tetragonal unit cell. An extended view of the crystal structure is also presented in Figure 11. Note that the local coordination of the cation in **17** is tetrahedral. Since K is less known for this type of coordination, we have used copper(I) as the counterion. Cu(I) compounds such as chlorides and oxides are known to exhibit a similar coordination for the metal center. In each unit cell, there are two $Al_5O_4^-$ and two Cu(I) atoms. The $Cu-Al_t$ distance is 2.24 Å, approximately 0.3 Å shorter than the sum
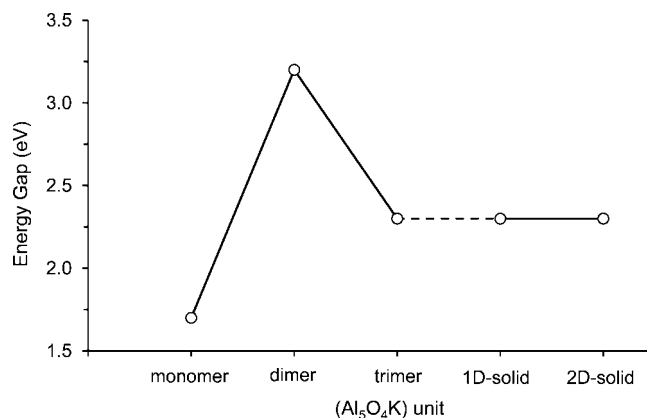
Figure 10. Variation in energy gap (HOMO−LUMO gap for clusters and band gap for solids).

Al₅O₄: Potential for New Materials Design

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2017**

17, *P4̄m2*

3.6 Å

7.2 Å
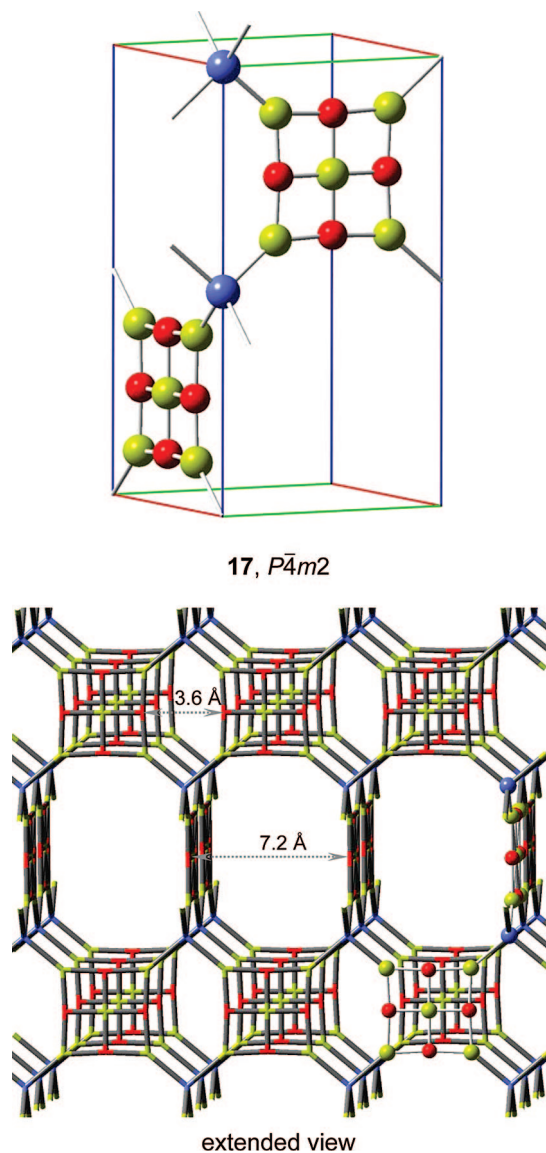
extended view

Figure 11. Unit cell (top) and an extended view (bottom) of a 3-D network of Al_5O_4Cu.

of their atomic radii. This confirms the ionic type of interaction between the species. The average computed lengths for Al_c−O and Al_t−O bonds are 1.78 Å and 1.87 Å, respectively. These values can be compared to the corresponding lengths of 1.78 Å and 1.90 Å in the bare cluster, which suggests very little structural change in the extended system. The angles around the Cu atom deviate somewhat from the ideal tetrahedral value of 109.5°. The Al_t−Cu−Al_t angles along the unit cell axes are in general shorter (99.7°) than the same angles between two different axes (114.6°).

The effective diameter of the larger pores in **17** is 0.72 nm (7.2 Å), while the diameter of the smaller pores is about half of this size (0.36 nm). The molecular diameter of H_2 (σ_{eff} = 0.30 nm) and N_2 (σ_{eff} = 0.36 nm) are comparable to the size of the small pores. On the other hand, hydrocarbon molecules such as propane, butane, and pentane are suitable to fit into the large pores. Therefore, the polar interior of the three-dimensional lattice may be selectively able to bind inorganic and organic molecules in these pores. This opens
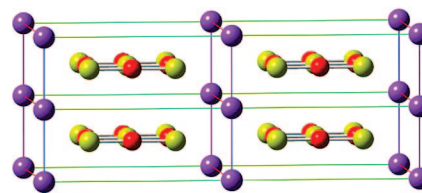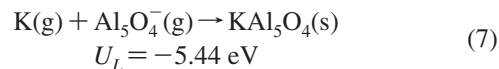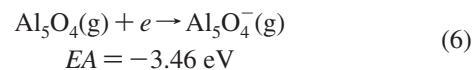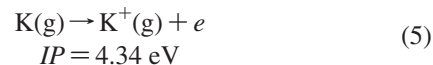

18, *P4/mmm*

Figure 12. Unit cell of a 3-D network of Al_5O_4K.

up the possibility that these zeolite-shaped structures can find important applications as storage materials and molecular sieves.

An alternative 3-D network, **18**, has been constructed based on structure **8**. **18** belongs to the *P4/mmm* space group, and the unit cell represents a body-centered tetragonal lattice (Figure 12). The alkali metals are placed at eight corners of the tetragonal lattice, while the anion occupies the body-center position. Thus each unit cell maintains a 1:1 ratio of the counterions. The unit cell parameters are presented in Table 2. The Al_t−K (3.49 Å), Al_c−O (1.79 Å), and Al_t−O (1.88 Å) distances do not change much from the discrete molecular systems. The K−Al_t−K angles are 53°.

Given that the ionization potential (IP) of K (4.34 eV) is higher than the electron affinity (EA) of Al_5O_4 (3.46 eV), the formation of K^+ and $Al_5O_4^-$ from individual molecules is slightly endothermic. However, the substantial Madelung energy resulting from the ionic lattice is expected to make the Al_5O_4K lattice formation to be highly exothermic. The lattice energy (U_L) can be estimated using the Kapustinskii equation (eq 8).[47] In this equation, ν is the number of ions in the empirical formula, z is the anionic and cationic charge, respectively, and r is the radius of the anion (2.69 Å) or cation (1.51 Å). It is important to remember that the unit of the lattice energy in this equation is given in kJ/mol, which is finally converted into electronvolt (eV). As shown in eq 9, overall the Al_5O_4K lattice formation is a thermodynamically favorable process.

$$K(g) \rightarrow K^+(g) + e$$
$$IP = 4.34 \text{ eV} \tag{5}$$

$$Al_5O_4(g) + e \rightarrow Al_5O_4^-(g)$$
$$EA = -3.46 \text{ eV} \tag{6}$$

$$K(g) + Al_5O_4^-(g) \rightarrow KAl_5O_4(s)$$
$$U_L = -5.44 \text{ eV} \tag{7}$$

$$U_L = -1202 \times \frac{\nu|z^+| \cdot |z^-|}{y^+ + y^-} \times \left(1 - \frac{0.345}{y^+ + y^-}\right) \tag{8}$$

$$\Delta H = IP + EA + U_L = -4.56 \text{ eV} \tag{9}$$

Finally, we would like to add some comments on the reactivity of Al_5O_4K. In particular, it is preferable for such predicted extended systems to be chemically stable especially in the presence of common solvent systems such as water. We have previously seen that small polar molecules such as water or methanol dissociatively add to the center of $Al_5O_4^-$.[36,46] Although surprisingly, ammonia, being a similar type of molecule, does not react. Since $Al_5O_4^-$ and Al_5O_4K have similar electronic structures, the latter may also show

similar chemical reactivity with the above molecules. However, this might not be the case for the extended systems. In the case of the monomer reactions, all these reactions are induced by initial charge-dipole type interactions between a lone pair on water and the central aluminum of the cluster anion. We have shown earlier in this paper that such charge-dipolar interaction becomes less dominant when the individual cluster units start to agglomerate. In addition, the less open access to the central aluminum due to the presence of the counterions may also lead to a less reactive environment. This gives some indication that the extended lattice may show chemical resistance against water or methanol. However, we note that, though preferable, such chemical inertness is not a requirement for the viability of formation of such materials.

Conclusions

In summary, we have shown that the aluminum oxide cluster anion $Al_5O_4^-$ exhibits a highly symmetric planar structure and magic cluster stability. The electron affinity of neutral Al_5O_4 is similar to that of chlorine, the element with the largest electron affinity in the periodic table. When interacting with electropositive metals such as Na and K, the neutral Al_5O_4 extracts an electron from the metal to form a binary salt Al_5O_4M consisting of $Al_5O_4^-$ and M^+ ions. The electronic and structural integrity of bare $Al_5O_4^-$ is completely retained in the salt structure. After a detailed structural analysis of dimers and trimers of the Al_5O_4M unit, we have designed different extended periodic networks using Al_5O_4M as building blocks. The predicted 3-D networks exhibit zeolite-shaped structure and contain pores of different diameters. This opens up the possibility that such new nanomaterials, if synthesized, may find potential application as molecular sieves.

Supporting Information Available: Coordinates of the discrete molecules and periodic systems. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kroto, H. W.; Heath, J. R.; Obrien, S. C.; Curl, R. F.; Smalley, R. E. *Nature* **1985**, *318*, 162.

(2) Guo, B. C.; Kerns, K. P.; Castleman, A. W. *Science* **1992**, *255*, 1411.

(3) Li, J.; Li, X.; Zhai, H. J.; Wang, L. S. *Science* **2003**, *299*, 864.

(4) Bergeron, D. E.; Castleman, A. W.; Morisato, T.; Khanna, S. N. *Science* **2004**, *304*, 84.

(5) Li, X.; Kiran, B.; Li, J.; Zhai, H. J.; Wang, L. S. *Angew. Chem., Int. Ed.* **2002**, *41*, 4786.

(6) Pyykko, P.; Runeberg, N. *Angew. Chem., Int. Ed.* **2002**, *41*, 2174.

(7) Zhai, H. J.; Li, J.; Wang, L. S. *J. Chem. Phys.* **2004**, *121*, 8369.

(8) Zhai, H. J.; Kiran, B.; Li, J.; Wang, L. S. *Nat. Mater.* **2003**, *2*, 827.

(9) Asmis, K. R.; Santambrogio, G.; Brummer, M.; Sauer, J. *Angew. Chem., Int. Ed.* **2005**, *44*, 3122.

(10) Zhai, H. J.; Dobler, J.; Sauer, J.; Wang, L. S. *J. Am. Chem. Soc.* **2007**, *129*, 13270.

(11) Khanna, S. N.; Jena, P. *Phys. Rev. B* **1995**, *51*, 13705.

(12) Khanna, S. N.; Jena, P. *Phys. Rev. Lett.* **1992**, *69*, 1664.

(13) Geske, G. D.; Boldyrev, A. I. *Inorg. Chem.* **2002**, *41*, 2795.

(14) Pancharatna, P. D.; Mendez-Rojas, M. A.; Merino, G.; Vela, A.; Hoffmann, R. *J. Am. Chem. Soc.* **2004**, *126*, 15309.

(15) Castleman, A. W.; Khanna, S. N.; Sen, A.; Reber, A. C.; Qian, M.; Davis, K. M.; Peppernick, S. J.; Ugrinov, A.; Merritt, M. D. *Nano Lett.* **2007**, *7*, 2734.

(16) Reber, A. C.; Khanna, S. N.; Castleman, A. W. *J. Am. Chem. Soc.* **2007**, *129*, 10189.

(17) Roach, P. J.; Reber, A. C.; Woodward, W. H.; Khanna, S. N.; Castleman, A. W *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14565.

(18) Yang, L. M.; Ding, Y. H.; Sun, C. C. *J. Am. Chem. Soc.* **2007**, *129*, 1900.

(19) Yang, L. M.; Ding, Y. H.; Sun, C. C. *J. Am. Chem. Soc.* **2007**, *129*, 658.

(20) Yang, L. M.; Ding, Y. H.; Sun, C. C. *Chem. Eur. J.* **2007**, *13*, 2546.

(21) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(22) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(23) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(24) Kudin, K. N.; Scuseria, G. E. *Chem. Phys. Lett.* **1998**, *289*, 611.

(25) Kudin, K. N.; Scuseria, G. E.; Schlegel, H. B. *J. Chem. Phys.* **2001**, *114*, 2919.

(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.; Scalmani, G.; Mennucci, B.; Barone, V.; Petersson, G. A.; Caricato, M.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Li, X.; Hratchian, H. P.; Peralta, J. E.; Izmaylov, A. F.; Kudin, K. N.; Heyd, J. J.; Brothers, E.; Staroverov, V. N.; Zheng, G.; Kobayashi, R.; Normand, J.; Sonnenberg, J. L.; Ogliaro, F.; Bearpark, M.; Parandekar, P. V.; Ferguson, G. A.; Mayhall, N. J.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Burant, J. C.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; 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.; Chen, W.; Wong, M. W.; Pople, J. A. *Gaussian DV, revision G.01*; Gaussian Inc: Wallingford, CT, 2007.

(27) Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, B864.

(28) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47*, 558.

(29) Kresse, G.; Hafner, J. *Phys. Rev. B* **1994**, *49*, 14251.

Al₅O₄: Potential for New Materials Design

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2019**

(30) Kresse, G.; Furthmuller, J. *Comput. Mater. Sci.* **1996**, *6*, 15.

(31) Perdew, J. P.; Yue, W. *Phys. Rev. B* **1986**, *33*, 8800.

(32) Perdew, J. P. In *Electronic Structure of Solids*; Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, Germany, 1991; p 11.

(33) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. *Rev. Mod. Phys.* **1992**, *64*, 1045.

(34) Kresse, G.; Joubert, D. *Phys. Rev. B* **1999**, *59*, 1758.

(35) Monkhorst, H. J.; Pack, J. D. *Phys. Rev. B* **1976**, *13*, 5188.

(36) Das, U.; Raghavachari, K.; Jarrold, C. C. *J. Chem. Phys.* **2005**, *122*, 014313.

(37) Wang, L. S.; Boldyrev, A. I.; Li, X.; Simons, J. *J. Am. Chem. Soc.* **2000**, *122*, 7681.

(38) Zheng, W. J.; Thomas, O. C.; Lippa, T. P.; Xu, S. J.; Bowen, K. H. *J. Chem. Phys.* **2006**, *124*, 144304.

(39) Meloni, G.; Ferguson, M. J.; Neumark, D. M. *Phys. Chem. Chem. Phys.* **2003**, *5*, 4073.

(40) Many of the planar discrete molecular systems reported in this study contain imaginary frequency mode(s) that corresponds to a minor out-of-plane distortion of the cluster oxygens. See the text for a brief discussion of this mode for the parent cluster anion. The number of imaginary modes

varies depending on the symmetry of the system. In most such cases, we have optimized a nonplanar structure taking into account the out-of-plane distortions that lead to lower energies. The energy difference between the planar and nonplanar configuration is always less than 1 kcal/mol. When zero-point energy corrections are considered, this energy difference is even smaller. This suggests that the vibrationally averaged structures of the reported molecules are planar, and, therefore, we have not discussed such imaginary modes in the text.

(41) Clouser, P. L.; Gordy, W. *Phys. Rev. A* **1964**, *134*, A863.

(42) Leeuw, F. H. D.; Vanwache, R.; Dymanus, A. *J. Chem. Phys.* **1970**, *53*, 981.

(43) Ashman, C.; Khanna, S. N.; Pederson, M. R. *Phys. Rev. B* **2002**, *66*, 193408.

(44) Khanna, S. N.; Rao, B. K.; Jena, P. *Phys. Rev. B* **2002**, *65*, 125105.

(45) Guevara-Garcia, A.; Martinez, A.; Ortiz, J. V. *J. Chem. Phys.* **2007**, *127*, 234302.

(46) Das, U.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127*, 154310.

(47) Huheey, J. E.; Keiter, E. A.; Keiter, R. L. In *Inorganic Chemistry: Principles of Structure and Reactivity,* 4th ed.; Pearson Education: Delhi, India, 2002; p 111.

JCTC Journal of Chemical Theory and Computation

Fragment-Localized Kohn−Sham Orbitals via a Singles Configuration-Interaction Procedure and Application to Local Properties and Intermolecular Energy Decomposition Analysis[†]

Peter Reinhardt,* Jean-Philip Piquemal, and Andreas Savin

UPMC University Paris 06, Laboratoire de Chimie Théorique, 4 place Jussieu,
F-75252 Paris, France, and CNRS UMR 7616, Laboratoire de Chimie Théorique,
4 place Jussieu, F-75252 Paris, France

Abstract: As for generating localized Hartree−Fock orbitals, we propose a potentially linear-scaling singles-CI scheme to construct fragment-localized density functional theory (DFT) orbitals for molecular systems as water clusters. Due to the use of a deformation step instead of a localization step, the influence of the environment on each separate molecule can be studied in detail. The generated orbital set for the whole molecular system is strictly equivalent to a set of canonical orbitals and is a subsequent energy decomposition of intermolecular interactions into electrostatic, exchange repulsion, and orbital interaction, well beyond dimer systems. Beyond this, the correspondence of the individual orbitals to the initial monomer orbitals permits to assess how an interaction deforms an electron density. We show this for dipole moments, which may be decomposed into monomer contributions, polarization, and charge-transfer contribution. Applications to a water and an ammonia dimer and chains of water molecules show possible further developments toward multipolar expansions and other orbital-based schemes for parametrizing force fields.

1. Introduction

Density functional theory (DFT) is now a widely used tool for calculating molecular properties due to a favorable scaling with system size, permitting calculations at high precision even for medium-sized molecules, inaccessible to perturbation- or configuration-interaction (CI) based quantum chemical methods.

Another advantage of DFT lies in the fact that a single Hartree product or Slater determinant with Kohn−Sham orbitals is usually employed for describing the wave function and not a long expansion in reference and excited determinants. The connection between the density and orbitals is therefore straightforward as in Hartree−Fock, nevertheless including correlation through the functional.

This Kohn−Sham wave function—the orbitals generating the density, assembled in a Hartree product or in a Slater determinant—is invariant under unitary transformations of the occupied orbitals. This property, although used, may deserve more attention for molecular systems, as it is a recurrent topic in modern wave function-based quantum chemistry, notably in the quest for linear-scaling algorithms.[1]

Orbital localization has a longstanding tradition in quantum chemistry (see ref 2 and references therein) and is associated with the names of Boys,[3] Edmiston-Ruedenberg,[4] Pipek−Mezey,[5] or even the natural bond orbital (NBO) analysis of Weinhold et al.[6,7] The Boys localization is known in the physicists community as Marzari−Vanderbilt[8] scheme. These methods optimize a localization criterion minimizing bielectronic repulsion (Boys) or the fragmentation of the orbitals on atomic centers (Pipek−Mezey). Commonly the starting point for these localization methods is a set of completely delocalized canonical orbitals, diagonalizing the Fock or Kohn−Sham matrix. The use of delocalized orbitals as

---

starting point becomes rapidly a handicap for larger systems, motivating the development of intrinsic localization procedures without the necessity to generate completely delocalized orbitals.

Recently, several approaches have been proposed in order to use localization criteria to unravel the origin of intermolecular interactions in weakly interacting molecular systems. For example, Rob et al.[9] presented a approach based on Boys' localization for the computation of intermolecular electrostatic energies including correlation showing promising results which could be competitive with density fitting approaches[10,11] as cutoff approximations lead to errors below 0.5 kcal/mol per dimer interaction. At the same time, a full energy decomposition analysis (EDA) was achieved by Khaliullin et al.[12,13] using "absolutely localized orbitals" demonstrating the advantage of the localization process for computations on large systems. Such results are important as EDA are essential to the development and the calibration of advanced force fields.[11,14]

For constructing intrinsically localized DFT molecular orbitals, we take up an old idea,[15] further developed and published by Daudey[16] and still cited recently,[2,17,18] without the need to have delocalized canonical orbitals. The central point is that in Hartree−Fock theory a CI of singly excited Slater determinants lowers the total energy toward a set of Hartree−Fock orbitals, reaching convergence when satisfying Brillouin's theorem: interactions between occupied and virtual orbitals via the Fock matrix vanish. The singles-CI matrix is locally concentrated, as excitations on fragments with large spatial separation are much less important than those within close neighborhood. That permits the local correction of the monomer orbitals, disregarding the actual extent of the complete system. No canonical (completely delocalized) molecular orbitals are created or needed.

The method becomes operational and potentially linear scaling when taking for the elements of the configuration interaction matrix between singly excited determinants only the part generated from Fock matrix elements as

$$\langle \Phi_i^a | \mathbf{H} | \Phi_j^b \rangle \approx F_{ab}\delta_{ij} - F_{ij}\delta_{ab} \qquad (1)$$

Otherwise a partial four-index transformation would be necessary for each SCF iteration.[19]

The use of density-functional theory for intermolecular interactions may seem inappropriate, as the important dispersion part is not correctly treated in actual Kohn−Sham formalism. However, improvements are under way,[20] and the scheme we like to present here will be directly applicable to these improved functionals as well. On the other hand, we may cite several successful DFT applications for hydrogen-bonded systems,[21] despite this inherent default.

We should stress again that the application of the present orbital localization scheme does not help to overcome basic defaults of current density functional theory by inclusion of missing parts of the correlation energy. Even though we employ a CI-based optimization scheme, nothing is for the moment included to go beyond a common Kohn−Sham energy, obtainable via standard procedures. Nevertheless, the generated local orbitals and moments developed here may be helpful in a subsequent step to develop a proper dispersion

scheme like in the recent DFT based SAPT (symmetry adapted perturbation theory), where the Kohn−Sham monomer orbitals act as starting point for a perturbative intermolecular scheme.[22,23]

In the SCF MI approach (SCF for molecular interaction) of Khaliullin et al.,[12,13] a similar scheme for orbital optimization is presented, replacing the iterative singles-CI step by one single calculation of fourth-order perturbation theory in monomer orbitals without explicit orthogonalization of the complete orbital set.

For the sake of completeness, we may cite the work of Adams and Gilbert with a similar objective,[24,25] constructing local fragment orbitals in the sense of building blocks for a molecule or solid, further developed by Kunz[26] and Seijo and Bandiaran.[27]

The paper is further organized as follows: in the first part, we recall the iteration scheme based of the CI step and our particular use for treating weakly interacting multimonomer or segmented systems. We show how the procedure allows the decomposition of any monoelectronic property into well-defined monomer contributions and how the scheme fits into common energy decomposition schemes like those of Ziegler and Rauk,[28] Bickelhaupt and Baerends,[29] or the constrained space orbital variations approach (CSOV) of Bagus et al.[30]

In the second part, results are shown for the dipole moments (local and global) of an ammonia dimer, the energy decomposition for a linear water dimer,[31,32] and an orbital analysis and an energy decomposition for small, helixlike water clusters (denoted longitudinal hydrogen-bonded chains (l-HBCs), see ref 33 and the Appendix) in order to identify the convergence and transferability of the orbital decomposition.

In the Appendix, we collect all technical details and the formal proof of the equivalence of the singles-CI scheme for obtaining Hartree−Fock or Kohn−Sham orbitals.

2. Methodological Details

We consider a system which may be decomposed intuitively into well-separated fragments. For each fragment, we construct canonical orbitals separately, in a set of atomic orbitals attached only to the particular fragment. From this calculation, we keep the virtual orbitals for spanning the virtual orbital space of the complete system. To obtain reasonable occupied orbitals of the monomers, we perform in a next step a calculation of canonical orbitals of the individual monomers in the atomic-orbital basis of the complete system. The advantage of such a procedure lies in the fact that the necessary integrals over atomic orbitals have to be calculated only once, and all iterations can be performed with the same set of integrals. At the end of the double series of monomer calculations, the orbitals are assembled to a starting set of occupied orbitals and orthogonalized employing Löwdin's $\mathbf{S}^{-1/2}$ orthogonalization.[34] This "democratic" orthogonalization insures that on average every orbital is corrected in a least-most perturbation, preserving as most as possible the characteristics of the monomer orbitals.[35] To this set of occupied starting orbitals, the previously generated set of virtual orbitals is orthogonalized through projection, and at last, the virtual orbitals are orthogonalized among themselves, again via Löwdin's procedure.

2.1. SCF Iterations through a Singles-CI. Now everything is set up for starting the SCF procedure toward the HF or DFT wave function of the whole system. A first determinant Φ_0 is constructed from the occupied starting orbitals, and the energy is lowered by variational inclusion of monoexcited determinants Φ_i^a

$$\Psi = \Phi_0 + \sum_{i,a} c_i^a \Phi_i^a \qquad (2)$$

from which we take the coefficients to correct in first-order the occupied and virtual orbitals $\phi_i(\vec{r})$ and $\phi_a(\vec{r})$, respectively, through

$$\phi_i'(\vec{r}) = \phi_i(\vec{r}) + \sum_a c_i^a \phi_a(\vec{r})$$
$$\phi_a'(\vec{r}) = \phi_a(\vec{r}) + \sum_i c_i^a \phi_i(\vec{r}) \qquad (3)$$

The correction of the virtual orbitals ensures that the occupied and virtual space remain well orthogonal with respect to each other. However, orthogonality within each of the two spaces is only ensured to first order, and the orbitals are orthogonalized again using Löwdin's procedure. With the orthogonal orbitals at hand, we construct again the singles-CI matrix, look for its lowest eigenvalue and corresponding eigenvector, and correct the orbitals, etc. The loop is controlled either by the changing of the total energy or by the smallness of the mean Fock (or Kohn–Sham) matrix element $\mathbf{F}_{ia}$, coupling occupied and virtual orbitals. If one insists (and making useless the effort of preserving locality), the generated (localized) orbital set can be transformed directly to a canonical one by a single diagonalization of the Fock (or the Kohn–Sham matrix in the present case). The final determinant (Hartree–Fock or Kohn–Sham) is indeed the lowest in energy as the energy evaluation is independent of the CI procedure, using in the Hartree–Fock case the Rayleigh quotient $\langle\Phi_0|\mathbf{H}|\Phi_0\rangle/\langle\Phi_0|\Phi_0\rangle$ or, respectively, in the Kohn–Sham case, the corresponding energy functional $E[\rho]$ with a particular exchange-correlation part $E_{XC}[\rho]$.

The cost of this CI procedure is equivalent to a straightforward iteration using a diagonalization of the Fock (or Kohn–Sham) matrix due to the use of an approximate CI matrix (eq 1) without an explicit four-index transformation of the bielectronic integrals at each iteration.

Linear scaling may be achieved because the orbital corrections due to the orthogonalization procedure and the Fock matrix elements are strongly localized in real space in the beginning and remain localized as unnecessary delocalizations due to any diagonalization are avoided.

As for all CI schemes, the question of size-consistency may be posed. Algorithms for the correction are well-known (Davidson correction, CEPA-like treatments,[36] etc.) but seem not to be of crucial importance in the present, medium-sized cases.[37] The final orbitals are the same, and only small improvements of convergence were observed.

2.2. Use of the Orbital Deformations for Analyzing the Wave Functions. As we have at the end orbitals associated to the monomer orbitals, we may try to expand an according density for example in multipolar moments or consider just the deformation in a linear decomposition of the form

$$\phi_i(\vec{r}) = \phi_i^M(\vec{r}) + \phi_i^S(\vec{r}) + \phi_i^O(\vec{r}) \qquad (4)$$

where the subscripts stand for "monomer", "same monomer" and "other monomers" to describe the effect of on-site polarization and of charge transfer. The monomer part is the projection of $\phi_i(\vec{r})$ on the corresponding, unperturbed monomer orbital $\phi_i^0(\vec{r})$, the second part is the expansion of the remainder on the basis functions on the same monomer, and the third one the expansion on the basis functions located elsewhere.

The three parts are not orthogonal but give rise to six different density matrices

$$P^{MM,i} + P^{MS,i} + P^{MO,i} + P^{SS,i} + P^{SO,i} + P^{OO,i} \qquad (5)$$

the sum of which is the partial density matrix created by the molecular orbital $\phi_i(\vec{r})$.

$$P^i = |\phi_i\rangle\langle\phi_i| \qquad (6)$$

In the spirit of a Mulliken analysis, we may attribute densities between different parts in halves to the proper parts MM, SS, and OO as The analysis may be done for each

$$P^i = \underbrace{P^{MM,i} + \frac{1}{2}\left(P^{MS,i} + P^{MO,i}\right)}_{\text{monomer}} + \underbrace{P^{SS,i} + \frac{1}{2}\left(P^{MS,i} + P^{SO,i}\right)}_{\text{polarization}} + \\ + \underbrace{P^{OO,i} + \frac{1}{2}\left(P^{MO,i} + P^{SO,i}\right)}_{\text{charge transfer}} \qquad (7)$$

orbital separately, showing thus more or less deformed orbitals of the cluster with respect to the individual monomers. Of course, this detailed decomposition depends on the form of the monomer orbitals and of that of the orbitals of the whole system. Nevertheless, the choice of canonical monomer orbitals and CI-generated multimer orbitals reduces this general ambiguity toward reproducible quantities.

$$P^i = P^i_{\text{Mono}} + P^i_{\text{Polarization}} + P^i_{\text{CT}} \qquad (8)$$

The sum of all of these partial density matrices gives the density matrix of the whole system.

$$P^{\text{system}} = \sum_i P^i \qquad (9)$$

On the other hand, summing these orbital-specific density matrices over the occupied orbitals of each monomer separately,

$$P^{\text{frag}} = \sum_{i \in \text{frag}} P^i \qquad (10)$$

we may decompose any monoelectronic operator into mono, polarization, and charge-transfer parts attributed to the individual monomers. We show this for the Mulliken populations and dipole moments in the next section. For instance with the dipole operator μ, we have

$$\vec{\mu}^{\text{frag}} = P^{\text{frag}}\vec{\mu} = (P^{\text{frag}}_{\text{Mono}} + P^{\text{frag}}_{\text{Polarization}} + P^{\text{frag}}_{\text{CT}})\vec{\mu} \\ = \vec{\mu}^{\text{frag}}_{\text{Mono}} + \vec{\mu}^{\text{frag}}_{\text{Polarization}} + \vec{\mu}^{\text{frag}}_{\text{CT}} \qquad (11)$$

Of course we have to remind that this decomposition is not unique, as it is based on orbitals. Nevertheless, starting from canonical monomer orbitals and employing the de-

scribed singles-CI procedure gives a recipe to construct decompositions with a well-defined procedure. Our dipole moments sum up to the global dipole moment of the whole system under study, and following them individually when assembling a system may help to identify characteristic regions of the fragments.

The same principle may serve for total energy decompositions, the (Kohn−Sham) kinetic energy and the electron−nuclear attraction also being monoelectronic properties. However, the bielectronic terms (electron−electron Coulomb repulsion and Hartree−Fock exchange as for the B3LYP functional) become very numerous, and the attribution of their cross-terms to individual monomers seems less reasonable than for the monoelectronic quantities. Therefore, we follow for the moment another proposition, as detailed below.

2.3. Intermolecular Energy Analysis. Intermolecular energy decompositions are often in the center of interest for singling out active sites of fragments and preponderant, chemically intuitive contributions to the overall stability of an assembly. One of the first attempts with great success was Morokuma's analysis of a Hartree−Fock interaction energy,[38,39] using parts of the Fock matrix in the self-consistent scheme in order to extract electrostatic, polarization, charge-transfer, and a sum of remaining terms. Instead of performing several independent calculations for obtaining the different contributions, we have all at hand for constructing the interaction energy step by step from the unrelaxed monomer orbitals (*electrostatic interactions*), the orthogonalized monomer orbitals (*Pauli repulsion*), and *orbital interaction* as relaxation of the orbitals from the starting orbitals of the singles CI step to the Kohn−Sham orbitals of the whole system. This scheme, initially proposed by Ziegler and Rauk,[28] has been popularized by Bickelhaupt and Baerends[29] and served recently for a comparison with the ab initio SAPT scheme.[40] The CSOV decomposition scheme, in its spirit closer to Morokuma's original scheme as subsequently more and more orbitals are allowed to relax in harmony, gives as well results close to those of the Bickelhaupt and Baerends decomposition.[41]

As these energy decompositions are completely invariant under orbital localization, we should obtain exactly the same results as for canonical orbitals, unless employing approximations within the orbital space as cutoffs or selection schemes. The advantage of the use of (fragment-)localized orbitals instead of canonical ones lies in the fact that approximations become distance dependent and can be chosen prior to a calculation (to be carried out more efficiently) without loosing the necessary precision.

Beyond a common decomposition of one single dimer interaction (like in SAPT or CSOV), we may decompose the *n*-fragment interactions in larger clusters without any additional effort. Such multifragment decompositions at the DFT level have only to be carried out at the moment by Khaliullin and Head-Gordon[12,13] and should be useful for obtaining reference data for force fields.
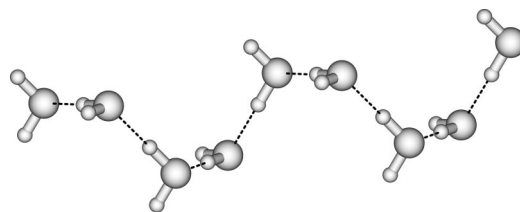
Figure 1. Water molecules in a helixlike linear structure. Pieces of one to five molecules from the left end of the chain have been considered in the present work.

3. Applications

The present approach has been applied to an ammonia dimer and a series of small water clusters. All technical details may be found in the Appendix.

The water cluster are parts of a helixlike arrangement (as depicted in Figure 1) of which we took pieces with two to five molecules.

3.1. Uniqueness of the Orbitals. The spread of the aforementioned decomposition (eq 7) gives us the possibility to measure the uniqueness of the generated orbitals for different conditions of their calculation. One can invoke the arbitrariness of the final orbitals to discredit the proposed generation scheme via the singles-CI step, as the final orbitals depend on the initial set of guess orbitals. Using canonical monomer orbitals as guess, this arbitrariness is already removed.

Here, we show that even different ways to achieve convergence, both through a constant Fock-matrix (or Kohn−Sham-matrix) mixing between the SCF cycles and a fixed scaling of the CI-vector, results only in minor differences for the resulting orbitals. For a system of three water molecules, we tested this dependence and did not find any significant changes in the occupations of the monomer, same-monomer, and other-monomer part of the individual orbitals. The variation (see Table 1) appears to be in the order of 10 μ-electrons both for the CI-coefficient scaling and the F-matrix mixing. For calculating the variation, we determine first the average population of each occupied orbital, and evaluate the sum of the quadratic deviations from these averages in each calculation. As the calculations converge in quite different numbers of iterations (see Table 1), this should be considered as reasonably stable.

3.2. Shape of the Orbitals. For the functionals LDA,[42] PW91,[43] BLYP,[44] and B3LYP[45] (and Hartree−Fock), we ran the series from two to five water molecules.

We may look (in Table 2) at the most diffuse (and overlapping) orbital at one extremity of our small water chains (the left one in Figure 1). From the five occupied orbitals of the first molecule in our series, this is the highest in orbital energy.

We recognize that in the Hartree−Fock case, the orbitals are slightly less deformed than DFT orbitals and that the different contributions seem stable for the 5-membered chain. The former parallels the usual observation that Hartree−Fock polarizabilities are smaller than DFT polarizabilities.[41]

Note that these orbitals are not Boys or Pipek−Mezey localized orbitals, producing lone-pairs or bond orbitals for instance. The spatial localization appears through the small-

Table 1. Convergence Acceleration through CI-Vector Scaling or Fock-Matrix Mixing between Iterations, for three Water Molecules in a standard Van Duijneveldt Basis[a]

			CI-Vector Scaling			
mixing/scaling	15%	20%	25%	30%	35%	40%
no. iterations	42	29	25	20	15	23
σ (monomer population)/10^{-5}	1.8	1.8	1.7	1.7	1.6	1.6
σ (same orb pop)/10^{-5}	1.6	1.5	1.5	1.5	1.4	1.4
σ (other orb pop)/10^{-5}	0.9	0.8	0.8	0.8	0.7	0.7
			Fock-Matrix Mixing			
mixing/scaling	75%	80%	85%	90%	95%	
no. iterations	40	28	35	52	94	
σ (monomer population)/10^{-5}	2.4	2.2	2.0	1.9	1.7	
σ (same orb pop)/10^{-5}	2.1	1.9	1.8	1.6	1.6	
σ (other orb pop)/10^{-5}	1.2	1.0	0.9	0.8	0.7	

[a] Outside the given range, the calculations did not converge. The next lines give the mean standard deviations of the populations from the average values of the 15 occupied orbitals in the three-molecule cluster. The data is for the BLYP functional.

Table 2. Decomposition of the Most Diffuse Occupied Orbital of One of the Two Extremities of the Water Chains into the Different Populations[a]
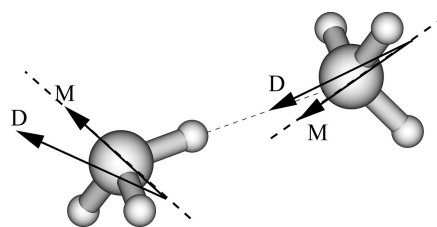
	$(H_2O)_2$	$(H_2O)_3$	$(H_2O)_4$	$(H_2O)_5$
		monomer orbital		
HF	0.9964	0.9962	0.9961	0.9961
LDA	0.9949	0.9945	0.9944	0.9943
PW91	0.9949	0.9946	0.9944	0.9944
BLYP	0.9950	0.9946	0.9945	0.9944
B3LYP	0.9944	0.9941	0.9940	0.9939
		same fragment		
HF	−0.0002	−0.0006	−0.0011	−0.0012
LDA	−0.0092	−0.0093	−0.0099	−0.0100
PW91	−0.0067	−0.0062	−0.0067	−0.0067
BLYP	−0.0050	−0.0044	−0.0047	−0.0048
B3LYP	−0.0173	−0.0166	−0.0169	−0.0170
		other fragments		
HF	0.0038	0.0045	0.0050	0.0052
LDA	0.0143	0.0148	0.0155	0.0157
PW91	0.0117	0.0117	0.0122	0.0124
BLYP	0.0100	0.0098	0.0102	0.0104
B3LYP	0.0229	0.0225	0.0229	0.0231

[a] Note that we treat Hartree−Fock on the same footing as the density-functional results.

Figure 2. Schematic drawing of the deformation of the molecular dipoles via the interaction of two NH_3 molecules. M indicates the monomer dipoles, and D, the dimer dipoles.

ness of the contributions S and O of eq 7. Small differences are observed between different functionals, reflecting the different importance of self-interaction in the functionals. This leads commonly to the slightly different literature values of, for instance, charge transfer terms[41] which we may identify as the other-fragment parts in our decomposition. In the case studied, we see that the B3LYP orbital does not present values intermediate between Hartree−Fock and pure exchange-correlation functionals without an explicit exchange term. Already for interaction energies of the three-molecule system B3LYP (1.15 kcal/mol) is lower than BLYP (2.86 kcal/mol) and Hartree−Fock (6.12 kcal/mol).

Differently than for the atoms-in-molecules (AIM) approach,[46] where atomic basins and their populations are studied, we may look at the density attributable to each molecule through the molecular orbitals. This decomposition scheme, based on orbitals rather than on the density, is extensively studied for instance by Fernández Rico et al.[47,48]

3.3. Local Properties As Dipole Moments. Apart from the question of linear scaling methods, we may employ the so-constructed orbitals for studying weakly interacting complexes by extracting monomer properties and interaction-

induced changes of these. As an example, we may look at the dipole moment of two interacting NH_3 molecules, for which we calculate the influence of the interaction of the size and the direction of the local dipole moments. The global dipole is a well-known quantity, and its measurement leads to an identification of the complex,[49,50] invalidating the common picture of a purely hydrogen-bonded system, with one of the hydrogen atoms lying in the N−N connection line. Nevertheless, we use that geometry for the present case, as we are not aiming at a complete study of this dimer system.

For each molecule, we may calculate a dipole moment within the full dimer basis separately and look for the deformation of the monomer-attributed orbitals when constructing the localized dimer orbitals via the described singles-CI procedure. We have a good trace for the deformation, as the iterations only deform in a minimal sense the starting guess orbitals. Figure 2 shows schematically the reorientation of the two moments toward an alignment. This is expected from basic physics for lowering the total energy in a dipole−dipole interaction. However, the contribution of these calculated dipoles to the interaction energy is rather small, being in the order of microhartrees (Table 3).

The global moment is varying from about 2 D to about 3 D, much too large compared to the experimental estimate of 0.75 D.[49]

3.4. Intermolecular Interactions. Energy decomposition schemes[29,51] furnish two first-order terms, an electrostatic interaction of the monomers E_{es} (or $E_{pol}^{(100)}$ in SAPT) and an exchange-repulsion term $E_{exch-rep}$ (or Pauli repulsion), the sum of which is called frozen-core contribution E_{FC} in the CSOV scheme of Bagus and Illas.[41,51] This E_{FC} is equivalent

Kohn−Sham Orbitals via a Singles-CI Procedure

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2025**

Table 3. Deformation of the Dipoles of the Individual, Separated Monomers (dipole (M)) toward the Attributed Dipole in the Interacting System (dipole (D)), Calculated with Different Functionals[a]

NH$_3$−NH$_3$	dipole (M)	dipole (D)	difference	angle
HF	1.56	1.81	0.25	15.3
	1.56	1.75	0.19	0.1

	total dipole (M)	total dipole (D)	angle (M)	angle (D)
	2.42	3.05	77.8	62.4

	dipole (M)	dipole (D)	difference	angle
LDA	1.45	1.66	0.20	13.5
	1.44	1.88	0.44	1.2

	total dipole (M)	total dipole (D)	angle (M)	angle (D)
	2.25	3.02	77.8	63.0

	dipole (M)	dipole (D)	difference	angle
PW91	1.42	1.62	0.20	14.8
	1.39	1.77	0.39	1.2

	total dipole (M)	total dipole (D)	angle (M)	angle (D)
	2.19	2.92	77.8	61.8

	dipole (M)	dipole (D)	difference	angle
BLYP	1.40	1.62	0.211596	15.5
	1.36	1.72	0.361039	1.2

	total dipole (M)	total dipole (D)	angle (M)	angle (D)
	2.16	2.88	77.3	60.7

	dipole (M)	dipole (D)	difference	angle
B3LYP	1.46	1.68	0.22	15.0
	1.43	1.75	0.32	0.8

	total dipole (M)	total dipole (D)	angle (M)	angle (D)
	2.25	2.95	77.5	61.7

[a] The phrase "total dipole" stands for the vector sum of the two individual dipoles, forming the indicated angle. Units are Debye for the dipole moments, and we give differences in length and orientation. The differences for the two monomer dipoles (M) within a calculation has to be ascribed to a non-negligible BSSE, as the basis sets were elaborated for Hartree−Fock orbitals.

to the Heitler−London energy, employing the unperturbed monomers orbitals, and corresponds to the antisymmetrized Hartree product of the isolated monomer wave functions.

Orbital relaxation toward the multimer orbitals yields the second-order term called orbital interaction,[29,40] corresponding to the sum of polarization and charge transfer in CSOV (named induction in the force-field terminology[33]). In that, scheme relaxation implying virtual orbitals is allowed in this step, in contrast to the first-order, where no mixing between occupied and virtual orbitals is permitted. The subtle question, at which instance a correction for a basis-set superposition error (BSSE) has to be included, is circumvented in our scheme as we calculate monomers in the monomer atomic basis and as well in the multimer atomic basis, and both can be compared to the final HF or DFT solution. Thus, results can be correlated to schemes respecting the BSSE in the very last step like CSOV or to schemes calculating intermolecular interactions in the full multimo-

lecular basis sets throughout like SAPT. The EDA approaches decompose the total interaction energy ΔE as

$$\Delta E = E_{es} + E_{exch-rep} + E_{OI} + E_{BSSE} = E_{FC} + E_{OI} + E_{BSSE} \quad (12)$$

with the terms defined above.

CSOV (using canonical orbitals) was used to provide accurate data for the development of a new generation of force fields based on the electron density.[14] The present approach in fragment-localized molecular orbitals permits to find corresponding terms, which will be detailed in a forthcoming paper. Here, we limit ourselves to the global terms discussed above (details about equivalence of terms can be found in refs 40 and 41).

We tested the accuracy of our approach on a previously investigated linear water dimer configuration.[14,31,32] As expected, we observe the invariance to orbitals localization of our decompositions scheme: localized and canonical orbitals lead to rigorously the same energy contributions, which is not the case for all decomposition schemes (due to projections or approximative orbital rotations).

Table 4 displays the comparison between our approach (fragment-localized orbitals) and CSOV (canonical orbitals) for different functionals, including Hartree−Fock. Differences between the two approaches appear to be around 0.1 kcal/mol and are due to a slightly smaller variational space, 6d or 5d orbitals in the AO basis, and different integration grids The already observed (refs 41 and 52) DFT exchange-repulsion-functional-specific behavior is observed. For example, PW91 exhibits a less repulsive value of the exchange-repulsion contribution compared to HF.

Table 4 shows as well the importance of taking into account the full dimer basis. Indeed the addition of the second monomer basis functions clearly affects the energy components by augmenting both the value of electrostatic and the exchange-repulsion energy. This "BSSE-like" effect is then clearly pronounced for frozen core (or first order in the SAPT terminology). Such a behavior, already observed (see refs 32 and 53) in the framework of SAPT should be carefully taken into account when developing force fields with a precision within this order of magnitude.

The decompositions may be easily extended beyond dimer systems, allowing the calculation of many-body contributions in contrast to SAPT or CSOV calculations, often restricted to the implementation of 2-body terms.[14] Another multi-monomer scheme, the reduced variational space scheme[54] (similar to CSOV) was applied for Hartree−Fock wave functions and may be extended easily to Kohn−Sham wave functions as well.

As it can be seen, once again the specific behavior of each functional for the exchange-repulsion component is confirmed: PW91 < B3LYP < HF < BLYP.

Table 5 gathers results only for the intermolecular contributions on three and five water-molecule chains. Such water chains were initially selected in order to enhance cooperative effects as the molecular dipole moments are parallel to the main axis of the helixlike chain.[33]

Such chains also confirmed the presence of an enhanced charge transfer at the DFT level since E_{OI} is strongly

Table 4. Energy Decomposition (kcal/mol) for the Linear Water Dimer[a]

functional	E_{es}	$E_{exch-rep}$	E_{FC}	$E_{OI+BSSE}$	total interaction
		monomers in the respective monomer basis			
HF (5D)	−8.26	6.88	−1.38	−2.16	−3.55
HF (6D)	−8.27	6.91	−1.36 (− 1.37)	−2.18 (− 2.19)	−3.55
BLYP	−8.00	7.42	−0.58	−3.28	−3.86
BLYP (6D)	−8.03	7.48	−0.55 (− 0.61)	−3.30 (− 3.26)	−3.86
PW91	−7.91	6.22	−1.69	−3.30	−4.99
PW91 (6D)	−7.93	6.27	−1.66 (− 1.50)	−3.33 (− 3.48)	−4.99
B3LYP	−8.05	6.67	−1.38	−2.97	−4.35
B3LYP (6D)	−8.07	6.71	−1.36 (− 1.45)	−2.99 (− 2.95)	−4.35 (4.40)
		monomers in the dimer basis			
HF (5D)	−8.30	7.02	−1.28	−2.27	−3.55
HF (6D)	−8.30	7.02	−1.28	−2.27	−3.55
BLYP	−8.19	7.68	−0.51	−3.35	−3.86
PW91	−8.10	6.52	−1.58	−3.42	−4.99
B3LYP	−8.18	6.87	−1.31	−3.04	−4.35

[a] Geometry and basis set are from ref 59. Results from CSOV are given in parenthesis. The Hartree−Fock results are given for a 5D and 6D AO basis set as CSOV is done in the latter. In the DFT case, our scheme uses 5D throughout. The effect on the different terms is small. For the second part taking 6D or 5D functions makes no difference for all functionals.

Table 5. Energy Decomposition Results (kcal/mol) for Three and Five Water Molecules in the Helix Structure[a]

	functional	E_{es}	$E_{exch-rep}$	E_{OI}	BSSE	total interaction
three water molecules	HF	−39.33	66.14	−21.19	0.51	6.12
	BLYP	−40.85	69.51	−26.88	1.08	2.86
	PW91	−40.03	64.49	−27.15	1.07	−1.62
	B3LYP	−40.24	65.82	−25.29	0.89	1.15
	HF	−80.88	132.08	−45.33	1.03	6.87
five water molecules	BLYP	−83.51	138.89	−57.16	2.19	0.41
	PW91	−81.86	128.74	− 57.75	2.16	−8.71
	B3LYP	−82.42	131.47	−53.86	1.74	−3.07

[a] Note the strongly different results of the total interaction energy for the different functionals.

increased in agreement with previous studies.[41] The (fragment-)localized approach offers as well an energy decomposition per orbital or molecule beyond the global terms.

4. Conclusion

We showed that the singles-CI scheme used for the construction of localized Hartree−Fock orbitals applies as well for the construction of (fragment-)localized Kohn−Sham orbitals. Through the use of canonical monomer orbitals, the localized orbitals of the whole, fragment-based system are very well defined and reproducible, even if an explicit localization criterion is absent in the procedure. Cutting the orbitals in three parts permits to decompose monoelectronic properties in a straightforward manner into monomer, polarization, and charge-transfer parts. This decomposition based on orbital contributions is, as any decomposition scheme of global observables, somehow arbitrary.

For the moment, the implementation gives the exact equivalence to canonical orbitals, but cutoffs, expansion of $\mathbf{S}^{-1/2}$ into a power series, and other selection criteria may be implemented easily, leading to a potentially linear-scaling procedure.

We do not doubt that the monomer-attributed properties (as dipole moments in the present case) will show useful for interpretation and the construction of modern force fields.

Note Added in Proof. During the processing of the manuscript, we had to learn the sad news that Jean-Pierre Daudey passed away in September 2008, without being able to see the final article.

Appendix A: Technical Details

For the calculations, we use our series of experimental programs,[55] which are linked to DALTON,[56] for the generation of integrals, and to Molpro,[57] for the evaluation of the density functional terms. For the CSOV calculation we employed an in-house implementation[41] of the method in HONDO 95.3.[58]

The basis set for the water molecules is a standard one of van Duijneveldt,[59] designed explicitly for the computation of intermolecular interactions in water aggregates.[60] The individual water molecules are in their ground-state geometry[31,32] and are assembled in a helixlike linear arrangement.

For the ammonia molecules, an augmented v. Duijneveldt basis and standard geometry are employed, in the same line as for our previous studies (e.g., ref 40).

Kohn−Sham Orbitals via a Singles-CI Procedure

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2027**

Calculations were carried out on a Pentium IV-based personal computer (local DFT) and on SP5 IBM processors (CSOV).

Appendix B: Equivalence of HF and DFT Singles-CI Procedure

The fact that the construction of Hartree−Fock orbitals through the singles-CI method is useful has been demonstrated several times since its proposition, in particular for periodic systems.[19,37,61] The derivation of the singles-CI equations through the minimization of the Rayleigh quotient with respect to the expansion coefficients in the space of determinant may be found in standard textbooks.[62]

What remains in this section is to show that the minimization of the DFT energy functional with respect to a multideterminantal wave function leads to similar expressions for the matrix elements of the CI matrix. It is of little importance whether we start with determinants or simple Hartree products of orbitals—the resulting densities are the same. For the sake of similarity to the derivation of the usual CI equations, we stick to determinants and the common Slater rules.

Our ansatz for the wave function as linear combination of singly excited determinants leads directly to a CI density as function of three-dimensional coordinates

$$\psi = c_0 \Phi_0 + \sum_I c_I \Phi_I \tag{13}$$

$$\rho(\vec{r}) = N \int \cdots \int d^3 r_1 \cdots d^3 r_{N-1} \psi^\dagger(\vec{r}_1, \cdots, \vec{r}_{N-1}, \vec{r}) \times$$

$$\psi(\vec{r}_1, \cdots, \vec{r}_{N-1}, \vec{r}) = c_0^2 \rho_{\Phi_0}(\vec{r}) +$$

$$\sum_I c_I^2 \rho_{\Phi_I}(\vec{r}) + 2N \sum_{I<J} c_I c_J \int \cdots \int d^3 r_1 \cdots$$

$$d^3 r_{N-1} \Phi_I^\dagger(\vec{r}_1, \cdots, \vec{r}_{N-1}, \vec{r}) \Phi_J(\vec{r}_1, \cdots, \vec{r}_{N-1}, \vec{r}) =$$

$$c_0^2 \rho_{\Phi_0}(\vec{r}) + \sum_I c_I^2 \rho_{\Phi_I}(\vec{r}) + 2 \sum_{I<J} c_I c_J \phi_k^I(\vec{r}) \phi_l^J(\vec{r}) \tag{14}$$

where k and l are the number of the single orbital by which the two determinants Φ_I and Φ_J differ (in the double sum $I < J$, we allow $\Phi_I = \Phi_0$). If there are more orbital differences, no contribution to the density is obtained, as the density operator is a monoelectronic operator. In a second quantization, we may write the density associated with the two determinants as

$$\rho_{IJ}(\vec{r}) = N \phi_k(\vec{r}) \phi_l(\vec{r}) \langle \Phi_I | a_k^\dagger a_l | \Phi_J \rangle$$

as the $(N-1)$-particle determinants $a_l|\Phi_J\rangle$ and $a_k|\Phi_I\rangle$ need to be equal to yield an integral different from zero.

With this expression at hand, we may write up the variation of the density-dependent exchange-correlation functional.[63,64] With the formal matrix element of the exchange-correlation potential $v^{XC}(\vec{r})$ as a one-electron operator $V^{XC} = \Sigma_i v^{XC}(\vec{r}_i)$ acting on determinants, we have

$$\int v^{XC}(r) \rho_I(\vec{r}) \, d^3 r = \langle \Phi_I | V^{XC} | \Phi_I \rangle \tag{15}$$

and

$$\int v^{XC}(r) \phi_K^I(\vec{r}) \phi_l^J(\vec{r}) \, d^3 r = \langle \Phi_I | V^{XC} | \Phi_J \rangle \tag{16}$$

and thus using

$$\frac{\partial \rho(\vec{r})}{\partial c_I} = 2 c_I \rho_{\Phi_I}(\vec{r}) + 2 \sum_{J \neq I} c_J \phi_k^I(\vec{r}) \phi_l^J(\vec{r}) \tag{17}$$

we obtain

$$\frac{\delta E^{XC}[\rho]}{\delta c_I} = \int \underbrace{\frac{\delta E^{XC}[\rho]}{\delta \rho}\Big|_{\vec{r}}}_{v^{XC}(\vec{r})} \frac{\partial \rho(\vec{r})}{\partial c_I} \, d^3 r$$

$$= 2 c_I \langle \Phi_I | V^{XC} | \Phi_I \rangle + 2 \sum_{I \neq J} c_J \langle \Phi_I | V^{XC} | \Phi_J \rangle \tag{18}$$

For the Coulomb matrix elements, the same construction holds: again with a one-electron operator **J**. All terms of the

$$E_J[\rho] = \frac{1}{2} \int \int \frac{\rho(\vec{r}_1) \rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d^3 r_1 d^3 r_2$$

$$\frac{\delta E_J}{\delta c_I} = \int \underbrace{\frac{\delta E_J[\rho]}{\delta \rho}\Big|_{\vec{r}}}_{\int \frac{\rho(\vec{r}_1)}{|\vec{r}-\vec{r}'|} d^3 r' = j(\vec{r}) \text{ and } \mathbf{J}=\Sigma_i \, j(\vec{r}_i)} \frac{\partial \rho(\vec{r})}{\partial c_I} d^3 r$$

$$\frac{\delta E_J[\rho]}{\delta c_I} = 2 c_I \langle \Phi_I | \mathbf{J} | \Phi_I \rangle + 2 \sum_{I \neq J} c_J \langle \Phi_I | \mathbf{J} | \Phi_J \rangle \tag{19}$$

DFT energy functional put together, we arrive at the same CI equations as for a standard Hartree−Fock case, with the monoelectronic Kohn−Sham matrix $\mathscr{K} = T + Z + J + V^{XC}$ instead of the Hamilton operator $H = T + Z + 1/r_{12}$ (with kinetic energy T and electron-nucleus attraction Z):

$$E = \langle \Phi_0 | \mathscr{K} | \Phi_0 \rangle + \sum_I c_I \langle \Phi_0 | \mathscr{K} | \Phi_I \rangle$$

$$c_I E = \langle \Phi_0 | \mathscr{K} | \Phi_I \rangle + c_I \langle \Phi_I | \mathscr{K} | \Phi_I \rangle + \sum_{J \neq 0, I} c_J \langle \Phi_I | \mathscr{K} | \Phi_J \rangle \tag{20}$$

As such, only determinants with one difference in the orbital occupation interact, and no approximation through the neglect of pure bielectronic integrals as for Hamilton matrix elements is necessary.

Thus, in the end exactly the same optimization scheme for the self-consistent orbitals may be employed, for Hartree−Fock or Kohn−Sham orbitals, including hybrid functionals with a nonzero exact-exchange term. This is made possible by the neglect of the pure bielectronic integrals, of which the exchange part is replaced in Kohn−Sham theory by the monoelectronic Kohn−Sham potentials.

References

(1) Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. Linear-Scaling Methods in Quantum Chemistry. *Reviews in Computational Chemistry*; Lipkowitz, K. B., Cundari, T. R., Eds.; Wiley-VCH: New York, 2007; Vol. 23, p 1.

(2) Maslen, P. E.; Ochsenfeld, C.; White, C. A.; Lee, M. S.; Head-Gordon, M. *J. Phys. Chem. A* **1998**, *102*, 2215.

(3) Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 296.

(4) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *34*, 457.

(5) Pipek, J.; Mezey, P. *J. Chem. Phys.* **1989**, *90*, 4916.

(6) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211.

(7) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736.

(8) Marzari, N.; Vanderbilt, D. *Phys. Rev. B* **1997**, *56*, 12847.

(9) Rob, F.; Podeszwa, R.; Szalewicz, K. *Chem. Phys. Lett.* **2007**, *445*, 315.

(10) Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A. *J. Chem. Phys.* **2005**, *123*, 044109.

(11) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Chem. Theo. Comput.* **2007**, *3*, 1960.

(12) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *124*, 204105.

(13) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. *J. Phys. Chem. A* **2007**, *111*, 8753.

(14) Piquemal, J.-P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *124*, 104101.

(15) The idea had been introduced and applied to model Hamiltonians in order to devise a general scheme for obtaining a solution to the closed and open-shell Hartree−Fock equations in a finite basis well before. (a) Lefebvre, R.; Moser, C. M. *J. Chim. Phys. (in French)* **1956**, *53*, 393. (b) Lefebvre, R. Une méthode de détermination des orbitales moléculaires self-consistentes. Thesis, University of Paris, 1959 (in french). (c) Lefebvre, R. *C.R. Acad. Sci.* **1955**, *240*, 1094.

(16) Daudey, J.-P. *Chem. Phys. Lett.* **1974**, *24*, 574.

(17) Gineityte, V. *J. Mol. Struc—THEOCHEM* **2007**, *810*, 91.

(18) Maynau, D.; Evangelisti, S.; Guihery, N.; Calzado, C. J.; Malrieu, J.-P. *J. Chem. Phys.* **2002**, *116*, 10060.

(19) Reinhardt, P.; Malrieu, J.-P.; Povill, A.; Rubio, J *Int. J. Quantum Chem.* **1997**, *70*, 167.

(20) Ángyán, J. G.; Gerber, I. C.; Savin, A. *Toulouse J. Phys. Rev. A* **2005**, *72*, 12510.

(21) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH: Weinheim, Germany, 2002.

(22) Heßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *357*, 464.

(23) Misquitta, A. J.; Podeszwa, B.; Jeziorski, K.; Szalewicz, J. *Chem. Phys.* **2005**, *123*, 214103.

(24) Adams, W. H. *J. Chem. Phys.* **1961**, *34*, 89.

(25) Gilbert, T. L. *J. Chem. Phys.* **1974**, *60*, 3835.

(26) Kunz, A. B. *Theor. Chem. Acta* **1993**, *84*, 353.

(27) Seijo, L.; Barandiaran, Z. *Int. J. Quantum Chem.* **1996**, *60*, 617.

(28) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1755.

(29) Bickelhaupt, F. M.; Baerends, E. J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.: Wiley-VCH: New York, 2000; Vol. 15, pp 1−86.

(30) Bagus, P. S.; Herrmann, K.; Bauschlicher, C. W. *J. Chem. Phys.* **1984**, *80*, 4378.

(31) Tschumper, G. S.; Leininger, M. L.; Hoffman, B. C.; Valeev, E. F.; Quack, M.; Schaefer, H. F., III *J. Chem. Phys.* **2002**, *116*, 690.

(32) van Duijneveldt-van de Rijdt, J. G. C. M.; Mooij, W. T. M.; van Duijneveldt, F. B. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1169.

(33) Piquemal, J.-P.; Chelli, R.; Procacci, P.; Gresh, N. *J. Phys. Chem. A* **2007**, *111*, 8170.

(34) Löwdin, P. O. *J. Chem. Phys.* **1950**, *18*, 365.

(35) Aiken, J. G.; Erdos, J. A.; Goldstein, J. A. *Int. J. Quantum Chem.* **1980**, *18*, 1101.

(36) Daudey, J.-P.; Heully, J.-L.; Malrieu, J.-P. *J. Chem. Phys.* **1993**, *99*, 1240.

(37) Rubio, J.; Povill, A.; Malrieu, J.-P.; Reinhardt, P. *J. Chem. Phys.* **1997**, *107*, 10044.

(38) Morokuma, K. *J. Chem. Phys.* **1971**, *55*, 1236.

(39) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325.

(40) Langlet, J.; Bergès, J.; Reinhardt, P. *J. Mol. Struct.—THEOCHEM* **2004**, *685*, 43.

(41) Piquemal, J.-P.; Marquez, A.; Parisel, O.; Giessner-Prettre, C. *J. Comput. Chem.* **2005**, *26*, 1052.

(42) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(43) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.

(44) Lee, C. L.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(45) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(46) Bader, R. F. W. In *Atoms in Molecules: A Quantum Theory*; Oxford Univ. Press: Oxford, Great Britain, 1990.

(47) Fernández Rico, J.; López, R.; Ema, I.; Ramíre, G. *J. Comput. Chem.* **2004**, *25*, 1347.

(48) Fernández Rico, J.; López, R.; Ramírez, G.; Ema, I.; Ludena, E. V. *J. Comput. Chem.* **2004**, *25*, 1355.

(49) Nelson, D. D.; Klemperer, W.; Fraser, G. T.; Lovas, F. J.; Suenram, R. D. *J. Chem. Phys.* **1985**, *87*, 6364.

(50) Lee, J. S.; Park, S. Y. *J. Chem. Phys.* **2000**, *112*, 230.

(51) Bagus, P. S.; Illas, F. *J. Chem. Phys.* **1992**, *96*, 8962.

(52) Langlet, J.; Bergès, J.; Reinhardt, P. *Chem. Phys. Lett.* **2004**, *396*, 10.

(53) van Duijneveldt-van der Rijdt, J. G. C. M.; van Duijneveldt, F. B. *J. Chem. Phys.* **1992**, *97*, 5019.

(54) Stevens, W. J.; Fink, W. H. *Chem. Phys. Lett.* **1987**, *139*, 15.

(55) Reinhardt, P. *ORTHO, Different Programs to Perform Ab-initio Calculations within Highly Localized Orbitals*, unpublished, 1998−2008.

(56) Helgaker, T.; Jensen, H. J. Aa.; Jørgensen, P.; Olsen, J.; Ruud, K.; Ågren, H.; Andersen, T.; Bak, K. L.; Bakken, V.; Christiansen, O.; Dahle, P.; Dalskov, E. K.; Enevoldsen, T.; Fernandez, B.; Heiberg, H.; Hettema, H.; Jonsson, D.; Kirpekar, S.; Kobayashi, R.; Koch, H.; Mikkelsen, K. V.; Norman, P.; Packer, M. J.; Saue, T.; Taylor, P. R.; Vahtras, O. *DALTON*, version 1.0; University of Oslo, Norway, 1997.

(57) Werner, H. J.; Knowles, P. J. with contributions from Almlöf, J.; Amos, R. D.; Bernhardsson A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Lindh, R.; Lloyd, A. W.; Meyer, W.; Mura, M. E.; Nicklass, A.; Peterson, K.; Pitzer, R.; Pulay, P.; Rauhut, G.; Schütz, M.; Stoll, H.; Stone, A. J.; Taylor, P. R.; Thorsteinsson, T. *Molpro*, version 2002.7; University of Stuttgart and Birmingham: Germany and Great Britain, 2003.

(58) Dupuis, M.; Marquez, A.; Davidson, E. R. *HONDO95.3*; Quantum Chemistry Program Exchange (QCPE), Indiana University: Bloomington, IN, 1995.

(59) van Duijneveldt, F. B. *Gaussian basis sets for the atoms H-Ne for use in molecular calculations*; IBM Technical Research Report No. RJ-945, IBM: San José, CA, 1971.

Kohn−Sham Orbitals via a Singles-CI Procedure

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2029**

(60) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-v.d.Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys.Chem. A* **1999**, *103*, 9872.

(61) Sano, T.; Matsuoka, O. *Bull. Chem. Soc. Jpn.* **1996**, *69*, 2195.

(62) Szabo, A.; Ostlund, N. In *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; MacMillan Publishing Co.: New York, 1982; p 33.

(63) Pople, J. A.; Gill, P. M. W.; Johnson, B. G. *Chem. Phys. Lett.* **1992**, *199*, 557.

(64) Parr, R. G.; Yang, W. In *Density-functional theory of atoms and molecules*; Oxford University Press: Oxford, Great Britain, 1989.

JCTC Journal of Chemical Theory and Computation

Implementation and Performance of DFT-D with Respect to Basis Set and Functional for Study of Dispersion Interactions in Nanoscale Aromatic Hydrocarbons

Roberto Peverati and Kim K Baldridge*

University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich

Received June 27, 2008

Abstract: The implementation, optimization, and performance of various DFT-D schemes have been tested on models for polar−π interactions between arenes spaced at van der Waals distances and on a series of functionalized corannulene derivatives and complexes. For DFT-D schemes involving a semiempirical correction, optimized parameters are proposed for several basis sets. Performance of the different DFT-D strategies is compared, where functionals include some of the most recently proposed, B97D, B2PLYP, BMK, and M06−2X functionals, together with several other well-known functionals. Semiempircally corrected dispersion functionals hold some promise as useful and affordable methods for studies involving large polynuclear aromatic molecules and molecules on metal surfaces.

Introduction

Aromatic carbon nanosystems display subtle dependencies among their structure, dynamics, and photophysical properties that make them challenging to model computationally. The challenges lie in the treatment of electron correlation, dispersion, polarization, and solvation, in a manner appropriate to the context of the application and within the limits of available computational resources. Investigations of polynuclear aromatic hydrocarbons have revealed weaknesses in commonly used computational theories regarding the treatment of delocalization and dispersion effects.[1−10] Delocalization lies at the heart of the concept of aromaticity, and the importance of van der Waals (vdW) interactions comes from the large surface areas of interaction and polarizable electron densities. Dispersion energies are a pure electron correlation effect,[11,12] whereas delocalization involves both short- and long-range effects. Accurate computations via higher-order *ab initio* based wave functions, such as CCSD(T), reach a good level of accuracy with large basis sets[13] but are prohibitively costly for even relatively small systems. Cheaper methods, such as Møller−Plesset perturbation (MP2),[14] still become costly when over 1000 basis functions are considered. In addition, they tend to overes-

timate the electron correlation contribution,[15,16] for example, that stemming from pure $\pi−\pi$ interactions.[17,18]

Density functional theory (DFT) offers the hope of tackling large systems in a broad context with reasonable effort and appears to be less basis set dependent than can be the more advanced wave function methods.[19,20] However, treatment of noncovalent interactions must be addressed on a functional and basis set level, significantly complicating general applicability. Conventional DFT techniques fail to treat dispersion effects completely. Recently, the DFT community has developed a variety of methods for the treatment of van der Waals (dispersion) interactions, including treatments with specialized functionals, such as BMK[21] and the M05[22]/M06[23] series, semiempirically dispersion corrected functionals, such as B97D,[24] and double hybrid functionals, such as B2PLYP.[25] The hierarchy of methods.[24−51] parallels the sophistication of the treatment of the exchange-correlation potential, $E_{xc} = aE_x + bE_{HF} + cE_c$, where the exchange and correlation terms each may be a mixture of terms with different weights. Proper balance of exchange, accounting for antisymmetry caused by the Pauli exclusion principle, and correlation accounting for many-bodied effects are necessary to adequately describe the delocalization phenomenon in these aromatic carbon systems. The sheer size of systems involving corannulene-based molecular recognition chemistry would be prohibitively expensive using conven-

* To whom correspondence should be addressed. Tel: +41 44 635 4201. Fax: +41 44 635 6888. E-mail: kimb@oci.uzh.ch.

tional wave function strategies, and therefore, DFT tends to be the most practical choice. As such, aromatic carbon nanosystems provide a good 'test bed' for implementation and optimization of DFT-based dispersion (DFT-D) models. Our goals in this work include, (a) contribution to the development of semiempirically corrected density functionals, emphasizing functionals and basis sets that are necessary for polynuclear aromatic carbon nanosystems, (b) implementation of several dispersion corrected functionals (via both semiempirical corrections as well as double-hybrid functionals) into our computational chemistry software, GAMESS,[52] and (c) development of a hybrid model for extended materials and corannulene-functionalized metallic surfaces, using GAMESS together with the materials software, SIESTA,[53] the latter of which is more limited in basis set and density functional representation.

Computational Methods

All calculations reported here were carried out using a locally modified version of the GAMESS electronic structure program,[52] and the SIESTA electronic simulations software,[54,55] running on the Baldridge group Linux cluster at UZH, and the latter also running CSCS supercomputer. Here, we consider several classes of DFT functionals including, M06−2X,[23] BMK,[21] several empirically corrected conventional functionals, PBE,[47] revPBE,[56] BP86,[57−59] B3LYP,[60,61] B97D,[24] and the double-hybrid functional B2PLYP.[25] In addition, some comparisons are made with the second-order Møller−Plesset perturbation theory (MP2).[14] In the case of the semiempircally corrected functionals, given the importance and dependence of the associated parameters on the choice of functional and basis set, we have carried out parameter optimization for several basis sets, including TZV2P,[62] with (2d,2p), and Dunning's correlation consistent basis sets,[63] denoted cc-pVnZ, where n = D for double with [3s2p1d] contraction, T for triple with [4s3p2d1f] contraction, and Q for quadruple with [5s4p3d2f1g] contraction. Finally, we have constructed a double-ζ quality (3d,4s,4p) basis set supplemented with extra diffuse (5s) functions for the corannulene on Cu(111) surface computations, as per the localized atomic orbital basis set specifications in SIESTA. The quality of the basis set was checked against surface energy, work function, and interlayer relaxation for the clean Cu(111) surface. The S22 reference set of data provided by Jurecka et al.[64] has been used as a validation test set, in addition to several other small dimers.

Theoretical Approach and Discussion

Empirical Correction. Perhaps motivated by the possibility of a more simplistic approach, several methods for correcting DFT for failures involving noncovalent interactions have involved addition of an empirical correction to the final DFT energy,[42,65] typically of the form $C_6 R^{-6}$, where R represents the interatomic distances and C_6 the dispersion coefficients. Such a strategy has been formulated and well established by several research groups.[24,39,40,66]

We have also implemented such a formulation into our computational software, GAMESS,[52] as well as in the SIESTA software in our collaborative effort, as

$$E_{\text{disp}} = -s_6 \sum_{i=1}^{N_{\text{nat}}-1} \sum_{j=i+1}^{N_{\text{nat}}} \frac{C_6^{ij}}{R_{ij}^6} f_{\text{dmp}}(R_{ij}) \quad (1)$$

with

$$f_{\text{dmp}} = 1/(1 + \exp(-d(r_{ij}/s_R R_{ij}^0) - 1)) \quad (2)$$

This scheme is general and can be applied not only to DFT but also to all mean field methods that lack a sufficient treatment of dispersion energy.[67,68] The main parameters in this scheme are (a) the C_6^{ij} coefficients, (b) the R_0 vdW radii, and (c) all of the scale factors. Differences in implementations are mainly in the values of these parameters. For example, R_0 is typically calculated with wave function based methods and scaled by an appropriate factor; C_6 has been calculated from atomic hybridization states and also determined from LDF calculated IPs and static dipole polarizabilities, α, and the damping factor, d, typically takes on values between 20−23. C_6^{ij} can be determined via formulas such as $C_6^{ij} = \sqrt{(C_6^i C_6^j)}$, which has been found by Grimme to be consistent for problems of interest in this work, and the functional dependent scale factors, s_6, typically determined via parameter fitting. The final mean-field (MF) DFT-D energy is the result of the direct addition of this dispersion energy to the computed DFT energy:

$$E_{\text{MF-D}} = E_{\text{MF}} + E_{\text{disp}} \quad (3)$$

Employing this strategy, three parameters are of concern: the s_6 prefactor to the C_6 atomic coefficients, the d damping factor, and the s_R prefactor to the vdW radii. The nonlinear parameter, d, is considered optimal and is fixed to the optimized value of Grimme, $d = 20$.[24]

In this work, we consider dependencies of both s_6 and s_R parameters simultaneously, with respect to basis sets and common functionals. Parameters for the several functionals are proposed for the first time, including those for the revised-PBE (revPBE) functional,[47,69] which has particular importance for our extensions of GAMESS with SIESTA[53] for computations involving materials.

Double Parameter Optimization. There has been substantial discussion in the recent DFT-D literature concerning the issue of the optimization of the parameters for the semiempircal correction (eq 1).[39,42,70] Many consider a fixed value of s_R and optimize the linear parameter s_6 (as in the original work of Grimme).[24,40] The atomic dispersion coefficient prefactors s_6 were originally optimized by Grimme using the TZV2P (also referred as TZV(2d,2p)) basis set, suggesting a fixed value of $s_R = 1.1$. Optimization of the nonlinear vdW factor s_R, keeping s_6 at one (as in the work of Jurecka), has also been considered,[39] because it has a more complete optimization of all empirical parameters of a damping function for $C_6 R^{-6}$ for three specific functionals (PBE, BP86, B3LYP) and several basis sets for application to π-stacking interactions in nucleic acids.[71] Given the importance of these parameters on the ultimate predictability of DFT-D, as well as the need to add the appropriate dispersion correction term for the specific functional and basis set being used, we have considered a simultaneous double optimization of the two parameters, showing the behavior of both parameters across various functionals and
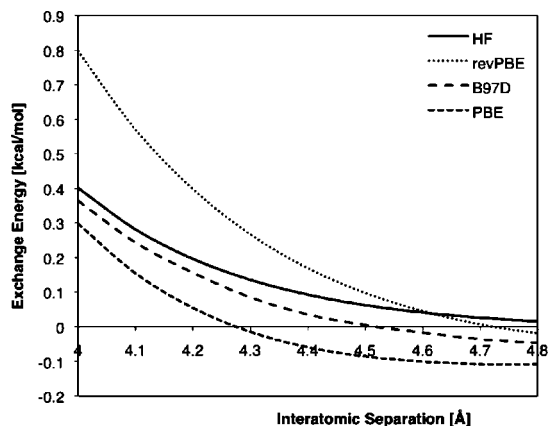
Figure 1. Plot of the exchange energy of the functionals, PBE, revPBE, and B97D, compared to the exact Hartree−Fock exchange in the vdW region, for the Kr_2 dimer.

basis sets. We propose optimal values of both parameters for each case.

The S22 reference set of data provided by Jurecka et al.[64] has been used for determination of the optimized parameters, a common choice for studies of the present type. As in previous works, [24,64] the two parameters were optimized by variation of their values so that the difference between reference single point energy and the DFT-D for the same geometry is minimized. Accurate geometries were taken for the S22 set from the original paper. All DFT-D calculations were performed with the newly modified GAMESS software.

For consideration of double optimization, we have implemented and tested the recently developed B97D exchange-correlation functional,[24] into GAMESS. This functional is a special reparametrization of the original Becke 1997 functional together with a semiempircal correction. Grimme paid particular attention to avoid a so-called "double-counting" especially in large molecules, where many electron correlations at intermediate distances are accounted for both by the correction term and the functional itself. In the B97D functional, power expansion series coefficients of the original functional description were optimized by Grimme to restrict the density functional description to the shorter electron correlation ranges, while the medium to long-range descriptions are handled by the semiempirical correction term. In this case, $a = 1$ (function of 3 parameters), $b = 0$, and $c = 1$ (function of 6 parameters), using the notation presented in the introduction, $E_{xc} = aE_x + bE_{HF} + cE_c$.

B97D tends to be, in fact, much more insensitive to spurious contamination caused by parametrizations in the exchange component, that in some functionals mimics the dispersion effects. This effect can be more clearly observed by plotting the exchange energy of the functionals together with the exact Hartree−Fock exchange in the vdW region for a simple dimer of noble gas atoms, for example Kr, as shown in Figure 1.

The Zhang and Yang functional, revPBE,[56] employs a parametric fit to exact exchange data, and therefore is another candidate for addition of semiempirical dispersion effects. In some cases, revPBE mimics the HF result, however tends to result in more expanded intramolecular geometries and loosely held complexes. While this is partially compensated

for by the addition of the attractive dispersion term, the overall effect is still lacking. In addition, one observes a rather large value for s_6 for revPBE because the correction is taking into account some effects that are typically handled with a correlation terms, which are not present in this functional.
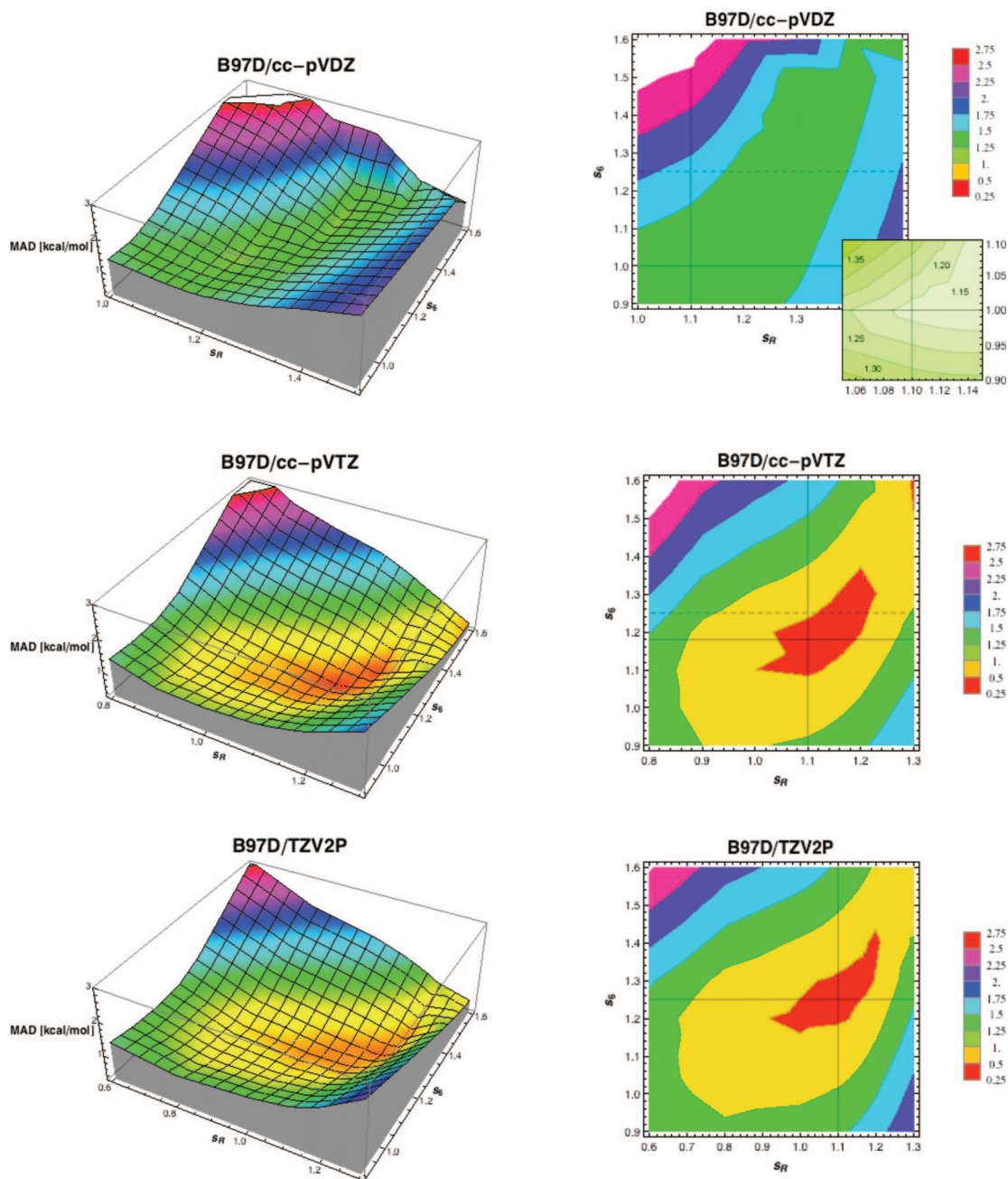
Optimization of the empirical dispersion function parameter across the Dunning cc-pVnZ family of basis sets with $n = 2,3$, and the TZV(2d,2p) basis set used by Grimme in his original paper, was carried out for the B97D functional. Results are shown in Figure 2, where the median absolute deviation (MAD) is reported as a surface function of the two parameters s_R and s_6 for the S22 system of complexes.

In these diagrams, the lowest energy is given in red; the highest energy is in blue, and all plots are on the same relative scale. All plots in Figure 2 show a large region of red corresponding to optimal s_6 and s_R parameters where the MAD associated with the S22 set of complexes is at a minimum. For the cc-pVDZ basis set, the minimum functional region is 1.0 kcal/mol deep, while for both of the triple-ζ basis sets, the minimum functional region is 0.5 kcal/mol. One can find the intersection point between horizontal (s_6) and vertical (s_R) within this minimum region, as shown in the right most picture of each set of plots in Figure 2. A value of $s_R = 1.1$ always intercepts the minimum-regions, showing that optimization of s_R is not necessarily required, if combined with a good optimized value of the linear s_6 parameter. The intersection with an optimized value of s_6 (solid lines in the associated plots) always lies in the minimum-region. The values of s_6, as optimized by Grimme for TZV(2d,2p), are indicated as dashed lines for the first two sets of plots, to indicate the resulting difference when considering other basis sets.

A similar analysis was performed for the revPBE functional. Our particular interest in this functional is its suitability for our hybrid computations of corannulene on surfaces, using the SIESTA software. The revPBE is a good choice for such computations because it is found to be superior in the description of energetics of atomic and molecular bonding to surfaces, as compared to experimental findings. However, up to this point, the attractive dispersion term for this functional was lacking in both SIESTA and GAMESS.

Optimization results for the revPBE functional for one basis set is shown in Figure 3, with very similar characteristics for higher-order basis sets. As expected, the optimized s_6 value is quite large, at 1.66, because the correction is taking into account some effects that are typically handled with correlation terms, which are not present in this functional.

Double Hybrid Functional. In contrast to the strategy of including the effects of dispersion via an empirical function as discussed so far, an alternative strategy proposed in the literature for improving density functional theory even beyond the so-called fourth rung functionals, is via the "double-hybrid" functionals. [72] These functionals, as initially proposed, [73,74] involve a multilevel approach including more advanced *ab initio* components, within the density functional formalism, for example, Møller−Plesset theory. Double-hybrid functionals were then realized by Grimme in his
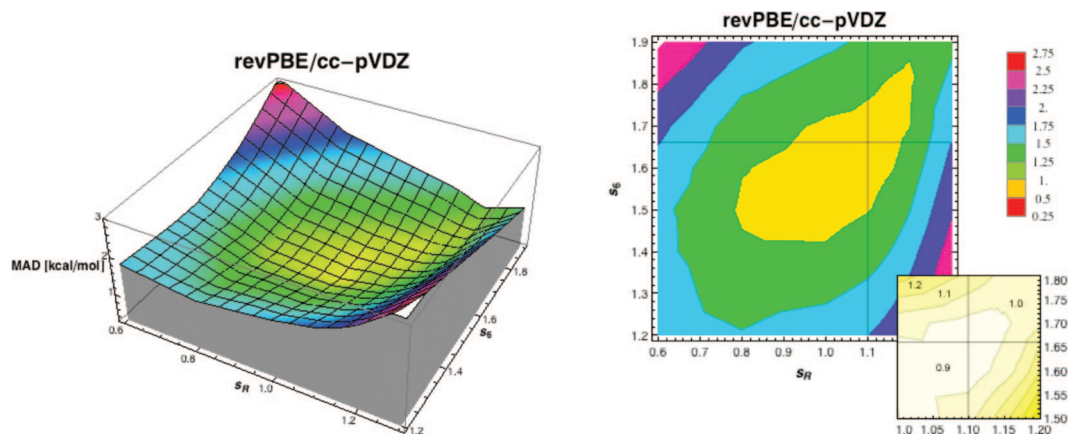
Figure 2. Optimized s_6 and s_R empirical dispersion function parameters for the B97D functional, with the cc-pVDZ, cc-pVTZ, and TZV(2d,2p) basis sets, respectively. All plots are given on the same relative scale, and a more detailed view of the minimum region is provided in the case of cc-pVDZ.

proposed B2PLYP set of functionals,[25,32] and later by others. In this type of treatment, typically either a hybrid density functional or meta-GGA functional is used together with exact HF exchange, but with a damped correlation functional. The remaining correlation is then treated with a more advanced *ab initio* method. In the case of B2PLYP, the meta-GGA functionals are LYP correlation[61] and B88 exchange,[57] with 53% exact HF exchange and 27% MP2 energy.

We have now implemented the B2PLYP functional into GAMESS, and considered optimal s_6 and s_R values for both the cc-pVDZ and cc-pVTZ basis sets. The resulting correlation plots are shown in Figure 4.

We observed results that are quite different from that found for the GGA and hybrid functionals. The s_6

parameter for the B2PLYP functional has a smaller optimized value (0.55), in part because of the dispersion contribution being taken into account with the PT2 component of the functional. However, the values are higher than zero because the PT2 component is scaled and takes into account only 27% of the dispersion. The s_R parameter will have a higher value for similar reasons. The differences found between the double-ζ and triple-ζ basis sets appear to be smaller than in the other functionals considered. In fact, the same pair of optimized s_6/s_R values can be utilized with good accuracy with either basis set. Therefore, we suggest values of $s_6 = 0.55$ and $s_R = 1.3$, which represent the best compromise between accuracy and general applicability. Again, the optimized values are

Figure 3. Optimized s_6 and s_R empirical dispersion function parameters for the revPBE functional, with the cc-pVDZ basis set. Plots are shown using the same relative scale as in Figure 2, and an inset showing a more detailed view of the minimum is provided.
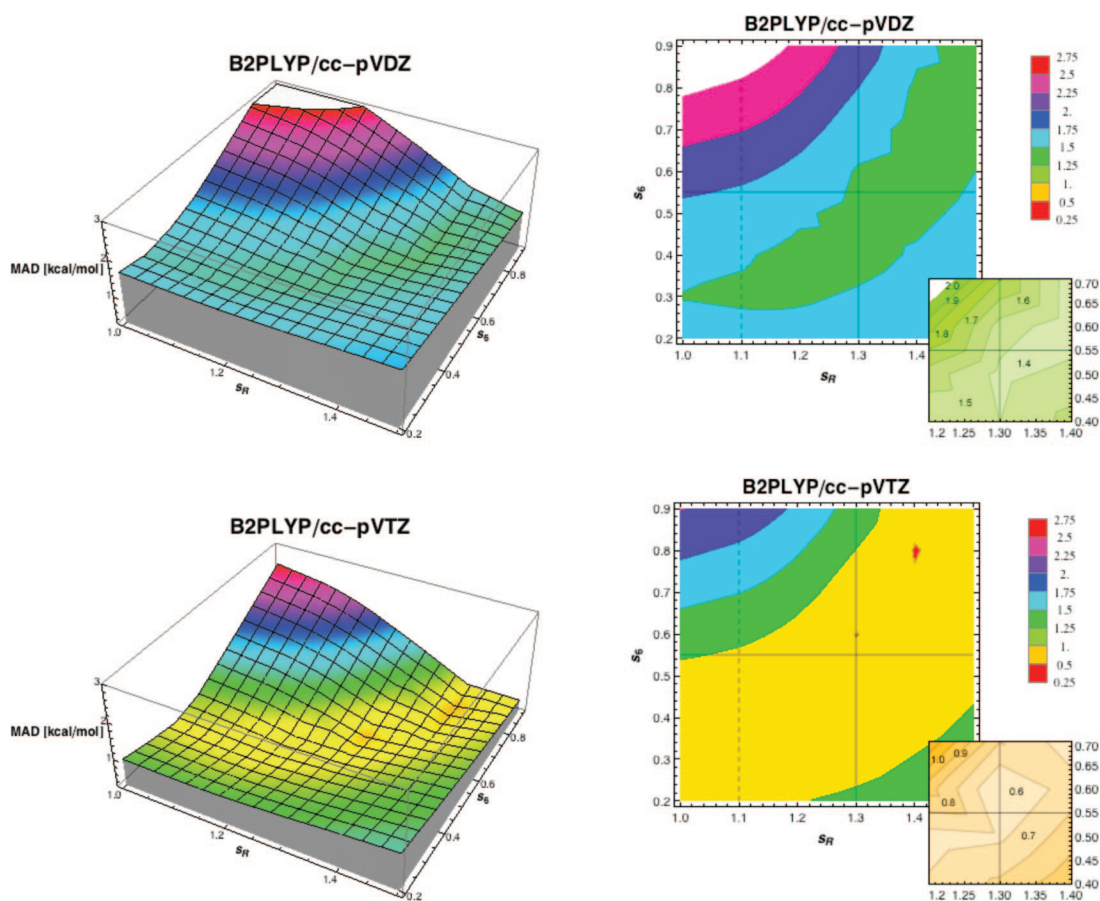
Figure 4. Optimized s_6 and s_R empirical dispersion function parameters for the B2PLYP double-hybrid functional, with the cc-pVDZ (top) and cc-pVTZ (lower) basis sets. Plots are shown using the same relative scale as those shown in Figures 2 and 3. Inset graphs are provided showing the minimum region in more detail.

shown as solid lines and the optimized values of Grimme are shown as dotted lines in the graphics.

Further Considerations of Basis Sets and BSSE. Following the above analysis, we proceeded to optimize the s_6 parameter for several additional functionals and basis sets, keeping the s_R value fixed to 1.1, as also suggested in the original paper of Grimme, given that this choice looks consistent also for the basis sets and functional types considered here. The first functional considered was the B97D exchange-correlation functional.[24] The cc-pVnZ basis

sets (with $n = 2-4$) were used to investigate the behavior of the optimized s_6 parameter for increasing value of $\zeta(n)$. In addition, the reference TZV(2d,2p) optimized parameter of 1.25 was also considered.

Figure 5 shows the median absolute deviation (MAD) for the S22 set of molecules, with respect to increasing s_6 value, using the B97D functional. The optimized s_6 parameter increases with the addition of every $\zeta(n)$ to the basis set, as expected. With small basis sets ($n = 2$), one expects the basis set superposition error, BSSE, to be significant. The triple-ζ

Optimization of DFT-D

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2035**
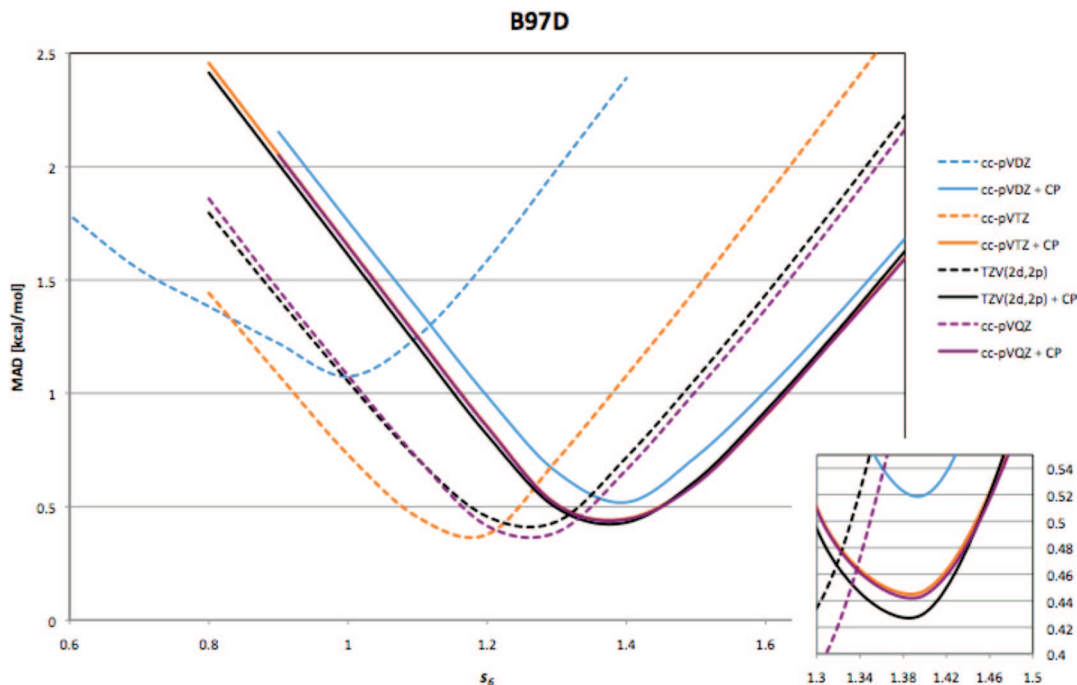
Figure 5. Median absolute deviation (MAD) with respect to increasing s_6 value, for the B97D functional. A detailed view of the minimum region is also provided for the CP corrected curves.

basis sets have almost the same overall performance as the higher-order quadruple-ζ basis set, with MAD lower than 0.5. The cc-pVTZ basis set, with optimized s_6 parameter of 1.18, has a slightly better overall performance than does the TZV(2d,2p) basis set. Indeed, the original TZV(2d,2p)-optimized parameter of 1.25 of Grimme provides acceptable errors of around 0.5 with the cc-pVTZ basis set. This fact suggests that a reoptimization of the s_6 parameters with basis sets that contain the same number of split shells is not really necessary, while the reduction of the number of split shells (i.e., from triple to double-ζ) can involve a significant increase in BSSE that must be compensated by a corresponding reduction of the optimized s_6 parameters.

The basis set superposition error (BSSE) plays a key role in the understanding of weakly bounded complexes interaction energies. It is well-known that basis sets that are too small will have large BSSE, resulting in poor binding energies and intermolecular distances.[40] Although it has been shown that DFT is much less affected by BSSE than other wave function types, a basis set of at least triple-ζ quality is necessary to significantly reduce the BSSE. The counterpoise correction (CP) is the standard method to correct for BSSE, and while the procedure itself has an associated error (typically results in an overestimation of BSSE) and requires additional effort, it typically provides good results for the vast majority of cases where it is used.

When the semiempirical-corrected DFT methodologies are employed, optimized parameters inherently account for part of the BSSE, as pointed out in the literature.[24,71] In fact, for small basis sets, BSSE can be of the same order of magnitude as the dispersion corrections; however the two corrections have different asymptotic behavior. To better understand the behavior of dispersion corrections and BSSE with respect to basis set and functional variations, we have performed parameter optimization with and without the CP correction.

In Figure 5, the optimization of the s_6 parameter for the B97D functional including counterpoise correction is shown for several basis sets. Interestingly, the CP curves for all basis sets considered have a minimum at $s_6 = 1.4$ (general value that we suggest to use when CP corrections are applied with B97D), a quite large value compared to the ones for non CP-corrected curves. Since we can reasonably expect BSSE for large basis (e.g., cc-pVQZ) to be almost completely reduced to zero, one would expect the optimal s_6 value for the CP-corrected curves to be in the same region, but this is not the case.

Looking at the global performance in terms of MAD of the S22 set, one can see that the MAD is correctly reduced on going from cc-pVnZ to cc-pVnZ + CP for $n = 2$, but slightly increases for $n > 2$ (where the BSSE should be very small also for the noncorrected basis). The loss of accuracy within the same basis set has to be attributed to the error associated with the CP corrections, since no other source of errors are introduced in the calculations. This difference should actually be a reasonable energetic estimate of the error associated with the CP procedure within the S22 set of molecules. Given a much larger database of molecules, this procedure should be a way to quantify the global performance of the CP methods, something that will be considered in future work.

Similar considerations can be made for the B3LYP, PBE, and revPBE functionals. Parameter optimizations were performed and results collected in Figures 6−8.

The optimal s_6 parameters show the same behavior for all three functionals, resulting in an increase of this parameter value with dimension of basis set. The revPBE functional actually shows a relatively strange overall performance: its accuracy is in fact reduced with increase in basis set, with the double-ζ basis set having the lowest MAD value. To better analyze the s_6 optimization process,
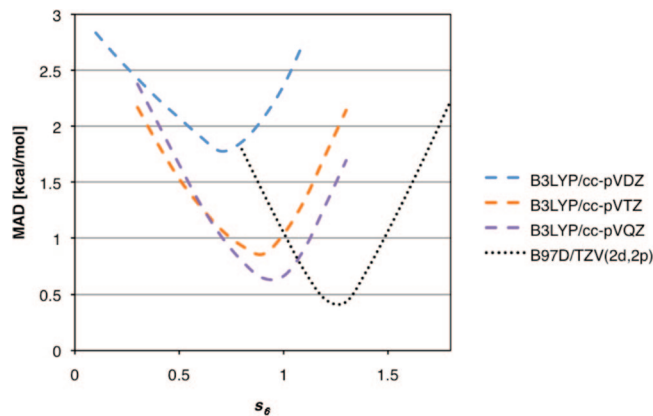
Figure 6. Optimization of the s_6 and s_R parameters for the B3LYP functional at cc-pVnZ, $n = 2$–4, compared to the B97D functional at the TZV(2d,2p) basis set with optimized $s_6 = 1.25$.
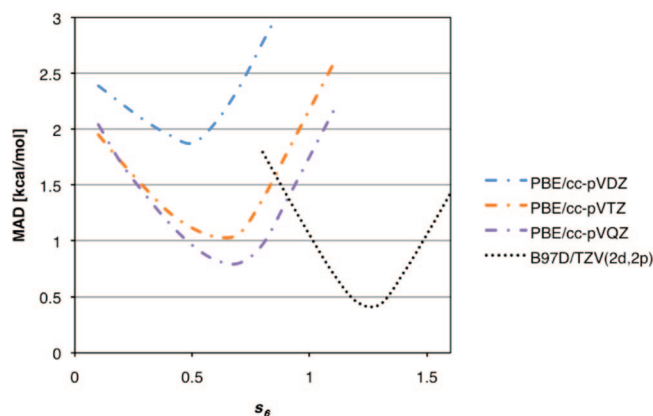
Figure 7. Optimization of the s_6 and s_R parameters for the PBE functional at cc-pVnZ, $n = 2$–4, compared to the B97D functional at the TZV(2d,2p) basis set with optimized $s_6 = 1.25$.

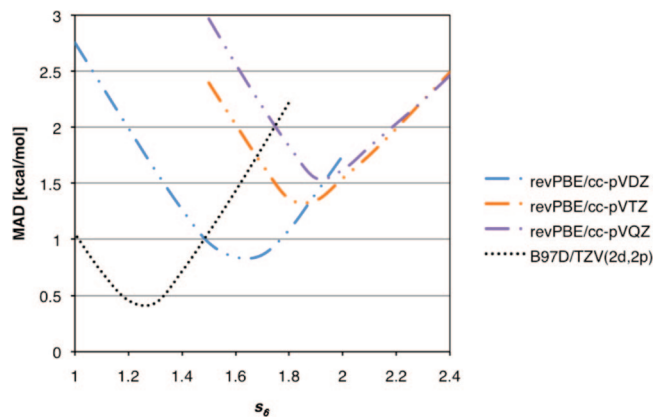**Figure 8.** Optimization of the s_6 and s_R parameters for the rev-PBE functional at cc-pVnZ, $n = 2$–4, compared to the B97D functional at the TZV(2d,2p) basis set with optimized $s_6 = 1.25$.

Figure 9. Deviation of DFT+D calculated and CCSD(T) reference at $s_6 = 1.25$ vs $s_6 = 1.00$ for the S22 set of complexes for B97D/cc-pVDZ.
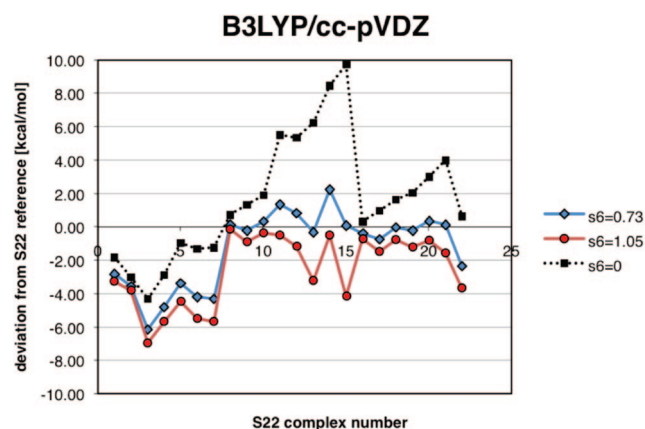
Figure 10. Deviation of DFT+D calculated and CCSD(T) reference at $s_6 = 1.05$ vs $s_6 = 0.73$ for the S22 set of complexes for B3LYP-D/cc-pVDZ.
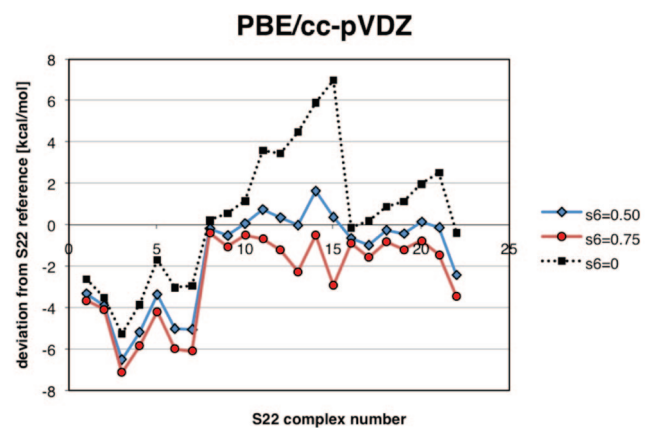
Figure 11. Deviation of DFT+D calculated and CCSD(T) reference at $s_6 = 0.75$ vs $s_6 = 0.5$ for the S22 set of complexes for PBE-D/cc-pVDZ.

the deviation from reference for cc-pVDZ with the nonoptimized and optimized s_6 values for the S22 set of complexes was investigated, in addition to results of DFT without dispersion (Figure 9–13).

As can be expected, one observes the general improvement of results with inclusion of the dispersion correction (DFT vs DFT-D results). The largest errors for DFT-D are found for the first 7 complexes in the S22 set, which are all hydrogen-bonded complexes. Complexes 8–15 have predominantly dispersion contributions, while complexes 16–22 involve more complicated diverse interactions. In the case
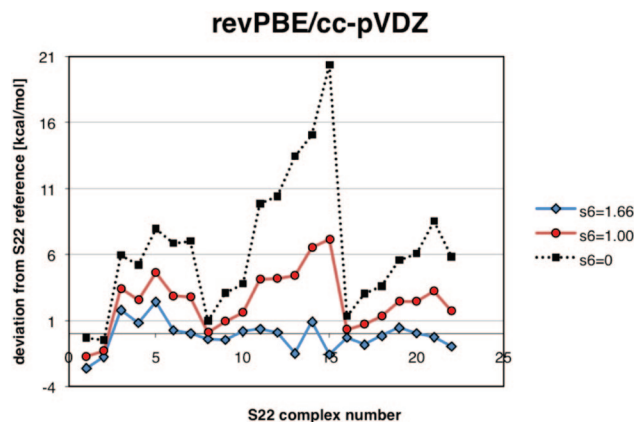
Figure 12. Deviation of DFT+D calculated and CCSD(T) reference at $s_6 = 1.66$ vs $s_6 = 1.00$ for the S22 set of complexes for rev-PBE-D/cc-VDZ.
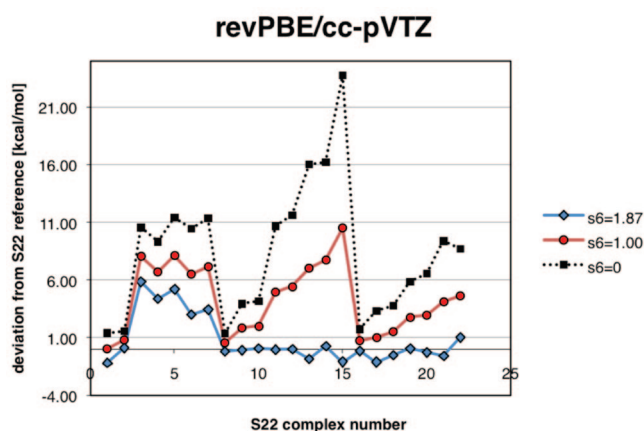
Figure 13. Deviation of DFT+D calculated and CCSD(T) reference at $s_6 = 1.87$ vs $s_6 = 1.00$ for the S22 set of complexes for rev-PBE-D/cc-VTZ vs rev-PBE/cc-pVTZ.

of PBE and revPBE, the optimized values sensibly flatten the curve in the pure dispersion and in the mixed regions, while maintaining acceptable results in the hydrogen-bonded area.

The apparently strange behavior of revPBE with respect to basis set increase can be rationalized through a comparison of the plot in Figure 12 for cc-pVDZ with a corresponding plot for cc-pVTZ shown in Figure 13. While errors for complexes 8−22 are satisfactorily reduced for the triple-ζ basis set, errors associated with the hydrogen bond complexes (complexes 1−7) are actually increased, reducing the global performance (MAD) for this functional. The MAD of complexes from 8−22 using revPBE/cc-pVTZ and revPBE/cc-pVQZ is on the same order of that obtained for B97D using the same respective basis sets, as can be seen in Table 1.

Figure 14 shows all results together, illustrating the performances of the different functionals. The B97D with triple-ζ basis set appears to be in general the best choice for accurate calculations. Interestingly, B97D with cc-pVDZ provides essentially the same performance as B3LYP-D with cc-pVTZ. Oddly, the results of revPBE/cc-pVDZ appear to provide the overall best of the other double-ζ basis sets (Figure 15) and, additionally, of nearly all triple-ζ basis sets and even some of the quadruple-ζ basis sets. However, as

Table 1. Mean Absolute Deviation of S22 Complexes for the Series of Functionals and Corresponding Basis Sets for the Optimized s_6 Values Shown

DFT functional	basis set	optimized s_6 value	MAD (kcal/mol)
B97D	cc-pVDZ	1.00	1.075
	cc-pVDZ + CP	1.39	0.518
	cc-pVTZ	1.18	0.337
	cc-pVTZ + CP	1.41	0.454
	cc-pVQZ	1.26	0.330
	cc-pVQZ + CP	1.39	0.441
	TZV(2d,2p)	1.25	0.375
	TZV(2d,2p) + CP	1.38	0.425
B3LYP	cc-pVDZ	0.73	1.709
	cc-pVTZ	0.88	0.853
	cc-pVQZ	0.96	0.612
PBE	cc-pVDZ	0.50	2.579
	cc-pVTZ	0.64	1.030
	cc-pVQZ	0.65	0.798
revPBE	cc-pVDZ	1.66	0.826
	cc-pVTZ	1.87	1.326
	cc-pVQZ	1.90	1.536
	cc-pVTZ (8−22)	1.87	0.393
	cc-pVQZ (8−22)	1.90	0.355
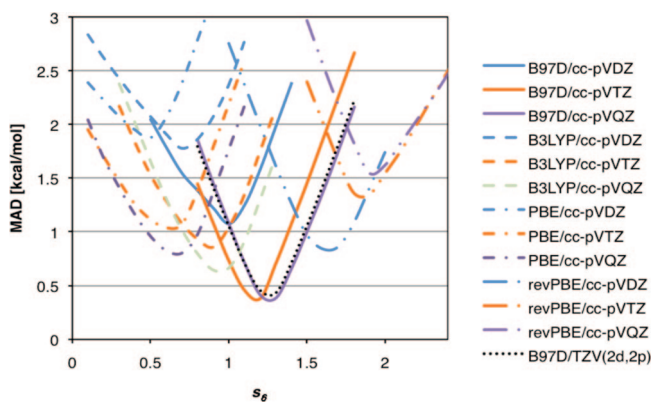
Figure 14. Mean absolute deviation at optimal s_6 values for the functionals and basis sets considered in this study.
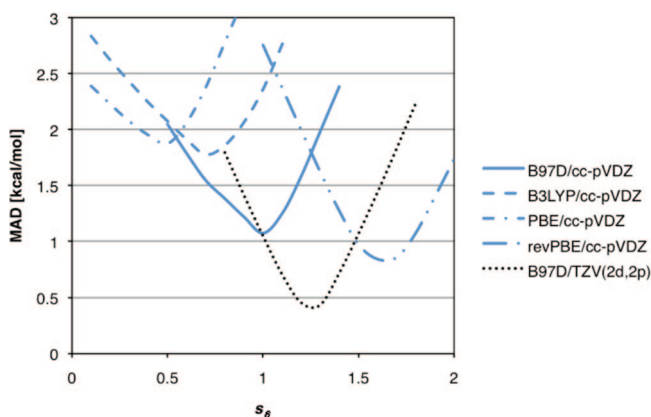
Figure 15. Mean absolute deviation at optimal s_6 values for functionals using the cc-pVDZ basis set as compared to B97D/TZV(2d,2p).

we observe when applying this functional, these results are a bit misleading because results for structure and properties of molecules outside the S22 test set are found to be not as good as with, for example, the B97D functional. The high values of the optimal s_6 for the revPBE functional are

Table 2. Summary of DFT-D Results As a Function of Basis Set and CP Correction for Methane Dimers

		$(CH_4)_2$ D_{3h} (eclipsed)		$(CH_4)_2$ D_{3d} (staggered)	
method	s_6	ΔE (kcal/mol)	R_e (Å)	ΔE (kcal/mol)	R_e (Å)
B97D/TZV(2d,2p)	1.25			−0.567	3.792
B97D/TZV(2d,2p) + CP	1.25			−0.558	3.792
B97D/cc-pVDZ	1.25	−0.743	3.600	−0.782	3.606
B97D/cc-pVDZ + CP	1.25	−0.527	3.600	−0.558	3.606
B97D/cc-pVDZ opt	1.00	−0.521	3.747	−0.548	3.792
B97D/cc-pVDZ + CP opt	1.39	−0.882	3.595	−0.924	3.518
B3LYP-D/cc-pVDZ	1.25	−0.778	3.411	−0.830	3.409
B3LYP-D/cc-pVDZ + CP	1.25	−0.518	3.411	−0.545	3.409
B3LYP/cc-pVDZ		unbound		unbound	
B3LYP/cc-pVDZ + CP					
Grimme[a]				−0.56	3.78
experiment[b]				−0.33 to −0.46	3.84 to 4.27

[a] Ref 24 [b] Experimental range of values from refs 76−81. Counterpoise (CP).[82,83]

evidence that the exchange part of this functional correctly provides little to no bond in the vdW region, forcing the vdW interactions to be treated by the semiempirical dispersion-like term. For this reason and because of its simple GGA implementation and large use in physical and surface science, revPBE/cc-pVDZ appears to be one of the best reference methods for large systems. The B97D functional is still a very good choice using either double-ζ or triple-ζ basis sets, given its rational balance of exchange and correlation terms, and over all good behavior in general. For use with the cc-pVDZ basis set, a sensible reduction of the optimized s_6 parameter is necessary.

To further validate results, we also looked at small molecule benchmark dimers that have been used in many other investigations of this type, some of our larger models for aryl/aryl interactions, as well as corannulene constructs, given the large focus on these latter systems in our research. The first two small but nontrivial tests include methane dimers and benzene dimers, the latter of which has been the subject of many experimental and theoretical investigations for which there is considerable benchmark data. In addition to all of the above-mentioned functionals, for the research applications, we include results from a recent implementation of the M06 series of functionals of Zhao and Truhlar into GAMESS by Sok.[75]

Methane Dimers. $(CH_4)_2$ in D_{3d} (staggered) conformations was first considered using B97D/TZV(2d,2p) to verify in accord with literature[24] and experimental results.[76−81] Counterpoise (CP)[82,83] corrections for basis set superposition error (BSSE) were also investigated for both, the summary of which is in Table 2.

Results for the staggered conformation of the methane dimer are in good agreement with those of Grimme, as well as experimental data. Small differences from that of Grimme can be attributed to (a) Grimme using a resolution of the identity RI approximation for two-electron integrals, and (b) a slight difference in basis set: Grimme uses TZV2P basis set[62] (discarded d- and f- function on first and 2−5-row atoms, respectively); here TZV(2d,2p)[84] as implemented in GAMESS is used.

As previously mentioned, counterpoise corrections need to be treated with caution because such effects are partially accounted for within the semiempirical dispersion correction itself. At the double-ζ, such considerations may still be required, but one should be aware that the semiempircal correction might influence the effectiveness of the CP treatment. One immediately sees that B3LYP+CP is very poor, actually predicting methane dimer to be unbound. In contrast, the addition of the semiempirical dispersion correction overshoots the mark, predicting the methane monomers to be too close and too strongly bound. B97D/cc-pVDZ actually does a comparable job to the larger basis set with the inclusion of CP, albeit again, slightly over bound. Ultimately, the study indicates the need for at least triple-ζ whenever feasible, but that the B97D functional does a reasonable job even for double-ζ.
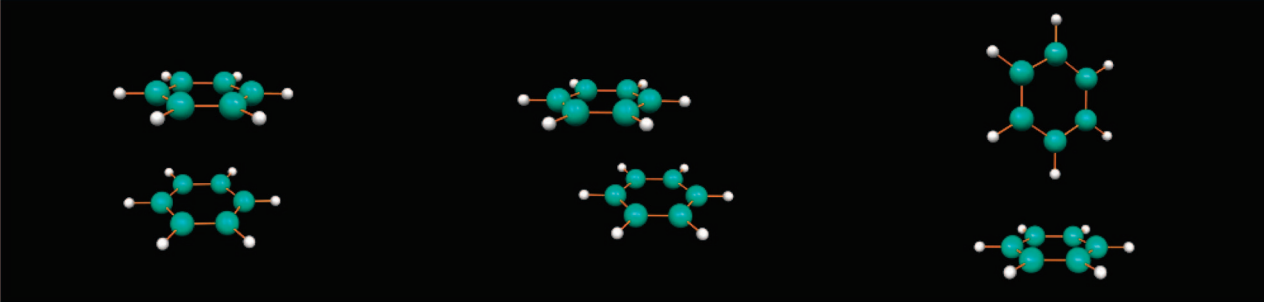
Benzene Dimers. Although benzene dimer structures can adopt a continuum of different relative orientations, three prototype structures are typically considered, parallel-stack (PS), parallel-offset (PO), and edge-to-face (EF) (Table 3). Whereas the PO and EF are stable minima, the PS is a transition state structure. Two important considerations in these dimers is the balance between contact surface (CS) and polar moment (PM). The polar moment overlap in PS is much larger than in the other two cases, resulting in maximum repulsive polar moment. In the PO structure, the two contributions are quite similar and offer moderate attraction between the two benzenes, while the EF structure has minimal attraction with the contact surface compared to the much larger polar moment attraction.

It is well-known that nondispersive DFT functionals, in particular the popularly exploited B3LYP, give a purely repulsive interaction for the PO configuration of the benzene dimer. Our goal in this work is to see how the DFT-D methods discussed above perform for this complex. The best theoretical estimate of the interaction energy for the PS conformation has been provided by Sinnokrot and Sherrill, at the CCSD(T)/aug-cc-pVQZ* (modified basis set) level of theory,[85] enabling us to compare complete dissociation curves.

Here we have investigated the B3LYP-D, B97D, PBE-D, and revPBE-D, with optimized s_6 parameters for the cc-pVDZ basis set. While it is true that one would choose a more substantial basis set for this small system, our goal is to determine feasibility of double-ζ level basis sets for the

Table 3. Relative Relationship between Contact Surface and Polar Moment in Benzene Dimer Structures

| | PS | PO | EF |

conformation	contact surface (CS)	polar moment (PM)	relative relationship
PS	maximum attraction	repulsive	CS<PM
PO	medium attraction	attractive	CS≈PM
EF	minimum attraction	attractive	CS≪PM
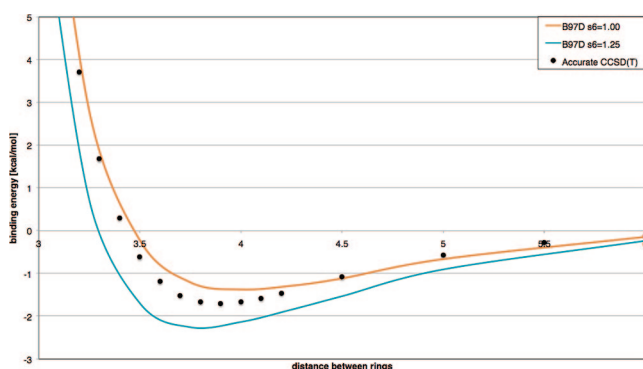
Figure 16. B97D-D/cc-pVDZ energy as a function of intermolecular distance for the benzene face-to-face dimer complex, with $s_6 = 1.0$ and $s_6 = 1.25$, compared to the CCSD(T) reference curve, as explained in the text.

purpose of extending these methods to significantly larger aryl/aryl systems shown below. Full dissociation curves have been computed for each of the four functionals. These dissociation curves, in comparison to the high quality curve of Sherrill and Sinnokrot, provide a better understanding of how the methods appropriately reproduce the strength of the $\pi-\pi$ interactions as a function of distance and orientation of the rings. All curves have been constructed using single-point energy calculations of the dimer at variant intermolecular benzene-benzene distance, using experimental benzene internal geometry. Results are shown Figures 16−19 for the different functionals.

Of the functionals investigated, the B97D functional appears to be one of the better functionals for prediction of the interaction energy profile for the benzene dimer, keeping in mind that the optimization of the s_6 parameter plays a key role when moving from double-ζ to higher-order basis sets. For the popularly used B3LYP functional, we note the anomaly that the nonoptimized curve ($s_6 = 1.06$) has a better behavior than the optimized curve, primarily because of a cancelation of errors. The s_6 optimization for the PBE functional does not have a particularly large influence on single-point energies at experimental equilibrium geometries. Even if the single-point energy looks worse, the minimum
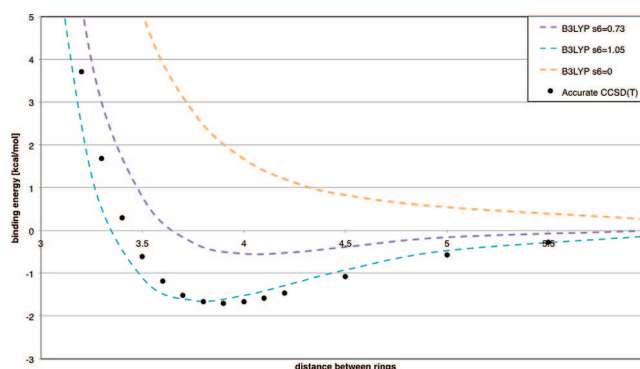
Figure 17. B3LYP-D/cc-pVDZ energy as a function of intermolecular distance for the benzene face-to-face dimer complex, with $s_6 = 1.05$ and $s_6 = 0.73$, and no dispersion correction, compared to the CCSD (T) reference curve, as explained in the text.
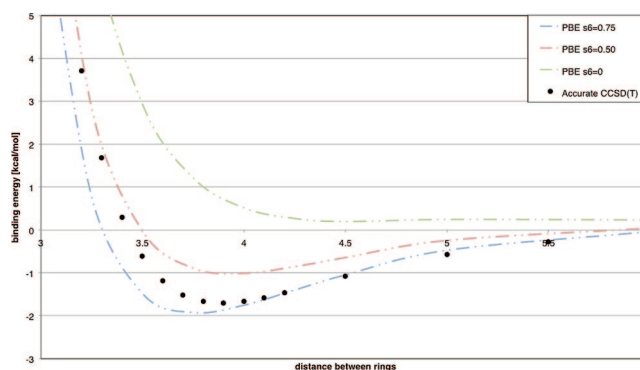
Figure 18. PBE-D/cc-pVDZ energy as a function of intermolecular distance for the benzene face-to-face dimer complex, with $s_6 = 0.75$ and $s_6 = 0.50$, and no dispersion correction, compared to the CCSD(T) reference curve, as explained in the text.

of the optimized curve is nearly predictive of the experimental optimal distance, while the nonoptimized curve predicts overbinding. The two PBE-based functionals have globally acceptable results, with errors reduced from 6 kcal/mol using unoptimized s_6 to almost 2 kcal/mol when used

Figure 19. revPBE-D/cc-pVDZ energy as a function of intermolecular distance for the benzene face-to-face dimer complex, with $s_6 = 1.66$ and $s_6 = 1.0$, and no dispersion correction, compared to the CCSD(T) reference curve, as explained in the text.

with the optimized s_6 parameters. This supports considerations of its general use for significantly large systems.

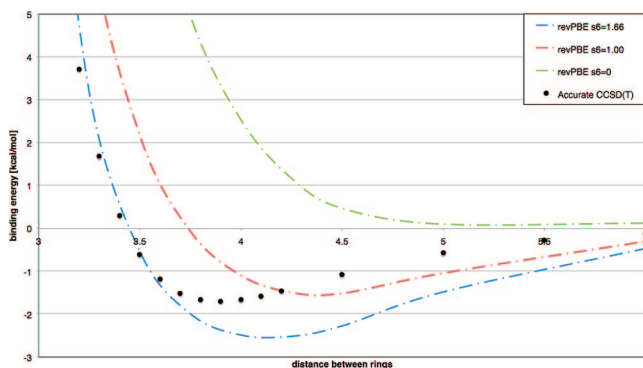The best estimate of the interaction energy for the PS conformation, evaluated at the equilibrium intermonomer distance, is provided by the CCSD(T)/aug-cc-pVQZ* (modified basis set) potential energy curve of Sinnokrot and Sherrill, at -1.70 kcal/mol. For reference, their MP2/aug-cc-pVDZ and CCSD(T)/aug-cc-pVDZ numbers are -2.83 and -1.33 kcal/mol, respectively. The experimental dissociation of the benzene dimer, determined by Grover in 1987, is found to be in the range of -2.0 to -2.8 kcal/mol.[86] Here, as seen above with the methane dimers, the B3LYP-D functional performs quite poorly. However, using the B97D functional, we obtain -1.70 kcal/mol with TZV(2d,2p) and optimized s_6 values of Grimme (-1.559 including CP) and -1.34 kcal/mol using our optimized s_6 values for cc-pVDZ, in very close agreement to that obtained with the CCSD(T)/aug-cc-pVDZ, -1.33.

Research Applications. There is considerable interest, both from an experimental and a theoretical perspective, of alkyl$-\pi$ and aryl$-\pi$ interactions because of their importance in many applications from materials to biological systems. Many such studies predict (a) that dispersive forces dominate the interaction, (b) a repulsive nature of the PS and an attractive nature of the PO and EF, and (c) that electron-donating (ED), as well as electron-withdrawing (EW), groups enhance stacking interactions.[87−91] Experimental evidence for the repulsive interaction in the PS conformation has been demonstrated by Siegel and Cozzi through a number of investigations[92−95] of substituted 1,8-diarylnapthalenes and confirmed by others,[96,97] which aid in the rationalization of a variety of other experimental results in different fields. Experimental evidence for the EF conformation, involving the interaction between the CH groups of the edge ring and the π electron density of the face ring, has been less conclusive, some indicating sensitivity to changes in the local charge distribution on the rings and others finding no particular sensitivity.

Experimental studies, again by Cozzi and Siegel, together with our theoretical contributions, focused on the PO conformation in two conformationally restricted polycyclic systems[98] and showed that through-space interactions be-

tween PO oriented arenes is strongly influenced by electrostatic effects. Therefore, because these same interactions are important in EF and PS structures, polar$-\pi$ interactions are seen to be a decisive factor for understanding arene$-$arene interactions.

In more recent theoretical investigations, together with the experimental groups of Siegel and Cozzi,[99] we proposed a model for systematic investigation of a PO geometry polar$-\pi$ interactions between arenes spaced at vdW distances, enabling us to provide a direct comparison of experiment and theory across different functional group modifications. A set of 1,8-diarylbiphenylene comprising two Hammett series was investigated (Figure 20). The model provides an excellent test to compare the sensitivity of DFT models to pick up appropriate polar$-\pi$ interactions between the two arenes. In particular, compare the following functionals: B97D, BMK, M06$-$2X, and B3LYP. Results are shown for the cc-pVDZ basis set, however, triple-ζ, additional polarization and diffuse functionality were also investigated, with no significant change in the following trends. The following substituents were considered: OMe, H, Cl, and F. Two possible ground-state conformations are possible corresponding to a slight canting of the aryl groups such that the methyl of the probe ring is either endo or exo. All derivatives tend to adopt the endo conformation by $\sim$2$-$5 kJ/mol, consistent with crystallographic forms found for this series.

In the previous work, we found a distinctly better performance of BMK for predicting geometry in the entire series across ten geometric parameters used to describe this interarene interaction. Table 4 reports calculated results across these 10 geometric parameters for the substituents X = OMe, H, Cl, and F, using our B97D implementation. In addition, the barriers to rotation of the probe aryl ring are determined and compared to experimental values that are derived from dynamic NMR data. The critical distances a, b, c, and the interplanar (ip) stacking distance, in particular, show how the different functionals respond to the polarizability and vdW effects.

A regression analysis of experiment versus theory for X = H, Cl, and F shows excellent prediction, with regression coefficients of $R = 0.9988$, $R = 0.9999$, and $R = 0.998$, respectively. The largest errors occur in the case of X = H and X = F, for the parameters ϕ and φ, which are associated with the canting of the rings with respect to each other. All barriers are predicted to be within 1.3 kcal/mol of the reported experimental values.

Results of a comparison across four density functionals for two of the substituted diarylbiphenylenes, X = H and X = Cl, was also carried out (Table 5a). The B3LYP functional shows the largest deviation from experimental values in both cases as expected. The interaction between the aryl rings is underestimated interaction as reflected in the greater distance found between the two aryl rings, and the small barrier to rotation of the probe ring. In the opposite extreme is M06$-$2X, which overestimates the interaction resulting in aryl rings that are too close and in barrier to rotation that is too large. The BMK functional provides very good agreement with experiment. The newest addition, B97D, is very close to the BMK results, the major difference being the amount

Optimization of DFT-D

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2041**
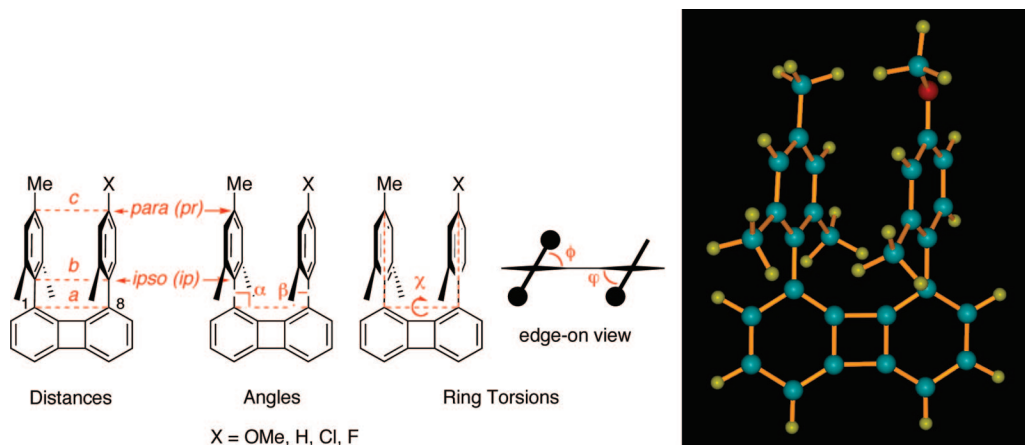
Figure 20. a) peri-Substituted biphenylene as a model for PO benzene dimers. (b) Representative OMe-substituted biphenylene.

Table 4. B97D/cc-pVDZ Structure and Rotational Barrier Results for the Set of 1,8-Diarylbiphenylenes with Substitutents X = OMe, H, Cl, and F

parameter	X = OMe calcd	X = H calcd	X = H exptl	X = Cl calcd	X = Cl exptl	X = F calcd	X = F exptl
a (Å)	3.8168	3.8251	3.80(1)	3.8262	3.773(6)	3.8260	3.797(5)
b (Å)	3.7962	3.8276	3.89(1)	3.8312	3.815(6)	3.8263	3.812(5)
c (Å)	3.6768	3.7442	3.84(1)	3.8532	3.776(6)	3.7146	3.715(5)
α (deg)	89.3	89.7	90.7(4)	89.8	90.3(2)	89.8	90.3(1)
β (deg)	89.7	90.3	92.3(4)	90.3	91.2(3)	90.1	90.3(1)
ϕ (deg)	60.1	62.8	61.0(9)	62.8	62.1(6)	62.7	65.8(3)
ψ (deg)	56.9	59.4	65.1(9)	59.5	59.2(7)	58.6	65.8(3)
χ (deg)	8.1	2.8	3.0(2)	2.9	3.7(1)	2.8	2.4(1)
ip$_\perp$ (Å)[a]	3.42	3.43	3.595(6)	3.40	3.456(4)	3.42	3.544(2)
pr$_\perp$ (Å)[a]	3.46	3.39	3.607(7)	3.46	3.481(4)	3.43	3.490(3)
barrier[b]	11.6	10.9	9.7	10.9	10.1	11.9	10.6

[a] ip$_\perp$ and pr$_\perp$ = orthogonal distance from *ipso*- or *para*-carbon of reference ring to mean plane of probe ring, respectively. [b] Experimental determinations were performed in solvent.

of canting of one ring with respect to each other, and the predicted barriers for the former are slightly higher. The interplanar distances ip$_\perp$ and pr$_\perp$ are predicted to be on average 3.4 for both BMK and B97D, which is what is found experimentally, because this parameter is relatively constant over the series of functional modifications. Figure 21 illustrates this phenomenon in an overlay of all four functional predictions for the X = Cl substituted compound. Here, one can clearly see the difference in association of the two aryl rings using the different functionals. Table 5b shows the deviations from experimental results across the entire set. Across all parameters, the B97D provides a uniform agreement with experimental results for all substituents, providing a high degree of predictability. The largest error is, again, associated with the tilt angle of the two rings with respect to each other.

Corannulene. Corannulene as a shallow bowl structure undergoes a dynamical bowl inversion process. The bowl inversion relates two symmetry-equivalent minima through a common flat structure transition state. Our previous investigations[9] involving systematic study of the differences in bowl-inversion barriers and bowl depth for a variety of substituted corannulene derivatives established a predictive structure–energy correlation, wherein deeper bowl depth gives rise to higher inversion barrier and vice versa. This correlation, $\Delta E = -a(x_{eq})^4$, leads to a quartic relationship

between the inversion energy (ΔE) and equilibrium bowl depth (x_{eq}) and holds for a large subset of corannulene derivatives.

As is well established, the B3LYP/cc-pVDZ level of theory happens to provide highly accurate predictions of structural parameters of corannulene and substituted derivatives as compared to experiment; however, it significantly underestimates the barrier to inversion. Møller–Plesset perturbation theory of order 2 (MP2) carried out at the optimized B3LYP/cc-pVDZ geometry enables accurate predictions of the barrier. This protocol is reliable for many functionalized corannulene derivatives and appears to be only compromised when secondary structural interactions in specific derivatives, such as additional vdW interactions, can occur.

We have investigated the sensitivity of various density functionals with regard to structure and properties of corannulene, including the simple local density approximation (LDA) of SVWN, several GGAs, such as BP86, BLYP, PBE, PBE0, several meta-GGA's and hybrid methods, such as X3LYP, PW91, and most importantly, for the structures with secondary interactions, the series of B97 functionals and the BMK and M06–2X functionals. We have added dispersion corrections to many of these functionals (indicated here with the appended "-D" notation), using optimized scale-factors (Table 6), and tested them out on the corannulene structure and dynamics.

Table 5. (a) Structural and Rotational Barrier Results and (b) Structural and Rotational Deviations from Experimental Results for a Subset of 1,8-Diarylbiphenylenes with Substitutents X = H and Cl

(a)

parameter	X = H				X = Cl			
	B3LYP	M06−2X	BMK	B97D	B3LYP	M06−2X	BMK	B97D
a (Å)	3.8699	3.7948	3.8357	3.8251	3.8637	3.7949	3.8291	3.8262
b (Å)	4.0547	3.7677	3.8915	3.8276	4.0177	3.7620	3.8483	3.8312
c (Å)	4.5661	3.6614	4.0101	3.7442	4.3911	3.6316	3.7800	3.8532
α (deg)	93.3	88.6	90.9	89.7	92.7	89.3	90.4	89.8
β (deg)	93.8	90.3	91.2	90.3	93.2	89.4	90.3	90.3
ϕ (deg)	71.1	63.5	65.4	62.8	71.1	63.7	65.5	62.8
ψ (deg)	68.1	61.1	61.6	59.4	67.8	60.2	60.7	59.5
χ (deg)	0.2	1.4	2.1	2.8	0.6	2.0	1.6	2.9
$ip_\perp$ (Å)	3.76	3.51	3.38	3.43	3.75	3.47	3.37	3.40
$pr_\perp$ (Å)	4.22	3.63	3.30	3.39	4.08	3.40	3.27	3.46
barrier	8.8	12.9	9.8	10.9	8.7	13.4	10.0	10.9

(b)

parameter	3b (X = H)				3d (X = Cl)			
	B3LYP	M06−2X	BMK	B97D	B3LYP	M06−2X	BMK	B97D
a (Å)	0.0689	−0.0062	0.0347	0.0241	0.0901	0.0213	0.0555	0.0526
b (Å)	0.1637	−0.1233	0.0005	−0.0634	0.2021	−0.0536	0.0327	0.0156
c (Å)	0.7251	−0.1796	0.1691	−0.0968	0.6145	−0.1450	0.0034	−0.0766
α (deg)	2.6	−2.1	0.16	−1.0	2.4	−1.0	0.10	−0.50
β (deg)	1.5	−2.0	−1.14	−2.0	2.0	−1.9	−0.94	−0.91
ϕ (deg)	10.0	2.4	4.3	1.7	9.0	1.6	3.3	0.68
ψ (deg)	2.9	−4.1	−3.6	−5.8	8.9	0.9	1.5	0.24
χ (deg)	−2.8	−1.6	−0.92	−0.22	−3.1	−1.8	−2.1	−0.77
$ip_\perp$ (Å)	0.16	−0.08	−0.21	−0.16	0.29	0.01	−0.09	−0.05
$pr_\perp$ (Å)	0.61	0.02	−0.31	−0.22	0.60	−0.08	−0.21	−0.01
barrier	−0.90	3.2	0.10	1.2	−1.3	3.3	−0.10	0.80

The high degree of symmetry of corranulene makes it difficult to obtain a precise experimental value for the barrier
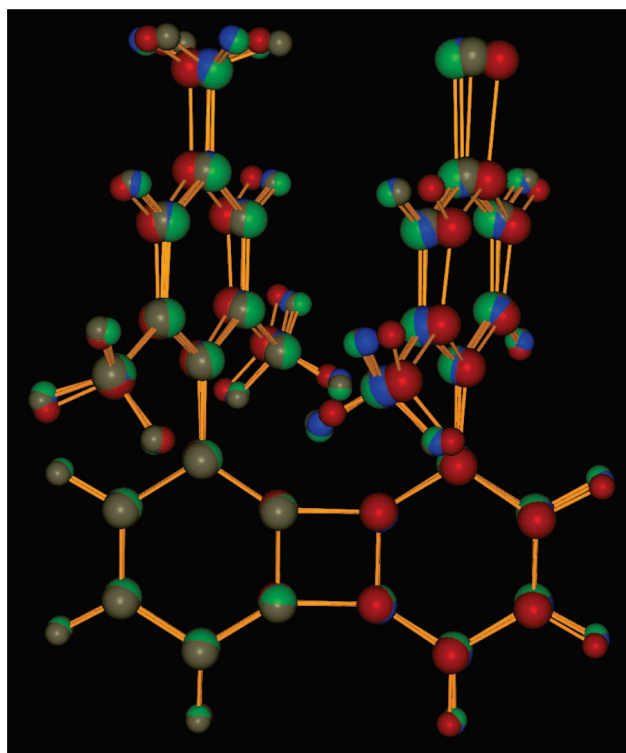
Figure 21. peri-Substituted biphenylene (X = Cl) for functionals B97D (brown), BMK (blue), M06−2X (green), and B3LYP (red).

to the bowl-to-bowl inversion process, but it has been estimated from a series of functionalized structures to be about 11.5 kcal/mol. The best theoretical estimate to date is 11.1 kcal/mol, at MP2/cc-pVDZ//B3LYP/cc-pVDZ, including ZPE. The ability to accurately estimate the barrier is directly related to the ability to accurately predict the structural parameters because of the known structure/energy correlation of bowl depth and inversion barrier. While the bond lengths and angles are quite well reproduced with respect to the experimental predictions, the more difficult geometric feature seems to be the bowl depth and curvature.

As expected, the general functionals can be highly improved using dispersion corrections. The most noticeable difference in the results are a slight increase in bowl depth, resulting in an increase in barrier, as would be predicted from the established bowl/depth functional relationship. The best results are obtained with the B97D functional, which reaches the performance of the MP2 single point calculations for both reaction barrier and dipole moment predictions. Across all functionals, the barrier is still lower than the experimental value, although again, the experimental value is only an estimate because of the high symmetry of the molecule. The BMK functional gives surprisingly good performance across many functionalized corranulenes, as well as this unsubstituted case, likely, because of the high percentage of HF exchange with terms dependent on the kinetic energy density, suitable for describing the aromatic nature. The M06−2X functional provides quite good agreement with respect to the

Optimization of DFT-D

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2043**

Table 6. Barrier to Interconversion, Bowl Depth, and Dipole Predictions for Corannulene Using Various Density Functionals, and the cc-pVDZ Basis Set
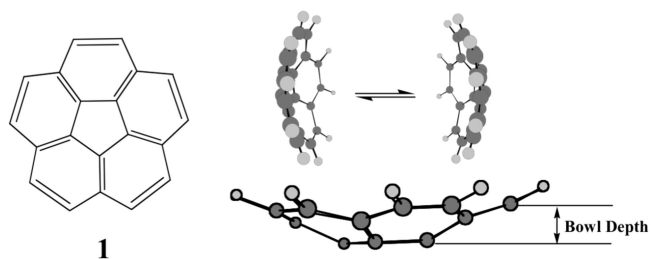
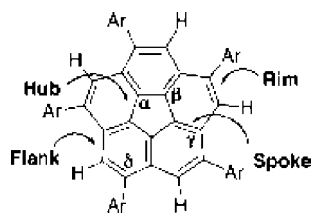| functional	ΔE^a (kcal/mol)	TS frequency (cm^{-1})	bowl depth (Å)	dipole (debye)	$-$(HOMO)^c (eV)
B3LYP	9.33 (9.21)	−109.1	0.87	1.88	6.0 (8.2)
MP2//B3LYP	11.12 (10.99)			2.29	
BLYP	8.74 (8.63)	−105.2	0.88	1.76	5.2 (6.7)
BP86	9.10 (9.09)	−107.9	0.89	2.00	5.6 (7.0)
PBE	9.41 (9.44)	−109.6	0.89	1.97	5.5 (7.0)
SVWN	9.98 (9.71)	−108.3	0.90	2.10	5.7 (7.2)
PBE0	9.50 (9.68)	−112.9	0.88	2.09	6.4 (8.8)
PW91	9.33 (9.31)	−108.9	0.89	1.94	5.5 (7.0)
X3LYP	9.09 (9.13)	−109.5	0.87	1.89	6.2 (8.4)
B97	9.19 (9.29)	−109.7	0.88	2.00	6.1 (8.2)
B97-1	9.25 (9.33)	−109.7	0.88	2.02	6.1 (8.3)
B97-2	9.56 (9.76)	−112.9	0.88	2.06	6.2 (8.4)
B3LYP-D	(9.96)		0.90	1.94	6.0 (8.2)
BLYP-D	(9.67)		0.91	1.82	5.2 (6.7)
BP86-D	(10.05)		0.92	2.07	5.6 (7.1)
PBE-D	(10.19)		0.91	2.00	5.5 (7.9)
revPBE-D opt	(9.55)		0.92	1.67	4.3 (5.8)
B97D	(10.39)		0.92	2.03	5.4 (6.9)
B97D opt	(10.10)		0.91	2.00	5.4 (6.8)
BMK	8.66 (8.84)	−104.0	0.88	2.14	6.9
M06−2X	9.98 (10.01)	−118.2	0.89	2.19	7.5
MP2/cc-pVDZ	9.11	−115.4	0.91	2.45	8.0
experiment	est. 11.5		0.87	2.071(18) ^{b}	8.37^c

^{a} Values in parenthesis are without ZPE. Hessians not yet available for dispersion functionals. ^{b} Ref 100 ^{c} While it has been shown that there is a correlation between HOMO eigenvalue and ionization potential, the absolute estimate is not a reliable measure of I.P. Values in parenthesis have been scaled according to the 3*HOMO - LUMO. Experimental IP is 8.37 eV and MP2/cc-pVDZ Koopmans Theorem=7.99 eV, ΔS CF=8.77 eV.[101]

experiment, although noticeably higher in many of the properties with respect to the other B97D functionals and BMK.

sym-1,3,5,7,9-Pentamanisylcorannulene. Motivated by the deeper bowl depth and higher barrier to inversion, *sym*-1,3,5,7,9-pentamanisylcorannulene has been thoroughly investigated both experimentally as well as computationally. We have previously shown that, without proper treatment of dispersion, one can not adequately predict important structural features or dynamic properties, for example, bowl depth, barrier, and kinetic isotope effects.[1] Further investigations revealed that vdW attractive forces among the peripheral *endo*-methyl groups in this case could contribute to a stabilization of the bowl (ground state) relative to the flat form (transition state), thereby increasing the barrier height. The structure represented a computational challenge not only for DFT methods but also for conventional methods because of its large size.

Geometry optimizations carried out using several functionals are reported in Table 7. While B3LYP/cc-pVDZ quite accurately predicts basic structural parameters, as well as bowl depth and curvature, for essentially all previously investigated corannulene derivatives, the functional comes short for *sym*-pentamanisylcorannulene. Of the bond lengths, the edge bonds (leyward flank and rim bonds) have the largest error compared to experiment, both overestimated by 1 ppm. M06−2X does much better but is still is off by nearly 1 ppm for the C−C hub bond.

The prediction of bowl depth and associated methyl/methyl distances and torsions are more challenging. B3LYP/cc-pVDZ would predict a bowl depth of 0.85 Å, compared to the experimental value of 0.91 Å. Given a structure−correlation formula of $E = ax^4 - bx^2$, this difference in bowl depth would translate into a large underestimation of the barrier to inversion, as is verified in the computation at 8.84 kcal/mol. Correction with the MP2 single point then overshoots

Table 7. Computed Geometry of *sym*-1,3,5,7,9-Pentamanisylcorannulene Compared to Experimental Geometry for a Variety of Density Functionals

| parameter	B3LYP	M06−2X	B97D	B97D opt	revPBE opt	MP2	exptl
C−C hub (Å)	1.418	1.420	1.4264	1.4252	1.4539	1.4245	1.413(4)
C−C spoke (Å)	1.385	1.381	1.3989	1.3975	1.4233	1.4051	1.382(4)
C−C flank (Å)	1.461	1.455	1.4580	1.4601	1.4908	1.4527	1.452(4) leyward
	1.450	1.446	1.4476	1.4496	1.4802	1.4442	1.446(4) wayward
C−C rim (Å)	1.399	1.388	1.4016	1.4033	1.4279	1.4099	1.381(8)
angle α (sym) (deg)	108.0	108.0	108.0	108.0	108.0	108.0	108.0(4)
angle β (sym) (deg)	123.4	122.8	122.5	122.0	122.0	122.4	124.3(4)
angle γ (sym) (deg)	114.8	115.4	115.9	115.6	115.7	115.8	111.9(4)
angle δ (sym) (deg)	120.3	120.6	120.3	120.4	120.3	120.7	122.2(4)
bowl depth (Å)	0.85	0.94	1.02	0.98	1.01	1.01	0.91
$Me_{endo}(1)-Me_{endo}(2)$ (Å)	4.62	4.35	4.24	4.37	4.4	4.2	4.4
$Me_{endo}(1)-Me_{endo}(3)$ (Å)	7.47	7.04	6.86	7.08	7.14	6.76	7.1
Ar−Cor torsion (deg)	79.6	72.2	82.4	83.7	83.4		79
inversion barrier (kcal/mol)	8.84(14.09) [a]	12.65	16.6	14.2	14.2		12.2−12.5

[a] MP2/cc-pVDZ//B3LYP/cc-pVDZ.
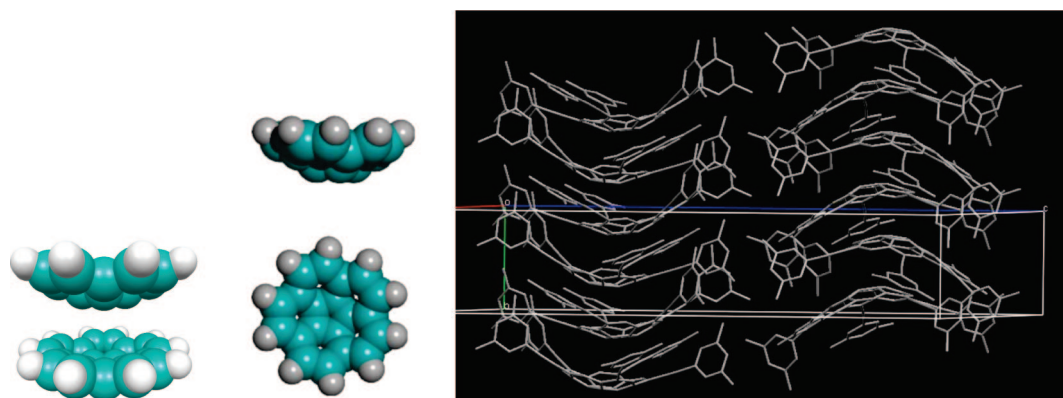
Figure 22. Example of stacking motifs and packing of corannulene and functionalized corannulene.

the barrier by a significant amount at 14.09 kcal/mol, but this can be expected because the depth of the bowl is not correct, and only the dispersion effects at the B3LYP/cc-pVDZ predicted geometry are included. M06−2X predicts the bowl depth to be slightly deeper (0.03 Å) than experimental predictions, and thus the barrier to inversion is also about 1 kcal/mol too high. In all of the semiempircally corrected functionals, B97D, B97D with optimized s_6, and revPBE with optimized s_6, the geometry of the base corannulene nucleus is not as good as with our standard, B3LYP, predicting bonds that are longer than experiment in most cases, although B97D is quite good. The revPBE with optimized s_6 parameters is quite far off in the bond lengths of the basic corannulene nucleus, and overestimates the bowl depth. In terms of bowl depth and barrier prediction, the B97D with optimized s_6 parameter is slightly better than MP2/cc-pVDZ, at a significant reduction in computational cost.

Higher-Order Constructs of Corannulene. The ability to model intercorannulene interactions, as associated with packing in the crystal or complexes of substituted stacked corannulenes, is very much associated with proper modeling of vdW interactions, (Figure 22). For example, interesting packing characteristics have been observed in the experimental crystal structures depending on functionalization, resulting in important materials properties.[4,102−106] Even though corannulene itself is not a $4n + 2$ Hückel system, the electronic structure would suggest a configuration that allows for a 6-electron cyclopentadienyl anion in the center, and a 14-electron cation as the outermost ring, each of which does satisfy a $4n + 2$ cycle ($n = 2$ and $n = 3$, respectively). In addition, experimentally, corannulene behaves like other aromatic compounds, for example undergoing electrophilic substitution reactions. As such, one might expect stability for dimers, trimers, and higher-order stacks, dominated by vdW interactions. We are therefore in the process of considering larger corannulene dimers and stacks with the new dispersion models. As an example, the dimer computation on the PS conformation of corannulene, with B3LYP/cc-pVDZ and B97D/cc-pVDZ, predict interstack distances
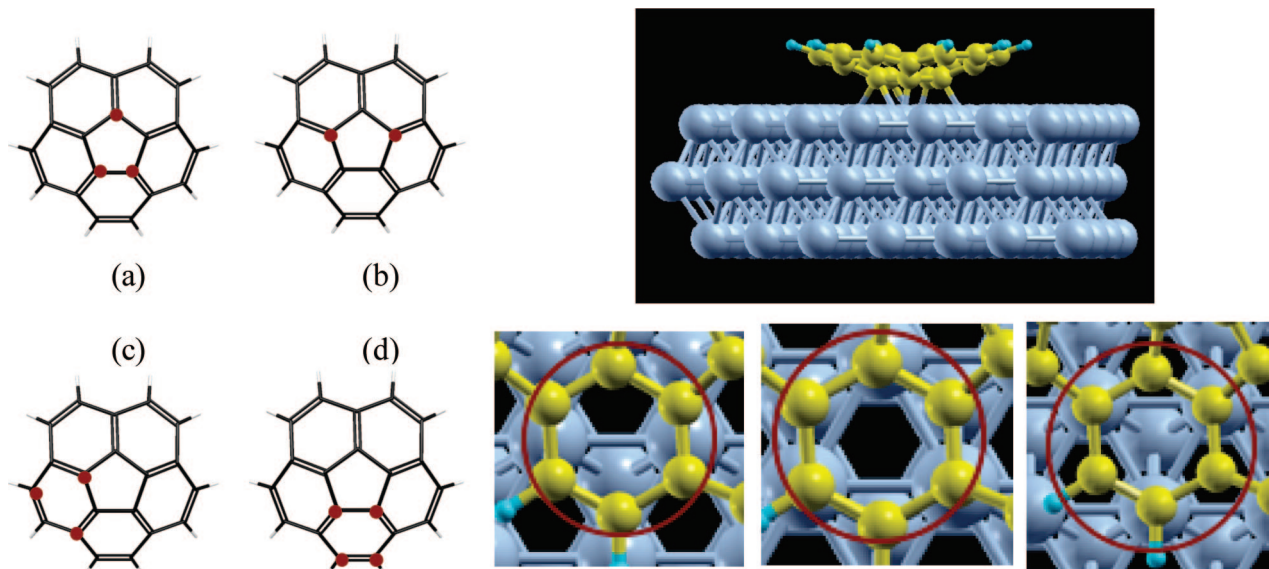
Figure 23. (Left) Possible binding motifs of corannulene on Cu(111): (a) η^5-hollow site, (b) η^5-bridge site, (c) η^6-hollow site, (d) η^6-bridge site. (Right, top) One conformation of corannulene on Cu(111) surface. (Right, bottom) hcp hollow site, bridge site, and fcc-hollow site.

of 4.63 and 3.55 Å, respectively. Because of the much larger surface area of corannulene over benzene, we see a much larger effect of dispersion, further illustrating the inadequacy of B3LYP. In addition, only for the B97D is the stack predicted to be bound, by 6.23 kcal/mol, significantly more than found for the benzene dimer. Further investigations of stacks and functionalized stacks are now ongoing.

Alternatively, it is of interest to investigate binding characteristics of metal−organic complexes composed of (substituted) corannulenes with a planar metal surface. [107,108] Such studies enable investigation of molecular self-organization processes of adsorbate molecules on the surface that can be responsible for formation of supramolecular aggregates. Two-dimensional assembly into supramolecular structures is of great interest for study of fundamental processes, such as molecular[109,110] and chiral recognition.[111−116] Moreover, functionalization of surfaces by adsorption of large organic molecules appears to be an important route toward new materials or templates for heterogeneous catalysis[117] and molecular electronics. [118]

Attempts to incorporate solutions to failures involving nonlocal, long-range electron correlation directly into a density functional can result in rather complicated formulations, leading to expressions at least as complex as some wave function methods. Examples include the vdW-DFT method that has just recently been implemented in the SIESTA software[54,55] and the first principles approach suggested by Kohn,[119] which breaks the Coulomb interaction into a short and long-range (adiabatic connection formula) component of which only the long-range contributes to the vdW energies. Ongoing work with the SIESTA group involves further testing the vdW-DFT method in the SIESTA software and considerations of practicality for the ongoing corannulene on surface computations.

Motivated by the work involving the semiempircally corrected DFT fucntionals, we have in the meantime proceeded to implement a protocol for investigations involv-

Table 8. Computed Geometry of Corannulene Conformations on Cu(111) Surface, Comparing the revPBE Functional with and without Empirical Dispersion Correction

parameter[a]	revPBE	revPBE-D
η^6-bridge fcc site	2.46 Å	2.23 Å
η^6-hollow fcc site	2.37 Å	2.16 Å
η^6-hollow hcp site	2.38 Å	2.18 Å
η^5-bridge site	detaches	
η^5-hollow hcp site	2.5 Å	2.36 Å

[a] Average C−C bond length for ring.

ing complex Cu(111) + corannulene surface interactions, using GAMESS, together with the surface modeling software, SIESTA, the latter of which also now has the semiempirical dispersion correction implemented. We have tested this method out on various conformational possibilities of corannulene on a 3 layer Cu(111) surface, as depicted in Figure 23.

While results suggest that functionals, such as B3LYP, revPBE, or BLYP, all predict the corannulene + Cu(111) to be repulsive or only weakly bound, depending on the particular conformation of corannulene on the surface, the semiempirical correction enables more realistic modeling of the complexation process (Table 8). Using a double-ζ quality basis set supplemented with extra diffuse functions for the copper surface, we are able to perform computation of corannulene in several different conformational possibilities on a 3-layer slab of Cu(111) plus corannulene. These preliminary results show promise for realistic modeling on a metallic surface, as compared to known STM results on the same system, enabling a more detailed research investigation including an increase to a 6-layer slab, which will be the subject of a future publication.

Conclusions

At the onset of this work, our major concerns were to implement more optimal DFT functionals for the purpose

Table 9. Summary of Density Functional Plus s_R/s_6 Coefficient Combinations Proposed for a Variety of Basis Sets, As Determined from Predictions of S22 Complexes

DFT functional	basis set	s_R value	optimized s_6 value	MAD (kcal/mol)
B97D	cc-pVDZ	1.1	1.00	1.075
	cc-pVDZ+CP	1.1	1.39	0.518
	cc-pVTZ	1.1	1.18	0.337
	cc-pVTZ+CP	1.1	1.41	0.454
	cc-pVQZ	1.1	1.26	0.330
	cc-pVQZ+CP	1.1	1.39	0.441
	TZV(2d,2p)	1.1	1.25	0.375
	TZV(2d,2p)+CP	1.1	1.38	0.425
B3LYP	cc-pVDZ	1.1	0.73	1.709
	cc-pVTZ	1.1	0.88	0.853
	cc-pVQZ	1.1	0.96	0.612
PBE	cc-pVDZ	1.1	0.50	2.579
	cc-pVTZ	1.1	0.64	1.030
	cc-pVQZ	1.1	0.65	0.798
revPBE	cc-pVDZ	1.1	1.66	0.826
	cc-pVTZ	1.1	1.87	1.326
	cc-pVQZ	1.1	1.90	1.536
	cc-pVTZ (8−22)	1.1	1.87	0.393
	cc-pVQZ (8−22)	1.1	1.90	0.355
B2PLYP	cc-pVDZ	1.3	1.55	1.394
	cc-pVTZ	1.3	1.55	0.517

of accurate predictions of large molecular polynuclear carbon-based complexes and corannulene on copper surfaces, all of which have dominant dispersion interactions. We have implemented several additional functionals into GAMESS, together with a semiempirical dispersion correction that can be applied with these and other functionals within GAMESS. We have also implemented and tested the B2PLYP double-hybrid functional. For functionals with a semiempircal correction, we have carried out a double parameter optimization to understand the variation in the s_R and s_6 coefficients with respect to basis set, functional type, and BSSE, and have proposed optimal values, a summary of which is shown in Table 9. We find overall good performance in particular of the B97D functional. In addition, we find a convenient convergence of optimal s_6 parameter of 1.4 when BSSE is considered for the S22 set of molecules using any of the series cc-pVnZ, $n = 2−4$, as well as TZV2P, a suggested value to use when CP corrections are applied with B97D, even though it is a relatively large value compared to non-CP-corrected values. The global performance of B97D+CP in terms of MAD of the S22 set shows a slight loss of accuracy within the same basis sets, which is attributed to the error associated with the CP corrections, providing a reasonable estimate of the error associated with CP within the S22 set of molecules. Future studies will look at this phenomenon more globally.

Comparative results across functionals were given for several applications with high degree of dispersive interactions, illustrating the strengths and weaknesses in these functionals. Consistently reliable results are found for several of the functional strategies, including the exchange-correlation functional provided by BMK, the M06−2X functional, and the semiempirically corrected B97D functional. The double hybrid functional, B2PLYP functional is also seen to provide very good performance, but with an associated higher cost for the MP2 evaluation.

A second purpose of this work was to create a reliable protocol for investigations of polynuclear aromatic carbon species, such as corannulene, on metallic surface. Our investigations here were initially limited to the PBE and revised-PBE functionals in the SIESTA software, a linear-scaling density functional package designed for studies involving materials. Including the semiempirical dispersion correction was seen to greatly enhance predictability of the functionals, as demonstrated in the GAMESS/SIESTA hybrid method for corannulene on Cu(111) surface results, particularly considering other strategies are too costly. Our initial investigations are quite promising, and with the results obtained in our investigations shown here, we are now actively pursuing these new capabilities with a larger Cu(111) surface.

References

(1) Hayama, T.; Baldridge, K. K.; Wu, Y.-T.; Linden, A.; Siegel, J. S. *J. Am. Chem. Soc.* **2008**, *130*, 1583.

(2) Petrukhina, M. A. *Angew. Chem., Int. Ed. Engl.* **2008**, *47*, 1550.

(3) Tiwary, A. S.; Mukherjee, A. K. *THEOCHEM* **2008**, *859*, 107.

(4) Wu, Y. T.; Bandera, D.; Maag, R.; Linden, A.; Baldridge, K. K.; Siegel, J. S. *J. Am. Chem. Soc.* **2008**, *120*, 1.

(5) Betowski, L. D.; Enlow, M.; Riddick, L.; Aue, D. H. **2007**, *111*, 3672.

(6) Mack, J.; Vogel, P.; Jones, D.; Kaval, N.; Sutton, A. *Org. Biomol. Chem.* **2007**, *5*, 2448.

(7) Hayama, T.; Wu, Y.-T.; Linden, A. L.; Baldridge, K. K.; Siegel, J. S. *J. Am. Chem. Soc.* **2007**, *129*, 12612.

(8) Yamaji, M.; Teakehira, K.; Mikoshiba, T.; Sachiko, T.; Okada, Y.; Fujitsuka, M.; Majima, T.; Tobita, S.; Nishimura, J. *Chem. Phys. Lett.* **2006**, *425*, 53.

(9) Seiders, T. J.; Baldridge, K. K.; Grube, G. H.; Siegel, J. S. *J. Am. Chem. Soc.* **2001**, *123*, 517.

(10) Seiders, T. J.; Baldridge, K. K.; Siegel, J. S. *J. Am. Chem. Soc.* **1999**, *121*, 7439.

(11) Kutzelnigg, W. Pair correlation theories.In *Modern Theoretical Chemistry*; Schaefer, H. F. ,Ed.; Plenum Press: New York, 1977; pp 129.

(12) Kutzelnigg, W. *Einführung in die Theoretische Chemie*; Verlag Chemie : Weinheim, Germany, 1978; Vol. 2

(13) Helgaker, T. ; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; J. Wiley: New York, 2000.

(14) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(15) Bates, D. M.; erson, J. A.; Oloyede, P.; Tschumper, G. S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2775.

(16) Bachorz, R. A.; Bischoff, F. A.; Hoefener, S.; Klopper, W.; Ottiger, P.; Leist, R.; Frey, J. A.; Leutwyler, S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2758.

(17) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210.

(18) Nishio, M. *CrystEngComm* **2004**, *6*, 130.

(19) Bauschlicher, C. W. , Jr.; Ricca, A.; Partridge, H.; Langhoff, S. R. Chemistry by density functional theory. In *Recent Advances in Density Functional Methods, Part II*; Chang, D. P. , Ed.; World Scientific: Singapore, 1997.

(20) Martin, J. M. L. Some observations and case studies on basis set convergence in density functional theory. In *Density Functional Theory: A Bridge between Chemistry and Physics*; Geerlings, P., De Proft, F., Langenaeker, W. , Eds.; VUB Press: Brussels, Belgium, 2000.

(21) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.

(22) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103/1.

(23) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

(24) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.

(25) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.

(26) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(27) Benighaus, T.; DiStasio, R. A.; Lochan, R. C.; Chai, J.-D.; Head-Gordon, M. *J. Phys. Chem. A.* **2008**, *112*, 27002.

(28) Hill, J. G.; Platts, J. A.; Werner, H.-J. *Phys. Chem. Chem. Phys.* **2008**, *8*, 4072.

(29) Silvestrelli, P. L. *Phys. Rev. Lett.* **2008**, *100*, 053002.

(30) Tarnopolsky, A.; Karton, A.; Sertchook, R.; Vuzman, D.; Martin, J. M. L. *Phys. Chem. A* **2008**, *112*, 3.

(31) Riley, K. E.; Vondrasek, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5555.

(32) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397.

(33) Grimme, S.; Antony, J.; Schwabe, T.; Mück-Lichtenfeld, C. *Org. Biomol. Chem.* **2007**, *5*, 741.

(34) Grimme, S.; Mück-Lichtenfeld, C.; Antony, J. *J. Phys. Chem. C* **2007**, *111*, 11199.

(35) Neese, F.; Schwabe, T.; Grimme, S. *J. Chem. Phys.* **2007**, *126*, 124115.

(36) Grimme, S.; Neese, F. *J. Chem. Phys.* **2007**, *127*, 154116.

(37) Grimme, S.; Steinmetz, M.; Korth, M. *J. Chem. Theor. Comput.* **2007**, *3*, 42.

(38) Zhao, Y.; Truhlar, D. G. *J. Chem. Theor. Comput.* **2006**, *2*, 1009.

(39) Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2006**, *28*, 555.

(40) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.

(41) Lilienfeld, O. A.; Tavernelli, L.; Rothlisberger, U. *Phys. Rev. Lett.* **2004**, *93*, 153004.

(42) Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *116*, 515.

(43) Wu, X.; Vargas, M. C.; Nayak, S.; Lotrich, V.; Scoles, G. *J. Chem. Phys.* **2001**, *116*, 8748.

(44) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijnevelt-van de Rijdt, J. G.C.M.; van Eijck, B. P *J. Phys. Chem. A.* **1999**, *103*, 9872.

(45) Becke, A. D. *J. Chem. Phys.* **1997**, *107*, 8554.

(46) Meijer, E. J.; Sprik, M. *J. Chem. Phys.* **1996**, *105*, 8684.

(47) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(48) Hobza, P.; Zahradnik, R. *Chem. Rev.* **1988**, *88*, 871.

(49) Kristyan, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175.

(50) Hobza, P.; Sponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315.

(51) Pérez-Jordá, J. M.; Becke, A. D. *Chem. Phys. Lett.* **1995**, *233*, 134.

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

(53) Soler, J. M.; Artacho, E.; Gale, J. D.; Garcia, A.; Junquera, J.; Ordejón, P.; Sánchez-Portal, D. *J. Phys.: Condens. Matter* **2002**, *14*, 2745.

(54) Dion, M.; Rydberg, H.; Schroeder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246.

(55) Langreth, D. C.; Dion, M.; Rydberg, H.; Schroeder, E.; Hyldgaard, P.; Lundqvist, B. I *Int. J. Quantum Chem.* **2004**, *101*, 599.

(56) Zhang, Y.; Yang, W. *Phys. Rev. Lett.* **1998**, *80*, 890.

(57) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(58) Perdew, J. P. *Phys. Rev. B.* **1986**, *33*, 8822.

(59) Perdew, J. P. *Phys. Rev. B.* **1986**, *23*, 7046.

(60) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(61) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785.

(62) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.

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

(64) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.

(65) Zimmerli, U.; Parrinello, M.; Koumotsakos, P. *J. Chem. Phys.* **2004**, *120*, 2693.

(66) Kubar, T.; Jurecka, P.; Cerny, J.; Rezáe, J.; Otyepka, M.; Valdés, H.; Hobza, P. *J. Phys. Chem. A* **2007**, *111*, 5642.

(67) Ahlrichs, R.; Penco, R.; Scoles, G. *Chem. Phys.* **1977**, *19*, 119.

(68) Hepburn, J.; Scoles, G. *Chem. Phys. Lett.* **1975**, *36*, 451.

(69) Zhang, Y.; Yang, W. *Phys. Rev. Lett.* **1997**, *80*, 890.

(70) Williams, R.; Malhotra, D. *Chem. Phys.* **2006**, *327*, 54.

(71) Ducere, J.-M.; Cavallo, L *J. Phys. Chem. B* **2007**, *111*, 13124.

(72) Perdew, J. P.; Schmidt, K. Density functional theory and its application to materials. Presented at the AIP Conference, 2000, Antwerp, Belgium.

(73) Zhao, Y.; Lynch, B. J.; Truhlar, D. G *J. Phys. Chem. A.* **2004**, *108*, 4786.

(74) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43.

(75) GAMES M06 series of functionals implemented by Sarom Sok, Ames, Iowa, unpublished work.

(76) Schamp, J., H.W.; Mason, E. A.; Richardson, A. C. B.; Altman, A *Phys. Fluids* **1958**, *1*, 329.

(77) Dymond, J. H.; Rigby, M.; Smith, E. B. *J. Chem. Phys.* **1965**, *42*, 2801.

(78) Snook, I. K.; Spurling, T. H. *J. Chem. Soc., Faraday Trans. 2* **1972**, *68*, 1359.

(79) Hanley, H. J. M.; Klein, M. *J. Phys. Chem.* **1972**, *76*, 1743.

(80) Pope, G. A.; Chappelear, P. S.; Kobayashi, R. *J. Chem. Phys.* **1973**, *59*, 423.

(81) Matthews, G. P.; Smith, E. B. *Mol. Phys.* **1976**, *32*, 1719.

(82) Janseb, H. B.; Ross, P. *Chem. Phys. Lett.* **1969**, *3*, 140.

(83) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(84) Dunning, T. H. *J. Chem. Phys.* **1971**, *55*, 716.

(85) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200.

(86) Grover, J. R.; Walters, E. A.; Hui, E. T. *J. Phys. Chem.* **1987**, *91*, 3233.

(87) Lee, E. C.; Kim, D.; Jurecka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. *J. Phys. Chem. A* **2007**, *111*, 3446.

(88) Ringer, A. L.; Sinnokrot, M. O.; Lively, R. P.; Sherrill, C. D. *Chem.—Eur. J.* **2006**, *12*, 3821.

(89) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656.

(90) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690.

(91) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2003**, *107*, 8377.

(92) Cozzi, F.; Cinquini, M.; Annunziata, R.; Dwyer, T.; Siegel, J. S. *J. Am. Chem. Soc.* **1992**, *114*, 5729.

(93) Cozzi, F.; Cinquini, M.; Annuziata, R.; Siegel, J. S. *J. Am. Chem. Soc.* **1993**, *115*, 5330.

(94) Cozzi, F.; Ponzini, F.; Annunziata, R.; Cinquini, M.; Siegel, J. S. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1019.

(95) Cozzi, F.; Siegel, J. S. *Pure Appl. Chem.* **1995**, *67*, 683.

(96) Zoltewicz, J. A.; Maier, N. A.; Fablan, W. M. F. *J. Org. Chem.* **1998**, *63*, 4985.

(97) Kaneta, N.; Mitamura, F.; Uemura, M.; Murata, Y.; Komatsu, K. *Tetrahedron* **1996**, *37*, 5835.

(98) Cozzi, F.; Annunziata, R.; Benaglia, M.; Cinquini, M.; Raimondi, L.; Baldridge, K. K.; Siegel, J. S. *J. Org. Biomol. Chem.* **2002**, *1*, 1.

(99) Cozzi, F.; Annunziata, R.; Benaglia, M.; Baldridge, K. K.; Aguirre, G.; Sritana-Anat, Y.; Siegel, J. S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2686.

(100) Lovas, F. J.; McMahon, R. J.; Grabow, J.-U.; Schnell, M.; Mack, J.; Scott, L. T.; Kuczkowski, R. L. *J. Am. Chem. Soc.* **2004**, *127*, 4345.

(101) Seiders, T. J.; Baldridge, K. K.; Gleiter, R.; Siegel, J. S. *Tetrahedron Lett.* **2000**, *41*, 4519.

(102) Zhao, M.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2813.

(103) Filatov, A. S.; Petrukhina, M. A. *J. Organomet. Chem.* **2008**, *693*, 1590.

(104) Wu, Y.-T.; Siegel, J. S. *Chem. Rev.* **2006**, *106*, 4843.

(105) Prinzbach, H.; Wahl, F.; Weiler, A.; Landenberger, P.; Woerth, J.; Scott, L. T.; Gelmont, M.; Olevano, D.; Sommer, F.; von Issendorff, B. *Chem.—Eur. J.* **2006**, *12*, 6268.

(106) Diudea, M. V. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3626.

(107) Parschau, M.; Fasel, R.; Ernst, K.-H.; Groening, O.; Brandenberger, L.; Schillinger, R.; Greber, T.; Seitsonen, A. P.; Wu, Y.-T.; Siegel, J. S. *Angew. Chem., Int. Ed.* **2007**, *46*, 8258.

(108) Xiao, W.; Passerone, D.; Ruffieux, P.; Aiet-Mansour, K.; Groening, O.; Tosatti, E.; Siegel, J. S.; Fasel, R. *J. Am. Chem. Soc.* **2008**, *130*, 4767.

(109) Theobald, J. A.; Oxtoby, N. L.; Phillips, M. A.; Champness, N. R.; Beton, P. H. *Nature* **2003**, *424*, 1029.

(110) Yokoyama, T.; Yokoyama, S.; Kamikado, T.; Okuno, Y.; Mashiko, S. *Nature* **2001**, *413*, 619.

(111) Fasal, R.; Parschau, M.; Ernst, K.-H. *Angew. Chem., Int. Ed.* **2003**, *43*, 2853.

(112) Parschau, M.; romer, S.; Ernst, K.-H. *J. Am. Chem. Soc.* **2004**, *126*, 15398.

(113) Chen, Q.; richardson, N. V. *Nat. Mater.* **2003**, *2*, 324.

(114) Fasel, R.; Parschau, M.; Ernst, K.-H. *Angew. Chem., Int. Ed.* **2003**, *42*, 5177.

(115) Greber, T.; Sljivancanin, Z.; Schillinger, R.; Wider, J.; Hammer, B. *Phys. Rev. Lett.* **2006**, *96*, 056103.

(116) Schillinger, R.; Sljivancanin, Z.; Hammer, B.; Greber, T. *Phys. Rev. Lett.* **2007**, *98*, 136102.

(117) Wan, T. A.; Davies, M. E. *Nature* **1994**, *370*, 449.

(118) Ratner, M. *Nature* **2005**, *435*, 575.

(119) Kohn, W.; Meir, Y.; Makarov, D. E. *Phys. Rev. Lett.* **1998**, *80*, 4153.

JCTC Journal of Chemical Theory and Computation

Local Hybrid Divide-and-Conquer Method for the Computation of Medium and Large Molecules

Guo-Liang Song,[†] Zhen Hua Li,[†] Zhi-Pan Liu,[†] Xiao-Ming Cao,[†] Wenning Wang,[†] Kang-Nian Fan,*,[†] Yaoming Xie,[‡] and Henry F. Schaefer III*,[‡]

Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Department of Chemistry, and Center for Theoretical Chemical Physics, Fudan University, Shanghai 200433, China, and Center for Computational Chemistry, University of Georgia, Athens, Georgia 30602

Abstract: A local hybrid divide-and-conquer method (LHDC) which combines the high accuracy of sophisticated wave function theory (WFT) methods and the low cost of density functional theory (DFT) has been proposed for computational studies of medium and large molecules. In the method, a large system is divided into small subsystems for which the coefficients of the exchange functional in a hybrid functional are first optimized according to the energy calculated by an accurate WFT method. The hybrid coefficients are then used to evaluate the energy of the whole system. The method not only can reproduce the total energies of the chosen WFT method in good accuracy but also provides electronic structure information for the entire system.

1. Introduction

With rapid developing theoretical techniques, it is becoming straightforward to perform electronic structure calculations for many complicated systems. Wave function theory (WFT) methods, such as Hartree−Fock (HF) and post-HF methods, in principle are capable of systematic convergence to the exact solution of the Schrödinger equation for a given system. For the small molecules, high level WFT methods in conjunction with a large basis set have proved to be capable of achieving chemical accuracy.[1] However, for medium and large systems consisting of several hundreds or thousands of atoms conventional WFT methods are still too expensive. To reduce computational costs, new strategies such as linear scaling methods,[2−6] have attracted much attention. By utilizing the locality of the physical properties[5,6] or the nearsightedness of electrons,[7] the computational costs of these methods promise to scale almost linearly with the size of the system instead of the conventional $O(N^3)$ or higher. Among these linear scaling methods, the divide-and-conquer

(DC) method originally proposed for density functional methods[8] has become popular in the chemistry community.[2]

Recently, a simple and easy to implement DC method, an energy-based DC method (EDC)[9−14] has been proposed. The basic idea of the method is to divide a large system (entire system, ES) into small subsystems. The energy of the ES is then expanded into a sum of many-body terms where the two-, three-, and higher-body correction terms are coupling terms between the subsystems. Since a subsystem within the large system is different from the isolated one, it is important to include the effect of the surrounding environment (i.e., the coupling between the subsystems) when considering the subsystems. The higher accuracy desired, the higher-body correction terms are needed. However, including more higher-body correction terms will greatly increase computational cost, especially for highly accurate WFT methods such as the CCSD(T) method, and the linear-scaling of the method no longer holds. One drawback of the EDC method is that it does not provide information on the electronic structure of the ES since it operates only on the energy of the subsystems.

Density functional theory (DFT) methods can be applied to medium and large systems with moderate computational costs.[14] Many useful functionals have been developed in the past two decades.[16,17] Hybrid DFT methods, which are

developed based on the adiabatic connection formula[18−22] and incorporate a fraction of exact exchange, are the most successful DFT methods to date. However, the challenge to DFT methods is that the exact exchange-correlation functional is unknown and has to be approximated. Unlike WFT methods, there is currently no systematic way to guarantee that DFT methods can converge to right answer. DFT methods still have some major drawbacks compared to conventional WFT methods, such as in the treatment of left-right correlation, van der Waals interactions, and π-π stacking.[16]

Since both approaches, the WFT and DFT methods, have their merits and drawbacks, it is desirable to combine the merits of the two approaches. Ab initio DFT[23−25] and Görling-Levy perturbation theory[26−28] are the two such approaches. In the present research, we propose a novel DC approach which combines the accuracy of high-level correlation methods with the low cost of DFT methods based on the adiabatic connection formula. In this approach, an ES is first divided into several subsystems. Then the fraction (η) of the local DFT exchange (E_x^{LSD}) in a hybrid DFT functional is optimized with respect to the energy of each subsystem (local region) evaluated by a highly accurate WFT method. These η values (the local hybrid (LH) coefficients) are then further refined, by incorporating couplings between the subsystems, and then used to compute the energy of the ES. The new approach not only can reproduce energy profiles of the highly accurate WFT method to high accuracy but also can provide electronic structure information such as charge distributions, where such analyses with highly accurate WFT methods are prohibitive.

2. Methodology

The design of hybrid DFT functionals is based on consideration of the adiabatic connection formula,[18−22] which connects the Kohn−Sham noninteracting reference system to the full-interacting physical system

$$E_{xc} = \int_0^1 U_{xc}^\lambda d\lambda \tag{1}$$

where E_{xc} is the exchange correlation energy, λ is an interelectronic coupling-strength parameter, and U_{xc}^λ is the exchange-correlation potential energy at an intermediate coupling strength λ. In Becke's half-and-half theory,[22] E_{xc} is approximated as the average of the E_{xc} of the noninteracting system ($E_{xc}^{\lambda=0} = E_x^{Exact}$) and that of the full interacting system ($E_{xc}^{\lambda=1} \approx E_{xc}^{LSD}$), i.e.

$$E_{xc} \cong \frac{1}{2}(E_x^{Exact} + E_{xc}^{LSD}) \tag{2}$$

The B3LYP hybrid functional,[29] perhaps the most popular DFT functional, as implemented in Gaussian 03,[30] has the following form:[16]

$$E_{xc} \cong a_0 E_x^{Exact} + \eta E_x^{LSD} + a_x \Delta E_x^{B88} + (1 - a_c)E_c^{VWN-RPA} + a_c \Delta E_c^{LYP} \tag{3}$$

In eq 3, a_0 is the coefficient of the exact exchange energy E_X^{Exact}; η is the coefficient of the exchange energy under the local spin density (LSD) approximation E_X^{LDA}; η is usually

taken as $(1-a_0)$; a_x is the coefficient of the gradient correction part (E_X^{B88}) of the B88 exchange functional; $(1-a_c)$ and a_c are the coefficients of VWN[31] LSD correlation energy $E_C^{VWN-RPA}$ and the gradient correction part of LYP correlation functional ΔE_C^{LYP},[32] respectively. The parameters a_0, a_x, and a_c are those used in the B3PW91 functional,[33] in which they are determined empirically by fitting to a set of accurate thermodynamics data. Similar to the B3LYP functional, in most hybrid functionals only a fraction of E_x^{Exact} is included (i.e., mixed with E_x^{LSD}) in E_{xc}. Usually, the mixing coefficient a_0 is fitted empirically to a set of accurate thermodynamics data and is then fixed for applications to all systems. Thus, the coefficient a_0 is universal in this limited sense. However, because E_x^{LSD} and other exchange correlation functionals used in the hybrid functionals are all approximate functionals, if we know the exact solution of the Schrödinger equation and thus the exact E_{xc}, one would immediately see that the mixing coefficient and also other parameters such as a_x and a_c in the B3LYP functional must be system dependent and are probably different in the various regions of a system, i.e. dependent on $\boldsymbol{r}$.[34] Thus, the mixing coefficient is not truly universal. Ideally, the mixing coefficient should be determined according to the property of each system.[35]

In principle, if the exact E_{xc} of a system is known, a coefficient in the hybrid functionals can be solved by fixing other coefficients. Since E_{xc} is unknown except for a few model systems, one can use E_{xc} obtained from highly accurate WFT methods, which can be improved by incorporating more correlation and enlarging basis set. Therefore, in the present work, an ES is divided into several subsystems for which a highly accurate WFT method, for example CCSD(T),[36,37] is applicable. The energy of each isolated subsystem (called monomer hereafter) is then computed. We fix all but one coefficient (typically a_0), by solving, if the B3LYP functional is selected, the following equation

$$E_{xc}^{WFT} = a_0 E_x^{Exact} + \eta E_x^{LSD} + a_x \Delta E_x^{B88} + (1 - a_c)E_c^{VWN-RPA} + a_c \Delta E_c^{LYP} \tag{4}$$

In (4) E_{xc}^{WFT} is the exchange-correlation energy calculated by the accurate WFT method, and an a_0 value for a monomer is then obtained. It should be noted that since electron density depends on a_0 and E_{xc}^{Exact} and other energy terms thus depend implicitly on a_0, a_0 has to be solved iteratively.

In the current study we choose B3LYP as the example functional in our local hybrid DC method (LHDC), since B3LYP is widely used and it is well-known that B3LYP (and most other DFT functionals) fails for long-range interactions such as π-π stacking, which is very important in biomolecules. For the accurate WFT method, the CCSD(T) method is employed. In the B3LYP functional, we choose η to be varied since the exchange energy is much more important than the correlation energy.[38,39] Conventionally η is fixed to be $1 - a_0$ to fulfill the uniform electron gas (UEG) limit, but in molecular systems it was found in practice that violation of such constraint still gives promising results, for example in the O3LYP functional.[40] On the other hand, usually 20%−25% (a_0) of the exact exchange is included in hybrid functionals.[33,41] Therefore, we will fix a_0 to be 0.20, a_x to be 0.72, a_c to be 0.81, and vary η using

0.80 as an initial guess. Our preliminary tests show that varying η instead of varying a_0 with the constraint of $\eta = 1 - a_0$ actually gives slightly better results. Further, better numerical stability is found when iteratively solving for the mixing coefficient (i.e., η or a_0) using eq 4.

In the EDC method,[9−14] the total energy of an ES is expanded as

$$E_{ES} \equiv \sum_i^M E_i = \sum_i^M (E_i^{(0)} + \Delta E_i^{(2)} + \Delta E_i^{(3)} + \ldots)$$

$$\Delta E_i^{(2)} = \frac{1}{2} \sum_{j \neq i}^M [E_{ij}^{(0)} - (E_i^{(0)} + E_j^{(0)})]$$

$$\Delta E_i^{(3)} = \frac{1}{3} \sum_{k \neq j \neq i}^M [E_{ijk}^{(0)} - (E_{ij}^{(0)} + E_{ik}^{(0)} + E_{jk}^{(0)}) - (E_i^{(0)} + E_j^{(0)} + E_k^{(0)})] \quad (5)$$

where E_i is the energy of a subsystem in the ES environment, $E_i^{(0)}$ is the energy of a monomer, $\Delta E_i^{(2)}$ and $\Delta E_i^{(3)}$ are second-order and third-order corrections to $E_i^{(0)}$ (due to the couplings between two monomers and three monomers, respectively), and M is the number of subsystems. By analogy to eq 5, the η value of a subsystem (local η) embedded in the ES environment may be expressed as

$$\eta_i = \eta_i^{(0)} + \Delta \eta_i^{(2)} + \Delta \eta_i^{(3)} + \ldots \quad (6)$$

where $\eta_i^{(0)}$ is the η value of a monomer determined iteratively as described above. As energy is an extensive thermodynamic variable while the hybrid coefficient η is an intensive one, the expressions for $\Delta \eta_i^{(2)}$ and $\Delta \eta_i^{(3)}$ are slightly different from those for $\Delta E_i^{(2)}$ and $\Delta E_i^{(3)}$

$$\Delta \eta_i^{(2)} = \sum_{j \neq i}^M (\eta_{ij}^{(0)} - \eta_i^{(0)}) \quad (7)$$

$$\Delta \eta_i^{(3)} = \sum_{k \neq j \neq i}^M [\eta_{ijk}^{(0)} - (\eta_{ij}^{(0)} + \eta_{ik}^{(0)}) + \eta_i^{(0)}] \quad (8)$$

where $\eta_{ij}^{(0)}$ is the η value of two subsystems determined in the same way as for a monomer. For example, in eq 4 all the energy terms used are calculated for two subsystems in a whole as a bigger isolated subsystem (named dimer hereafter). The use of the above expressions for $\Delta \eta_i^{(2)}$ and $\Delta \eta_i^{(3)}$ is to guarantee that if no truncation is applied to eq 6, the η_i given by eq 6 is exactly the η of the ES to fulfill the condition that η is an intensive variable.

Once all the local η values are determined, a DFT calculation is performed for the ES to solve the Kohn−Sham equation with these new η values. Since the matrix elements of the exchange-correlation potential $\mathbf{v}_{xc}$ are evaluated in finite basis sets, when calculating element $<\phi_u|\mathbf{v}_{xc}|\phi_v>$ where ϕ_u and ϕ_v are atomic orbitals belonging to different subsystems i and j, an average value of η, i.e. $(\eta_i + \eta_j)/2$ is used. Finally, the total energy and the Kohn−Sham wave function of the system are obtained.

Eq 6 is exact in reproducing the energy calculated by the accurate WFT method if all correction terms up to M^{th} order are included. In practical usage, eq 6 has to be truncated. To further reduce computational costs, contributions from
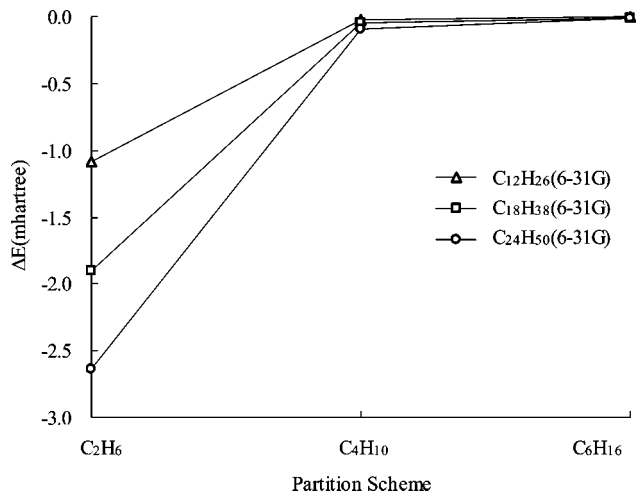
Figure 1. Error (relative energy to the full CCSD(T) energy, in mhartree) of the LHDC2(B3LYP:CCSD(T)) method for the longer straight-chain alkanes under three partition schemes (see text for details).

some dimers, trimers (three subsystems calculated in a whole),..., and m-mers consisting of far apart subsystems may be neglected. In the present work, the sum in eq 6 only includes contributions from those dimers, trimers,..., and m-mers in which subsystems are connected through chemical bonds (neighboring subsystems). Three different truncation schemes have been investigated, where η is corrected to zeroth order, second order, and third order by eq 6, respectively. It should be noted that for the conventional EDC method, there is no zeroth order method for systems that must be divided by cutting through a chemical bond, since the contributions from cap atoms cannot be canceled out in eq 5.

The model systems selected to test our LHDC methods are three straight-chain alkanes, two fully conjugated straight-chain alkenes, and the benzene dimer. The nomenclature of the new method is LHDCn(DFT:WFT) in which n ($= 0, 2, 3\ldots$) represents the truncation order in eq 6, DFT represents the hybrid functional used, and WFT represents the WFT method used. In the current research, DFT is B3LYP and WFT is CCSD(T). To divide ES into subsystems, one may cut through chemical bonds. In such a case, the broken bond is capped by a hydrogen atom along the direction of the original bond with a fixed C−H bond length of 1.100 Å. All the WFT calculations and the DFT calculations except the final one for ES are performed with the Gaussian 03 software package.[30] The iterative solution for the mixing coefficient is done utilizing the IOP option of Gaussian 03 to change the coefficients of a hybrid functional. The final DFT calculation with local η values is performed with our own program using integrals calculated from the Gaussian 03 program.

3. Results and Discussion

3.1. Straight-Chain Alkanes. Three straight-chain alkanes with increasing sizes, $C_{12}H_{26}$, $C_{18}H_{38}$, and $C_{24}H_{50}$, have been tested as prototypes of saturated systems. For the purpose of testing the accuracy of our method on total energies, the geometries of these molecules are not optimal.

Table 1. Total Energies for the $C_{16}H_{18}$ Molecule by the LHDCn(B3LYP:CCSD(T)) (n = 0, 2, 3) and CCSD(T) Methods

basis set	CCSD(T) E (a.u.)	subsystem[a]	η[b]			ΔE (mh)[d]		
			LHDC0[c]	LHDC2	LHDC3	LHDC0	LHDC2	LHDC3
6−31G	−617.53317	terminal C_4H_6	0.76689	0.76697	0.76698	6.82	−2.78	−0.54
		middle C_4H_6		0.76705	0.76699			
6−31G(d)	−618.35579	terminal C_4H_6	0.77542	0.77564	0.77569	23.00	−2.72	−0.57
		middle C_4H_6		0.77585	0.77574			
cc-pVDZ	−618.53343	terminal C_4H_6	0.77770	0.77753	0.77745	−23.20	−2.70	−0.62
		middle C_4H_6		0.77736	0.77738			
cc-pVTZ[e]	−619.02579	terminal C_4H_6	0.78188	0.78187	0.78196	−3.32	−2.30	−0.49
		middle C_4H_6		0.78187	0.78183			

[a] The $C_{16}H_{18}$ molecule is divided into 4 C_4H_6 subsystems. [b] η is the coefficient of the exchange energy under the local spin density (LSD) approximation E_X^{LDA} and is optimized in the LHDC method. [c] The terminal and middle C_4H_6 have the same η value for the LHCD0 method since they have the same geometry. [d] The error ΔE (in mhartree) is the energy relative to the full CCSD(T) energy. [e] The cc-pVTZ basis set is for carbon, and the cc-pVDZ basis set is used for hydrogen.

In all the alkanes the C−C bonds are fixed at 1.523 Å; C−H bonds are fixed at 1.115 Å; CCH, CCC, and HCH bond angles are fixed at 110.0, 124.9, and 109.3 degrees, respectively; and CCCC dihedral angles are fixed at 180.0 degrees. The 6−31G(d) basis set is used for all the calculations.

In the DC methods, the accuracy of the methods is greatly affected by the choice of the partition scheme. Usually, the larger the subsystems, the higher the accuracy and the computational cost are. It is thus essential to find a balance between accuracy and cost. In the present study we have tested three partition schemes. In the first scheme, the molecules are divided into 6, 9, and 12 C_2H_6 molecules (capped by hydrogen atoms), respectively. In the second scheme, the molecules are divided into 3 C_4H_{10}, 4 C_4H_{10} + 1 terminal C_2H_6, and 6 C_4H_{10} molecules, respectively. In the third scheme, the molecules are divided into 2, 3, and 4 C_6H_{14} molecules, respectively. The three partition schemes here are designated C_2H_6, C_4H_{10}, and C_6H_{14} partition schemes.
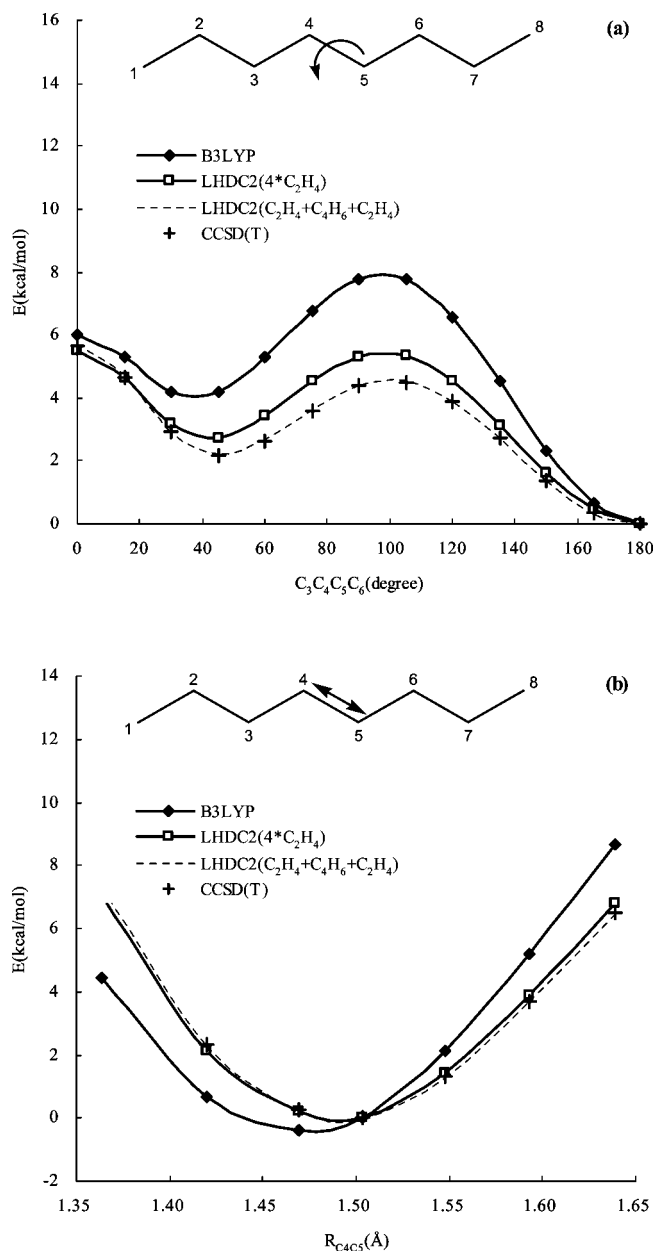
In Figure 1 the error (the energy difference between the LHDC and the full CCSD(T) methods) of the second-order LHDC method is plotted as a function of the size of the subsystems. It can be seen that the LHDC2(B3LYP: CCSD(T)) method with the C_2H_6 partition scheme can already reproduce the full CCSD(T) energy within 3 mhartree (1.9 kcal mol^{-1}). As the size of the subsystem enlarged from C_2H_6 to C_4H_{10}, the error is greatly decreased, down to below 0.1 mhartree, which is just 0.06 kcal mol^{-1}. For the C_4H_{10} partition scheme, the most timing consuming step is the calculation of dimers (i.e., C_8H_{14}) at the CCSD(T) level, which is still affordable on a common personal computer (PC) with a moderate basis set. For the C_6H_{14} partition scheme, although the error is very small (0.013 mhartree), it requires the calculation of a large $C_{12}H_{26}$ molecule, which is perhaps a little expensive for PC computation and is not recommended. It is worth noting that the error of the LHDC method increases with the size of the alkane and is always negative, indicating the energy calculated by the LHDC method is systematically lower than that of the WFT method.

3.2. Straight-Chain Alkenes. To test our method on conjugated systems, two fully conjugated straight-chain alkenes, $C_{16}H_{18}$ and C_8H_{10} ((3E,5E)-octa-1,3,5,7-tetraene), have been tested. In both alkenes, C−C single and double bonds are arbitrarily fixed at 1.500 Å; C−H bonds are fixed at 1.100 Å; and CCC and CCH bond angles are both fixed at 120 degrees. For $C_{16}H_{18}$, we have also tested the effect of

basis set. Four basis sets have been tested: 6−31G, 6−31G(d), cc-pVDZ, and cc-pVTZ*. In the latter the cc-pVTZ basis set is used for carbon, while the smaller cc-pVDZ basis set is used for hydrogen to reduce computational times. The results for $C_{16}H_{18}$ are presented in Table 1, where $C_{16}H_{18}$ is divided into 4 C_4H_6.

From Table 1, it may be seen that energies given by the LHDC0 method can already reproduce the full CCSD(T) energy to an accuracy of about 14 kcal mol^{-1} (23 mhartree). For the LHDC2 and LHDC3 methods, the energies calculated are systematically lower than the full CCSD(T) energies, by −2.78 to −2.30 mhartree for the LHDC2 method and by −0.62 to −0.49 mhartree for the LHDC3 method. However, for the LHDC0 method including no higher-order corrections, the error is not systematic. For the LHDC2 and LHDC3 methods, the basis-set dependence of the error is rather weak, and a moderate decrease is observed when the basis set varies from simple to complex. Compared with saturated hydrocarbons, the error in the total energy for the fully conjugated hydrocarbons is at least an order of magnitude larger when alkene and alkane are both divided into subsystems containing four carbon atoms. Therefore, to achieve a similar accuracy in total energy as that for saturated hydrocarbons, either higher-order corrections should be included or larger subsystems should be used for conjugated systems. This is not unexpected since electrons in conjugated systems are more delocalized.

For C_8H_{10}, we have investigated the potential energy profiles with respect to rotation around and stretching along the central C4C5 bonds, while keeping all other geometrical parameters fixed. The resulting energy profiles are presented in Figure 2. Since one of the major deficiencies of the B3LYP method is the calculation of long-range interactions, which play an important role in the C_8H_{10} rotation barrier, including second-order corrections is necessary to give good results. Therefore, the results of the LHDC0 method are not presented here. Figure 2 indicates that LHDC methods greatly improve the results of the original B3LYP method. The results indicate that although the error in the total energy computed by the LHDC2 method is large (see also Table 1), the error in the calculated relative energies is much smaller (less than 1 kcal mol^{-1}) even when C_8H_{10} is divided into subsystems as small as C_2H_4. On the other hand, the error for the original B3LYP method is as large as 3.4 kcal mol^{-1}. When enlarged subsystems are used, the energy

(a)

Figure 3. Potential energy profiles for the parallel displaced configuration of the benzene dimer, with horizontal separation R_2 fixed at 1.6 Å. The zero of the energy is that for two isolated benzenes, each with the geometry in the dimer.

(b)

Figure 2. Potential energy profiles of (3E,5E)-octa-1,3,5,7-tetraene: a) rotating around the central C4C5 bond, where the zero of the energy is that of the structure with a C3C4C5C6 dihedral angle of 180 degrees and b) stretching along the central C4C5 bond, where the zero of the energy is that of the structure with a C4C5 bond length of 1.5 Å. Two partition schemes have been investigated: dividing into 4 C_2H_4 and 2 terminal-C_2H_4 + 1 C_4H_6.
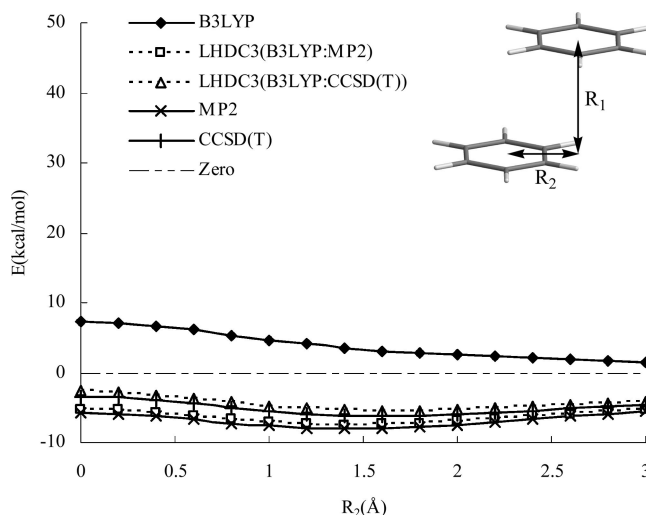
Figure 4. Potential energy profiles for the parallel displaced configuration of the benzene dimer, with vertical separation R_1 fixed at 3.4 Å. The zero of energy is that for two isolated benzenes, each with the geometry in the dimer.

profiles calculated by the LHDC2 method nearly overlap with those calculated by the full CCSD(T) method. Therefore, the error in the total energies calculated by the LHDC methods except LHDC0 is most likely systematic. Since relative energies are more meaningful than total energies, it seems that even for fully conjugated systems, the LHDC2 method with a rather fine partition scheme is acceptable.

3.3. Benzene Dimer. The benzene dimer is a notorious system for most DFT functionals.[42] For example, it is unbound at the B3LYP level.[43] In the present LHDC calculations, benzene is divided into three C_2H_2 subsystems, and the third-order LHDC method is used since benzene

contains three C_2H_2 units and its π electrons are highly delocalized. It should be noted that all trimers have been included for this special case of benzene dimer. The energy profiles of the parallel configuration of the benzene dimer are studied, where the CC and CH bonds are fixed at 1.391 Å and 1.080 Å,[44] respectively, and the R_1 and R_2 geometrical parameters as indicated in Figures 3 (varying R_1 with R_2 fixed at 1.6 Å) and 4 (varying R_2 with R_1 fixed at 3.4 Å) are varied. The purpose of this test is to see if our LHDC method can successfully overcome the deficiency of the B3LYP method and how accurate it is in reproducing the energy profiles predicted by the sophisticated WFT method. Therefore, the energy profiles presented in Figures 3 and 4 are calculated with a moderate aug-cc-pVDZ basis set and are not corrected by basis set superposition error (BSSE). The zero of energy in Figures 3 and 4 is the two isolated benzene monomers with the geometry found for the dimer.
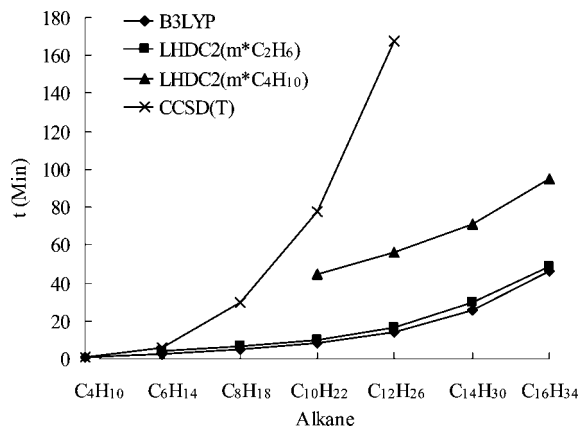
Figure 5. Timings (t) in minutes for the calculation of straight-chain alkanes on a single 3.0-GHz Xeon CPU. Two partition schemes for the LHDC2(B3LYP:CCSD(T)) methods are investigated: the molecule is divided into C_2H_6 and C_4H_{10} subsystems.
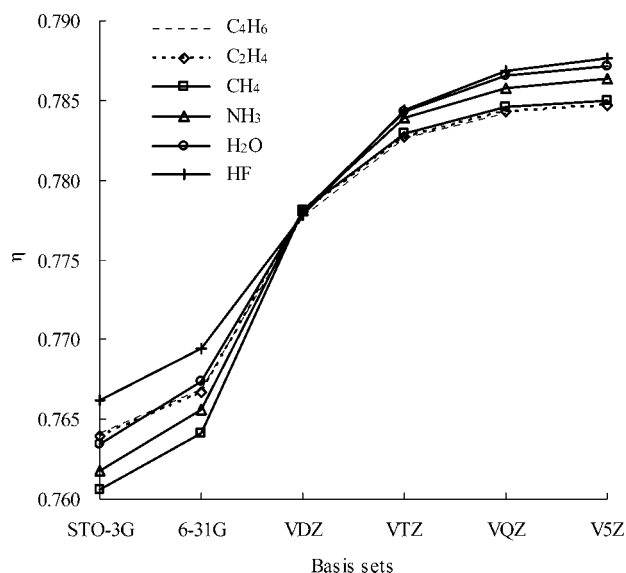
Figure 6. Basis-set dependence of the hybrid coefficient η for six small molecules. Basis sets considered are STO-3G, 6−31G(d), VDZ (cc-pVDZ), VTZ (cc-pVTZ), VQZ (cc-pVQZ), and V5Z (cc-pV5Z).

Two LHDC methods have been tested, LHDC3(B3LYP:MP2) and LHDC3 (B3LYP:CCSD(T)) methods, to reproduce the energy profiles calculated by the full MP2 and CCSD(T) methods, respectively. The results indicate that both methods can accurately reproduce the energy profiles calculated by the WFT methods. Except for the first point in Figure 3, the differences between the energies calculated by the LHDC3 methods and the corresponding full WFT methods are small, all below 1 kcal mol−1. It can also be seen that benzene dimer is slightly over bound under the LHDC treatments compared to the full WFT methods.

3.4. Cost and Efficiency. The computational requirements of the LHDC method are greater than a conventional DFT calculation since in addition to the calculation of ES at the DFT level it also requires additional computations to obtain local η values. The additional cost consists of two parts: the first part is the energy calculation of m-mers ($m = 0, 2, 3,..., n$) by the WFT method, and the second part is the iterative

solution for local η values by the DFT method. The iterative procedure usually finishes in less than 4 iterations. It is known that the DFT method is much faster than correlation WFT methods, and thus the costs for the DFT calculations for the m-mers are negligible compared to the expensive WFT methods. In addition, for the demanding WFT methods, the calculation of an m-mer ($m > 1$) is much more expensive than that of an ($m-1$)-mer. Therefore, the cost for the LHDC method mainly comes from two calculations, the calculation of the largest m-mers by the WFT method and the final DFT calculation for the ES. In the LHDC2 methods, the calculation of $M(M − 1)/2$ dimers is required if all second-order corrections are included. However, this number can be greatly reduced by neglecting contributions from dimers in which two monomers are far apart. For example, in the present computations for straight-chain alkanes and alkenes, this number is reduced to $M − 1$ by excluding those dimers in which monomers are not connected through chemical bonds.

In Figure 5 we have compared the computation time for the straight-chain C_nH_{2n+2} ($n = 4−16$) alkanes which are divided into C_2H_6 or C_4H_{10} subsystems. The 6−31G basis set is used, and the calculations are performed on a single 3.0-GHz XEON CPU. The comparison indicates that the costs of the LHDC2(B3LYP:CCSD(T)) method with an C_2H_6 partition scheme are just slightly higher than those of the full B3LYP method. The results for the alkenes indicate that the LHDC2(B3LYP:CCSD(T)) method with a fine partition scheme in which alkanes and alkenes are divided into subsystems containing two carbon atoms is probably accurate enough to study the energy profiles of common molecules. For highly conjugated systems containing aromatic rings, one may need to use the more expensive LHDC3 method for satisfactory accuracy. However, the LHDC3(B3LYP:CCSD(T)) method is still much less expensive than the full CCSD(T) method.

3.5. Basis-Set Dependence of the Hybrid Coefficient η. Six sample molecules including methane, ethene, 1,3-dibutene, ammonia, water, and hydrogen fluoride are selected to the test basis set dependence of η. The geometries of these molecules are all optimized using the B3LYP/6−31G(d,p) method. The results are presented in Figure 6. Two conclusions may be drawn from our results. First, η for each molecule appears systematically convergent to the basis set limit when the basis set is enlarged up to cc-pV5Z, and the changes from the smallest basis set to the largest basis set are only about 3%. Though the basis set dependence of η is not significant, we have found the change of the total energy for a given molecule is significant. However, since the dependence of η on the basis set is systematic, this will not introduce nonsystematic errors for relative energies, which are more meaningful than total energies. Second, the η values for different molecules are close to each other. For a given basis set, the difference between the η values for the molecule with the largest η and the molecule with the smallest η is only 0.006.

Local Hybrid Divide-and-Conquer Method

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2055**

4. Concluding Remarks

We have presented a novel LHDC method, which combines highly accurate WFT methods with low-cost DFT methods, to compute the energy and electronic structure of medium and large molecules. The method requires one to divide an entire system into small subsystems and to optimize the local hybrid coefficient η of a hybrid DFT functional for each subsystem according to the energy of the subsystem calculated by a high level WFT method. Further refinement to η (local hybrid coefficient) is made by incorporating couplings between the subsystems. The local hybrid coefficients are then used to calculate the energy and electronic structure of the entire system with the hybrid DFT functional. The new method can accurately reproduce the energy profiles of the highly accurate WFT methods.

Since for small subsystems, highly accurate WFT methods may be applied, our LHDC method provides a route to systematically converge DFT results to "exact" results for medium and large molecules which are too large to apply the accurate WFT methods to. The cost of the LHDC method equals the cost of a usual DFT calculation for the entire system plus additional efforts to obtain the local hybrid coefficients. Our results indicate that a second-order LHDC method which is just slightly more expensive than a DFT method can already give satisfactory results on potential energy profiles for highly delocalized systems such as fully conjugated alkenes.

From a WFT viewpoint, our method is a DC method using DFT as "glue" to bind subsystems but with the capability of reproducing the WFT energy and providing electronic structure information for the entire system that is not addressed by conventional EDC methods. From a DFT viewpoint, the hybrid coefficients of subsystems (local regions) are optimized according to the energy of the accurate WFT method, and thus our method can be viewed as a new local hybrid DFT method.[34] Thus the "local hybrid (LH)" in the name LHDC has a double meaning, the *hybrid* of DFT and WFT through the optimization of local *hybrid* coefficient η of the DFT exchange.

References

(1) Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.

(2) Galli, G. *Curr. Opin. Solid State Mater. Sci.* **1996**, *1*, 864.

(3) Morokuma, K. *Phil. Trans. R. Soc. London A* **2002**, *360*, 1149.

(4) Bowler, D. R.; Miyazaki, T.; Gillan, M. J. *J. Phys.: Condens. Matter* **2002**, *14*, 2781.

(5) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085.

(6) Goedecker, S.; Scuseria, G. E. *Comput. Sci. Eng.* **2003**, *5*, 14.

(7) Kohn, W. *Phys. Rev. Lett.* **1996**, *76*, 3168.

(8) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.

(9) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.

(10) Li, S. H.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7215.

(11) Stoll, H.; Paulus, B.; Fulde, P. *J. Chem. Phys.* **2005**, *123*, 144108.

(12) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.

(13) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.

(14) Fedorov, D. G.; Ishida, T. *J. Comput. Chem.* **2007**, *28*, 1476.

(15) Kohn, W. *Mod. Phys.* **1999**, *71*, 1253.

(16) Scuseria, G. E.; Staroverov, V. N. Progress in the Development of Exchange- Correlation Functionals In *Theory and Application of Computational Chemistry: The First 40 Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 669–724.

(17) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(18) Harris, J.; Jones, R. O. *J. Phys. F* **1974**, *4*, 1170.

(19) Gunnarsson, O.; Lundqvist, B. I. *Phys. Rev. B* **1976**, *13*, 4274.

(20) Langreth, D. C.; Perdew, J. P. *Phys. Rev. B* **1977**, *15*, 2884.

(21) Harris, J.; Jones, R. O. *Phys. Rev. A* **1984**, *29*, 1684. Yang W. *J. Chem. Phys.* **1998**, *109*, 10107.

(22) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1327.

(23) Bartlett, R. J.; Lotrich, V. F.; Schweigert, I. V. *J. Chem. Phys.* **2005**, *123*, 062205.

(24) Bartlett, R. J.; Schweigert, I. V.; Lotrich, V. F. *J. Mol. Struct.* **2006**, *771*, 1.

(25) Grabowski, I.; Lotrich, V.; Bartlett, R. J. *J. Chem. Phys.* **2007**, *127*, 154111.

(26) Görling, A.; Levy, M. *Phys. Rev. A* **1993**, *47*, 13105.

(27) Görling, A.; Levy, M. *Phys. Rev. A* **1994**, *50*, 196.

(28) Görling, A.; Levy, M. *Int. J. Quantum Chem. Symp.* **1995**, *29*, 93.

(29) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 03, Revision C. 02*; Gaussian, Inc.: Pittsburgh, PA, 2003.

(31) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys* **1980**, *58*, 1200.

(32) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(34) Jaramillo, J.; Scuseria, G. E.; Ernzerhof, M. *J. Chem. Phys.* **2003**, *118*, 1068.

(35) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.

(36) Pople, J. A.; Krishnan, R.; Schlegel, H. B; Binkley, J. S. *Int. J. Quantum Chem.* **1978**, *14*, 545.

(37) Bartlett, R. J.; Purvis, G. D. *Int. J. Quantum Chem.* **1978**, *14*, 516.

(38) Ivanov, S.; Bartlett, R. J. *J. Chem. Phys.* **2001**, *114*, 1952.

(39) Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673.

(40) Cohen, A. J.; Handy, N. C. *Mol. Phys.* **2001**, *99*, 607.

(41) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1997**, *274*, 242.

(42) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656.

(43) Ye, X. Y.; Li, Z. H.; Wang, W. N.; Fan, K. N.; Xu, W.; Hua, Z. Y. *Chem. Phys. Lett.* **2004**, *397*, 56.

(44) Gauss, J.; Stanton, J. F. *J. Phys. Chem. A* **2000**, *104*, 2865.

CT800265P

JCTC Journal of Chemical Theory and Computation

Accurate Spin-State Energies for Iron Complexes

Marcel Swart*

Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain, and Institut de Química Computacional and Departament de Química, Universitat de Girona, Campus Montilivi, 17071 Girona, Spain

Abstract: A critical assessment of the OPBE functional is made for its performance for the geometries and spin-states of iron complexes. In particular, we have examined its performance for the geometry of first-row transition-metal (di)halides (MnX_2, FeX_2, CoX_2, NiX_2, CuX, X=[F, Cl]), whose results were previously [*J. Chem. Theory Comput.* **2006,** *2,* 1282] found to be representative for a much larger and more diverse set of 32 metal complexes. For investigating the performance for spin ground-states of iron complexes, we examined a number of small iron complexes ($Fe(II)Cl_4^{2-}$, $Fe(III)Cl_4^{1-}$, $Fe(II)Cl_6^{4-}$, $Fe(III)Cl_6^{3-}$, $Fe(II)CN_6^{4-}$, $Fe(III)CN_6^{3-}$, $Fe(VI)O_4^{2-}$, $Fe(III)(NH_3)_6^{3+}$), benchmark systems ($Fe(II)(H_2O)_6^{2+}$, $Fe(II)(NH_3)_6^{2+}$, $Fe(II)(bpy)_3^{2+}$), and several challenging iron complexes such as the $Fe(II)(phen)_2(NCS)_2$ spin-crossover compound, the monopyridylmethylamine $Fe(II)(amp)_2Cl_2$ and dipyridylmethylamine $Fe(II)(dpa)_2^{2+}$, and the bis complex of Fe(III)-1,4,7-triazacyclononane ($Fe(III)(^9aneN_3)_2^{3+}$). In all these cases OPBE gives excellent results.

Introduction

The reliable prediction of the spin ground-state of transition-metal complexes remains a challenging task,[1] both for theory and experiment. On the experimental side, the situation may be complicated by ligand-exchange reactions,[2] dimerization processes, oxidation/reduction, impurities, etc. Moreover, the structural characterization of the complexes may be hampered by problematic crystallization, reduced lifetimes of transient species (such as the elusive complex I in the catalytic cycle of cytochrome P450),[3–6] or temperature-dependences of the population of the different spin-states. The latter is for instance observed in spin-crossover compounds[7] whose structures may not be resolved at all temperatures. In principle, theory should be able to help with the interpretation of the experimental data, predict the spin ground-state, and help to determine reaction mechanisms. However, theory is not without its own problems. The most accurate *ab initio* theoretical methods (CCSD(T), MR-CI) are too demanding for everyday use and in some cases (such as CASPT2) need expert knowledge of the methodology. More efficient are calculations based on density functional theory (DFT),[8] but the results are shown to depend largely on the choice of DFT functional that is being used.

This is in particular true for the calculation of spin-state splittings,[1] where standard pure functionals (like LDA,[9] BLYP,[10,11] or PBE[12]) systematically overstabilize low-spin states, while hybrid functionals (e.g., B3LYP,[13,14] PBE0[15]) overstabilize high-spin states due to the inclusion of a portion of Hartree−Fock exchange. The problems with B3LYP in correctly describing the relative energies of the spin states of iron complexes led Reiher and co-workers to propose a new functional (dubbed B3LYP*),[16] in which the amount of HF exchange was lowered to 15% (instead of 20% in B3LYP). For many systems, this reduction indeed improves the B3LYP results,[17] apparently without sacrificing the good performance for organic systems.[17] However, it was not successful for all iron complexes,[18,19] as for instance is the case for the $Fe(phen)_2(NCS)_2$ spin-crossover compound. A further reduction to 12% of HF exchange seems necessary[18] to give good results for this particular iron complex. Therefore, with B3LYP and B3LYP*, it is *a priori* unknown if the amount of HF exchange is appropriate for the transition-metal complex under study, which is an undesirable situation.

* Corresponding author fax: +34-972-183241; e-mail: marcel.swart@icrea.es.
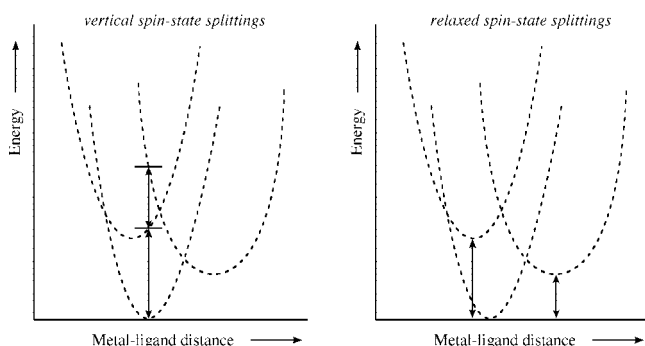
Figure 1. Schematic representation of *vertical* (left-side) and *relaxed* (right-side) spin-state splittings.

Recently, we made a systematical investigation of the influence of the functional[1] and of the basis set[20] on the relative spin-state energies (i.e., the spin-state splittings[21]) for a number of iron complexes. The influence of the basis set was found to be substantial. In principle, with an infinitely large basis set, both Slater-type orbital (STO) and Gaussian-type orbital (GTO) series should converge to the same final answer, which is indeed what we observed for both vertical and relaxed spin-state splittings (see Figure 1 for the difference between vertical and relaxed spin-state splittings). However, we found that the STO basis sets give consistent and rapidly converging results, while the convergence with respect to the basis set size is much slower for the GTO basis sets. Very demanding basis sets like Dunning's correlation consistent (cc-pVTZ, cc-pVQZ) were needed to achieve good results, especially for relaxed spin-state splittings.[20] Furthermore, the use of basis sets containing effective core potentials (ECPBs) resulted in spin-state splittings that are systematically different from the STO-GTO results.

From these and related studies,[1,22,23] it became clear that recent and improved functionals provide more accurate results. This is in particular true for the functionals containing Handy and Cohen's optimized exchange (OPTX) functional,[24] such as OLYP or OPBE.[25] The latter OPBE functional provided the correct spin ground-state for vertical spin-state splittings of a number of Fe(II) and Fe(III) complexes,[1] which is corroborated by good results obtained more recently in studies from other groups.[6,19,22,23,26−35] This good performance of OPBE (and the related OLYP) concurs with recent benchmark studies on the energy profiles of nucleophilic substitution reactions[36] and on the NMR chemical shifts of organic molecules.[27,28] In the former, it was shown that the underestimation of reaction barriers by standard pure DFT functionals is dramatically reduced by using OPBE, while for the NMR chemical shifts its performance is significantly better than other DFT functionals (including B3LYP) and surpasses many times even the *ab initio* MP2 method.

Han and Noodleman recently used a series of small iron complexes to obtain the Mössbauer isomer shift parameters for OPBE and OLYP,[37] which were subsequently used to study the intermediate Q of the hydroxylase component of soluble methane monooxygenase (MMOH).[38] They showed that OPBE (and OLYP) do not overestimate the Fe-ligand covalency for these model structures, in contrast to the PW91 functional. Moreover, OPBE and OLYP correctly predicted

the high-spin antiferromagnetically (AF) coupled Fe^{4+} sites. Furthermore, in a series of papers Ghosh and co-workers[23,39−45] have clearly demonstrated that OPBE (and OLYP) seems to be giving good results not only for iron complexes but also for other transition metals.

In the present contribution, we make an assessment of the OPBE functional for the spin-states of iron complexes. Apart from studying its accuracy for the geometries of transition-metal complexes, we investigate the spin ground-states of small iron complexes like $Fe(CN)_6^{3−}$, and challenging iron complexes such as the $Fe(phen)_2(NCS)_2$ spin-crossover compound, and compare the results of OPBE with other functionals. Moreover, a comparison is made with benchmark *ab initio* data obtained with high-level CASPT2 methods, where available. Finally, using an energy decomposition analysis the origin of the spin ground-state of these complexes is explained in terms of a compromise between metal−ligand bonding and Hund's rule of maximum multiplicity.

Computational Details

Most DFT calculations were performed with the Amsterdam Density Functional (ADF) suite of program.[46,47] MOs were expanded in an uncontracted set of Slater type orbitals (STOs)[48] of triple-ζ quality containing diffuse functions and one (TZP) or two (TZ2P) sets of polarization functions. Core electrons (1s for second period, 1s2s2p for third-fourth period) were not treated explicitly during the geometry optimizations (frozen core approximation[46]), as it was shown[49] to have a negligible effect on the obtained geometries. An auxiliary set of s, p, d, f, and g STOs was used to fit the molecular density and to represent the Coulomb and exchange potentials accurately for each SCF cycle.

Energies and gradients were calculated using the local density approximation (LDA; Slater exchange and VWN correlation[9]) with gradient-corrections (GGA) for exchange (OPTX[24]) and correlation (PBE[50,51]) included self-consistently, i.e. the OPBE functional. Geometries were optimized with a locally adapted version of the QUILD program[52,53] using adapted delocalized coordinates[52] until the maximum gradient component was less than $1.0 \cdot 10^{-4}$ a.u. Single point energies for all other DFT functionals were obtained (post-SCF within the METAGGA scheme) at the OPBE optimized geometries using the corresponding all-electron basis sets.

As the gradients for hybrid functionals are not yet available within the ADF program, the geometries of the first-row transition-metal (di)halides were also obtained with the NWChem program (version 5.0)[54] for a number of DFT functionals (OPBE, B3LYP, BP86, PBE, B3LYP*) using the cc-pVTZ[55] basis set of Gaussian-type orbitals (GTOs).

For some of the iron complexes in this paper, we included solvent effects through the use of a dielectric continuum (COSMO[56,57]) model, with the appropriate dielectric constants (ϵ) and solvent radii (R_{solv})[58] for the solvent used. A nonempirical approach[59] to including solvent effects in QM calculations has been used, which works well for solvation processes.[58,59]

Energy Decomposition Analysis. The total energy ΔE_{total} for the heterolytic association[19] reaction between the iron(II)

Accurate Spin-State Energies

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2059**

cation and n ligands L with charge q ($Fe^{2+} + n \cdot L^q \rightarrow FeL_n^{nq+2}$) results directly from the Kohn–Sham molecular orbital (KS-MO) model[60] and is made up of two major components (eq 1):

$$\Delta E_{total} = \Delta E_{prep} + \Delta E_{int} \qquad (1)$$

In this formula, the preparation energy ΔE_{prep} is the energy needed to prepare the (ionic/neutral) fragments and consists of three terms (eq 2):

$$\Delta E_{prep} = \Delta E_{deform} + \Delta E_{lig-lig} + \Delta E_{valexc} \qquad (2)$$

The first is the energy needed to deform the separate molecular fragments (in this case only for the ligands) from their equilibrium structure to the geometry that they attain in the overall molecular system (ΔE_{deform}). The second ($\Delta E_{lig-lig}$) is the interaction energy between the ligands when they are placed at the geometry of the molecule (but without the iron present) to make one fragment file that contains all ligands. This interaction results mainly from electrostatic repulsion in case of negatively charged ligands. The third term (ΔE_{valexc}) is the valence-excitation energy needed to prepare the metal from its atomic spin-unrestricted (polarized) ionic ground-state to the spin-restricted (polarized) ionized form. The valence-excitation energy consists of two terms: the first (positive, e.g. destabilizing) term is the energy difference between the spin-polarized metal cation in its ground state (e.g., the quintet ^5D state for Fe^{2+}) and the spin-restricted, nonpolarized (singlet) cationic form used for the metal cation fragment (the fragments need to be spin-restricted). For the ground-state of the cation, we use the "average of configuration" approach,[61] which gives an approximate single-determinant description of the true atomic spin ground-state. Note also that the metal cation fragment is prepared with the occupation of the orbitals it attains in the molecule, i.e. it does not necessarily, and usually does not, correspond to the isolated metal cation. As a result, the ΔE_{valexc} values cannot be compared directly with experimental excitation energies for the metal cation (see also refs 61 and 62). The second term results from preparing (polarizing) the cation fragment with the multiplet state it attains in the metal complex; this term is negative (stabilizing) for triplet and quintet and zero for singlet states. It is achieved by changing the occupations of the fragment orbitals. For instance for iron(II), the spin-restricted cationic fragment would be prepared with 3 α and 3 β d-electrons; within the molecule calculation, the occupations of the iron-fragment are changed to make 4 α and 2 β d-electrons for a triplet state or to 5 α and 1 β d-electrons for a quintet state. There is a discrepancy (of ca. 2 kcal mol^{-1}) between the interaction energy thus obtained from these fragments (*vide infra*) and the change in energy when going from the isolated ligands and (spin-unrestricted, polarized) metal cation to the metal complex. This difference results from the fact the α and β orbitals are kept the same in the former (fragment) approach, while they are allowed to relax in the latter. There are two possibilities to deal with this discrepancy, either to make this energy difference part of the preparation energy (as done here) or to scale the interaction energy components accordingly (values reported in the Supporting Information, Tables

S1 and S2). However, this energy difference is generally negligible compared to the interaction energy components and is therefore of no consequence for the importance of the components of the interaction energy. For the interested reader, the EDA analysis for the $Fe(amp)_2Cl_2$ complex has been performed also with the BP86 functional (see the Supporting Information). Although it induces a slight change in the values for the different energy components, this does not influence the importance of the compromise between Hund's rule of maximum multiplicity and metal–ligand covalent interactions for the determination of the spin ground-state of these molecules (*vide infra*).

The interaction energy ΔE_{int} is the energy released when the prepared fragments (i.e., $Fe^{2+} + n \cdot L^q$) are brought together into the position they have in the overall molecule. It is analyzed for our model systems in the framework of the KS-MO model[60] using a Morokuma-type[63] decomposition into electrostatic interaction, Pauli repulsion (or exchange repulsion), and (attractive) orbital interactions (eq 3).

$$\Delta E_{int} = \Delta V_{elstat} + \Delta E_{Pauli} + \Delta E_{orbint} \qquad (3)$$

The term ΔV_{elstat} corresponds to the classical electrostatic interaction between the unperturbed charge distributions of the prepared (i.e., deformed) fragments and is usually attractive. The Pauli-repulsion, ΔE_{Pauli}, comprises the destabilizing interactions between occupied orbitals and is responsible for the steric repulsion. The orbital interaction ΔE_{orbint} in any MO model, and therefore also in Kohn–Sham theory, accounts for electron-pair bonding, charge transfer (i.e., donor–acceptor interactions between occupied orbitals on one fragment with unoccupied orbitals of the other, including the HOMO–LUMO interactions), and polarization (empty-occupied orbital mixing on one fragment due to the presence of another fragment). In the case of metal complexes with symmetry, the orbital interaction energy can be further decomposed into the contributions from each irreducible representation Γ of the interacting system (eq 4) using the extended transition state (ETS) scheme developed by Ziegler and Rauk.[64,65]

$$\Delta E_{orbint} = \sum_{\Gamma} \Delta E_{\Gamma} \qquad (4)$$

Results

Here we report a critical assessment of the OPBE functional for its performance for the geometries and spin-states of iron complexes. Spin contamination is in all cases negligible, as shown by the expectation values for S^2 that are very close to the pure spin-state values. No attempt at spin-projection[66,67] has therefore been made, as these corrections would not alter the energies significantly.

Geometry Optimization of (Di)halides. We have examined the performance of OPBE for the structure of first-row transition-metal (di)halides (MnX_2, FeX_2, CoX_2, NiX_2, CuX, $X=[F, Cl]$), whose results were found[68] to be representative for a much larger and more diverse set of 32 metal complexes. For the geometry optimization of these (di)ha-

***Table 1.*** Metal-Halide Distances (Å) and Mean Absolute Deviations (MAD, Å) for a Set of 10 (Di)halides

compd (mult[a])	exp.	OPBE TZ2P[b]	OPBE cc-pVTZ[c]	BP86 TZ2P[b]	BP86 cc-pVTZ[c]	PBE TZ2P[b]	PBE cc-pVTZ[c]	B3LYP cc-pVTZ[c]	B3LYP* cc-pVTZ[c]
MnF$_2$ (6)	1.797	1.793	1.788	1.790	1.790	1.794	1.789	1.796	1.793
FeF$_2$ (5)	1.755	1.757	1.750	1.752	1.752	1.755	1.751	1.757	1.755
CoF$_2$ (4)	1.738	1.720	1.713	1.725	1.716	1.728	1.715	1.720	1.717
NiF$_2$ (3)	1.715	1.719	1.710	1.711	1.710	1.714	1.709	1.730	1.728
CuF (1)	1.745	1.759	1.758	1.747	1.753	1.752	1.751	1.761	1.757
MnCl$_2$ (6)	2.184	2.162	2.166	2.169	2.173	2.168	2.171	2.194	2.188
FeCl$_2$ (5)	2.128	2.117	2.121	2.123	2.127	2.123	2.125	2.132	2.126
CoCl$_2$ (4)	2.090	2.072	2.076	2.077	2.082	2.077	2.080	2.104	2.098
NiCl$_2$ (3)	2.056	2.045	2.047	2.050	2.054	2.051	2.053	2.073	2.067
CuCl (1)	2.052	2.044	2.061	2.050	2.066	2.051	2.064	2.089	2.083
MAD		0.011	0.011	0.007	0.008	0.006	0.009	0.014	0.011

[a] Experimentally (and theoretically) observed multiplicity for this complex. [b] STO basis set. [c] GTO basis set.

lides, we used both a STO (TZ2P) and a GTO (cc-pVTZ) basis set to be able to compare directly with literature data that have been obtained with GTO basis sets. The dihalide molecules were treated as linear molecules ($D_{\infty h}$ symmetry), as observed experimentally. The obtained distances with both basis sets and a number of functionals are reported in Table 1. In the original paper by Bühl and Kabrede,[68] the best results were obtained by pure GGA (BPW91, BP86) and metaGGA (TPSS) functionals, with mean absolute deviations (MAD) from experimental data of 0.008–0.009 Å when using the AE1 basis. This GTO basis set consists of the augmented all-electron Wachters basis set on the metal and 6–31G* on the halides. For BP86 and B3LYP, the MAD values obtained here with the cc-pVTZ basis set are similar to the ones obtained with this AE1 basis, with values for BP86 of 0.008 Å (cc-pVTZ) and 0.009 Å (AE1), while for B3LYP they are 0.014 Å (cc-pVTZ) and 0.015 Å (AE1).

The OPBE functional is shown to be less accurate than BP86 but more accurate than B3LYP, with MAD values of 0.011 Å for both the TZ2P and the cc-pVTZ basis set (see Table 1). The same value is observed for the B3LYP* functional with the cc-pVTZ basis (see Table 1). Also the maximum error is significantly smaller for OPBE (0.025 Å with the cc-pVTZ basis, 0.022 Å with TZ2P) compared to B3LYP* (0.032 Å cc-pVTZ) and B3LYP (0.038 Å cc-pVTZ) and again only slightly larger than the most accurate functionals BP86 (0.022 Å with both cc-pVTZ and AE1) and TPSS (0.022 Å, AE1 basis). Therefore, the OPBE functional seems to provide a good description for the geometries of transition-metal complexes, with an accuracy that is close to that of the best functionals and significantly better than other functionals, such as BLYP, B3LYP, or VS98.[68]

Benchmark Systems for Spin-State Splittings. Recently, Pierloot and co-workers reported a benchmark study[69,70] on a set of three iron complexes (Fe(H$_2$O)$_6^{2+}$, Fe(NH$_3$)$_6^{2+}$, Fe(bpy)$_3^{2+}$, see Scheme 1), which were investigated with high-level CASPT2 calculations and, for comparison, with Hartree–Fock (HF) and a number of DFT functionals (LDA, BP86, B3LYP, PBE0), following up on earlier studies on these complexes.[62,71–73] Because of the absence of electron correlation for electrons with unlike spins in HF, this method unduly favors high-spin states too much,[19] which shows up clearly in their results. For all three complexes HF predicts a high-spin state, mistakenly also for the low-spin bipyridyl

Scheme 1. Benchmark Iron Complexes Fe(H$_2$O)$_6^{2+}$, Fe(NH$_3$)$_6^{2+}$, and Fe(bpy)$_3^{2+}$[a]

Fe(H$_2$O)$_6^{2+}$ **Fe(NH$_3$)$_6^{2+}$**   **Fe(bpy)$_3^{2+}$**
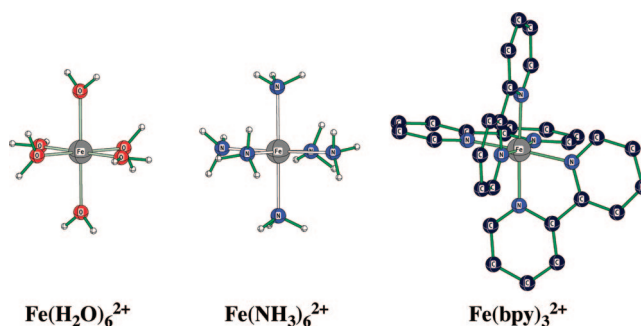
[a] Hydrogens were omitted for clarity.

complex, and with a large deviation from the reference CASPT2 ΔE_{HL} values (see Table 2). The overstabilization of low-spin states by standard pure functionals[1] is reconfirmed by their data for LDA and BP86[69] and for RPBE in another study by Deeth and Fey.[71] Interestingly, the BP86 functional still predicts the correct spin ground-state for all three molecules, i.e. a high-spin state for Fe(H$_2$O)$_6^{2+}$ and Fe(NH$_3$)$_6^{2+}$ and a low-spin state for Fe(bpy)$_3^{2+}$, albeit with a large deviation (ca. 15 kcal mol^{-1}) from the reference CASPT2 data. Although this deviation is smaller for the hybrid B3LYP and PBE0 functionals with a value of respectively 11 and 9 kcal mol^{-1}, these latter two functionals fail to predict the correct spin ground-state for the bipyridyl complex.[69]

These systems have been investigated with the OPBE functional using the TZP and TZ2P (STO) basis sets (see Table 2). The calculations were performed within D_3 symmetry for the ammonia and bipyridine complexes and within C_i with water as ligand, in order for a fair comparison with the best available CASPT2 data that were obtained with the same symmetry constraints. For the quintet (and triplet) state of the water complex this leads to Jahn–Teller distortions, similar to the CASPT2 study by Pierloot and co-workers. The symmetry constraints for the triplet states, for which no CASPT2 data are available to compare with, are for all three complexes the same as those of the singlet and quintet states. One of the reviewers pointed out that these triplet and quintet states are formally Jahn–Teller active, and lower energies might be obtained by allowing Jahn–Teller

Accurate Spin-State Energies

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2061**

Table 2. Spin-State Splittings ΔE (kcal mol^{-1})^a for Benchmark Iron Complexes

	Fe(H$_2$O)$_6^{2+}$			Fe(NH$_3$)$_6^{2+}$			Fe(bpy)$_3^{2+}$			
	sing	trip	quin	sing	trip	quin	sing	trip	quin	MAD^b
CASPT2^{c,d}	46.6	n/a^e	0	20.3	n/a^e	0	0	n/a^e	13.2	-
HF^c	81.1	n/a^e	0	72.8	n/a^e	0	0	n/a^e	−70.5	56.9
LDA^c	9.6	n/a^e	0	−22.4	n/a^e	0	0	n/a^e	61.9	42.8
BP86^c	28.4	n/a^e	0	5.1	n/a^e	0	0	n/a^e	23.2	14.5
RPBE ^{f}	34.3	n/a^e	0	6.3	n/a^e	0	0	n/a^e	29.9	14.3
B3LYP^c	33.1	n/a^e	0	14.1	n/a^e	0	0	n/a^e	−0.6	11.2
PBE0^c	46.0	n/a^e	0	24.7	n/a^e	0	0	n/a^e	−9.0	9.1
OPBE^g	48.6	36.8	0	19.5	35.2	0	0	25.9	13.3	1.0
OPBE^h	49.3	37.0	0	19.0	35.1	0	0	26.5	14.9	1.9

^{a} Relative to experimental spin ground-state, using D_3 symmetry for the ammonia and bipyridine complexes, and C_i symmetry with water as ligand. ^{b} Mean absolute deviation (MAD) of ΔE_{HL} with respect to reference CASPT2 data. ^{c} From ref 69. ^{d} CAS[10,12] space with atomic natural orbitals (ANO) basis sets, contracted to [7s6p5d3f2g1h] on Fe, [4s3p2d1f] on N and O, and [3s1p] on H. ^{e} Not available, triplet state was not considered in refs 69 and 71. ^{f} From ref 71. ^{g} This work, obtained with TZP (STO) basis set. ^{h} This work, obtained with TZ2P (STO) basis set.

Table 3. Spin-State Energies (ΔE, kcal mol^{-1})^a and Iron-Ligand Distances (R, Å) for Small Iron Complexes

	low spin		interm spin		high spin	
compd	ΔE	R	ΔE	R	ΔE	R
Fe(II)F$_4^{2-b}$	81.9	1.980	42.3	2.015	0	2.021
Fe(II)Cl$_4^{2-b}$	80.6	2.310	39.3	2.363	0	2.381
Fe(II)Cl$_4^{2-c}$	80.4	2.237, 2.321	36.1	2.331, 2.361	0	2.383, 2.383
Fe(II)Br$_4^{2-b}$	71.0	2.446	37.8	2.526	0	2.545
Fe(II)(CN)$_6^{4-d}$	0	1.919	51.8	2.126	38.7	2.419
Fe(III)F$_4^{1-b}$	75.4	1.798	53.4	1.828	0	1.842
Fe(III)Cl$_4^{1-b}$	60.1	2.144	40.8	2.195	0	2.218
Fe(III)Br$_4^{1-b}$	57.4	2.306	37.3	2.370	0	2.386
Fe(III)(CN)$_6^{3-d}$	0	1.925	48.2	2.092	45.9	2.234
Fe(III)(NH$_3$)$_6^{+3e}$	7.4	2.055	18.8	2.173	0	2.239
Fe(VI)O$_4^{2-b}$	17.2	1.663	0	1.663	50.1	1.728

^{a} Relative to experimental spin ground-state. ^{b} Using T_d symmetry. ^{c} Using C_{2v} symmetry. ^{d} Using O_h symmetry. ^{e} Using D_3 symmetry.

distortions to take place. For the bipyridine complex, this indeed does seem to be the case, but the energy gain is only about 100−200 cm^{-1} (ca. 0.3−0.6 kcal·mol^{-1}) for the quintet state of the bipyridyl complex.[69] For the ammonia complex, the Jahn−Teller distortions seem to lead to similarly small energy differences; however, in this case the energy of the quintet state goes up in energy at the CASPT2[10,12]//PBE0 level.[69] In any case, here the comparison is made with the best available reference CASPT2 energies, for which Jahn−Teller distortions were not taken into account. The conformations of the ammonia and water ligands in these calculations correspond to those used in the CASPT2 study by Pierloot and co-workers, which therefore enables a fair comparison between the OPBE and CASPT2 results. Furthermore, the vibrational frequencies of these complexes were calculated, which resulted in all-positive frequencies. For all three complexes OPBE predicts the correct spin ground state, i.e. a quintet for Fe(H$_2$O)$_6^{2+}$ and Fe(NH$_3$)$_6^{2+}$ and a singlet for Fe(bpy)$_3^{2+}$. Moreover, the deviation from the reference CASPT2 data is rather small with values of 1.0 kcal mol^{-1} (TZP basis) and 1.9 kcal mol^{-1} (TZ2P basis), i.e. a significant reduction by an order of magnitude compared to the values from the other functionals. The OPBE deviations with both basis sets fall well within the estimated accuracy of the CASPT2 data ($\sim$1000 cm^{-1} or 2.9 kcal mol^{-1}) and therefore clearly show the excellent performance of the OPBE functional.

Spin States of Small Iron Complexes. Next, we investigated a set of small molecules that had been studied previously (partly) by Noodleman,[37] Deeth,[71] Neese,[74] and Filatov[75] (among others) for the description of Mössbauer spectroscopy of these complexes. The predicted spin-state splittings for some of the small complexes by OPBE are reported in Table 3, which confirms the correctness[37] of the OPBE functional for providing spin ground states of iron complexes. Surprisingly, the correct description of the spin ground-state of most of these molecules is given by all DFT functionals, even though some are high-spin and other low-spin. The only exception is the Fe(III)(NH$_3$)$_6^{3+}$ complex, which has a high-spin ground-state experimentally and with OPBE, but for which other functionals were shown to fail by Deeth and Fey.[71] Not surprisingly, it was the standard pure DFT functionals, which tend to favor low-spin states, that failed. We will return to this issue (*vide infra*) when looking at which factors determine the actual spin ground-state of these iron complexes.

For these small and highly symmetric molecules, there are several spin-states that are formally Jahn−Teller active, which means that symmetry-lowering may result in more favorable energies for these spin-states. When using the QUILD program, this symmetry-lowering can consist of two parts, because one can separate the geometric symmetry from the electronic (orbitals) symmetry. For instance for Fe(II)-Cl$_4^{2-}$, the spin-states can be studied using (i) T_d symmetry for geometry and orbitals, (ii) T_d symmetry for geometry and C_{2v} symmetry for orbitals, or (iii) C_{2v} symmetry for both geometry and orbitals. Note that this symmetry-lowering is in particular beneficial for the intermediate (triplet) state,

Scheme 2. Challenging Iron Complexes Fe(phen)$_2$(NCS)$_2$, Fe(amp)$_2$Cl$_2$ (amp = 2-pyridylmethylamine), Fe(dpa)$_2^{2+}$ (dpa = di(2-pyridylmethyl)amine), and Fe([9]aneN$_3$)$_2$[a]

Fe(phen)$_2$(NCS)$_2$ **Fe(amp)$_2$Cl$_2$**

Fe([9]aneN$_3$)$_2$ **Fe(dpa)$_2^{2+}$**
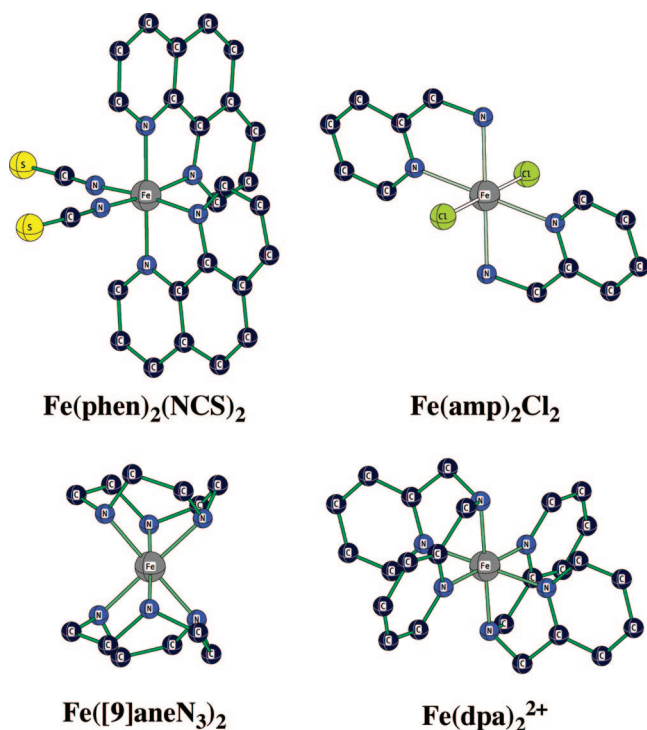
[a] Hydrogens were omitted for clarity.

which is spin-contaminated within T_d symmetry and a pure spin-state within C_{2v} symmetry. However, the energy that is gained by symmetry-lowering is only ca. 3 kcal·mol^{-1} (see Table 3).

Spin-State Splittings of Challenging Iron Complexes: Fe(phen)$_2$(NCS)$_2$. Reiher and co-workers[16] reparameterized the B3LYP functional to include only 15% of HF exchange (dubbed B3LYP*), based on the spin-state splittings for a number of Fe(II) complexes. Although this reduction of the amount of HF exchange does not seem to affect the performance for organic molecules,[17] and seems to be an improvement over B3LYP,[17] it still fails for some molecules.[18,19] Previously, it was already shown that B3LYP completely fails for spin-crossover compounds,[76] i.e. it usually predicts a high-spin ground-state, although at low temperatures a low-spin is observed experimentally. One typical example for which B3LYP (and B3LYP*) was shown to fail is the spin-crossover compound Fe(phen)$_2$(NCS)$_2$ (see Scheme 2).[18] The experimental value for the energy splitting between the low and high-spin state (ΔE_{HL}) is estimated to be of the order of 3 kT (ca. +1.8 kcal mol^{-1}) at 0 K, in order for the thermal spin transition to be viable. B3LYP and B3LYP* wrongly predict a high-spin state, with a ΔE_{HL} splitting of −8.0 kcal mol^{-1} (B3LYP) and −1.5 kcal mol^{-1} (B3LYP*).[18] The OPBE calculations, on the other hand, correctly show a low-spin ground-state for this complex, with the high-spin state higher in energy by 2.1 kcal mol^{-1}. This OPBE ΔE_{HL} splitting of 3.5 kT is therefore in perfect agreement with the estimated experimental value (*vide supra*).

Table 4. Spin-State Splittings (kcal mol^{-1})[a] for Challenging Iron Complexes

compd	low spin	interm spin	high spin
Fe(phen)$_2$(NCS)$_2$	0	15.2 (19.3)	2.1 (6.7)
Fe(amp)$_2$Cl$_2$[b]	8.5 (10.7)	13.3 (7.9)	0
Fe(dpa)$_2^{2+c}$	0	11.8 (15.0)	2.3 (6.5)
Fe([9]aneN$_3$)$_2^{3+}$	0	13.8 (17.0)	2.9 (9.4)

[a] Relative to experimental spin ground-state, in parentheses the values when the solvent (methanol) is included. [b] amp = 2-pyridylmethylamine. [c] dpa = di(2-pyridylmethyl)amine.
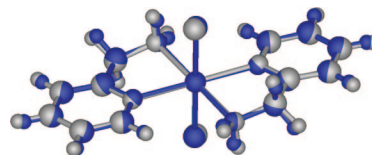
Figure 2. Overlay of experimental X-ray (in blue) and OPBE optimized (in gray) structure of Fe(amp)$_2$Cl$_2$.

Apart from the B3LYP and B3LYP* functionals, which were shown by Reiher[18] to fail for this complex, we were also interested in the performance of other recent functionals such as X3LYP[77] and M06.[78] Therefore, single-point energy calculations were performed where the energies of a large number of DFT functionals are calculated simultaneously (see Table S3 in the Supporting Information). These calculations showed that all hybrid functionals, except TPSSh[79,80] (with a ΔE_{HL} of +5.6 kcal mol^{-1}, i.e. ca. 10 kT), failed to provide the low-spin ground-state. Most standard (pure) functionals do provide the correct spin ground-state, but because they overstabilize low-spin states, they predict a too large energy separation between the low- and high-spin. In other words, they fail to describe the molecule as a spin-crossover compound. It should also be noted that the failure to predict the correct spin ground-state is not limited to hybrid functionals only. Also pure functionals like OLYP, HCTH, VS98, Becke00, OLAP3, and M06-L wrongly predict a high-spin ground-state (see Table S3).

Spin-State Splittings of Challenging Iron Complexes: Pyridylmethylamines. Westerhausen and co-workers[81] reported a series of iron complexes based on pyridylmethylamine ligands, which readily form complexes with iron halides in methanol. Here, we focus on two of the complexes reported in that study, Fe(amp)$_2$Cl$_2$ (amp = 2-pyridylmethylamine) and Fe(dpa)$_2^{2+}$ (dpa = di(2-pyridylmethyl)amine) (see Scheme 2). It was shown that Fe(amp)$_2$Cl$_2$ has a high-spin ground-state while Fe(dpa)$_2^{2+}$ has a low-spin ground-state. Both complexes are structurally highly similar, i.e. they have approximately an octahedral arrangement of the ligands around iron, and the only change in going from the monopma to the dipma complex is the replacement of two chloride ligands by pyridines.

We have optimized the geometries of the three spin-states for both these complexes, which result in the experimentally observed spin ground-state, i.e. a singlet for the dpa-complex and a quintet for the amp-complex (see Table 4). Also the obtained structures are in good agreement with the crystal structures (see Figure 2 for an overlay of the experimental and computed structures). For the amp-complex in the gas-phase, the Fe−Cl distances are a bit shorter (2.37 Å) and

Accurate Spin-State Energies

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2063**

the Fe−N distances a bit longer (2.27−2.31) than observed in the crystal structure, where they are found at 2.50 and 2.19 Å respectively. However, by including the methanol solvent in the calculations (through a dielectric continuum model, COSMO) the iron-ligand distances match the ones from the crystal structure perfectly with values of 2.53 Å (Fe−Cl), 2.20 Å (Fe−N$_{pyr}$), and 2.23 Å (Fe−N$_{am}$). I.e., the molecule is severely Jahn−Teller distorted but in a similar fashion within the experiments and as optimized by OPBE. In the optimizations, the magnitude and orientation of the Jahn−Teller distortions came out the same, irrespective of the starting structure and initial distortions applied to it (see also Figure 2). The Fe−N distances in the dpa-complex are computed to be almost independent of the presence of the solvent, with values of 1.99 Å (Fe−N$_{pyr}$) and 2.01 Å (Fe−N$_{am}$) in the gas-phase and 1.98 Å (Fe−N$_{pyr}$) and 2.00 Å (Fe−N$_{am}$) in methanol. These distances are in excellent agreement with the crystal structure that puts them at 1.99 Å (Fe−N$_{pyr}$) and 2.03 Å (Fe−N$_{am}$). The origin of the change from high- to low-spin when going from the amp-complex to the dpa-complex is discussed below.

For these two complexes we also investigated the performance of a number of DFT functionals, with different results for each one (see Table S3 in the Supporting Information). For the high-spin complex Fe(amp)$_2$Cl$_2$, the standard DFT functionals like LDA, BP86, BLYP, or PBE fail and predict a low-spin ground-state instead. On the other hand, for the low-spin complex Fe(dpa)$_2^{2+}$, the hybrid functionals and some other ones (HCTH, OLAP3) wrongly predict a high-spin ground-state. In fact, apart from OPBE, OPerdew, and TPSSh, no other functional is able to correctly predict the spin ground-state of these challenging iron complexes. Most notably is the failure of XLYP, X3LYP (which were claimed to be the best functionals available for studying spin-state splittings[77]), and the M06 functionals, which predict the wrong spin ground-state for these complexes.

Spin-State Splittings of Challenging Iron Complexes: Cyclononanes. Wieghardt and co-workers[82] reported some time ago the first examples of low-spin iron complexes with all six donors as saturated nitrogens, which was based on the 1,4,7-triazacyclononane ligand [9]aneN$_3$. Here, we investigate the Fe(III) complex with two [9]aneN$_3$ ligands (see Scheme 2). As expected, OPBE correctly predicts a low-spin ground-state for this complex with the intermediate and high spin-state higher in energy by 14 and 3 kcal mol^{-1} in the gas-phase, which increase to 17 and 10 kcal mol^{-1} with the solvent present. The optimized structure with Fe−N distances of 2.03 Å in the gas phase and with 2.00 Å in methanol is also in perfect agreement with the crystal structure, that shows Fe−N distances of 1.98−2.01 Å. Therefore, like all other complexes, we see a consistent trend for the study on these challenging iron complexes, where the OPBE functional gives reliable and accurate results.

Rationalization of the Factors That Determine the Spin Ground-State of Iron Complexes. Now that the reliability of the OPBE functional for providing the spin ground-state of transition-metal complexes is established, the question remains to determine the factors that govern the spin ground-state of these transition-metal complexes. For

Table 5. Energy Decomposition Analysis (kcal mol^{-1}) for Small Complexes[a]

	Fe(II)Cl$_4^{2-b}$			Fe(II)CN$_6^{4-c}$		
	singlet	triplet[d]	quintet	singlet	triplet	quintet
ΔE_{prep}	618.4	567.7	493.0	1346.1	1217.9	1047.0
ΔE_{deform}	-	-	-	0.3	0.1	0.0
$\Delta E_{lig-lig}$	504.8	494.7	490.9	1232.3	1145.1	1044.9
ΔE_{valexc}	113.6	73.0	2.1	113.5	72.7	2.1
ΔE_{int}	−1134.5	−1109.6	−1090.1	−1722.2	−1541.6	−1383.6
ΔE_{Pauli}	160.6	138.7	127.1	320.5	173.9	73.2
ΔE_{elstat}	−1045.8	−1027.0	−1019.7	−1367.5	−1287.2	−1197.7
ΔE_{orbint}	−249.3	−221.1	−197.6	−675.3	−428.2	−259.1
A1	−117.0	−109.2	−91.1	−470.6	−278.0	−151.5
A2	−14.1	−18.1	−13.1	−30.4	−18.0	−10.8
B1	−82.2	−47.1	−46.7	−87.1	−64.8	−48.4
B2	−36.0	−46.7	−46.7	−87.1	−67.4	−48.4
ΔE_{total}	−516.1	−541.9	−597.1	−376.1	−323.7	−336.6

[a] Using C_{2v} symmetry for orbitals. [b] Using T_d symmetry for geometry. [c] Using O_h symmetry for geometry. [d] Spin-projection applied (see the Supporting Information).

Table 6. Energy Decomposition Analysis (kcal mol^{-1}) for Challenging Complexes[a]

	Fe(amp)$_2$Cl$_2^b$			Fe(dpa)$_2^{2+c}$		
	singlet	triplet	quintet	singlet	triplet	quintet
ΔE_{prep}	265.9	209.0	125.6	196.5	134.6	48.7
ΔE_{deform}	15.4	9.8	8.1	29.5	20.3	14.6
$\Delta E_{lig-lig}$	136.9	126.5	115.4	53.4	41.6	32.0
ΔE_{valexc}	113.6	72.7	2.1	113.6	72.7	2.1
ΔE_{int}	−868.0	−806.5	−736.5	−558.4	−484.6	−408.3
ΔE_{Pauli}	251.7	205.2	136.3	266.5	250.8	133.7
ΔE_{elstat}	−651.5	−623.7	−596.7	−336.1	−309.1	−258.6
ΔE_{orbint}	−468.3	−387.9	−276.2	−488.9	−426.4	−283.4
A$_g$	−333.3	−262.5	−168.1	−345.5	−296.7	−169.5
A$_u$	−135.0	−125.5	−108.1	−143.4	−129.7	−113.9
ΔE_{total}	−602.1	−597.5	−610.9	−361.9	−350.0	−359.6

[a] Using C_i symmetry for orbitals. [b] amp = 2-pyridylmethylamine. [c] dpa = di(2-pyridylmethyl)amine.

instance, why does the spin ground-state of the pyridylmethylamine complexes change from high-spin to low-spin if a chloride-ligand is changed by pyrimidine? And why do standard functionals have problems with predicting the high-spin ground-state of the complicated systems yet not with those for the small complexes?

In order to gain further insight, the bonding mechanism of a number of typical iron complexes from this study was analyzed in terms of an energy decomposition analysis of the Kohn−Sham molecular orbitals. The energy decomposition is reported in Table 5 for two small complexes, with a high-spin (Fe(II)Cl$_4^{2-}$) and a low-spin (Fe(II)CN$_6^{4-}$) ground-state, and in Table 6 for the pyridylmethylamine complexes. For the two small complexes, there is a striking difference in the preparation energy that is more than twice as large for Fe(II)CN$_6^{4-}$ than for Fe(II)Cl$_4^{2-}$. However, this energy difference results only from the ligand-ligand interactions ($\Delta E_{lig-lig}$), i.e. the electrostatic repulsion between the negatively charged ligands. Because the cyanide complex has six of them, their mutual repulsion is much larger. More importantly, this repulsive interaction is largest for the low-spin state, by 14 kcal·mol^{-1} compared to the quintet state for the chloride complex and by ca. 187 kcal·mol^{-1} for the cyanide complex. It results from the shorter metal−ligand

bond distances in the low spin-states, in which antibonding d-orbitals on iron are not occupied, in contrast to the higher spin-states where they are partially occupied.

The larger preparation energy for low spin-states is reinforced by Hund's rule of maximum multiplicity, which shows up in the valence excitation energy (ΔE_{valexc} in Tables 56). Hund's rule says that for atoms the state with the highest multiplet is lowest in energy, i.e. for the isolated iron(II) cation the quintet state is most favorable. Therefore, in order to make the low-spin iron complex, the iron(II) cation has to be changed from its favorable quintet state to the unfavorable singlet state. This change in multiplet state of the isolated cation (but with the occupation of the orbitals it attains in the molecule, see the Computational Details section) costs around 114 kcal·mol^{-1} (see Table 5). Taken together with the (repulsive) ligand-ligand interactions, the low spin-state has a much larger preparation energy than higher spin-states, of ca. 125 kcal·mol^{-1} for the chloride complex and ca. 300 kcal·mol^{-1} for the cyanide complex.

The interaction between the iron cation and its ligands on the other hand favors the low spin-states considerably. For Fe(II)Cl$_4$$^{2-}$, the interaction energy (ΔE_{int}) is 44 kcal·mol^{-1} larger for the singlet than for the quintet state (see Table 5), which results almost entirely (52 kcal·mol^{-1}) from covalent interactions (ΔE_{orbint}). The sum of the Pauli repulsion (ΔE_{Pauli}) and electrostatic interactions (ΔE_{elstat}) contributes only ca. 8 kcal·mol^{-1} to this energy difference. However, the difference in interaction energy of ca. 44 kcal·mol^{-1} in favor of the low spin-state is not sufficient to overcome the difference in preparation energy (125 kcal·mol^{-1}), and as a result this chloride complex has a high-spin ground-state. For the Fe(II)CN$_6$$^{4-}$ complex, the interaction energy (ΔE_{int}) is significantly larger than that for the chloride complex (ca. 588 kcal·mol^{-1} for the singlet state, see Table 5) and is ca. 338 kcal·mol^{-1} larger for the low spin-state compared to the high spin-state. Again, this difference results mainly from the covalent interactions (ΔE_{orbint}) which are 416 kcal·mol^{-1} more favorable for the singlet state, with a much smaller contribution (78 kcal·mol^{-1}) from the steric interactions (sum of ΔE_{Pauli} and ΔE_{elstat}). Therefore, for the cyanide complex the favorable difference in interaction energy is sufficiently large to overcome the unfavorable difference in preparation energy between the singlet and quintet, and hence a low-spin ground-state is observed with the quintet higher in energy by ca. 40 kcal/mol.

The interplay between the interaction energy and the preparation energy is for these small complexes severely unbalanced, which results in a clearly defined spin ground-state for either of these two small complexes. Consequently, the other DFT functionals do provide the correct ground-state preferences. This however changes for the challenging complexes where the interplay between interaction energy versus preparation energy is more subtle.

For the pyridylmethylamine complexes Fe(amp)$_2$Cl$_2$ and Fe(dpa)$^{2+}$, both the absolute values of the interaction energies as well as the differences of them for the three spin states is much smaller than for the small complexes (see Table 6). Similar to these latter complexes, also for the pyridylmethylamine complexes is the preparation energy the largest for

the low-spin singlet state. For all three components of the preparation energy (ΔE_{deform}, $\Delta E_{lig-lig}$, ΔE_{valexc}) is the largest value observed for the low spin-state. This larger preparation energy for the low spin-state is counteracted by a larger interaction energy, in a similar fashion to the small complexes (*vide supra*). The difference in interaction energy ($\Delta\Delta E_{int}$) results again mainly from the covalent interactions ($\Delta\Delta E_{orbint}$, 192 and 206 kcal·mol^{-1}, see Table 6), with a smaller contribution from the steric interactions (sum of $\Delta\Delta E_{Pauli}$ and $\Delta\Delta E_{elstat}$) of ca. 55–60 kcal·mol^{-1}. The more favorable orbital interactions occur mainly in those irreps (A$_1$ for the small complexes, A$_g$ here) that contain the iron d-orbitals, which are unoccupied in the low-spin and partially occupied in high-spin states.

The balance between the preparation of the fragments, and the subsequent interaction between these, then determines the actual spin state observed. Thus, for Fe(amp)$_2$Cl$_2$ the interaction energy difference ($\Delta\Delta E_{int}$, see Table 6) between singlet and quintet (131 kcal·mol^{-1}) is too small to overcome the difference in preparation energy ($\Delta\Delta E_{prep}$, 140 kcal·mol^{-1}), and hence this molecule has a high-spin ground-state. On the other hand, for Fe(dpa)$^{2+}$ the interaction energy difference (150 kcal·mol^{-1}) is sufficiently large to overcome the unfavorable preparation energy ($\Delta\Delta E_{prep}$, 148 kcal·mol^{-1}), and the molecule has a low-spin ground-state. Therefore, the spin ground-state of these iron complexes is determined completely by a delicate compromise[19] between Hund's rule of maximum multiplicity, which favors high spin-states, and metal–ligand bonding that favors low spin-states.

Conclusions

The performance of the OPBE functional has been checked for the geometries and relaxed (adiabatic) spin-state energies of transition-metal complexes. The performance for the geometry was checked for the bond-lengths of transition-metal (di)halides, which was shown[68] to be representative for a much larger and more diverse set of 32 metal complexes. The accuracy of the OPBE bond lengths is close to that of the best-performing DFT functionals and significantly better than most others such as B3LYP.

The performance of OPBE for the relaxed spin-state energies has been checked for a set of small complexes, a set of benchmark systems where highly accurate CASPT2 energies are available as reference, and a set of challenging complexes such as the spin-crossover compound Fe(phen)$_2$-(NCS)$_2$ and pyridylmethylamine complexes. The failure of hybrid functionals such as B3LYP and B3LYP* for these low-spin complexes has been reported before[18] and is confirmed here, including for other (newer) functionals such as X3LYP and M06. In contrast, the OPBE functional gives excellent results in all cases. For the set of benchmark systems, the difference between the reference CASPT2 data and the OPBE energies (1–2 kcal·mol^{-1}) is an order of magnitude smaller than those of other functionals (9–15 kcal·mol^{-1}), and it lies well within the estimated accuracy (3 kcal·mol^{-1}) of the reference CASPT2 data.

In order to gain insight in the factors that determine the spin ground-state of transition-metal complexes, the chemical bonding for a number of iron complexes is analyzed in terms

Accurate Spin-State Energies

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2065**

of an energy decomposition analysis. From this analysis, it becomes clear that two opposing forces act on the metal, one that prefers a high spin-state (Hund's rule of maximum multiplicity) which is counteracted by the metal−ligand bonding that prefers low spin-states. The interplay between these two opposing effects then determines the spin ground-state of the metal-complex.

Supporting Information Available: Details of the spin-projection for the triplet state of $Fe(II)Cl_4^{2-}$, interaction and preparation energies corrected for self-consistency of the metal cation (see computational details), additional energy decomposition analyses, and Cartesian coordinates of the challenging complexes. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Swart, M.; Groenhof, A. R.; Ehlers, A. W.; Lammertsma, K. *J. Phys. Chem. A* **2004**, *108*, 5479.

(2) Sellmann, D.; Soglowek, W.; Knoch, F.; Ritter, G.; Dengler, J. *Inorg. Chem.* **1992**, *31*, 3711.

(3) Limberg, C. *Angew. Chem., Int. Ed.* **2003**, *42*, 5932.

(4) Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. *Chem. Rev.* **2005**, *105*, 2253.

(5) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279.

(6) Groenhof, A. R.; Ehlers, A. W.; Lammertsma, K. *J. Am. Chem. Soc.* **2007**, *129*, 6204.

(7) *Spin Crossover in Transition Metal Complexes I-III*; Gütlich, P., Goodwin, H. A., Eds.;Springer: Berlin/Heidelberg, Germany, 2004; Vols. 233−235 of Topics in Current Chemistry series.

(8) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, Germany, 2000.

(9) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys* **1980**, *58*, 1200.

(10) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(11) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(12) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(13) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *45*, 11623.

(14) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(15) Perdew, J. P.; Emzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982.

(16) Reiher, M.; Salomon, O.; Hess, B. A. *Theor. Chem. Acc.* **2001**, *107*, 48.

(17) Salomon, O.; Reiher, M.; Hess, B. A. *J. Chem. Phys.* **2002**, *117*, 4729.

(18) Reiher, M. *Inorg. Chem.* **2002**, *41*, 6928.

(19) Swart, M. *Inorg. Chim. Acta* **2007**, *360*, 179.

(20) Güell, M.; Luis, J. M.; Solà, M.; Swart, M. *J. Phys. Chem. A* **2008**, *112*, 6384.

(21) Harvey, J. N. *Struct. Bonding (Berlin)* **2004**, *112*, 151.

(22) Conradie, J.; Ghosh, A. *J. Chem. Theory Comput.* **2007**, *3*, 689.

(23) Conradie, J.; Ghosh, A. *J. Phys. Chem. B* **2007**, *111*, 12621.

(24) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403.

(25) Swart, M.; Ehlers, A. W.; Lammertsma, K. *Mol. Phys.* **2004**, *102*, 2467.

(26) Zhang, Y.-Q.; Luo, C.-L. *J. Phys. Chem. A* **2006**, *110*, 5096.

(27) Zhang, Y.; Wu, A.; Xu, X.; Yan, Y. *Chem. Phys. Lett.* **2006**, *421*, 383.

(28) Wu, A.; Zhang, Y.; Xu, X.; Yan, Y. *J. Comput. Chem.* **2007**, *28*, 2431.

(29) Wasbotten, I.; Ghosh, A. *Inorg. Chem.* **2006**, *45*, 4910.

(30) Derat, E.; Kumar, D.; Neumann, R.; Shaik, S. *Inorg. Chem.* **2006**, *45*, 8655.

(31) Romo, S.; Fernández, J. A.; Maestre, J. M.; Keita, B.; Nadjo, L.; de Graaf, C.; Poblet, J. M. *Inorg. Chem.* **2007**, *46*, 4022.

(32) Rong, C.; Lian, S.; Yin, D.; Shen, B.; Zhong, A.; Bartolotti, L.; Liu, S. *J. Chem. Phys.* **2006**, *125*, 174102.

(33) Liao, M.-S.; Watts, J. D.; Huang, M.-J. *J. Comput. Chem.* **2006**, *27*, 1577.

(34) Liao, M.-S.; Watts, J. D.; Huang, M.-J. *J. Phys. Chem. A* **2007**, *111*, 5927.

(35) Zein, S.; Borshch, S. A.; Fleurat-Lessard, P.; Casida, M. E.; Chermette, H. *J. Chem. Phys.* **2007**, *126*, 014105.

(36) Swart, M.; Solà, M.; Bickelhaupt, F. M. *J. Comput. Chem.* **2007**, *28*, 1551.

(37) Han, W. G.; Noodleman, L. *Inorg. Chim. Acta* **2008**, *361*, 973.

(38) Han, W. G.; Noodleman, L. *Inorg. Chem.* **2008**, *47*, 2975.

(39) Ghosh, A. *J. Biol. Inorg. Chem.* **2006**, *11*, 671.

(40) Ghosh, A. *J. Biol. Inorg. Chem.* **2006**, *11*, 712.

(41) Tangen, E.; Conradie, J.; Ghosh, A. *J. Chem. Theory Comput.* **2007**, *3*, 448.

(42) Conradie, J.; Quarless, D. A.; Hsu, H. F.; Harrop, T. C.; Lippard, S. J.; Koch, S. A.; Ghosh, A. *J. Am. Chem. Soc.* **2007**, *129*, 10446.

(43) Wasbotten, I. H.; Ghosh, A. *Inorg. Chem.* **2007**, *46*, 7890.

(44) Conradie, J.; Ghosh, A. *J. Chem. Theory Comput.* **2007**, *3*, 689.

(45) Conradie, J.; Wondimagegn, T.; Ghosh, A. *J. Phys. Chem. B* **2008**, *112*, 1053.

(46) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931.

(47) Baerends, E. J.; Autschbach, J.; Bérces, A.; Berger, J. A.; Bickelhaupt, F. M.; Bo, C.; de Boeij, P. L.; Boerrigter, P. M.; Cavallo, L.; Chong, D. P.; Deng, L.; Dickson, R. M.; Ellis, D. E.; van Faassen, M.; Fan, L.; Fischer, T. H.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Groeneveld, J. A.; Gritsenko, O. V.; Grüning, M.; Harris, F. E.; van den Hoek, P.; Jacob, C. R.; Jacobsen, H.; Jensen, L.; Kadantsev, E. S.; van Kessel, G.; Klooster, R.; Kootstra, F.; van Lenthe, E.; McCormack, D. A.; Michalak, A.; Neugebauer, J.; Nicu, V. P.;

2066   *J. Chem. Theory Comput., Vol. 4, No. 12, 2008*

Osinga, V. P.; Patchkovskii, S.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Romaniello, P.; Ros, P.; Schipper, P. R. T.; Schreckenbach, G.; Snijders, J. G.; Solà, M.; Swart, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visscher, L.; Visser, O.; Wang, F.; Wesolowski, T. A.; van Wezenbeek, E. M.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; Yakovlev, A. L.; Ziegler, T. *ADF 2007.01*; SCM: Amsterdam, The Netherlands, 2007.

(48) van Lenthe, E.; Baerends, E. J. *J. Comput. Chem.* **2003**, *24*, 1142.

(49) Swart, M.; Snijders, J. G. *Theor. Chem. Acc.* **2003**, *110*, 34.

(50) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(51) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533.

(52) Swart, M.; Bickelhaupt, F. M. *Int. J. Quantum Chem.* **2006**, *106*, 2536.

(53) Swart, M.; Bickelhaupt, F. M. *J. Comput. Chem.* **2008**, *29*, 724.

(54) Bylaska, E. J.; de Jong, W. A.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Apra, E.; Windus, T. L.; 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.; 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*; Pacific Northwest National Laboratory: Richland, WA, 2006.

(55) Environmental and Molecular Sciences Laboratory. Basis set order form. http://www.emsl.pnl.gov/forms/basisform.html (accessed Feb 8, 2008).

(56) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.

(57) Pye, C. C.; Ziegler, T. *Theor. Chem. Acc.* **1999**, *101*, 396.

(58) Bon, R. S.; van Vliet, B.; Sprenkels, N. E.; Schmitz, R. F.; de Kanter, F. J. J.; Stevens, C. V.; Swart, M.; Bickelhaupt, F. M.; Groen, M. B.; Orru, R. V. A. *J. Org. Chem.* **2005**, *70*, 3542.

(59) Swart, M.; Rösler, E.; Bickelhaupt, F. M. *Eur. J. Inorg. Chem.* **2007**, 3646.

(60) Bickelhaupt, F. M.; Baerends, E. J. In *Reviews in Computational Chemistry, Vol. 15*; Wiley-VCH: New York, 2000; Vol. 15, p 1.

(61) Baerends, E. J.; Branchadell, V.; Sodupe, M. *Chem. Phys. Lett.* **1997**, *265*, 481.

(62) Fouqueau, A.; Mer, S.; Casida, M. E.; Daku, L. M. L.; Hauser, A.; Mineva, T.; Neese, F. *J. Chem. Phys.* **2004**, *120*, 9473.

(63) Morokuma, K. *Acc. Chem. Res.* **1977**, *10*, 294.

(64) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1558.

(65) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1755.

(66) Groenhof, A. R.; Swart, M.; Ehlers, A. W.; Lammertsma, K. *J. Phys. Chem. A* **2005**, *109*, 3411.

(67) Wittbrodt, J. M.; Schlegel, H. B. *J. Chem. Phys.* **1996**, *105*, 6574.

(68) Bühl, M.; Kabrede, H. *J. Chem. Theory Comput.* **2006**, *2*, 1282.

(69) Pierloot, K.; Vancoillie, S. *J. Chem. Phys.* **2006**, *125*, 124303.

(70) Pierloot, K.; Vancoillie, S. *J. Chem. Phys.* **2008**, *128*, 034104.

(71) Deeth, R. J.; Fey, N. *J. Comput. Chem.* **2004**, *25*, 1840.

(72) Fouqueau, A.; Casida, M. E.; Daku, L. M. L.; Hauser, A.; Neese, F. *J. Chem. Phys.* **2005**, 122.

(73) Lawson Daku, L. M.; Vargas, A.; Hauser, A.; Fouqueau, A.; Casida, M. E. *ChemPhysChem* **2005**, *6*, 1393.

(74) Neese, F. *Inorg. Chim. Acta* **2002**, *337*, 181.

(75) Filatov, M. *J. Chem. Phys.* **2007**, *127*, 084101.

(76) Paulsen, H.; Duelund, L.; Winkler, H.; Toftlund, H.; Trautwein, A. X. *Inorg. Chem.* **2001**, *40*, 2201.

(77) Xu, X.; Goddard III, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673.

(78) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

(79) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(80) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129.

(81) Malassa, A.; Görls, H.; Buchholz, A.; Plass, W.; Westerhausen, M. *Z. Anorg. Allg. Chem.* **2006**, *632*, 2355.

(82) Boeyens, J. C. A.; Forbes, A. G. S.; Hancock, R. D.; Wieghardt, K. *Inorg. Chem.* **1985**, *24*, 2926.

CT800277A

JCTC Journal of Chemical Theory and Computation

Description of Phosphate Hydrolysis Reactions with the Self-Consistent-Charge Density-Functional-Tight-Binding (SCC-DFTB) Theory. 1. Parameterization

Yang Yang,[†] Haibo Yu,[†] Darrin York,[‡] Marcus Elstner,*,[§] and Qiang Cui*,[†]

Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, 1101 University Avenue, Madison, Wisconsin 53706, Department of Chemistry, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455, and Department of Physical and Theoretical Chemistry, TU Braunschweig, Hans-Sommer-Strasse 10, D-38106 Braunschweig, Germany

Abstract: Phosphate chemistry is involved in many key biological processes, yet the underlying mechanism often remains unclear. For theoretical analysis to effectively complement experimental mechanistic analysis, it is essential to develop computational methods that can capture the complexity of the underlying potential energy surface and allow for sufficient sampling of the configurational space. To this end, we report the parametrization of an approximate density functional theory, the Self-Consistent-Charge Density-Functional Tight-Binding (SCC-DFTB) method for systems containing phosphorus. Compared to high-level density functional theory and ab initio (MP2 and G3B3) results, the standard second-order parametrization is shown to give reliable structures for a diverse set of phosphate compounds but inaccurate energetics. With the on-site third-order terms included, referred to as SCC-DFTBPA, calculated proton affinities of phosphate compounds are substantially improved, although it remains difficult to obtain reliable proton affinity for both phosphates and compounds that do not contain phosphorus, indicating that further improvement in the formulation of SCC-DFTB is still a challenge to meet. To make SCC-DFTB applicable to phosphate reactions in the current (on-site-third-order-only) formulation, a "reaction-specific" parametrization, referred to as SCC-DFTBPR, is developed based on hydrolysis reactions of model phosphate species. Benchmark calculations in both the gas phase and solution phase indicate that SCC-DFTBPR gives reliable structural properties and semiquantitative energetics for phosphate hydrolysis reactions. Since the number of reaction-specific parameters is small, it is likely that SCC-DFTBPR is applicable to a broad set of phosphate species. Indeed, for 56 reaction exothermicities and 47 energy barriers related to RNA catalysis model reactions collected from the QCRNA database, which involve molecules rather different from those used to parametrize SCC-DFTBPR, the corresponding root-mean-square difference between SCC-DFTBPR and high-level DFT results is only 5.3 kcal/mol. We hope that the parametrized SCC-DFTB models will complement NDDO based reaction-specific models (e.g., AM1-d/PhoT) and high-level ab initio QM/MM methods in better understanding the mechanism of phosphate chemistry in condensed phase, particularly biological systems.

I. Introduction

Phosphorus is the one of the most abundant elements on earth. It is part of many essential biological components such

* Corresponding author e-mail: cui@chem.wisc.edu (Q.C.) and m.elstner@tu-bs.de (M.E.).

† University of Wisconsin.

‡ University of Minnesota.

§ TU Braunschweig.

as lipids, bones, genetic materials, energy rich molecules (e.g., ATP), and signaling molecules (e.g., GTP).[1,2] Most of the phosphorus in living systems exists in the form of phosphate, and the hydrolysis of phosphate is a key reaction involved in many fundamental life processes such as energy production and signal transduction. In molecular motors, for example, regulation of ATP hydrolysis by the conformational dynamics of the system is the key to the mechanochemical coupling in these amazing "nanomachines".[3−6] Revealing

Table 1. Different Sets of Parameterizations of the SCC-DFTB Approach for Phosphate[a]

notation[b]	reference data[c]	$U^d_{P, O, C, H}$	D_0, Γ_0, Q_0
		Second-Order Parametrizations	
2nd-order	13 small molecules (see main text)	---	---
		Third-Order Parametrizations	
3rd-order	---	−0.07, −0.17, −0.16, −0.16	---
SCC-DFTBPA	proton affinities, 18 P-compounds	−0.07, −0.20, −0.22, −0.23	−0.06, 17.9, 0.86
mix-optimized	proton affinities, 5 P- and 11 non-P compounds	−0.10, −0.15, −0.24, −0.15	−0.08 37.5, 0.83
SCC-DFTBPR	37 phosphate reaction energetics	−0.07, −0.22, −0.24, −0.08	−0.09, 16.1, 0.75

[a] U^d_α is the Hubbard derivative defined in eq 3; $D_0, \Gamma_0,$ and Q_0 are Gaussian parameters defined in eq 4. [b] The notations are used in all tables. [c] "P-compounds" indicate phosphorus containing compounds; "non-P compounds" indicate compounds that do not contain phosphorus. For the list of compounds, see Tables 3 and 4.

the mechanism of phosphate hydrolysis and factors that regulate the hydrolysis activity, therefore, is crucial to the understanding of many essential biological processes.

Unfortunately, phosphate reactions are, in general, fairly complex, and there are many possible reaction pathways.[7−15] Which pathway dominates is expected to depend rather sensitively on the environment.[16] Experimental investigations in this context are complicated by the fact that the interpretation of typical data, such as kinetic isotope effects and linear free energy relations, is often not straightforward.[11,12] This explains why the precise mechanism of phosphate hydrolysis, especially in biomolecules such as phosphatases[17] and ribozymes,[18] remains controversial after many decades of studies and debates. Theoretical studies, in principle, are powerful in complementing experimental work for detailed mechanistic analysis. However, phosphate chemistry poses a major challenge to theory due to the intrinsic complexity and sensitivity to the environment. Employing a hybrid quantum mechanical/molecular mechanical (QM/MM) framework[19−23] is promising but only with a sufficiently reliable QM method, an adequate treatment of the QM/MM interactions, and a sufficient amount of conformational sampling.[24,25]

The importance of conformational sampling makes approximate QM methods such as semiempirical methods based on the Neglect-of-Diatomic-Differential-Overlap (NDDO) approximation uniquely valuable in this context. Unfortunately, the popular NDDO based methods, such as MNDO,[26] AM1,[27] and PM3[28] in general give rather poor results for phosphate reactions and therefore cannot be used without improvements.[29,30] Considering the importance of *d* orbitals in describing the structure and energetics of phosphate compounds, extensions have been made for MNDO[31−33] and AM1[34,30] to include *d* orbitals in the corresponding Hamiltonians; the extension to AM1 was done largely for specific reactions involving phosphoryl transfers rather than a general parametrization. Despite notable improvements[35,36] and successful applications,[37−39] the general results indicate that these methods are still not sufficiently robust for general mechanistic studies.

One important issue relevant to the mechanistic study of phosphate chemistry concerns the prediction of proton affinities. In many elementary steps in the phosphate hydrolysis, for example, protons are transferred between the nucleophile, the phosphate, and the leaving group; proton transfers involving molecules in the nearby environment (such as water molecules that may act as "proton relays")[40−42] have been proposed to play a catalytic role. Therefore,

predicting accurate, or at least balanced, proton affinities for different reactive motifs is essential. In this regards, the popular NDDO approaches require major improvements. In a recent benchmark study by Range et al.,[29] 16 model molecules representing the nucleophiles, phosphate compounds, and leaving groups involved in biologically important phosphoryl transfer reactions have been studied with high level ab initio, density functional theories (DFT) and several semiempirical methods. It was found that all semiempirical methods, which include AM1, PM3, MNDO, MNDO/d, and SCC-DFTB[43] (see below), all have rather large errors in the calculated proton affinities; the root-mean-square errors (RMSE) are 19.1, 13.8, 27.4, 31.0, and 17.4 kcal/mol, respectively. If it is only the proton affinity that is of interest, simple correction schemes can be developed.[29] For the purpose of analyzing reaction mechanisms, however, more sophisticated modifications have to be introduced.

In the past few years, our groups have been actively pursuing the development and application of an approximate density functional theory, the self-consistent-charge density-functional-tight-binding (SCC-DFTB) method, originally proposed by one of us.[43] This was driven by its reasonable balance in computational efficiency (comparable to AM1 and PM3) and accuracy, which is essential to condensed-phase studies. The SCC-DFTB method has been applied successfully to a range of problems involving biomolecules, such as conformational energies of peptides[44−47] and catalysis in several enzymes[25,48−52] Furthermore, the SCC-DFTB approach has been benchmarked for reaction energies, geometries and vibrational frequencies for small molecules in comparison to the G2 approach,[53] and a large set of experimental data for organic molecules.[54,55] An empirical dispersion correction has also been developed,[56] which was found crucial for predicting reliable nucleic acid base-stacking interactions[56] and the relative stability of α and 3_{10} helices in proteins.[57]

Considering those attractive features and the fact that it is straightforward to include *d* orbitals in the method (as been done for sulfur),[58] it seems natural to pursue the parametrization of SCC-DFTB for phosphorus. Another important motivation in this regard is that SCC-DFTB has been recently extended to include specific third-order terms,[25,59−61] which were found to dramatically improve the calculated proton affinities. Since reliable proton affinities are important in the mechanistic analysis of phosphate chemistry, as discussed above, the advantage of SCC-DFTB over other semiempirical methods becomes apparent.

In this work, we report the parametrization of SCC-DFTB for phosphorus. Two sets of parameters have been developed, which work better for proton affinities of phosphate compounds and the hydrolysis reactions of phosphates, respectively, as compared to high-level ab initio calculations. With the current SCC-DFTB model, which includes on-site third-order contributions (vide infra), it seems difficult to describe both classes of properties with high accuracy using a single set of parameters, and the precise reason is under investigation. In the next section, we briefly summarize the SCC-DFTB methodology and the procedures for parametrization. We then present data from the parametrization process and discuss trends in the results, which are followed by additional benchmark calculations of gas-phase models, which have been studied by York et al.[62] using high-level ab initio calculations as well as solution calculations with a QM/MM framework. Finally, we draw a few conclusions. In a separate publication, we further test the robustness of the model in the context of phosphate hydrolysis in solution and enzymes using SCC-DFTB/MM simulations. A brief summary on the performance of the parametrization and the application to ATP hydrolysis in the molecular motor myosin has been reported recently.[5]

II. Computational Methods

In this section, we first briefly review the formulation of SCC-DFTB as used in the parametrization for phosphorus. We then present details regarding the parametrization procedure and additional benchmark systems for further validating the fitted parameters.

A. Theory: SCC-DFTB. As described in detail in several previous publications,[43,63] the standard SCC-DFTB approach is based on a second-order expansion of the density functional theory energy around a reference density, ρ_0

$$E = \sum_i^{occ} \langle \Psi_i | \hat{H}^0 | \Psi_i \rangle + \frac{1}{2} \int \int' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'} \Big|_{\rho_0} \right) \delta\rho\delta\rho' - \frac{1}{2} \int \int' \frac{\rho_0 \prime \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0]\rho_0 + E_{cc} \quad (1)$$

where $\hat{H}^0 = \hat{H}[\rho_0]$ is the effective Kohn–Sham Hamiltonian evaluated at the reference density ρ_0, and the Ψ_i are Kohn–Sham orbitals. E_{xc} and V_{xc} are the exchange-correlation energy and potential, respectively, and E_{cc} is the core–core repulsion energy. With a minimal basis set, a monopole approximation for the second-order term and the two-center approximation to the integrals, the SCC-DFTB total energy is given in the following form

$$E = \sum_{i\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + \frac{1}{2} \sum_{\alpha\beta} U[R_{\alpha\beta}; \rho_0^\alpha, \rho_0^\beta] \quad (2)$$

where $c_{\mu/\nu}^i$ are orbital coefficients, $\Delta q_{\alpha/\beta}$ are the Mulliken charges on atom α/β, and $\gamma_{\alpha\beta}$ is the approximate second-order kernel derived based on two interacting spherical charges. The last pairwise summation gives the so-called repulsive potential term, which is the core–core repulsion plus double counting terms and defined relative to infinitely separated atomic species.

As discussed in our recent work,[25,59−61] it was found that further including the third-order contribution can substantially improve calculated proton affinity for a set of biologically relevant small molecules, even with only the on-site terms included. Since proton affinity is of great relevance to phosphate chemistry, as emphasized above, we choose to adopt the same formulation. The corresponding expression for the SCC-DFTB total energy is[25,59,61]

$$E = \sum_{i\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + \frac{1}{2} \sum_{\alpha\beta} U[R_{\alpha\beta}; \rho_0^\alpha, \rho_0^\beta] + \frac{1}{6} \sum_\alpha U_\alpha^d \Delta q_\alpha^3 \quad (3)$$

where U_α^d is the derivative of the Hubbard parameter of atom α with respect to atomic charge. In our recent study,[61] U_α^d is regarded as a fixed parameter for each element type; in other words, the Hubbard parameter is taken to be linearly dependent on the atomic charge. For phosphorus containing compounds, a complicating factor is that the oxygen atoms on the phosphorus tend to be highly charged, thus the linear charge dependence of the Hubbard parameter may no longer be valid. To take this deviation from the linear behavior into account, we add an additional charge dependent term to the Hubbard charge derivative; i.e.

$$U_\alpha^d(q) = U_{0\alpha}^d + D_0 \exp[-\Gamma_0(\Delta q_\alpha - Q_0)^2] \quad (4)$$

where the charge-independent parameter ($U_{0\alpha}^d$) is dependent on the element type, whereas the three parameters associated with the Gaussian (D_0, Γ_0, Q_0) are taken to be independent of element type to minimize the number of parameters. The choice of the Gaussian functional form is entirely empirical and designed to avoid undesired behavior of the Hubbard derivative for large charges.

B. Reference Systems and Parameter Fitting. To parametrize an effective SCC-DFTB approach for phosphorus compounds and phosphate chemistry, the parametrization procedure is divided into several stages. First, the atomic properties (basis functions, zero-order Hamiltonian matrix elements, Hubbard parameter) for P are derived based on a set of rather well-defined protocols involving atomic calculations.[43,63] The pairwise repulsive potentials between P and O, N, C, H for a second-order SCC-DFTB approach are then fitted based on small molecule compounds and B3LYP calculations. With these parameters held fixed, the Hubbard derivatives are then fitted based on more specific properties such as proton affinity or phosphate hydrolysis reaction energetics. With the current third-order formulation, it seems difficult to develop a single set of Hubbard derivatives to simultaneously produce reliable proton affinities and phosphate hydrolysis energetics (see below). Therefore, two sets of Hubbard derivative parameters have been developed based on proton affinity and phosphate hydrolysis reactions, respectively; for clarity, they are referred to as SCC-DFTBPA and SCC-DFTBPR, respectively (see Table 1). The phosphate hydrolysis set of parameters is further tested with additional benchmark calculations using results from the QCRNA database established by the York group.[62] These gas-phase calculations are finally supplemented with potential of mean force calculations for monophosphate ester hydrolysis in solution with SCC-DFTBPR/MM simulations.

2070 *J. Chem. Theory Comput., Vol. 4, No. 12, 2008*

Yang et al.

Table 2. Errors in Structural Properties of Different Parameterizations of the SCC-DFTB Approach for Common Phosphate Compounds[a]

geometrical parameter[b]	second-order	SCC-DFTBPA[c]	SCC-DFTBPR
P$-$O (Å)	0.03	0.04	0.03
O$-$P$-$O (°)	1.8	1.9	1.7

[a] The root-mean-square (rms) errors are calculated relative to those from B3LYP/6$-$31+G(d,p) calculations. [b] In total, 10 common phosphate compounds are included in the analysis (including MMP/DMP-water complexes, pentavalent intermediate structures for the hydrolysis of MMP/DMP, different protonation states for phosphoric acid, models for ADP and ATP). For selected structures, see Figure 1. [c] With SCC-DFTBPA, the pentavalent intermediate for MMP hydrolysis is not stable as an intermediate, and geometry optimization leads to the reactant state.

1. Atomic Properties and Repulsive Potentials. Since the parametrization procedure for SCC-DFTB has been reported in detail in previous articles for several elements,[43,63] we only include a very short description here. The atomic properties include the Slater basis functions, the reference density (ρ_0), and the chemical hardness (Hubbard) parameter; these are determined based on atomic DFT calculations with an in-house program TWOCENT. Once these are established, other quantities such as the second-order kernel γ_{AB} can be determined.[43,63] The matrix elements for the effective Kohn$-$Sham Hamiltonian with the reference density can also be calculated and tabulated; the exchange-correlation functional used is the one of Perdew, Burke, and Ernzerhof (PBE).[64]

For the repulsive potential, $U[R_{\alpha\beta}; \rho_0^\alpha, \rho_0^\beta]$, five different pairs need to be derived (P$-$P and P$-$O/N/C/H). Accordingly, several small molecules are chosen, and full DFT (B3LYP[65$-$67]/6$-$311G**) calculations are calculated as a function of specific bond distances. The repulsive potential is then calculated as the difference between the full DFT potential energy curve and the electronic contribution from SCC-DFTB at the same structures

$$U[R_{\alpha\beta};\rho_0^\alpha, \rho_0^\beta] = [E^{DFT}(R_{\alpha\beta}) - E^{DFT}(\infty)] -$$
$$\sum_{i\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0(R_{\alpha\beta}) - \frac{1}{2}\sum_{\alpha\beta} \gamma_{\alpha\beta}\Delta q_\alpha \Delta q_\beta \quad (5)$$

The repulsive potential is fitted into a cubic spline and truncated to zero in the 2.2$-$3.2 Å range. The specific molecules used for parametrizing different repulsive potentials are as follows: PH$_3$, PCH, HPCH$_2$, H$_2$PCH$_3$, PN, HPNH, H$_2$PNH$_2$, P$_2$, HPPH, H$_2$PPH$_2$, OPH, H$_3$PO$_4$, and H$_4$PO$_5^-$.

2. Hubbard Derivative Related Parameters. Similar to our recent work on using third-order terms to improve SCC-DFTB proton affinities,[25,59] the Hubbard derivative related parameters are optimized using a Genetic Algorithm[68,69] by minimizing the penalty function defined as

$$\chi = \frac{\sum_i w_i (Y_i^{ref} - Y_i^{SCC})^2}{\sum_i w_i} \quad (6)$$

where the summation is over all properties of interest in a particular optimization set (see below), w_i is the weight of a specific property, and Y_i^{ref}/Y_i^{SCC} are the values of the i-th property from reference calculation (see below) and SCC-DFTB calculation with a specific set of Hubbard derivative related parameters ($U_{0\alpha}^d$, D_0, Γ_0, Q_0), respectively. During the Genetic Algorithm (GA) optimization, the properties of interest include proton affinities/reaction energetics and the root-mean-square gradient (GRMS) of the molecule at the reference geometry, addressing both energetic and structural information; the corresponding weights in χ are 10 and 1, respectively. The micro-GA technique[68] is applied with a population of 10 chromosomes for 100 generations with uniform crossovers.

SCC-DFTBPA: Proton Affinity of Phosphate Compounds. Eighteen phosphate species of biological relevance (see Table 3) are chosen as the reference system to optimize the Hubbard derivative related parameters. A subset of these molecules was used as benchmark systems in the previous work of Range et al.,[29] who found that all semiempirical methods including the standard SCC-DFTB have significant systematic errors. As the high-level reference, geometries are optimized at the B3LYP/6$-$31G(d) level, and the energetics are obtained at the G3B3[70] level. Benchmark calculations by Range et al.[29] showed that this combination gives systematically reliable energetics compared to experimental values. For the purpose of making a comparison and establishing a less expensive reference level for the subsequent parametrization involving larger molecules, MP2 with the G3Large basis set is also carried out; the G3Large was modified based on the 6$-$311+G(3df,2p) basis set for G3[71] calculations.

Rigorously speaking, the proton affinity of molecule A$^-$ is the negative of the *enthalpy* change for the gas-phase reaction A$^-$(g)+H$^+$ (g) → AH(g) at the room temperature, which involves thermal vibrational contributions. To avoid a large number of vibrational calculations in the parametrization process, we consistently consider only the potential energy contribution in both the reference calculations and SCC-DFTB calculations during the GA optimization. Another subtle point is that the energy of proton in SCC-DFTB is not zero due to the definition of the repulsive potential in the total energy expression;[63] however, once a value (141.8 kcal/mol) is selected, the results are consistent among all SCC-DFTB calculations.

To test if the parameters optimized based on phosphate proton affinities are transferrable to non-phosphate compounds, a set of 11 small molecules including water, alcohols, and carboxylic acids are also considered. A set of Hubbard derivative parameters are optimized based on the proton affinities of both phosphate and non-phosphate molecules, although the result is rather disappointing (see below), which suggests that further improvements in the SCC-DFTB formalism are needed to predict accurate proton affinities for a broad range of molecules that include both second and third row elements.

SCC-DFTBPR: Phosphate Hydrolysis. As discussed above, a balanced treatment for the proton affinities of phosphate, the nucleophile, and the leaving group is required (but not sufficient) for a reliable description of phosphate hydrolysis reactions. Since this seems difficult to achieve even with the third-order extension of SCC-DFTB, we choose to pursue a

Phosphate Hydrolysis Reactions

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2071**

Table 3. Comparison of Proton Affinities from Different Parameterizations of SCC-DFTB and ab Initio Methods for 18 Phosphorus Containing Molecular Systems[a]

molecules[b]	G3B3	MP2[c]	SCC-DFTB[d]		
			2nd-order	3rd-order	SCC-DFTBPA
H_3PO_4	334.0	−1.5	27.3	13.7	5.0
$H_2PO_4^-$	463.6	−1.1	36.7	15.2	1.9
DMPH[e]	336.2	−1.5	19.9	9.7	2.9
MMP[e]	336.7	−1.6	22.2	10.3	2.4
MMP$^-$[e]	460.5	−1.2	31.7	14.2	3.5
PH_3OH	201.6	−2.0	1.6	−2.0	−2.9
PH_2OHOH	201.6	−1.9	7.4	1.6	0.1
PHOHOHOH	200.8	−1.7	14.7	6.9	4.6
$PH_2(OH)=O$	336.6	−1.6	13.2	4.9	−1.9
$PH(OH)(OH)=O$	334.7	−1.5	20.9	9.9	2.0
$P(O)(OH)(-O-CH_2CH_2-O-)$	336.3	−1.7	18.0	7.5	1.0
$P(OH)(OH)(-O-CH_2CH_2-O-)(OH^*)$	359.0	−2.0	6.9	7.6	0.8
$P(OH^*)(OH)(-O-CH_2CH_2-O-)(OH)$	350.4	−1.7	16.8	5.7	−2.3
$P(OH^*)(OH)(-O-CH_2CH_2-O-)(OCH_3)$	351.2	−1.6	12.0	2.3	−5.4
$P(OH)(OCH_3)(-O-CH_2CH_2-O-)(OH^*)$	359.6	−1.9	6.3	−3.5	−0.4
$P(OH^*)(OCH_3)(-O-CH_2CH_2-O-)(OH)$	353.0	−1.7	13.7	3.9	−3.9
$P(OH)(OH)(OH)(OH^*)(OH)\_ax$	357.3	−1.8	14.2	10.8	4.2
$P(OH)(OH)(OH)(OH^*)(OH)\_eq$	347.0	−1.9	24.2	−18.5	−1.1
Error Analysis[f]					
MAXE		−2.0	36.7	−18.5	−5.4
RMSE		1.7	19.3	9.5	3.0
MUE		1.7	17.1	8.2	2.6
MSE		−1.7	17.1	5.6	0.6

[a] The proton affinity (PA) of A$^-$ is defined as the negative of the enthalpy change for the gas-phase reaction $A_{(g)}^- + H_{(g)}^+ \rightarrow AH_{(g)}$. In the current calculations, 0 K electronic energies instead of the room temperature enthalpy are used without zero point energy correction. All quantities are given in kcal/mol. For G3B3, such calculated PA values are given; for all other methods, the errors relative to G3B3 results are given (positive error indicates overestimation). [b] All molecules correspond to the protonated species (AH). Asterisk "*", "ax" (axial), and "eq" (equatorial) are used to identify the acidic proton. [c] The basis set used is G3Large, which is a basis set used in the G3 method (see http://chemistry.anl.gov/compmat/g3theory.htm). [d] See Table 1 for the notations used to label different parametrizations of the SCC-DFTB approach. As discussed in previous studies, a value of 141.8 kcal/mol is included to account for the self-interaction energy of the H atom. [e] "DMPH" refers to dimethyl hydrogen phosphate; "MMP" refers to $P(O)(OH)(OH)(OCH_3)$; "MMP$^-$" refers to $P(O)(O)(OH)(OCH_3)^-$. [f] MAXE: largest error; RMSE: root-mean-square error, defined as $<(err)^2>^{1/2}$; MUE: mean unsigned error; MSE: mean signed error.

more pragmatic avenue by fitting the Hubbard derivative parameters based on results for a set of representative phosphate hydrolysis reactions. The underlying assumption is that errors in the proton affinity of different species can cancel out to yield satisfactory reaction energies. An additional advantage of using reaction properties for parametrization is that both reaction energy and barrier can be taken into account. We emphasize that the number of parameters is rather small, which include five $U_{0\alpha}^d$ values for P, O, N, C, H and three element-*independent* Gaussian parameters, D_0, Γ_0, Q_0. Therefore, we hope that parameters optimized based on phosphate hydrolysis are reasonably transferrable to other phosphate reactions with similar characteristics, such as phosphoryl transfer reactions; this will be tested with the additional benchmark calculations discussed below.

The reference reactions include the hydrolysis of dimethyl monophosphate ester (DMP) and monomethyl monophosphate ester (MMP) with different protonation states (see Table 5). Several considerations account for this choice. First, the hydrolysis of MMP and DMP are basic models[8,72] for the hydrolysis of nucleotides (e.g., ATP, GTP) and the scission reactions in many ribozyme systems, respectively. Second, different protonation states for the reactant stress the effects due to pH changes or pK_a shifts induced by the macromolecular environment, which may have important biological implications.[42] For instance, the phosphorane intermediate is a short-lived species with −2 charge; for the neutral and monoanionic species, the lifetime is long enough

for pseudorotation.[73−75] Finally, the small size of DMP and MMP allows us to perform relatively high-level ab initio calculations as reference.

Regarding the reaction mechanism, both dissociative and associative mechanisms are considered. In addition, pathways that involve water-assisted proton transfers, which have been proposed to be important for at least the dissociative mechanism,[40] are also included. All together, 37 gas-phase reaction energies (18 of which are energy barriers) involving 47 structures are included as the reference set; 15 reaction energies are based on MMP model reactions (8 energy barriers), and the remaining 22 are from DMP reactions with 10 energy barriers. Despite the relatively small size of the DMP and MMP systems, it is impractical to carry out G3B3 calculations. Instead, B3LYP/6−31+G(d,p) geometries and MP2/G3Large single point energies are used as reference. For all energy calculations, no zero-point correction or vibrational contribution has been included.

C. Benchmark Calculations. *1. Geometrical Parameters.* To test the performance of the different SCC-DFTB parametrization on structural properties, a series of common phosphate compounds are studied, and the optimized gas-phase geometries are compared to those from B3LYP calculations with the 6−31+G(d,p) basis set. The list of compounds include inorganic phosphate with different protonation states (H_3PO_4, $H_2PO_4^-$, HPO_4^{2-}, and PO_4^{3-}), mono-

Table 4. Different Performances for SCC-DFTB Including on-Site Third-Order Terms for Non-Phosphorus Compounds and Phosphate Species[a]

molecules	G3B3	SCC-DFTB[b]		
		2nd-order	3rd-order	mix-optimized 3rd-order
		Non-Phosphorus		
water	398.4	26.5	−5.4 (1.3)	1.0 [−13.9]
methanol	392.6	4.5	−6.3 (−4.0)	−8.3 [−13.1]
ethanol	388.3	8.7	−2.8 (−0.5)	−4.3 [−9.3]
propanol	387.6	7.9	−3.5 (−1.1)	−4.8 [−9.8]
2-propanol	385.6	11.5	−0.5 (1.8)	−1.4 [−6.5]
formic acid	351.2	11.9	3.1 (4.2)	−0.9 [−6.8]
acetic acid	355.1	11.3	1.6 (2.9)	−2.8 [−8.4]
propanoic acid	354.5	11.2	1.9 (3.1)	−2.5 [−8.0]
phenol	356.7	5.5	0.2 (1.1)	−2.6 [−5.8]
p-methylphenol	357.9	4.6	−0.4 (0.4)	−3.1 [−6.2]
p-nitrophenol	334.6	0.9	−5.4 (−4.4)	−5.8 [−12.0]
$CH_2(OH)_2$	374.8	11.7	1.3 (3.4)	−4.7 [−10.1]
$CH_2(OH)(OCH_3)$	377.2	5.1	−1.1 (0.0)	−3.4 [−8.4]
$CH(OH)_3$	361.9	18.0	3.7 (5.9)	−1.8 [−7.2]
$CH(OCH_3)_2OH$	365.7	9.5	2.5 (3.6)	−0.4 [−5.7]
		Error Analysis		
MAXE		26.5	−6.3 (5.9)	−8.3 [−13.9]
RMSE		11.6	3.2 (3.0)	3.8 [9.1]
MUE		9.9	2.6 (2.5)	3.2 [8.7]
MSE		9.9	−0.7 (1.2)	−3.1 [−8.7]
		Phosphate		
H_3PO_4	334.0	27.3	13.7	8.5 [1.6]
$H_2PO_4^-$	463.6	36.7	15.2	8.7 [3.7]
DMPH	336.2	19.9	9.7	5.0 [1.7]
MMP	336.7	22.2	10.3	5.2 [0.1]
MMP⁻	460.5	31.7	14.2	10.1 [6.1]
PH_3OH	201.6	1.6	−2.0	−2.0 [−3.0]
PH_2OHOH	201.6	7.4	1.6	1.6 [−1.0]
PHOHOHOH	200.8	14.7	6.9	6.5 [3.0]
PH_2(OH)=O	336.6	13.2	4.9	0.1 [−3.6]
PH(OH)(OH)=O	334.7	20.9	9.9	4.7 [−0.8]
		Error Analysis		
MAXE		36.7	15.2	10.1 [6.1]
RMSE		22.1	10.0	6.1 [3.0]
MUE		19.6	8.8	5.2 [2.5]
MSE		19.6	8.4	4.8 [0.8]

[a] The PA values and errors (in kcal/mol) are defined in the same manner as in Table 3. [b] See Table 1 for the notations used to label different SCC-DFTB approaches. The values in parentheses are obtained using a third-order approach optimized in ref 59 based on 32 non-phosphorus compounds. For the "mix-optimized 3rd-order" set, both non-phosphorus and phosphate molecules (shown in italics) are included to attempt to optimize a set of third-order parameters that work for both classes of molecules, although the results are not satisfying (see text). The numbers in brackets are based on the phosphate hydrolysis reaction parameter set (SCC-DFTBPR).

phosphate esters (MMP, DMP), diphosphate ester (model ADP, see Figure 1), and triphosphate ester (model ATP, see Figure 1).

2. Additional Phosphate Hydrolysis Reactions. As additional benchmark systems for the newly parametrized SCC-DFTBPR, all (19) RNA model reactions with the overall charge of −1 are selected from the QCRNA database established by the York group;[62] these include 56 reaction energies and 47 barriers. These systems were consistently calculated at the B3LYP/6−311++G(3df,2p) level for energy and B3LYP/6−31++G(d,p) for structure by the York group. Both single point energy calculations and geometry optimizations are carried out at the SCC-DFTBPR level; for transition states, only single point energies are considered. Dipole moments are also compared. In a separate study,[76] selected pseudorotation barriers at the SCC-DFTBPR level are also compared to the QCRNA database, for which good agreement is found.

3. Explicit Solvent Simulations with SCC-DFTBPR/ MM. Finally, to supplement the gas-phase calculations, potential of mean force simulations are carried out with SCC-DFTBPR/MM to investigate if the parametrized model works in an explicit condensed-phase environment. This is an important test since the ultimate goal is to use SCC-DFTBPR in enzyme simulations (for an initial application, see ref 5).

The specific reaction studied is the first step for the hydrolysis of MMP in water. The stochastic boundary condition[77] is used as the simulation protocol. The MMP molecule is solvated with a water droplet of 22 Å radius. Only MMP and the lytic water are treated with SCC-DFTBPR, while the rest of the water molecules are described with the modified version of TIP3P[78,79] in CHARMM.[80] The van der Waals parameters for the atoms in MMP are taken from the CHARMM force field for lipid[81] without further optimization. The importance of QM van der Waals parameters in QM/MM simulations has been discussed in previous

Phosphate Hydrolysis Reactions

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2073**

Table 5. Comparison of Exothermicity and Barrier Height from SCC-DFTB and High-Level Ab Initio Calculations for 37 Elementary Steps in the Hydrolysis of MMP and DMP[a]

process[b]	ab initio	SCC-DFTB single point	SCC-DFTBPR optimization[i]	MP2//SCC-DFTBPR[j]
com1→ts1 (MMP,B)	31.0[c]/−1.7[d]	−0.9[e]/−3.0[f]/0.4[g]/3.1[h]	1.1	−0.9
com1→int1 (MMP,E)	30.6/−1.4	−2.1/−2.2/0.7/2.1	1.4	−0.8
com1→ts1_2 (MMP,B)	41.5/−2.1	1.2/−0.3/2.4/5.3	−3.4	−1.6
com1→int1_2 (MMP,E)	31.0/−1.1	−4.4/−0.6/1.6/3.1	1.9	−0.1
int1_2→ts2_1 (MMP,B)	11.9/−2.0	−2.5/−2.1/−3.3/−2.3	−[k]	
int1_2→ts2_2 (MMP,B)	3.6/0.1	−5.4/−5.0/−5.2/−5.0	0.3	6.1
int1_2→com2 (MMP,E)	−28.8/−0.9	2.5/0.6/0.2/0.4	−0.4	−0.8
com1→diss_tsa (MMP,B)	36.8/−4.2	4.7/4.0/2.4/4.8	2.6	0.2
com1→diss_int (MMP,E)	19.6/−6.4	−7.1/−6.0/−3.5/−2.8	−2.9	−1.0
com1_w2→ts1_2_w2 (MMP,B)	39.9/−2.1	−8.2/−9.4/−6.1/−3.7	−5.4	−2.5
com1_w2→int1_2a_w2 (MMP,E)	28.0/0.8	−5.4/−2.5/−1.2/0.8	0.2	−1.1
int1_2a_w2→int1_2_w2 (MMP,E)	0.4/−1.7	0.4/0.7/1.2/1.0	1.7	1.5
int1_2_w2→ts2_0_w2 (MMP,B)	11.4/−0.5	−3.7/−7.3/−5.2/−3.8	−7.3	−1.1
com1_da→ts1_da (MMP,B)	55.0/−8.4	−22.5/−12.3/−9.2/−10.1	−8.9	0.0
com1_da→int_da (MMP,E)	4.5/−2.0	−2.9/−0.7/−1.8/−0.4	−12.1	−1.5
com1→ts1 (DMP,B)	38.6/−1.4	−0.9/−4.1/−0.8/3.1	−1.6	0.8
com1→int1 (DMP,E)	35.4/−0.2	−5.6/3.1/−0.5/0.2	−0.5	0.6
int1→int1_2 (DMP,E)	1.3/−0.7	−3.0/−0.9/−0.9/0.1	−4.0	2.0
int1_2→ts2 (DMP,B)	0.6/−0.5	0.5/−0.1/−0.6/−0.6	−0.5	−1.6
int1_2→com2 (DMP,E)	−35.2/−0.7	7.1/4.6/4.9/4.2	7.0	−1.3
n_com1→n_ts3 (DMP,B)	33.6/−1.4	4.9/4.3/1.0/3.5	1.2	−0.4
n_com1→n_int1 (DMP,E)	13.2/0.4	−3.7/−0.8/0.4/1.1	0.1	0.1
n_int1→n_ts4 (DMP,B)	22.9/−1.6	6.4/4.9/2.0/4.2	0.9	1.0
n_int1→n_com2 (DMP,E)	−15.8/−1.9	2.6/0.5/0.9/0.6	0.0	0.0
DMP_P→diss_ts (DMP,B)	40.9/−2.9	11.8/9.4/5.5/7.2	6.1	−0.8
DMP_P→diss_prod (DMP,E)	28.2/−3.8	0.6/−2.1/−2.7/−2.9	−2.6	−1.2
diss_prod2→diss_ts2 (DMP,B)	13.5/0.7	13.4/13.0/7.5/11.7	8.4	−0.5
diss_prod2→MMP_P (DMP,E)	−29.8/3.6	0.8/2.8/2.6/3.6	2.8	0.2
diss_w_reac→diss_w_ts (DMP,B)	20.9/−2.3	5.9/3.4/−0.2/2.0	−0.1	−0.4
diss_w_reac→diss_w_prod (DMP,E)	18.4/−2.6	4.8/1.2/−2.5/−1.3	−2.0	−0.3
diss_w_prod2→diss_w_ts2 (DMP,B)	1.9/0.2	2.5/2.7/1.1/3.7	1.8	−0.8
diss_w_prod2→diss_w_reac2 (DMP,E)	−21.0/2.7	−2.9/0.5/0.9/1.5	0.4	0.3
n_w_com1→n_w_ts3 (DMP,B)	28.2/−1.8	−3.0/−2.3/−4.8/−2.0	−	−
n_w_com1→n_w_int1 (DMP,E)	13.1/1.0	−4.2/−1.3/0.0/0.7	−3.1	−0.5
n_w_int1→n_w_int2 (DMP,E)	−0.5/0.5	0.3/0.5/0.7/1.1	0.4	0.3
n_w_int2→n_w_ts4 (DMP,B)	15.1/−2.3	1.8/−0.5/−4.0/−2.3	−	−
n_w_int2→n_w_com2 (DMP,E)	−13.0/−2.0	1.2/−1.0/−0.4/−2.0	−1.3	0.6
Error Analysis[l]				
MAXE	−8.4	−22.5/13.0/−9.2/11.7	−12.1	6.1
RMSE	2.5	6.1/4.6/3.3/3.9	4.0	1.4
MUE	1.9	4.4/3.3/2.4/2.9	2.8	1.0
MSE	−1.4	−0.4/−0.4/−0.4/0.8	−0.5	−0.2

[a] No zero-point corrections are included in either exothermicity or barrier heights. All quantities are given in kcal/mol. [b] The processes are labeled in the following manner: e.g., "com1→ts1 (MMP,B)" refers to the reaction from the reactant "com1" to the transition state "ts1", "MMP" in the parentheses refers to the monomethyl monophosphate ester model system, "B" in the parentheses stands for "Barrier". Similarly, "com1→int1 (DMP,E)" refers to the reaction from the reactant "com1" to the intermediate "int1", "DMP" refers to the dimethyl diphosphate ester model system, and "E" stands for "Exothermicity". For the structures, see Figure S1 in the Supporting Information. [c] The number before the slash refers to the MP2/G3Large single point calculation based on the B3LYP/6−31+G** optimized structures. For more details about the G3Large basis set, see http://chemistry.anl.gov/compmat/g3theory.htm. [d] The number after the slash refers to the energy difference between the B3LYP result and the MP2/G3Large single point calculation at the B3LYP structure. [e] The difference between the standard (second-order) SCC-DFTB and MP2/G3Large single point energies at the B3LYP/6−31+G(d,p) structures. [f] The difference between the default 3rd-order SCC-DFTB and MP2/G3Large single point energies at the B3LYP/6−31+G(d,p) structures. [g] The difference between SCC-DFTBPR and MP2/G3Large single point energies at the B3LYP/6−31+G(d,p) structures. [h] The difference between the "mix-optimized 3rd-order" SCC-DFTB and MP2/G3Large single point energies at the B3LYP/6−31+G(d,p) structures. [i] The difference between fully geometry-optimized SCC-DFTBPR energies and MP2/G3Large single point energies at the B3LYP/6−31+G(d,p) structures. [j] The difference between MP2/G3Large single point energies at the SCC-DFTBPR structures and those at the B3LYP/6−31+G(d,p) structures. [k] As discussed in the main text and illustrated in Figure 2, only one transition state is found at the SCC-DFTBPR level. [l] The errors are defined in the same manner as in Table 3. For the entry for "ab initio", the errors are for the B3LYP/6−31+G(d,p) energies relative to the MP2/G3Large results.

studies,[82] and it was argued that in some cases allowing the van der Waals parameters to vary during the reaction can be important,[83] which we examine briefly here using the MMP system as an example (vide infrra).

Both the associative and dissociative mechanisms (see Scheme 1) are considered, and the according potentials of mean force surfaces are calculated with umbrella sampling.[84] For the associative mechanism, the P−O^{Nu} distance and the

antisymmetric stretch describing the proton transfer from the lytic (nucleophilic) water to phosphate are defined as the reaction coordinates. For the dissociative mechanism, the reaction coordinates include the P−O^{Lg} distance (where O^{Lg} is the oxygen atom of the leaving group, which is methanol in the current case) and the antisymmetric stretch that describes the intramolecular proton transfer between the protonated oxygen of MMP and O^{Lg}. In both cases, the
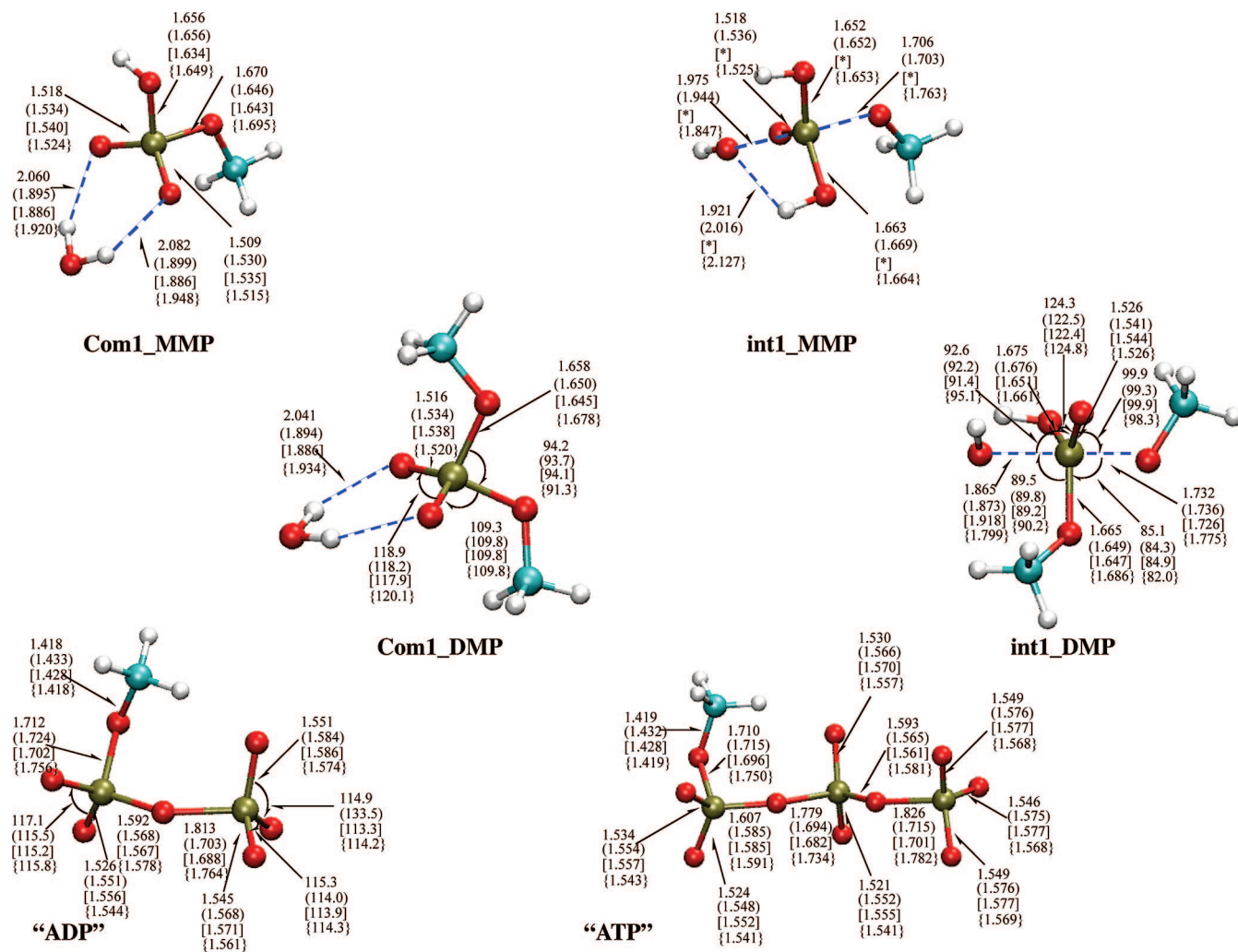
Figure 1. Selected structures included in the SCC-DFTB parametrization protocol optimized at different levels. The first four are involved in the hydrolysis of small phosphate molecules, and the last two are models for ADP and ATP, respectively. Values without parentheses are from B3LYP/6−31+G** calculations; values with parentheses are from SCC-DFTBPR; values with brackets are from SCC-DFTBPA; values with curly brackets are from second-order SCC-DFTB parametrization. See Table 1 for the notation of different SCC-DFTB parametrizations. Distances are in Å, angles in degrees.

Scheme 1
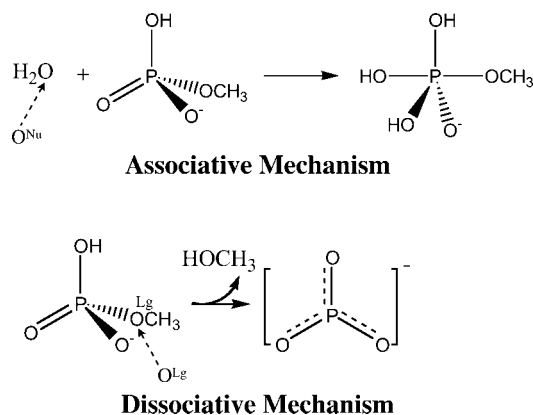
antisymmetric stretch is defined as the distance of donor-proton minus the distance of acceptor-proton. Based on the defined reaction coordinates, two-dimensional potential of mean force (PMF) in solution are generated using umbrella sampling; for comparison, two-dimensional potential energy surface (PES) in the gas phase are also calculated using

adiabatic mapping. A technical point is that the O−H bond in MMP needs to be constrained in calculations for the associative mechanism; otherwise, the proton of MMP is transferred back to the nucleophilic water as the latter transfers its proton to the unprotonated oxygen in MMP. In the umbrella sampling calculations, 122 and 108 windows are used for the associative and dissociative mechanisms, respectively, where each window includes 50 ps of MD simulations. The data from umbrella sampling are combined using the Weighted Histogram Analysis Method (WHAM) approach.[85]

Nonbonded interactions (electrostatic and van der Waals) are calculated without cutoffs. The bulk electrostatics are considered via GSBP[86,87] with an 24 Å inner region, which includes an additional 2 Å buffer region. The generalized reaction field matrix is evaluated using 400 spherical harmonics with an outer region dielectric constant of 80. Thermal collisions due to the bulk are included in the outer 2 Å of the 22 Å water sphere via Langevin dynamics,[88] and the Langevin atom list is heuristically updated during the simulation. A 1 fs time step is used, and the temperature is

Phosphate Hydrolysis Reactions

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2075**

maintained at 298 K. Bonds involving hydrogen are constrained with the SHAKE algorithm[89] except for those involved in the proton transfer.

III. Results and Discussions

In this section, we first briefly go over results for the structural properties of common phosphate compounds and then focus on the performance of the third-order parametrizations for proton affinities and hydrolysis reactions. Some of the gas-phase benchmark results are summarized in the first application of SCC-DFTBPR/MM although without detailed discussions.[5]

A. Structural Properties. As shown in Table 2, the structures of common phosphate compounds are well reproduced by all three parametrizations of the SCC-DFTB model. The bond distances and bond angles have an rms error in the range of 0.03 Å and 1.8 degrees, respectively. The results are equally impressive for species with rather complex electronic structures, such as "ADP/ATP" and pentavalent intermediate involved in the associative pathways of hydrolysis (see Figure 1). For the bridging P–O bonds in the model ADP and ATP species, the values from second-order SCC-DFTB agree better with B3LYP results, while both SCC-DFTBPR and SCC-DFTBPA tend to underestimate the distances. For most geometrical parameters, any one of the SCC-DFTB models can be used to give satisfying results. Hydrogen-bonding distances have larger errors, as illustrated in Com1_MMP and Com1_DMP in Figure 1, although this is not unique to phosphate species.[61,82]

B. SCC-DFTBPA: Proton Affinities. For the 18 phosphorus containing species, large errors are seen in the gas-phase proton affinities with the standard second-order SCC-DFTB parametrization (see Table 3). The largest error is 36.7 kcal/mol for $H_2PO_4^-$, which is not unexpected for a highly charged species in the gas phase. The root-mean-square-error (RMSE) is 19.3 kcal/mol, similar to those reported for the second-order SCC-DFTB and other semiempirical methods in the previous study.[29] With the on-site third-order terms, even without specific parametrization, the errors in the proton affinity reduce dramatically as also seen in our recent analysis for non-phosphate compounds (also see Table 4).[61] The largest error becomes −18.5 kcal/mol, and the RMSE is cut nearly in half to 9.5 kcal/mol. With further optimizations, the results become very good with the largest error of −5.4 kcal/mol and RMSE of merely 3.0 kcal/mol. Similar to the observations in previous analysis, the Mulliken charges on the phosphate oxygen atoms are substantially modified with the third-order terms, which is related to the reduced error in proton affinity. Take MMP as an example. The Mulliken charges on the acidic oxygen atom are −0.566 and −0.816 before and after deprotonation, respectively, with the second-order SCC-DFTB; the corresponding values are −0.710 and −0.895, respectively, with the optimized third-order SCC-DFTB.

Despite these encouraging results, difficulties arise when we attempt to optimize the current third-order formulation for both phosphorus-containing compounds and species without phosphorus. The situation is illustrated in Table 4. Although the RMSEs for both classes of compounds seem

to be reasonable, 3.8 and 6.1 kcal/mol for the non-phosphorus- and phosphorus-containing compounds, respectively, the signs of error are opposite. For non-phosphorus compounds, the proton affinity is typically underestimated by a few kcal/mol, while the trend is the opposite for phosphorus-containing species. This is alarming because error will accumulate for reactions between, for example, a non-phosphorus nucleophile and a phosphate compound, which suggests that the calculated reaction energetics are likely poor (see discussions below). With the SCC-DFTBPR parameter set, the errors in the proton affinity follow the same trend: the non-phosphorus- and phosphorus-containing compounds tend to have errors of opposite signs.

The origin for the different trends is not clear, and an interesting observation is that different optimal Hubbard derivatives for oxygen (U_O^d) apply to two classes of compounds. The computed U_O^d based on atomic calculations is −0.17; the optimized value based on the proton affinity for 32 non-phosphorus compounds and 18 phosphorus-containing compounds is −0.14 and −0.20, respectively. The variation toward different directions indicates some intrinsic differences between the two classes of compounds although the precise origin is not clear; further improvements in the SCC-DFTB formalism, such as including the off-site third-order terms and the treatment of polarization, are likely required.

C. SCC-DFTBPR: Phosphate Hydrolysis Reactions. The hydrolysis of MMP and DMP molecules have been studied by a number of groups using different ab initio and DFT methods,[11,16,40,90−95] and the role of additional water as proton relay has been discussed for the dissociative[40] pathway. Our present calculations, which also include the hydrolysis of DMP in different protonation states and the effect of additional water in associative pathways, are largely consistent with previous studies. Overall (see Table 5 for energetics and the Supporting Information for structures), the barriers for DMP and MMP hydrolysis are rather similar and show a weak dependence on the protonation state; e.g., the barrier for the protonated form of DMP (charge neutral) in the gas phase is lower than that of the anionic species by ∼5 kcal/mol along the associative pathway, and the additional water helps to further reduce the potential barrier (not free energy barrier) by another 5 kcal/mol. The role of the additional water on the dissociative potential barriers is much more significant, on the order of ∼20 kcal/mol. Therefore, when the entropic factor is taken into consideration, the effect of an additional water as proton relay is likely small for associative pathway; a detailed analysis in the condensed phase using a QM/MM framework will be reported in the near future (Yang and Cui, work in progress).

The reference energetics are taken to be MP2/G3-Large, which has been shown to give reliable proton affinities compared to the more elaborate CBS and G3B3 methods.[29] Compared to the MP2/G3-Large data, the energetics at the B3LYP/6−31+G(d,p) are generally rather close with a difference typically smaller than 3 kcal/mol; the RMSE is 2.5 kcal/mol. In certain cases, especially when the phosphate is highly charged, the difference between B3LYP/6−
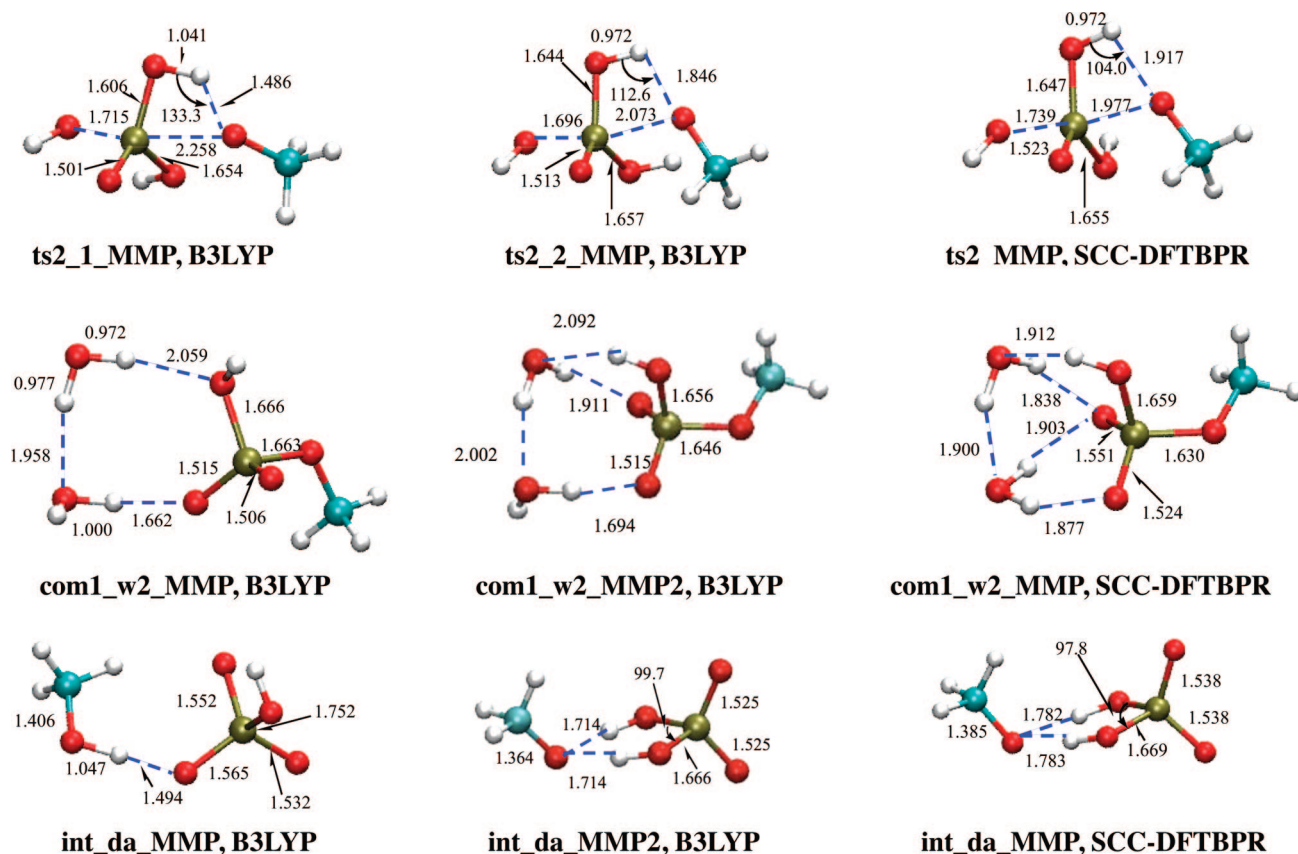
Figure 2. Structures for which significant discrepancy is found between independent B3LYP/6−31+G** and SCC-DFTBPR geometry optimizations. For com1_w2_MMP and int_da_MMP, B3LYP optimization starting from SCC-DFTBPR geometry located structure closer to the SCC-DFTBPR result and lower in energy than the original B3LYP structure. See main text (section 3.3) for discussions. Distances are in Å, angles in degrees.

31+G(d,p) and MP2/G3-Large can be rather substantial and as large as 8.4 kcal/mol (for com1_da → ts1_da, see Table 5).

When single point energies are calculated at the B3LYP/6−31+G(d,p) structures, the standard second-order SCC-DFTB gives rather large errors; the largest error is −22.5 kcal/mol and the RMSE is 6.1 kcal/mol. When the third-order terms are included, the errors decrease substantially even without specific parametrizations, especially for the dianionic cases; the largest error decreases to 13.0 kcal/mol, and the RMSE becomes 4.6 kcal/mol. Using the "mixed-optimized 3rd-order" parameter set based on PA comparisons, the reaction energies become slightly better, with largest error of 11.7 kcal/mol and a RMSE of 3.9 kcal/mol. With further parameter optimizations (SCC-DFTBPR), the largest error becomes −9.2 kcal/mol, and the RMSE is 3.3 kcal/mol, which are rather remarkable for a semiempirical method without many reaction-specific parameters. The errors for the stable states are overall smaller than that for the transition states; e.g., the RMSE is 1.9 kcal/mol when only the stable states are considered.

The reliability of the parametrized SCC-DFTBPR is further tested by full structure optimizations for both stable and transition states using the ABNR[80] and CPR[96] algorithms implemented in CHARMM; care is taken to ensure that the same local minima/saddle-points are used when compared to the B3LYP structures. The errors in energies are largely consistent with those for SCC-DFTBPR single-point energies

at B3LYP geometries, and the RMSE increases only slightly to 4.0 kcal/mol. As shown in Table 6, the SCC-DFTBPR structures are rather similar to the B3LYP/6−31+G(d,p) results, with RMSE for P−O distances typically of 0.02 Å and for O−P−O angles of ∼2−3 degrees. The errors in the transition state structures are slightly larger, especially for the bond distance associated with the leaving group (RMSE of 0.08 Å). The encouraging aspect is that MP2/G3-Large single point energies at the SCC-DFTBPR structures are overall close to the reference (MP2/G3-Large//B3LYP/6−31+G(d,p)) values, with a RMSE of merely 1.4 kcal/mol. Large errors more than 2 kcal/mol occur very rarely, which suggest that the SCC-DFTBPR structures are usually satisfactory.

There are, however, cases where notable differences between SCC-DFTBPR and B3LYP results are observed. For the second step of MMP hydrolysis along the associative pathway (int1_2 → com2), two pathways are obtained at the B3LYP/6−31+G(d,p) level with energy barriers of 9.9 and 3.7 kcal/mol, respectively, and differ in the orientation of the OH group in the equatorial plane of the phosphorane-like transition state (see Figure 2). When the OH forms a hydrogen-bond to the leaving group (-OCH₃), the corresponding transition state has a lower energy. At the SCC-DFTBPR level, however, only one transition state is obtained despite numerous attempts; the corresponding OH group is oriented in a position that is approximately the average of those in the two B3LYP

Phosphate Hydrolysis Reactions

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2077**

Table 6. Errors in the SCC-DFTBPR Optimized Structures for Species Involved in the Hydrolysis Reactions Included in the Parameterization[a]

structures	property[a]	error analysis[b]			
		MAXE	RMSE	MUE	MSE
stable states	P–O[c]	0.41/–0.12	0.05/0.02	0.02/0.02	0.00/0.00
	O–P–O[d]	–11.8/7.8	2.3/1.9	1.4/1.3	–0.1/0.0
	P–O[Nu e]	–0.40/–0.04	0.14/0.02	0.07/0.02	–0.03/0.01
transition states	P–O[Lg f]	–0.33/0.18	0.15/0.08	0.10/0.05	0.03/0.04
	P–O[g]	0.07	0.02	0.02	0.00
	O–P–O[h]	11.5/6.5	2.9/2.5	2.0/1.9	0.0/–0.2

[a] The reference structures are optimized at the B3LYP/6–31+G(d,p) level. Bond distances are in Å and bond angles are in degrees. [b] The errors are defined in the same manner as in Table 3. [c] The number before the slash is the overall performance; the number after the slash excludes one extreme deviation from int1_2 of DMP. [d] The number before the slash is the overall performance; the number after the slash excludes one extreme deviation from int1 of DMP. [e] The distance between P and the nucleophilic oxygen ("O[Nu]"); the number before the slash is the overall performance; the number after the slash excludes extreme deviations from ts1_2 and ts1_da of MMP. [f] The distance between P and the key oxygen in the leaving group ("O[Lg]"); the number before the slash is the overall performance; the number after the slash excludes one extreme deviation from ts2 and ts1_da of MMP. [g] The distances between P and oxygen atoms not involved in the nucleophilic attack or the leaving group. [h] The number before the slash is the overall performance; the number after the slash excludes one extreme deviation from ts1_2 of MMP.

Table 7. Benchmark Results for SCC-DFTB for the 19 Phosphate Reactions from the QCRNA Database[a]

state	property[c]	error analysis[b]			
		MAXE	RMSE	MUE	MSE
overall	energy SP (2nd-order)[e]	–24.6	10.3	8.2	–6.0
	energy SP (SCC-DFTBPR)	14.4	5.6	4.8	–0.2
stable states	energy SP (2nd-order)	–23.9	9.9	7.9	–4.6
	energy SP (SCC-DFTBPR)	14.4	5.3	4.3	1.0
	energy OPT (2nd-order)	–27.2	11.8	9.1	–6.1
	energy OPT (SCC-DFTBPR)	14.8	5.3	4.5	–0.8
	P–O (Å) OPT (SCC-DFTBPR)[d]	0.45/0.19	0.07/0.035	0.03/0.02	0.00/0.00
	O–P–O (°) OPT (SCC-DFTBPR)[e]	–9.3	2.2	1.5	0.0
transition states	energy SP (2nd-order)	–24.6	10.6	8.7	–7.5
	energy SP (SCC-DFTBPR)	12.8	6.0	5.4	–1.6
dipole moment[f]	2nd-order	–1.2/–1.2	0.6/0.6	0.5/0.5	–0.4/–0.4
	SCC-DFTBPR	0.8/1.1	0.3/0.5	0.2/0.3	0.1/0.1

[a] The reference values are from the B3LYP calculations included in the QCRNA database (http://theory.chem.umn.edu/QCRNA/). The energies are in kcal/mol and dipole moments in Debye. For specific values, see Supporting Information. [b] The errors are defined in the same manner as in Table 3. [c] "Energy" refers to reaction energy (energy relative to the reactant state); "SP" represents single point calculation at the QCRNA structures; "OPT" indicates full geometry optimization at the relevant SCC-DFTB level with the B3LYP reference structure as the initial guess. [d] The number before the slash is the performance including all the P–O distances (in total, 327 P–O distances are studied); the number after the slash excludes one parameter from C6H5OH...P(O)(O)(-O-sugar-O-)_min2, (CH3)2CHOH... P(O)(O)(-O-sugar-O-)_min3, CH3OH...P(O)(O)(-O-CH2CH2-O-)_min3, CH3OH...P(O)(O)(-O-sugar-O-)_min3, HOH...P(O)(O)(-O-2'methyl-sugar-O-)_min3a, HOH...P(O)(O)(-O-CH2CH2-O-)_min3; the notation scheme follows the QCRNA convention (http://theory.chem.umn.edu/QCRNA/Nomenclature.html). [e] In total, 568 O–P–O angles are analyzed. [f] Numbers before slashes are from single point calculations at the DFT geometries from the QCRNA database; those after the slashes are after geometry optimization at the respective SCC-DFTB levels. The reference values are from QCRNA database at the level of B3LYP/6–311++G(3df,2p)//B3LYP/6–31++G(d,p).

transition states. For the dianionic species, the product-like complex (int_da_MMP in Figure 2) at the B3LYP level is featured with a very strong and short (1.494 Å) hydrogen bond between CH_3OH and HPO_4^{2-}. At the SCC-DFTBPR level, however, the proton on CH_3OH is transferred to HPO_4^{2-}, which leads to a very different structure; interestingly, B3LYP optimization starting from the SCC-DFTBPR geometry led to a very similar structure, which is in fact lower than the original B3LYP structure by ~2.9 kcal/mol. The similar behavior is found for the molecular complex between two water and MMP (com1_w2), where B3LYP optimization starting from SCC-DFTBPR geometry led to a structure with lower energy (~2 kcal/mol) than the original B3LYP structure. These findings not only confirm the robustness of SCC-DFTBPR for structural properties but also highlight the value of a fast method in exploring conformational space.

D. Additional Benchmark Calculations with the QCRNA database. For the large number of energetics data points studied here (56 reaction energies and 47 barriers), single point values at the standard second-order SCC-DFTB level have larger errors (see Table 7 for a summary and the Supporting Information for details); the largest error is –24.6 kcal/mol, and the RMSE is 10.3 kcal/mol. Similar errors are found when the structures are optimized (only for stable states). With SCC-DFTBPR, the errors are substantially smaller; for single point energetics, the largest error is 14.4 kcal/mol, and the RMSE is only 5.6 kcal/mol. When the structures are optimized for the stable states, the corresponding values are 14.8 kcal/mol (see below) and 5.3 kcal/mol, respectively, which are rather encouraging. The RMSEs in the optimized P–O distances and O–P–O angles are 0.035 Å and 2.2 degrees, respectively. The dipole moment for the structures are well described with both the second-order
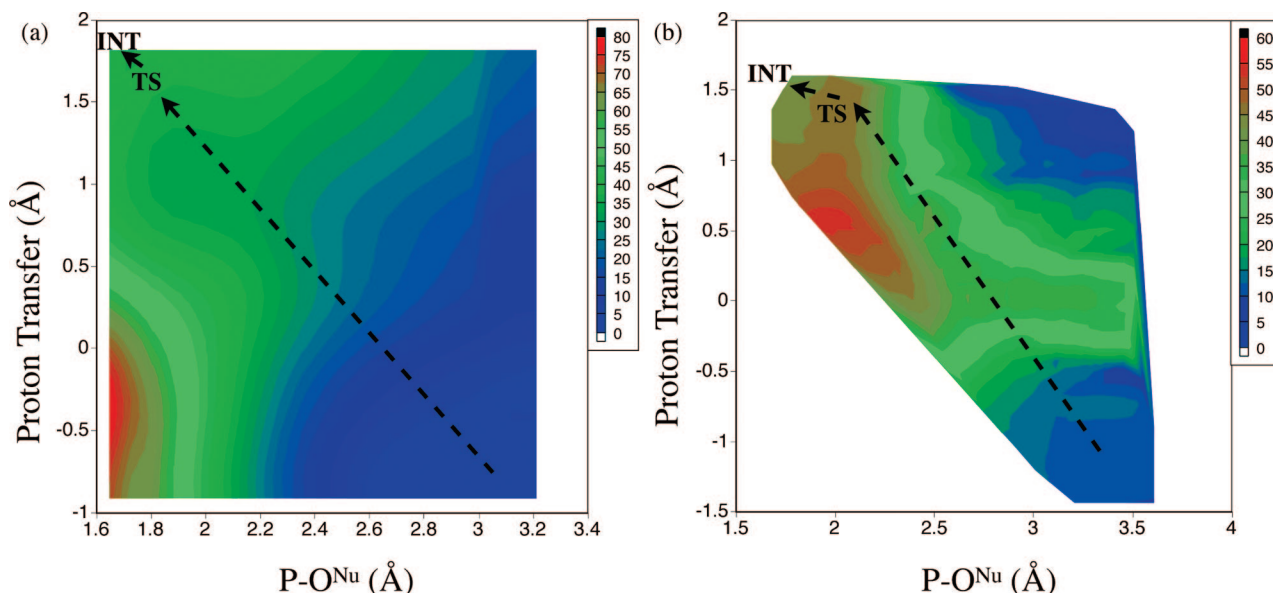
Figure 3. SCC-DFTBPR results (energies in kcal/mol) for the first step of the associative pathway for the hydrolysis of monomethyl phosphate (MMP). (a) Two-dimensional potential energy surface from adiabatic mapping calculations in the gas phase using SCC-DFTBPR. (b) Two-dimensional potential of mean force from umbrella sampling in solution using a SCC-DFTBPR/MM model. The O^{Nu} stands for the nucleophilic oxygen in water (see Scheme 1), and the proton transfer coordinate is the antisymmetric stretch that describes the proton transfer between the nucleophilic water and the basic oxygen in MMP.

SCC-DFTB and SCC-DFTBPR, with RMSEs on the order of 0.5 Debye.

Detailed analyses of results indicate that all larger errors (>15 kcal/mol) in the standard SCC-DFTB approach occur in reactions that involve OH$^-$ as a reactant, such as OH$^-$...P(O)(OCH$_3$)(-O-CH$_2$CH$_2$-O-) and OH$^-$...P-(O)(OCH$_3$)(OCH$_3$)(OCH$_3$). With the third-order terms, the errors in those reactions typically reduce significantly to merely a few kcal/mol (for details, see Table S1 in the Supporting Information). The largest error at the SCC-DFTBPR level, 14.8 kcal/mol, is found for the exothermicity for the reaction between OH$^-$ and P(O)(OH)(OH)(OCH$_3$), which has a very large value of -60.7 kcal/mol at the B3LYP/6$-$311++G(3df, 2p) level.

E. MMP Hydrolysis with SCC-DFTBPR/MM Simulations. *1. Associative Mechanism.* In the gas phase, the two-dimensional PES (Figure 3a) shows a high barrier of $\sim$33 kcal/mol, which is consistent with the results of transition state optimization at the same SCC-DFTBPR level (32.1 kcal/mol, com1 $\rightarrow$ ts1, Table 5). The transition state and intermediate regions are very flat on the two-dimensional PES, which is also seen in the optimization calculations; the intermediate is lower than the transition state by 0.1 kcal/mol at the SCC-DFTBPR level and 0.4 kcal/mol at the B3LYP/6$-$31+G** level. The shape of the two-dimensional PES makes it clear that a strictly stepwise mechanism, in which the proton is first transferred to MMP to generate the nucleophilic hydroxide, is not energetically favorable. This is confirmed by comparing the adiabatic energy profile along the proton transfer coordinate with the P$-$O^{Nu} fixed at 3.0 Å at the SCC-DFTBPR and B3LYP/6$-$311++G**//SCC-DFTBPR levels. The results are purely uphill and numerically close at the two levels of calculations (data not shown); for example, at the proton transfer (antisymmetric stretch) coordinate of -0.2, 0.2, and 0.4 Å, respectively, the B3LYP

result is lower than SCC-DFTBPR by 0.8, 3.2, and 4.0 kcal/mol, respectively.

In the solution phase, an equally late transition state is identified on the two-dimensional PMF (Figure 3b) with a barrier of $\sim$34 kcal/mol. Clearly, solvation does not stabilize the transition state over the reactant in any major way in the associative pathway, which is reasonable since there is no major difference between the dipole moment of the transition state and the reactant complex; at the SCC-DFTBPR level, the values are 6.9 and 6.2 Debye for the reactant and transition state, respectively. Our value is close to the finding of Florian et al., who obtained a barrier of $\sim$35 kcal/mol using a Langevin dipole model for solvation and structures from gas-phase reaction path optimizations.[16]

2. Dissociative Mechanism. Without the "proton relay" mediated by water molecules, the first step of the dissociative pathway involves an intramolecular proton transfer to the leaving group, generating a complex between metaphosphate and methanol. In the gas phase, the two-dimensional PES (Figure 4a) clearly shows a saddle point in which the proton is equally shared between the metaphosphate and the leaving group oxygen (proton transfer coordinate $\sim$$-0.1-0.0$ Å), and the distance between the leaving group and phosphorus significantly lengthens by $\sim$0.4 Å relative to MMP. The barrier is about 35 kcal/mol, which is close to the result of the transition state search (com1 $\rightarrow$ diss_tsa in Table 5) at both the SCC-DFTBPR (34.2 kcal/mol) and MP2 (36.8 kcal/mol) levels.

In solution, the two-dimensional PMF (Figure 4b) points to a transition state with a similar shared-proton feature but more compressed P$-$O^{Lg} distance of 1.7$-$1.8 Å. Physically speaking, the observed shift in the P$-$O^{Lg} distance in the transition state upon solvation is to be expected. Without breaking the P$-$O^{Lg} bond, the intramolecular proton transfer leads to a species with significant charge separation, which

Phosphate Hydrolysis Reactions

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2079**
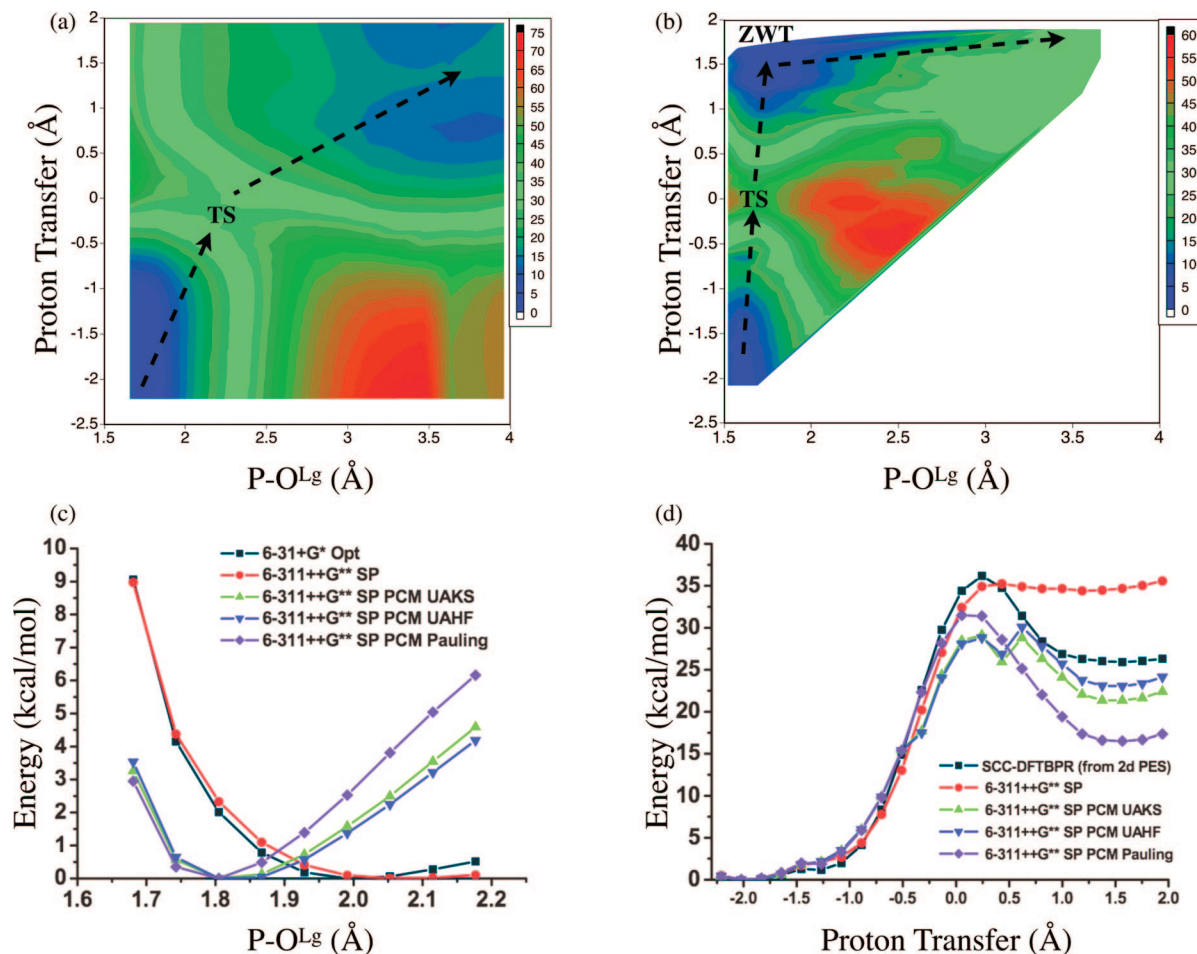
Figure 4. SCC-DFTBPR results (energies in kcal/mol) for the first step of the dissociative pathway for the hydrolysis of monomethyl phosphate (MMP). (a) Two-dimensional potential energy surface from adiabatic mapping calculations in the gas phase using SCC-DFTBPR. (b) Two-dimensional potential of mean force from umbrella sampling in solution using a SCC-DFTBPR/MM model. The O^{Lg} stands for the oxygen in the leaving group (see Scheme 1), which is methanol in this case; the proton transfer coordinate is the antisymmetric stretch that describes the intramolecular proton transfer between the protonated oxygen in MMP and O^{Lg}. (c) Adiabatic mapping along the P–O^{Lg} distance with the proton transfer coordinate fixed at 0.0; structures are partially optimized at the B3LYP/6−31+G(d) level in the gas phase, followed by single point continuum solvent calculations with the 6−311++G(d,p) basis and various sets of radii. (d) Adiabatic mapping along the proton transfer coordinate with the P–O^{Lg} distance fixed at ∼1.824 Å; here structures are optimized with SCC-DFTBPR, followed by single point energy calculations at the B3LYP level in both gas-phase and continuum solvent model.

is much better stabilized in solution than in the gas phase. Therefore, the transition state in the gas phase has a substantially longer P–O^{Lg} distance than in solution. To quantitatively verify the solvation effect on the transition state structure observed in the SCC-DFTBPR based calculations, an adiabatic mapping scan in the gas phase along the P–O^{Lg} distance is first performed with the antisymmetric stretch proton transfer coordinate constrained at 0.0 Å; all other degrees of freedom are optimized at the B3LYP/6−31+G* level. Single point B3LYP/6−311++G** energies at these partially optimized structures are then compared between gas-phase and implicit solvent (PCM) calculations. As shown in Figure 4c, an energy minimum is found around the P–O^{Lg} distance of 2.0−2.1 Å in the gas phase, which is consistent with the 2d-PES shown in Figure 4a. With the implicit solvent model, however, the position of the energy minima along P–O^{Lg} is shifted to be around 1.8 Å, regardless of the atomic radii used in the B3LYP-PCM calculations. Therefore, the SCC-DFTBPR based results are consistent with B3LYP

calculations in both the gas phase and solution, which is very encouraging.

The solution PMF barrier calculated at the SCC-DFTBPR/MM level is about 32 kcal/mol (Figure 4b), which is consistent with previous results obtained using various implicit solvent models. For example, Warshel and co-workers found a barrier of 34 kcal/mol by using MP2/6−31+G(d,p) energies and the Langevin dipole model for solvation;[16] Bianciotto et al.[90] obtained a barrier of 33.5 kcal/mol using B3LYP and a modified double-ζ plus polarization valence basis set along with the PCM solvation model.

Despite these encouraging aspects of the SCC-DFTBPR/MM calculations, an unexpected feature is observed in the PMF. As shown in Figure 4b, a very low-energy region is found in the left-upper corner of the 2d-PMF, which corresponds to a zwitterionic intermediate. In fact, the current PMF calculations predict that this zwitterionic species is even more stable than the reactant state by ∼3 kcal/mol. Such a zwitterionic intermediate has been discussed in the compu-
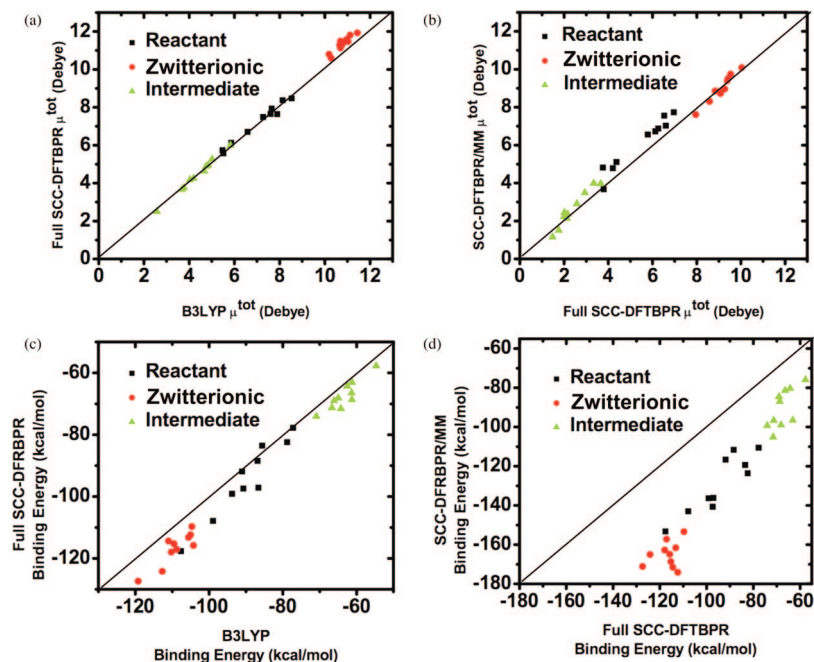
Figure 5. Additional analysis for the reliability of the SCC-DFTBPR approach for the dissociative pathway in MMP hydrolysis. (a-b) Dipole moment of the solute in different chemical states (MMP, zwitterionic and intermediate state) calculated at different levels. (c-d) Binding energy between the solute and nearby water molecules (within the first solvation shell, see Figure S10 in the Supporting Information for the computed solvent distribution function) at different levels; SCC-DFTBPR/MM refers to calculations in which the solute is treated with SCC-DFTBPR and the water with MM, while SCC-DFTBPR and B3LYP indicates calculations where both the solute and water molecules are treated as QM.

tational study of Bianciotto et al.,[90,91] who argued that the dissociative mechanism of the MMP hydrolysis in solution may follow a stepwise pathway that involves a stable zwitterionic intermediate produced by intramolecular proton transfer to the leaving group. Using B3LYP-PCM and a double-ζ plus polarization quality basis, Bianciotto et al. predicted that such an intermediate is ~21.2 kcal/mol higher than the reactant, which is at odds with the current SCC-DFTBPR/MM result.

Since the zwitterionic type of species has not been considered in the development of SCC-DFTBPR, it is possible that the stability of this structure is overestimated at the SCC-DFTBPR/MM level. To better understand the situation, several sets of calculations are carried out.

First, to explore the intrinsic performance of SCC-DFTBPR, adiabatic mapping calculations are carried out for MMP in the gas phase along the proton transfer coordinate while holding the $P-O^{Lg}$ distance fixed at 1.824 Å. At these partially optimized structures at the SCC-DFTBPR level, B3LYP/6−311++G** single-point energy calculations are carried out in both the gas phase and with the PCM model for solution. In the gas phase, as shown in Figure 4d, SCC-DFTBPR and B3LYP agree well for the antisymmetric proton transfer coordinate below 0.4 Å; the agreement deteriorates quickly as the PT coordinate further increases. For example, with the proton transfer coordinate of 2.0 Å, the SCC-DFTBPR energy is lower than the B3LYP value by as much as 9 kcal/mol. At the B3LYP level, with the PCM model describing solvation, the partially optimized zwitterionic species varies in the range of 18−24 kcal/mol above the reactant, depending on the radii used; these values

are largely consistent with the result of Bianciotto et al.,[90,91] indicating that the SCC-DFTBPR structures are quite promising.

The ~9 kcal/mol error in the SCC-DFTBPR result for the zwitterionic species in the gas phase does not explain the significantly overestimated stability in solution by the SCC-DFTBPR/MM simulations. In fact, this discrepancy suggests that the current SCC-DFTBPR/MM simulations have significantly unbalanced treatment of solute−solvent interactions between the reactant and the zwitterionic species, with the latter significantly overstabilized. This can be caused either by large errors in the predicted charge distribution (e.g., dipole moment) in the zwitterionic structure or by nonoptimal van der Waals parameters for the QM atoms used in the current simulations. To distinguish the two possibilities, we study the binding energies of the solute with nearby water molecules (~9−11) in the reactant (MMP), zwitterionic, and "intermediate" (metaphosphate with a weakly bound methanol) states; for each case, 10 snapshots are taken from SCC-DFTBPR/MM trajectories. As shown in Figure 5a, the dipole moment of the solute at the SCC-DFTBPR level correlates very well with B3LYP/6−311++G** result, regardless of the chemical state of the solute. The dipole moment of the solute-water cluster is not sensitive to whether the water molecules are treated as MM or QM (Figure 5b). We do note, however, that the magnitude of charge separation in the zwitterionic state is substantially higher at the SCC-DFTBPR/MM level; e.g., the average net Mulliken charge on the metaphosphate is −1.54 with SCC-DFTBPR/MM, as compared to the value of ~−1.15 with full SCC-DFTBPR, and the value of ~−0.95 based on B3LYP NBO charges.

Phosphate Hydrolysis Reactions

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2081**

Therefore, the electronic structure (or electron distribution) seems adequately described at the SCC-DFTBPR level even for the zwitterionic species, and the MM treatment of the nearby solvent causes a higher degree of polarization in the solute.

The total binding energy between the solute and nearby water molecules, however, is significantly overestimated at the SCC-DFTBPR/MM level, especially when the solute is in the zwitterionic state (Figure 5c). In the MMP state, the average interaction between the solute and nearby water molecules is −129.1 kcal/mol with SCC-DFTBPR/MM, which is significantly larger than the value of −94.3 (−89.7) kcal/mol when all atoms are described with SCC-DFTBPR (B3LYP/6−311++G**); for the zwitterionic state, the corresponding values are −165.1, −116.7, and −109.1 kcal/mol. In other words, the MM treatment of the nearby water molecules preferentially stabilizes the zwitterionic species by almost 20 kcal/mol [e.g., (165.1−129.1) − (109.1 ≤ 89.7) ∼ 17 kcal/mol!], which explains why the zwitterionic species is overstabilized in the SCC-DFTBPR/MM simulations. Therefore, these results vividly demonstrate that the quantitative accuracy of QM/MM simulations for reactions involving significant charge redistributions relies very sensitively on the treatment of QM/MM interactions. We emphasize that this is not a straightforward problem to "fix" by simply refitting the van der Waals parameters for the QM atoms based on simple solute-water dimers as commonly done in the literature;[21,82] in fact, such fitted QM van der Waals parameters are found to be very similar to the set used here in a recent QM/MM study of phosphoryl transfer reaction in water and *tert*-butyl alcohol solutions.[76] For highly charged solutes, the errors are due to the accumulative effects of a large number of solute−solvent interactions (e.g., more than 8 hydrogen bonds are involved in the MMP-water clusters studied here), thus a balanced treatment of QM/MM interactions over different chemical states of the solute likely calls for more sophisticated approaches that allow the QM nonbond parameters to vary during the reaction.[83] Another possibility is that the QM/MM interaction is better represented by a Klopman-Ohno form, which is substantially damped at the intermediate and short distances and more consistent with the way that charge−charge interactions are treated for the QM atoms in SCC-DFTB.[60] The simple Coulombic form is used in the integration of second-order SCC-DFTB with MM because this combination empirically compensates for the fact that SCC-DFTB Mulliken charges tend to be too low;[48,82] with modified Mulliken charges at the third-order level, however, it is possible that the Klopman-Ohno form becomes more appropriate. We leave the systematic analysis of these possibilities to future work (however, see the Supporting Information for some preliminary results). Along this line, it is encouraging to observe that the differential solute/solvent interaction between MMP and the zwitterionic species at the SCC-DFTBPR level (i.e., nearby water also treated as QM), −22.4(=94.3−116.7) kcal/mol, is very close to the value of −19.4(=89.7−109.1) kcal/mol at the B3LYP/6−311++G** level, which suggests that an attractive alternative is to treatment the first solvation shell at the QM level[97] and that SCC-DFTBPR seems adequate

in this regard. A technical challenge for such calculations, however, is that the QM/MM interface needs to be updated in an adaptive fashion during the simulation,[98−101] which in fact is not problematic if only thermodynamic quantities are of interest.

In summary, the combination of gas-phase adiabatic mapping, solution QM/MM PMF calculations, and comparison to relevant B3LYP calculations in the gas phase or with implicit solvent models indicate that the SCC-DFTBPR approach gives reliable structures, including transition states for both associative and dissociative pathways and the zwitterionic intermediate in the dissociative pathway. For structures with similar charge distributions (e.g., dipole moment), the SCC-DFTBPR/MM simulations give rather satisfying energetics as well; for describing the relative energetics of species with very different charge distributions, however, a simple treatment of the solute/solvent interaction at the traditional QM/MM level may not be sufficient.

IV. Conclusions

Despite extensive efforts from both experimental and theoretical studies, the precise mechanism of phosphate chemistry such as phosphate hydrolysis and phosphoryl transfer remains controversial. This is a major challenge to tackle because phosphate chemistry plays a key role in many essential biological processes such as energy/signal transduction and synthesis of protein and nucleic acids. From a theoretical stand point, the key is to develop effective computational methods that can balance accuracy and sampling efficiency, which, with the typical computational hardware, naturally points to the development of semiempirical methods. There have been several such models established in recent years for specific types of phosphate reactions[30,33,34] based on the traditional NDDO framework although their general applicability still remains to be fully explored.

In this work, we make an attempt to parametrize an approximate density functional theory, SCC-DFTB, as an alternative approach for studying the chemistry of phosphorus containing systems in solution and biological systems. This is motivated by the recent success of SCC-DFTB for studying the structural and energetics of biological systems.[25,52,59] It is found that although a standard second-order formulation of SCC-DFTB gives good geometries compared to high-level density functional theories, further extensions are needed to obtain reliable proton affinity and reaction energies. Including the on-site third-order terms is found to improve the proton affinity significantly, although it remains difficult to obtain accurate proton affinity for both phosphorus containing compounds and those that do not contain phosphorus; it is possible that off-diagonal third-order terms are needed to resolve such difficulty. As a pragmatic approach, we have developed two sets of "reaction specific" parametrizations, SCC-DFTBPA for proton affinity of phosphates and SCC-DFTBPR for phosphate hydrolysis. The number of "reaction specific" parameters, however, is small (7 in total for O, N, C, H, and P), and therefore SCC-DFTBPR is likely applicable to a broader set of phosphate reactions.

Yang et al.

Benchmark calculations in the gas phase and solution with a QM/MM framework indicate that the current parametrizations, particularly SCC-DFTBPR, generally give reliable structures and semiquantitative energetics (e.g., with a RMSE of ~3−5 kcal/mol compared to high-level calculations). Therefore, these methods are attractive choices for exploring the gross features of the potential energy surfaces of condensed-phase systems and for identifying amino acids and/or structural fluctuations that play an important role in controlling the chemical step.[5] Higher level QM/MM calculations are still required for more quantitative understanding, although the number of variables in these much more expensive calculations can be substantially reduced by SCC-DFTB(PR) based studies. The solution-phase benchmark study for the zwitterionic species in the dissociative pathway of MMP hydrolysis also underlines the importance of carefully handling solute/solvent interactions for reactions involving highly charged species, which may require sophisticated treatment of van der Waals parameters for the QM atoms or describing the first solvation shell of the solute with a QM approach. Quantitative analysis along these lines and *systematic* comparison of the new SCC-DFTB models to NDDO based models for phosphate chemistry[30,33,34,102] remain important goals for the near future.

Supporting Information Available: Structures for all molecules (optimized at the SCC-DFTBPR level) involved in the parametrization process, solvent radial distribution functions near MMP from SCC-DFTBPR/MM simulations, and summarization of the detailed comparison (energetics and dipole moment) of the second-order SCC-DFTB and optimized SCC-DFTBPR for selected molecules in the QCRNA database. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular biology of the cell*; Garland Publishing, Inc.: 1994.

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

(3) Vale, R. D.; Milligan, R. A. *Science* **2000**, *288*, 88–95.

(4) Yu, H.; Ma, L.; Yang, Y.; Cui, Q. *PLoS Comput. Biol.* **2007**, *3*, 199.

(5) Yang, Y.; Yu, H.; Cui, Q. *J. Mol. Biol.* **2008**, *381*, 1407–1420.

(6) Mueller-Planitz, F.; Herschlag, D. *J. Biol. Chem.* **2006**, *281*, 23395–23404.

(7) Herschlag, D.; Jencks, W. P. *J. Am. Chem. Soc.* **1987**, *109*, 4665–4674.

(8) Herschlag, D.; Jencks, W. P. *J. Am. Chem. Soc.* **1989**, *111*, 7579–7586.

(9) Hollfelder, F.; Herschlag, D. *Biochemistry* **1995**, *34*, 12255–12264.

(10) Admiraal, S.; Herschlag, D. *Chem. Biol.* **1995**, *2*, 729–739.

(11) Åqvist, J.; Kolmodin, K.; Florian, J.; Warshel, A. *Chem. Biol.* **1999**, *6*, R71–R80.

(12) Florian, J.; Aqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 11524–11525.

(13) Glennon, T. M.; Villa, J.; Warshel, A. *Biochemistry* **2000**, *39*, 9641–9651.

(14) Florian, J.; Warshel, A. *J. Am. Chem. Soc.* **1997**, *119*, 5473–5474.

(15) Friedman, J. M.; Freeman, S.; Knowles, J. R. *J. Am. Chem. Soc.* **1988**, *110*, 1268–1275.

(16) Florian, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 719–734.

(17) O'Brien, P. J.; Herschlag, D. *Biochem.* **2002**, *41*, 3207–3225.

(18) Doherty, E. A.; Doudna, J. A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 457–475.

(19) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.

(20) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solution*; Wiley: New York, 1991.

(21) Gao, J. In *Reviews in Computational Chemistry VII*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1995.

(22) Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173–290.

(23) Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.

(24) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.

(25) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458–6469.

(26) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.

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

(28) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209–220.

(29) Range, K.; Riccardi, D.; Elstner, M.; Cui, Q.; York, D. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3070–3079.

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

(31) Thiel, W.; Voityuk, A. A. *Theo. Chem. Acc.* **1992**, *81*, 391–404.

(32) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616–626.

(33) Arantes, G. M.; Loos, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 347–353.

(34) Lopez, X.; York, D. M. *Theo. Chem. Acc.* **2001**, *109*, 149–159.

(35) Xu, D.; Guo, H.; Liu, Y.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 13827–13834.

Phosphate Hydrolysis Reactions

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2083**

(36) Nam, K.; Gao, J.; York, D. M. Wiley: New York, 2007; chapter New QM/MM Models for Multi-scale Simulation of Phosphoryl Transfer Reactions in Solution, pp 201−218.

(37) Nam, K. H.; Gao, J. L.; York, D. M. *J. Am. Chem. Soc.* **2008**, *130*, 4680–4691.

(38) Lee, T. S.; Lopez, C. S.; Glambasu, G. M.; Martick, M.; Scott, W. G.; York, D. M. *J. Am. Chem. Soc.* **2008**, *130*, 3053–3064.

(39) Nam, K.; Gao, J.; York, D. M. RNA 2008, in press.

(40) Hu, C. H.; Brinck, T. *J. Phys. Chem. A* **1999**, *103*, 5379–5386.

(41) Li, G. H.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 3342–3357.

(42) Gerratana, B.; Sowa, G. A.; Cleland, W. W. *J. Am. Chem. Soc.* **2000**, *122*, 12615–12621.

(43) Elstner, M.; Porezag, D.; Jungnickel, G.; Elstner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.

(44) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2000**, *256*, 15–27.

(45) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2001**, *263*, 203–219.

(46) Hu, H.; Elstner, M.; Hermans, J. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 451–463.

(47) Zhu, X.; Yethiraj, A.; Cui, Q. *J. Chem. Theory Comput.* **2007**, *3*, 1538–1549.

(48) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569–585.

(49) Zhang, X.; Harrison, D.; Cui, Q. *J. Am. Chem. Soc.* **2002**, *124*, 14871–14878.

(50) Li, G.; Cui, Q. *J. Am. Chem. Soc.* **2003**, *125*, 15028–15038.

(51) Bondar, A. N.; Fischer, S.; Smith, J. C.; Elstner, M.; Suhai, S. *J. Am. Chem. Soc.* **2004**, *126*, 14668–14677.

(52) Elstner, M.; Frauenheim, T.; Suhai, S. *THEOCHEM* **2003**, *632*, 29.

(53) Kruger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. *J. Chem. Phys.* **2005**, *122*, 114110.

(54) Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. *J. Phys. Chem. A* **2006**, *110*, 13551–13559.

(55) Otte, N.; Scholten, M.; Thiel, W. *J. Phys. Chem. A* **2007**, *111*, 5751–5755.

(56) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149–5155.

(57) Liu, H. Y.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Hermans, J.; Yang, W. T. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 484–489.

(58) Neihaus, T. A.; Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct. Theochem* **2001**, *541*, 185–194.

(59) Elstner, M. *Theo. Chem. Acc.* **2006**, *116*, 316−325.

(60) Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5614–5621.

(61) Yang, Y.; Yu, H.; York, D. M.; Elstner, M.; Cui, Q. *J. Phys. Chem. A* **2007**, *111*, 10861–10873.

(62) Giese, T. J. et. al.; York, D. M. *J. Mol. Graphics Modell.* **2006**, *25*, 423–433.

(63) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565–581.

(64) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(65) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(66) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(67) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(68) Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley: Boston, 1989.

(69) Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618–4627.

(70) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650–7657.

(71) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764–7776.

(72) Herschlag, D.; Jencks, W. P. *J. Am. Chem. Soc.* **1989**, *111*, 7587–7596.

(73) Perreault, D. M.; Anslyn, E. *Angew. Chem., Int. Ed.* **1997**, *36*, 432–450.

(74) Lopez, C. S.; Faza, A. N.; Gregersen, B. A.; Lopez, X.; de Lera, A. R.; York, D. M. *ChemPhysChem* **2004**, *5*, 1045–1049.

(75) Lopez, C. S.; Faza, O. N.; de Lera, A. R.; York, D. M. *Chem. Eur. J.* **2005**, *11*, 2081–2093.

(76) Yang, Y.; Cui, Q. Manuscript in preparation.

(77) Brooks, C. L., III.; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312–6325.

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

(79) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902–1921.

(80) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(81) Schlenkrich, M.; Brickmann, J.; MacKerell, A., Jr.; Karplus, M. *Biological Membranes: A Molecular Perspective from Computation and Experiment*; Birkhauser: 1996.

(82) Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467–6478.

(83) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2007**, *127*, 194101.

(84) Torrie, G.; Valleau, J. *J. Comp. Phys.* **1977**, *23*, 187–199.

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

(86) Im, W.; Bernéche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924–2937.

(87) Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys.* **2005**, *123*, 014905.

(88) Brooks, C. L.; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159–181.

(89) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(90) Bianciotto, M.; Barthelat, J. C.; Vigroux, A. *J. Am. Chem. Soc.* **2002**, *124*, 7573–7587.

(91) Bianciotto, M.; Barthelat, J. C.; Vigroux, A. *J. Phys. Chem. A* **2002**, *106*, 6521–6526.

(92) Wang, Y.-N.; Topol, I. A.; Collins, J. R.; Burt, S. K. *J. Am. Chem. Soc.* **2003**, *125*, 13265−13273.

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

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

(95) Grigorenko, B. L.; Rogov, A. V.; Nemukhin, A. V. *J. Phys. Chem. B* **2006**, *110*, 4407–4412.

(96) Fischer, S.; Karplus, M. *Chem. Phys. Lett.* **1992**, *194*, 252–261.

(97) Yang, Y.; Cui, Q. *J. Phys. Chem. B* **2007**, *111*, 3999–4002.

(98) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. *Chem. Phys.* **1996**, *211*, 313.

(99) Hofer, T. S.; Pribil, A. B.; Randolf, B. R.; Rode, B. M. *J. Am. Chem. Soc.* **2005**, *127*, 14231.

(100) Kerdcharoen, T.; Morokuma, K. *Chem. Phys. Lett.* **2002**, *355*, 257.

(101) Heyden, A.; Lin, H.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 2231–2241.

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

JCTC Journal of Chemical Theory and Computation

Gas-Phase Reaction of NH$_2$$^+$ with Acetic Acid: Implications in Astrochemistry

Laura Largo, Víctor M. Rayón, Carmen Barrientos, Antonio Largo, and
Pilar Redondo*

*Departamento de Química Física, Facultad de Ciencias, Universidad de Valladolid,
47005 Valladolid, Spain*

Received July 18, 2008

Abstract: A theoretical study of the ion−molecule reaction, NH$_2$$^+$ with acetic acid that could lead to precursors of glycine in the interstellar medium, has been carried out on the triplet and singlet potential energy surfaces. All stationary points and transition states on the (NO$_2$C$_2$H$_6$)$^+$ triplet and singlet surfaces have been determined at the MP2(full) level with the cc-pVTZ basis set. Energetic data have been obtained at the CCSD(T) level employing the aug-cc-pVTZ basis set. The global minimum of the (NO$_2$C$_2$H$_6$)$^+$ system is predicted to be protonated glycine in its singlet state, ^{1}A′, and in general singlet states are more stable than the corresponding triplet ones. Formation of ionized glycine from this reaction is shown to be a feasible process under interstellar conditions, but the proton transfer channel and the formation of the compound CH$_2$COOH$^+$ seem to be more favorable processes on the triplet and singlet potential surfaces, respectively.

Introduction

To date more than 140 molecular species have been identified in space and in interstellar clouds as well as in circumstellar envelopes. In the future this list of molecules will undoubtedly continue to grow as more powerful telescopes and interferometer arrays are built and utilized. A large fraction of these molecules is organic or carbon-based. In this way a lot of attention is devoted to the quest for biomolecules, particularly interstellar amino acids which are the basic units of proteins and therefore key elements for the origin of life. In this context, a subject that has received great interest is the origin of amino acids in the prebiotic chemistry of early Earth.[1,2] One of the possibilities establishes that amino acids may have first originated in the interstellar medium and were then transported to the Earth by comets and meteorites.[3−7] This point of view is supported by the fact that more than 70 amino acids, including eight out of 20 present in living organisms,[8] were isolated in the Murchison meteorite fallen in 1969. However, amino acids have not yet been detected in the interstellar medium. The simplest amino acid, glycine, has long been searched for in space, but its presence still

remains to be confirmed.[9−15] The difficulties in the detection of glycine can be related to the fact that glycine is a relatively large molecule, and, therefore, its rotational spectrum has relatively weak lines.[16] In addition, amino acids are highly susceptible to UV photodestruction,[17] and only in shielded environments such as dense molecular clouds and hot cores (star-forming regions) might they be detectable. However, the recent detection of amino acetonitrile,[18] a possible precursor of glycine, opens new possibilities of identifying more complex organic molecules in the next years. In addition, theoretical studies of the formation of amino acetonitrile in the interstellar medium have been carried out.[19−21] In this context, an interesting point is to determine if there are efficient synthetic routes toward amino acids, in particular toward glycine the simplest one, under interstellar conditions. According to the Miller experiment,[1] amino acids and in particular glycine can be synthesized by mixtures of CH$_4$, NH$_3$, H$_2$O, and H$_2$. Maeda and Ohno in their theoretical studies have proposed a synthetic route of glycine from simple molecules (NH$_2$, CH$_2$, and CO$_2$) in two steps;[22] alternatively they show that the one step reaction of aceto-lactone with ammonia can produce glycine, and its activation barrier is only 1.6 kcal/mol.[23] In a recent work,[24] they also predict that glycine and its analogs can be generated without

* Corresponding author fax: 34-983-423013; e-mail: predondo@
qf.uva.es.
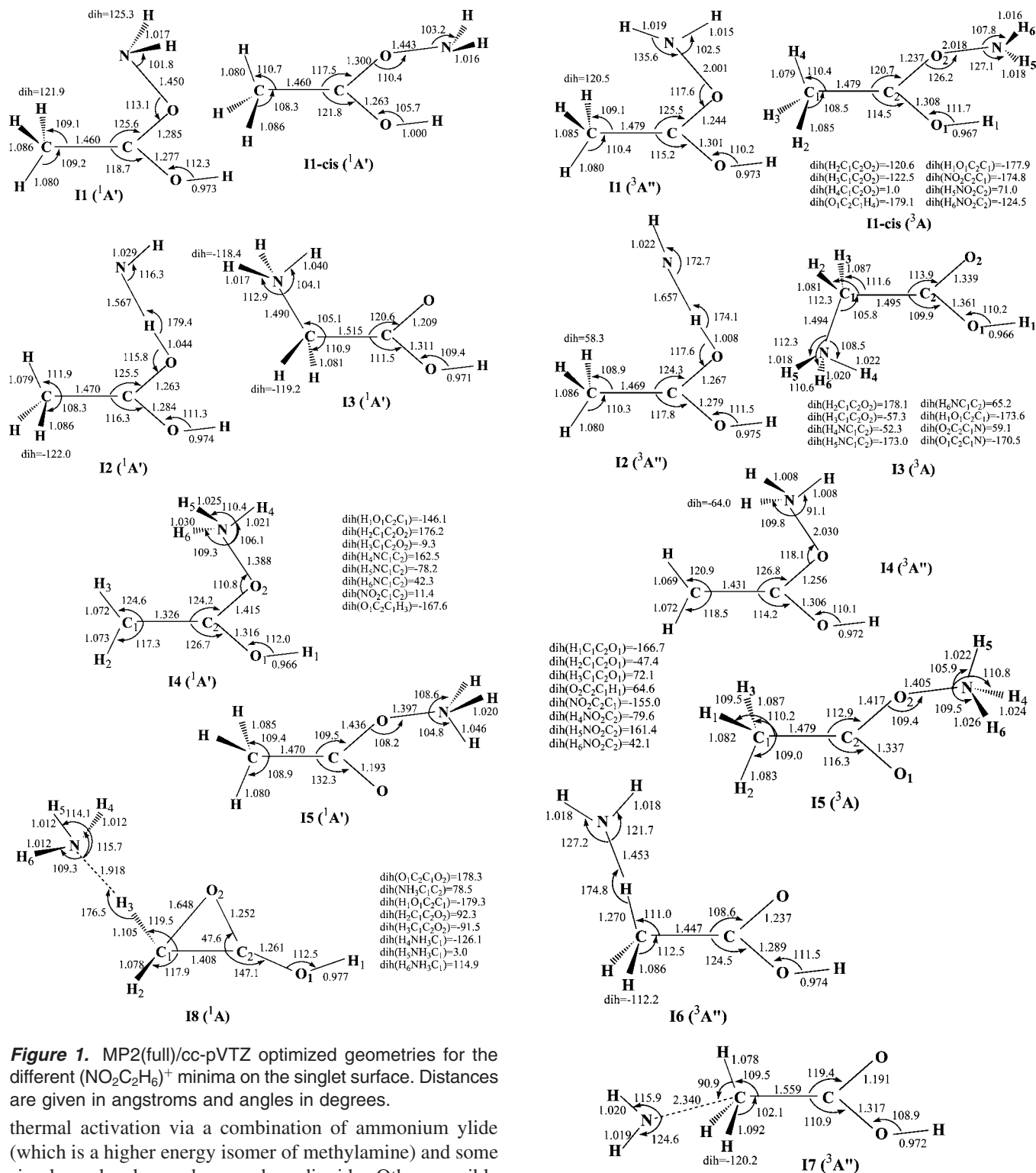
Figure 1. MP2(full)/cc-pVTZ optimized geometries for the different $(NO_2C_2H_6)^+$ minima on the singlet surface. Distances are given in angstroms and angles in degrees.

thermal activation via a combination of ammonium ylide (which is a higher energy isomer of methylamine) and some simple molecules such as carbon dioxide. Other possible routes of amino acids, which are very relevant in interstellar chemistry, are gas-phase ion−molecule reactions initiated by cosmic and ultraviolet radiation, since such reactions are known to be involved in the formation of many other molecules in the interstellar medium.[25] Recent theoretical and experimental studies[26,27] show that glycine and β-alanine could be formed in the interstellar medium by the gas-phase reactions of protonated and ionized hydroxylamine with acetic and propanoic acids. The importance of this route is not known because $NH_{2/3}OH^+$ has not yet been observed in the interstellar space. In a previous work,[28] we have carried

Figure 2. MP2(full)/cc-pVTZ optimized geometries for the different $(NO_2C_2H_6)^+$ minima on the triplet surface. Distances are given in angstroms and angles in degrees.

out a theoretical study of various ion−molecule reactions that could lead to precursors of interstellar glycine where their reaction enthalpies have been calculated at different levels of theory. The most favorable reaction from the thermochemical viewpoint is the reaction of the cation NH_2^+ with acetic acid. This is predicted to lead to ionized glycine $(NH_2CH_2COOH^+)$, but charge transfer reaction was predicted

Table 1. Relative Energies (kcal/mol) for the Relevant (NO$_2$C$_2$H$_6$)$^+$ Species in the Reaction of NH$_2$$^+$ with CH$_3$COOH[a]

	singlet surface			triplet surface	
isomer	ΔE (MP2 = full)	ΔE (CCSD(T))	isomer	ΔE (MP2 = full)	ΔE (CCSD(T))
CH$_3$COOHNH$_2$$^+$ **I1**(^{1}A′)	57.43	53.18	CH$_3$COOHNH$_2$$^+$ **I1**(^{3}A″)	110.66	101.62
CH$_3$COOHNH$_2$$^+$ **I1-cis**(^{1}A′)	51.49	47.99	CH$_3$COOHNH$_2$$^+$ **I1-cis**(^{3}A)	113.60	103.92
CH$_3$COOHHNH$^+$ **I2**(^{1}A′)	138.52	123.67	CH$_3$COOHHNH$^+$ **I2**(^{3}A″)	88.45	80.92
NH$_3$CH$_2$COOH$^+$ **I3**(^{1}A′)	0.00	0.00	NH$_3$CH$_2$COOH$^+$ **I3**(^{3}A)	111.29	99.93
CH$_2$COOHNH$_3$$^+$ **I4**(^{1}A′)	82.95	79.37	CH$_2$COOHNH$_3$$^+$ **I4**(^{3}A″)	148.32	139.17
CH$_3$COONH$_3$$^+$ **I5**(^{1}A′)	47.57	45.37	CH$_3$COONH$_3$$^+$ **I5**(^{3}A)	164.02	149.74
			NH$_2$HCH$_2$COOH$^+$ **I6**(^{3}A″)	124.37	122.20
			NH$_2$CH$_3$COOH$^+$ **I7**(^{3}A″)	128.65	124.01
NH$_3$HCHCOOH$^+$ **I8**(^{1}A)	64.24	63.57			

[a] Zero-point vibrational energy differences have been included.

Table 2. Relative Energies (kcal/mol) for the Possible Products of the Reaction NH$_2$$^+$ with CH$_3$COOH, on the Triplet and Singlet Potential Energy Surfaces[a]

triplet surface	ΔE (MP2 = full)	ΔE (CCSD(T))	singlet surface	ΔE (MP2 = full)	ΔE (CCSD(T))
NH$_2$$^+$(^{3}B$_2$) + CH$_3$COOH(^{1}A′)	0.00	0.00	NH$_2$$^+$(^{1}A$_1$) + CH$_3$COOH(^{1}A′)	0.00	0.00
NH$_2$(^{2}B$_1$) + CH$_3$COOH$^+$(^{2}A′)	−2.88	−11.39	NH$_2$(^{2}B$_1$) + CH$_3$COOH$^+$(^{2}A′)	−39.17	−41.46
CH$_3$COONH$_2$$^+$(^{2}A″) + H(^{2}S)	−2.97	2.19	CH$_3$COONH$_2$$^+$(^{2}A″) + H(^{2}S)	−39.26	−27.88
NH$_2$CH$_2$COOH$^+$(^{2}A″) + H(^{2}S)	−31.22	−26.93	NH$_2$CH$_2$COOH$^+$(^{2}A″) + H(^{2}S)	−67.51	−57.00
NH(^{3}Σ) + CH$_3$C(OH)$_2$$^+$(^{1}A′)	−36.92	−39.15	NH(^{1}Δ) + CH$_3$C(OH)$_2$$^+$(^{1}A′)	−19.08	−25.51
NH$_3$(^{1}A$_1$) + CH$_2$COOH$^+$(^{1}A)	−62.78	−58.10	NH$_3$(^{1}A$_1$) + CH$_2$COOH$^+$(^{1}A)	−99.07	−88.17

[a] ZPV energy corrections have been included.

as a competitive process. Both reactants are present in the interstellar medium.[29,30] It is well-known that to play a role in the interstellar chemistry reactions in addition to being exothermic must be barrier free. However, a full exploration of the potential energy surface is required to ascertain whether energetic barriers are present. On the other hand, ion−molecule reactions of ionized amine fragment compounds with neutral carboxylic acid and ester species have been experimentally investigated in a selected ion flow tube (SIFT) at 298 K.[31] These studies suggest that NH$_2$$^+$ reacts with acetic acid via proton transfer giving CH$_3$CO(OH)$_2$$^+$ + NH as products (the formation of these products are not been considered in our previous theoretical study). In order to explain this experimental result in this work we have carried out a detailed theoretical study of the reaction of NH$_2$$^+$ with acetic acid considering both the triplet and singlet potential energy surfaces. We will try to determine if under interstellar conditions the formation of glycine ionized is viable from this reaction and to provide a reasonable interpretation of the mass spectrometry experiments.[31] It should be noted that theoretical studies in this field are very appropriate because the conditions present in molecular clouds (low pressure and temperature) are almost ideal for the application of theoretical methods.

Computational Methods

The geometries of the different species involved in the reaction of NH$_2$$^+$ with acetic acid have been obtained at the second-order Moller−Plesset level including all the electrons in the calculations. In these optimizations the cc-pVTZ (correlation-consistent polarized valence triple-ζ) basis set developed by Dunning[32,33] has been employed. Each optimized structure was verified to be a stationary point or a transition state by vibrational analysis carried out at the same level (MP2(full)/cc-pVTZ).

In order to compute accurate relative energies CCSD(T) calculations (coupled-cluster single and double excitation model augmented with a noniterative triple excitation correction)[34] have been carried out on the MP2(full)/cc-pVTZ geometries. The aug-cc-pVTZ basis set, which also includes diffuse functions, has been employed for the energetic calculations. Zero-point vibrational (ZPV) energy corrections were included at the MP2(full)/cc-pVTZ level.

All calculations reported in this work were carried out with the *Gaussian-98* program package.[35]

Results and Discussion

In the present study the possible intermediate species as well as the relevant transition states of the reaction of NH$_2$$^+$ with acetic acid have been characterized. Given the multiplicity of the reactants (NH$_2$$^+$ (^{3}B$_1$) and CH$_3$COOH (^{1}A′)), this reaction should take place in principle on the triplet potential energy surface. In addition to the detailed study of the triplet surface, we have also analyzed the singlet potential energy surface to check a possible implication of this surface through an intersystem-crossing process.

1. (NO$_2$C$_2$H$_6$)$^+$ Isomers. In this section we will briefly describe the stationary points found on the (NO$_2$C$_2$H$_6$)$^+$ triplet and singlet surfaces. Only the minima that are relevant in the reaction of NH$_2$$^+$ with acetic acid will be reported. The optimized geometries of the singlet isomers are collected in Figure 1 and the corresponding to the triplet ones in Figure 2, whereas the relative energies of both singlet and triplet isomers are shown in Table 1. All the isomers considered in Figures 1 and 2, are tested to be true minima on their respective potential surfaces. Similar structures are implicated in both surfaces with the exception of I6 and I7, which correspond to the direct interaction of the nitrogen atom in the NH$_2$$^+$ ion with a hydrogen or the carbon atom, respectively, of the CH$_3$ group of acetic acid. Both species are true

TS$_{I1-I2}$ (^{3}A″)

TS$_{I1-I3}$ (^{3}A″)

TS$_{I1-B}$ (^{3}A″)

TS$_{I3-D}$ (^{3}A)
dih(C$_1$C$_2$O$_2$O$_1$)=-167.4
dih(H$_1$O$_1$C$_2$C$_1$)=171.6
dih(H$_2$C$_1$C$_2$O$_2$)=161.3
dih(H$_3$C$_1$C$_2$O$_2$)=25.5
dih(NC$_1$C$_2$O$_2$)=-87.6
dih(H$_4$NC$_1$C$_2$)=-177.9
dih(H$_5$NC$_1$C$_2$)=61.5
dih(H$_6$NC$_1$C$_2$)=-58.1

TS$_{I3-A}$ (^{3}A)
dih(O$_1$C$_2$O$_2$N)=-8.7
dih(H$_1$O$_1$C$_2$C$_1$)=-178.8
dih(H$_2$C$_1$C$_2$O$_2$)=-68.5
dih(H$_3$C$_1$C$_2$O$_2$)=47.5
dih(NC$_1$C$_2$O$_2$)=172.0
dih(H$_4$NC$_1$C$_2$)=-169.8
dih(H$_5$NC$_1$C$_2$)=11.4
dih(H$_6$NC$_1$C$_2$)=118.7

TS$_{I1-I4}$ (^{3}A)
dih(O$_2$C$_2$C$_1$H$_4$)=20.6
dih(O$_2$C$_2$C$_1$H$_2$)=-86.6
dih(O$_2$C$_2$C$_1$H$_3$)=133.5
dih(H$_4$C$_1$C$_2$O$_1$)=-111.1
dih(C$_1$C$_2$O$_1$H$_1$)=-170.5
dih(C$_1$C$_2$O$_2$N)=-30.1
dih(H$_5$NO$_2$C$_2$)=85.8
dih(H$_4$NO$_2$C$_2$)=141.6

TS$_{I1c-I5}$ (^{3}A)
dih(O$_1$C$_2$C$_1$H$_2$)=63.2
dih(H$_1$O$_1$C$_2$C$_1$)=174.9
dih(H$_2$C$_1$C$_2$O$_2$)=177.4
dih(H$_3$C$_1$C$_2$O$_2$)=-63.8
dih(H$_4$C$_1$C$_2$O$_2$)=56.7
dih(NO$_2$C$_2$C$_1$)=176.3
dih(H$_5$NO$_2$C$_2$)=-89.1
dih(H$_4$NO$_2$C$_2$)=157.1

TS$_{I5-B}$ (^{3}A″)

TS$_{I1-I1c}$ (^{3}A)
dih(O$_1$C$_2$C$_1$H$_2$)=18.8
dih(H$_1$O$_1$C$_2$C$_1$)=174.1
dih(H$_2$C$_1$C$_2$O$_2$)=-164.9
dih(H$_3$C$_1$C$_2$O$_2$)=-40.2
dih(H$_4$C$_1$C$_2$O$_2$)=76.2
dih(NO$_2$C$_2$C$_1$)=96.0
dih(H$_5$NO$_2$C$_2$)=-168.2
dih(H$_6$NO$_2$C$_2$)=-0.345

Figure 3. MP2(full)/cc-pVTZ optimized geometries for the relevant transition states involved in the reaction NH$_2^+$(^{3}B$_1$) with CH$_3$COOH(^{1}A′). Distances are given in angstroms and angles in degrees.
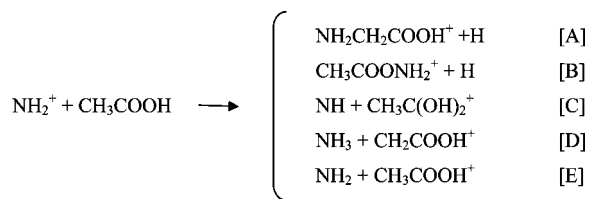
minima on the triplet surface and are not located on the singlet one. On the other hand, isomer I8, corresponding to

a van der Waals complex between NH$_3$ and CH$_2$OCOH$^+$, is only located on the singlet potential surface. In both surfaces the direct interaction of NH$_2^+$ ion through the nitrogen atom with the oxygen atom of the carboxylic group of acetic acid gives isomers I1 and I1-cis, corresponding to two relative positions of the NH$_2$ and OH groups in the molecule. The other isomers which are involved in the reaction are obtained from these two by different isomerization processes which will be explained in the next sections.

In general we can see from Figures 1 and 2 that the N—O bond distances are clearly longer for triplet than for singlet isomers. The C—C bond distance is in general similar to typical carbon—carbon single bond, with the exception of isomer I4 (^{1}A′) where the C—C bond distance has a certain double character. Finally, the C—N distances are found to be similar in triplet and singlet states.

Concerning the relative stabilities of the different intermediates collected in Table 1, we can observe that the energy values calculated at both levels of theory, MP2=full and CCSD(T), give similar stability order between the species considered. Only a change is found on the relative stability of triplet isomers I3 and I1, whose energy difference is only about 1 kcal/mol. The global minimum is predicted to be isomer I3 in its singlet state, ^{1}A′, which corresponds to protonated glycine. With the only exception of isomer I2, singlet states are more stable than the corresponding triplet ones. In a previous study of the (NO$_2$C$_2$H$_6$)$^+$ singlet potential surface,[36] the ion derived from protonation of glycine on the amine group is also found as the lowest-lying. On the singlet potential energy surface the stability order found at both levels of theory is (> means more stable than) as follows: I3 (^{1}A′) > I5 (^{1}A′) > I1-cis (^{1}A′) > I1 (^{1}A′) > I8 (^{1}A) > I4 (^{1}A′) > I2 (^{1}A′). Above the protonated glycine we found isomers I5, I1, and I1-cis, which are quite close in energy (about 45—53 kcal/mol above the global minimum at the CCSD(T) level) and have one nitrogen—oxygen bond. The van der Waals complex is located above isomer I4 that also has a N—O bond. The less stable isomer is I2, which could be seen as a result of the interaction of NH$_2^+$ ion through one hydrogen atom with the oxygen atom of the carboxylic group. If we observe the relative stabilities of the triplet isomers, we can see that the main difference is found for isomer I2, which is the most unstable on the singlet surface, whereas on the triplet one it corresponds to the most stable structure. Another significant difference is that structure I5 is found to be the most unstable one on the triplet surface. On the triplet surface protonated glycine, I3, is less stable than the global minimum, and it is located 19.01 kcal/mol above structure I2. Other common isomers have similar relative energies on both surfaces. The stability order on the triplet surface at the CCSD(T) level is the following: I2 (^{3}A″) > I3 (^{3}A) > I1 (^{3}A″) > I1-cis (^{3}A) > I6 (^{3}A″) > I7 (^{3}A″) > I4 (^{3}A″) > I5 (^{3}A).

2. Reaction of NH$_2^+$ with Acetic Acid. We will discuss in this section the different pathways of the reaction of NH$_2^+$ ion with acetic acid. One possible product of this reaction is ionized glycine; however, other possible channels could be competitive. Therefore, we should consider the following possibilities:
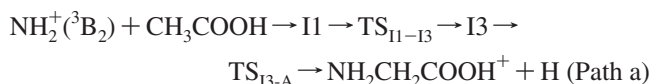
Gas-Phase Reaction of NH₂⁺ with Acetic Acid

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2089**

$$NH_2^+ + CH_3COOH \longrightarrow \begin{cases} NH_2CH_2COOH^+ + H & [A] \\ CH_3COONH_2^+ + H & [B] \\ NH + CH_3C(OH)_2^+ & [C] \\ NH_3 + CH_2COOH^+ & [D] \\ NH_2 + CH_3COOH^+ & [E] \end{cases}$$

The relative energies with respect to reactants of the different possible products for the reaction of NH_2^+ with acetic acid on the triplet and singlet surfaces are given in Table 2. Optimized geometries for reactants and products are given as Supporting Information in Figures S1 and S2. As can be seen in Table 2, with the only exception of channel C, the products obtained on both surfaces are identical. In channel C, on the triplet potential energy surface NH molecule is obtained in their ground state, $^3\Sigma$, whereas on the singlet surface an excited $^1\Delta$ state is reached. From a thermodynamical point of view, we can seen from Table 2 that on the triplet surface channel B, giving amide formation, is slightly endothermic at the high computational level. All other processes are exothermic, and the most favorable one is the formation of $NH_3 + CH_2COOH^+$ through channel D. On the singlet potential energy surface all the channels are clearly exothermic, and the most favorable one is predicted to be channel D, as on the triplet surface. The formation of ionized glycine (channel A) is an exothermic process on both surfaces. Nevertheless, production of ammonia and CH_2COOH^+ ion is clearly the most exothermic channel. On the other hand, in their SIFT experiments Jackson et al.[31] only observe one product, channel C.

In order to see the importance of this reaction as a source of ionized glycine in the interstellar medium and to explain

experimental results a detailed study of the possible energy barriers involved in the different channels should be carried out. Given the multiplicity of reactants the reaction would take place in principle on the triplet surface. However, we will analyze the results found on the triplet potential energy surface, and we will complete the study with a brief comment of the results obtained for the singlet potential energy surface.

2.1. Triplet Potential Energy Surface. The geometries of the relevant transition states are given in Figure 3, whereas the energy profile for the reaction $NH_2^+(^3B_2) + CH_3COOH(^1A')$ is shown in Figure 4. All transition states have one imaginary frequency associated with the corresponding reaction coordinate. Both MP2(full) and CCSD(T) levels give a similar energy profile.

When a $NH_2^+(^3B_2)$ ion approaches the CH_3COOH molecule, the most favorable interaction takes place between the nitrogen and the oxygen atom of the acetic acid giving isomer I1 or I1-cis. Both isomers are close in energy, and isomerization can take place through TS_{I1-I1c}. The exothermic formation of I1 creates an energy reservoir that is used as the reaction proceeds toward the products. Several processes may then occur. An exothermic and barrier free path has been identified leading to the formation of ionized glycine:

$$NH_2^+(^3B_2) + CH_3COOH \rightarrow I1 \rightarrow TS_{I1-I3} \rightarrow I3 \rightarrow$$
$$TS_{I3-A} \rightarrow NH_2CH_2COOH^+ + H \text{ (Path a)}$$

This path starts with isomerization of I1 into I3 upon insertion of the $NH_2(^3B_2)$ group into a C−H bond. The corresponding transition state TS_{I1-I3} lies below the reactants about 22.6 kcal/mol at the CCSD(T) level. Once I3 is obtained, hydrogen atom elimination through transition state
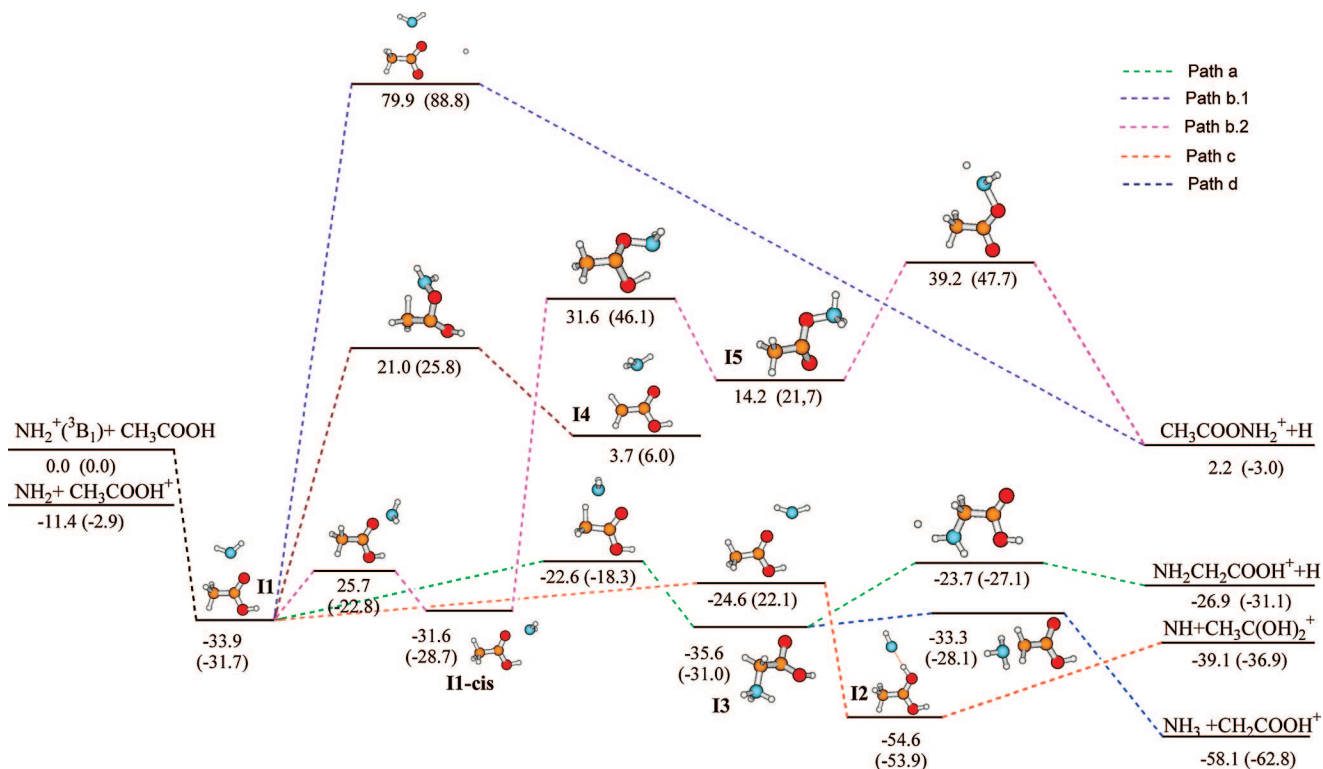
Figure 4. Energy profile, in kcal/mol, for the reaction of $NH_2^+(^3B_1)$ with $CH_3COOH(^1A')$ at the CCSD(T) and MP2(full) (in parentheses) levels (including zero-point vibrational energy differences).

TS$_{I1-I2}$ (^{1}A)

dih(O$_2$C$_2$C$_1$H$_4$)=3.3
dih(O$_2$C$_2$C$_1$H$_2$)=-118.2
dih(O$_2$C$_2$C$_1$H$_3$)=125.4
dih(H$_4$C$_1$C$_2$O$_1$)=-177.2
dih(C$_1$C$_2$O$_1$H$_1$)=-179.2
dih(H$_5$O$_2$C$_2$C$_1$)=4.3
dih(NH$_5$O$_2$C$_2$)=-95.8
dih(H$_6$NH$_5$O$_2$)=-78.9

TS$_{I1-I4}$ (^{1}A)

dih(H$_2$C$_1$C$_2$O$_2$)=-52.1   dih(C$_1$C$_2$O$_2$N)=-13.8
dih(H$_3$C$_1$C$_2$O$_2$)=161.0   dih(H$_5$NO$_2$C$_2$)=-134.6
dih(H$_2$C$_1$C$_2$O$_1$)=136.1   dih(H$_6$NO$_2$C$_2$)=105.9
dih(C$_1$C$_2$O$_1$H$_1$)=159.2   dih(C$_2$O$_2$NH$_4$)=-8.9

TS$_{I4-I3}$ (^{1}A)

dih(H$_2$C$_1$C$_2$O$_2$)=-75.6
dih(H$_3$C$_1$C$_2$O$_2$)=153.9
dih(H$_2$C$_1$C$_2$O$_1$)=103.2
dih(C$_1$C$_2$O$_1$H$_1$)=177.9
dih(NO$_2$C$_2$C$_1$)=-5.5
dih(H$_4$NO$_2$C$_2$)=163.3
dih(H$_5$NO$_2$C$_2$)=-83.4
dih(H$_6$NO$_2$C$_2$)=50.8

TS$_{I3-D}$ (^{1}A')

TS$_{I1-I1c}$ (^{1}A)

dih(O$_1$C$_2$C$_1$H$_2$)=-1.3
dih(H$_1$O$_1$C$_2$C$_1$)=-178.3
dih(H$_2$C$_1$C$_2$O$_2$)=-177.7
dih(H$_3$C$_1$C$_2$O$_2$)=-54.7
dih(H$_4$C$_1$C$_2$O$_2$)=-59.5
dih(NO$_2$C$_2$C$_1$)=95.3
dih(H$_5$NO$_2$C$_2$)=-130.1
dih(H$_6$NO$_2$C$_2$)=119.9

TS$_{I1c-I5}$ (^{1}A)

dih(O$_2$C$_2$C$_1$H$_4$)=10.2
dih(O$_2$C$_2$C$_1$H$_2$)=-110.4
dih(O$_2$C$_2$C$_1$H$_3$)=132.9
dih(H$_4$C$_1$C$_2$O$_1$)=-170.5
dih(C$_1$C$_2$O$_1$H$_1$)=-179.2
dih(NO$_2$C$_2$C$_1$)=179.2
dih(H$_5$NO$_2$C$_2$)=-120.9
dih(H$_6$NO$_2$C$_2$)=121.2

Figure 5. MP2(full)/cc-pVTZ optimized geometries for the relevant transition states involved in the reaction of excited NH$_2^+$(^{1}A$_1$) with CH$_3$COOH(^{1}A'). Distances are given in angstroms and angles in degrees.

TS$_{I3-A}$, which has a relative energy quite similar to TS$_{I1-I3}$, would lead to ionized glycine.

The CH$_3$COONH$_2^+$, which is quasi-isoenergetic with the reactants, can be obtained through two different paths. One of them, path b.1, implies hydrogen atom elimination through transition state TS$_{I1-B}$. The second path, b.2, involves two isomerization processes leading to isomers I1-cis and I5, followed by hydrogen atom elimination:

$$NH_2^+(^3B_2) + CH_3COOH \rightarrow I1 \rightarrow TS_{I1-B} \rightarrow$$
$$CH_3COONH_2^+ + H \text{ (Path b.1)}$$

$$NH_2^+(^3B_2) + CH_3COOH \rightarrow I1 \rightarrow TS_{I1-IIc} \rightarrow I1\text{-cis} \rightarrow$$
$$TS_{I1c-I5} \rightarrow I5 \rightarrow TS_{I5-B} \rightarrow CH_3COONH_2^+ + H \text{ (Path b.2)}$$

The transition states and isomer I5 involved in both paths are clearly located above the reactants. This barrier quite likely would prevent the formation of CH$_3$COONH$_2^+$ in space.

Another exothermic and barrier free process is isomerization of I1 into isomer I2, through transition state TS$_{I1-I2}$, lying 24.6 kcal/mol below the reactants. The intermediate formed I2, produces NH + CH$_3$C(OH)$_2^+$, through direct NH elimination and does not seem to involve any transition state. We carried out a scan for this process, performing optimizations at different N−H distances and found no sign of transition state:

$$NH_2^+(^3B_2) + CH_3COOH \rightarrow I1 \rightarrow TS_{I1-I2} \rightarrow I2 \rightarrow$$
$$NH + CH_2C(OH)_2^+ \text{ (Path c)}$$
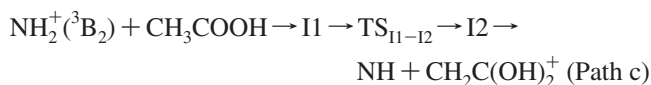
The formation of NH$_3$ + CH$_2$COOH$^+$ involves the same three phases as path a giving isomer I3. Once isomer I3 is obtained, products D are obtained through NH$_3$ abstraction. This process involves transition state TS$_{I3-D}$. The global process is exothermic and barrier free:

$$NH_2^+(^3B_2) + CH_3COOH \rightarrow I1 \rightarrow TS_{I1-I3} \rightarrow I3 \rightarrow TS_{I3-D} \rightarrow$$
$$NH_3 + CH_2COOH^+ \text{ (Path d)}$$

Finally, isomer I1 could give isomer I4 through hydrogen atom migration involving transition state TS$_{I1-I4}$. This process is not relevant in the context of interstellar chemistry, because it has a significant activation barrier. The transition state is clearly located above the reactants, and formation of isomer I4 is an endothermic process. Therefore, evolution of isomer I4 has not been considered.

Theoretical calculations on the triplet potential energy surface suggest that paths a, c, and d are feasible processes under interstellar conditions that preclude reaction that are endothermic or proceed through significant activation barriers. The most exothermic channel is d, production of NH$_3$ + CH$_2$COOH$^+$. However, the most favorable channel from kinetic arguments should be proton transfer process, which has lower activation barriers (notice that TS$_{I1-I2}$ lies lower in energy than TS$_{I1-I3}$). The less favorable one seem to be formation of ionized glycine. It should be remarked that our results agree with the results from SIFT experiments,[31] since production of NH + CH$_3$C(OH)$_2^+$ is observed.

2.1. Singlet Potential Energy Surface. The geometries of the transition states located on the singlet surface are collected in Figure 5, and the energy profile for the reaction NH$_2^+$(^{1}A$_1$) + CH$_3$COOH(^{1}A') is shown in Figure 6. As in the case of the triplet potential energy surface, all transition states have been checked to have one imaginary frequency associated with the corresponding reaction coordinate.
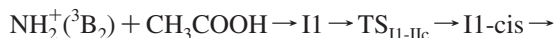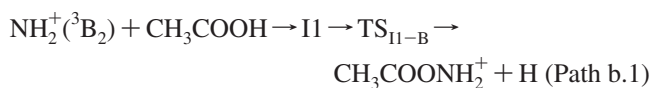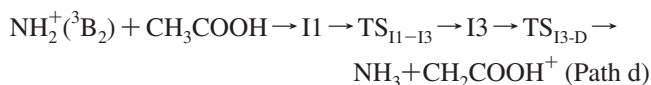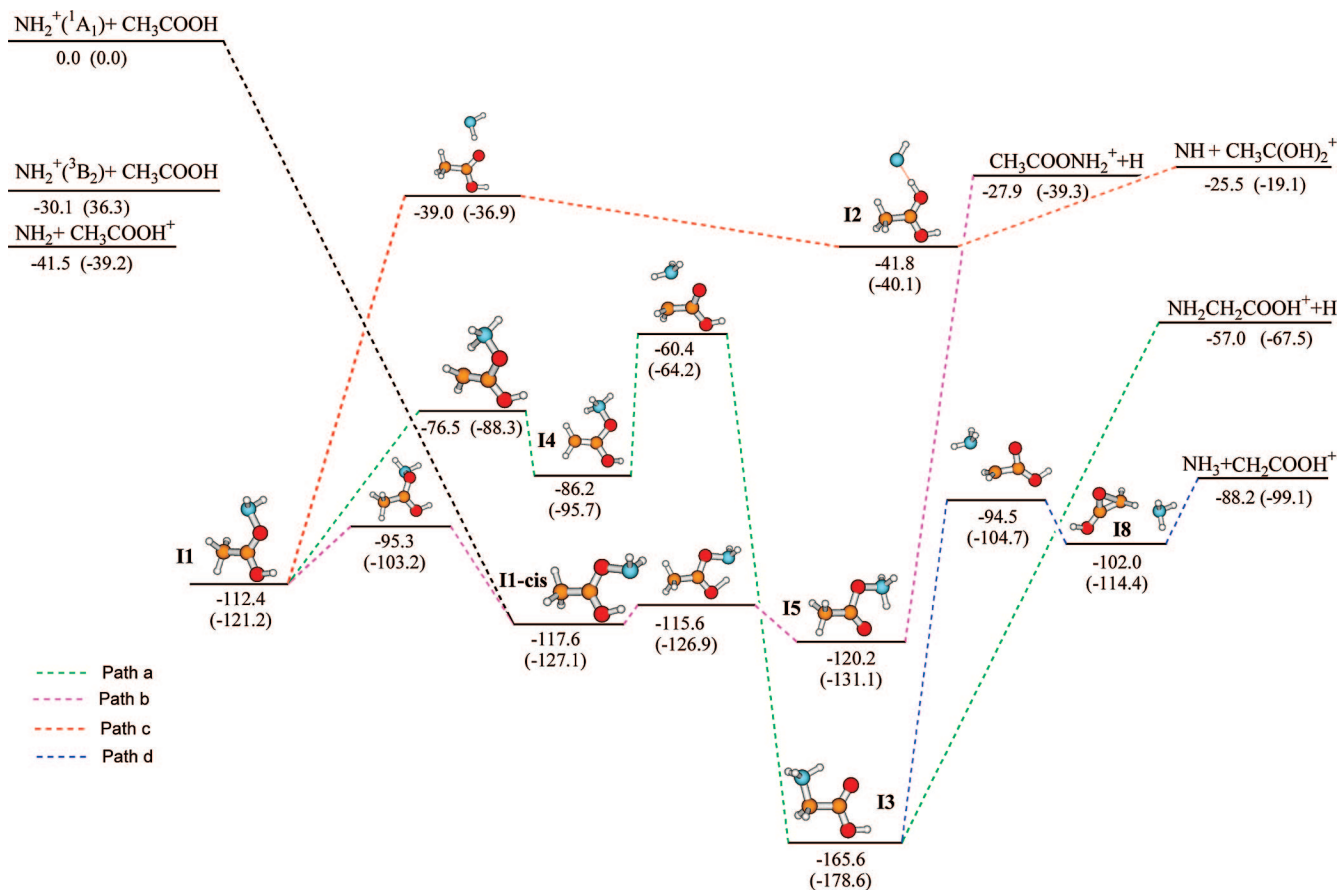
The most favorable interaction of NH$_2^+$(^{1}A$_1$) ion with acetic acid occurs between nitrogen and oxygen atoms giving I1 and I1-cis as in the case of the triplet surface. However, in this case the most stable process is the

Gas-Phase Reaction of NH$_2$$^+$ with Acetic Acid

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2091**

Figure 6. Energy profile, in kcal/mol, for the reaction of excited NH$_2$$^+$(^{1}A$_1$) with CH$_3$COOH(^{1}A′) at the CCSD(T) and MP2(full) (in parentheses) levels (including zero-point vibrational energy differences).

formation of isomer I1-cis. Both isomers can be reached and isomerization implies transition state TS$_{I1-I1c}$. From isomer I1, the process giving ionized glycine on the singlet potential energy surface has some differences with respect to that on the triplet one and proceeds through the following steps:

$$NH_2^+(^1A_1) + CH_3COOH \rightarrow I1 \rightarrow TS_{I1-I4} \rightarrow I4 \rightarrow$$
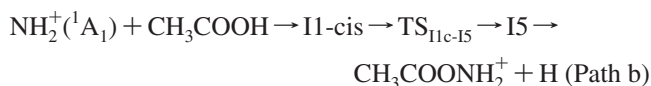$$TS_{I4-I3} \rightarrow I3 \rightarrow NH_2CH_2COOH^+ + H \text{ (Path a)}$$
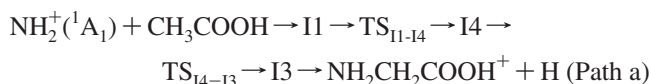
Isomerization of I1 into I4 implies a proton transfer from the carbon atom to the nitrogen one through transition state TS$_{I1-I4}$. The migration of the NH$_3$ group from oxygen to carbon gives protonated glycine, I3, and this process implies transition state TS$_{I4-I3}$. The direct elimination of one of the hydrogens bonded to nitrogen gives ionized glycine. The global process giving ionized glycine is exothermic and barrier free (both transition states are located below the reactants, about −76.5 and −60.4 kcal/mol, respectively). Protonated glycine is a very stable intermediate and could be a long-lived species given its energy difference with the reaction products.
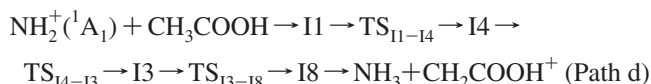
Path b, leading to CH$_3$COONH$_2$$^+$, shows similar steps to those found on the triplet surface for path b.2. In this case isomer I1-cis is formed directly from the reactants, and the only difference is the last step. On the singlet surface, elimination of hydrogen atom is a direct process, which does not seem to involve any transition state. On

the singlet surface the process is exothermic and barrier free. This process can be summarized as

$$NH_2^+(^1A_1) + CH_3COOH \rightarrow I1\text{-cis} \rightarrow TS_{I1c-I5} \rightarrow I5 \rightarrow$$
$$CH_3COONH_2^+ + H \text{ (Path b)}$$

The path for the formation of NH + CH$_3$C(OH)$_2$$^+$, path c, has the same steps as that on the triplet surface, but in this case isomer I2 is relatively less stable and the NH is formed in an excited state ($^1\Delta$).

Finally, another channel is originated from protonated glycine, I3, by fragmentation into NH$_3$ + CH$_2$COOH$^+$:

$$NH_2^+(^1A_1) + CH_3COOH \rightarrow I1 \rightarrow TS_{I1-I4} \rightarrow I4 \rightarrow$$
$$TS_{I4-I3} \rightarrow I3 \rightarrow TS_{I3-I8} \rightarrow I8 \rightarrow NH_3 + CH_2COOH^+ \text{ (Path d)}$$

When NH$_3$ is eliminated from I3 through TS$_{I3-I8}$, a van der Waals complex is formed before dissociation. It can be seen in Figure 6 that once I3 is formed path d seems to be more favorable than path a, from a thermodynamic point of view.

All the processes considered on the singlet potential energy surface are exothermic and barrier-free and therefore feasible under the conditions in the interstellar space. The last process, formation of NH$_3$ + CH$_2$COOH$^+$, seems to be in principle the most favorable one. Since in the experiments[31] only NH + CH$_3$C(OH)$_2$$^+$ are observed as products, it seems that the

reaction takes place only on the triplet surface, and the singlet one is not reached.

Conclusions

A theoretical study of the reaction of NH_2^+ ion with acetic acid has been carried out on the triplet and singlet potential energy surfaces. We have characterized the possible intermediate species as well as the relevant transition states for this reaction. The lowest-lying species of the $(NO_2C_2H_6)^+$ system is found to be protonated glycine in its singlet state, $^1A'$. With the only exception of isomer I2, singlet states are more stable than the corresponding triplet ones. For the $(NO_2C_2H_6)^+$ system the stability order found is the following: I3 ($^1A'$) > I5 ($^1A'$) > I1-cis ($^1A'$) > I1 ($^1A'$) > I8 (^1A) > I4 ($^1A'$) > I2 ($^3A''$) > I3 (^3A) > I1 ($^3A''$) > I1-cis (^3A) > I6 ($^3A''$) > I2 ($^1A'$) > I7 (^3A)'' > I4 ($^3A''$) > I5 (^3A).

Theoretical calculations on the triplet potential energy surface suggest that paths, a, c and d, giving ionized glycine, proton transfer products, and the CH_2COOH^+ compound, respectively, are feasible processes under interstellar conditions. The most favorable channel, from kinetic arguments, should be the proton transfer process, which has lower activation barriers. The less favorable one seems to be formation of ionized glycine. On the other hand, all the processes considered on the singlet potential energy surface are exothermic and barrier-free and therefore feasible under the conditions in the interstellar space. The formation of NH_3 + CH_2COOH^+ seems to be in principle the most favorable one. Experimental results observed from SIFT experiments[31] show that the products correspond to the proton transfer process ($NH + CH_3C(OH)_2^+$), in agreement with the theoretical results found on the triplet potential energy surface. Therefore the present theoretical study provides a reasonable interpretation of the experiments.

Our results suggest that even if the formation of ionized glycine from the reaction of NH_2^+ ion with acetic is a favorable process under interstellar conditions, there are other competitive channels that are most favorable.

Supporting Information Available: Optimized geometries of the reactants and products (Figures S1 and S2). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Miller, S. L. *Science* **1953**, *117*, 528.

(2) De Duve, C. *Origins Life Evol. Biospheres* **2003**, *33*, 559.

(3) Winnewisser, G.; Herbst, E. *Rep. Prog. Phys.* **1993**, *56*, 1209.

(4) Tielens, A. G. G. M.; Charnley, S. B. *Origins Life Evol. Biospheres* **1997**, *27*, 23.

(5) Combes, F.; Rieu, N. Q.; Wlodarczak, G. *Astron. Astrophys.* **1996**, *308*, 618.

(6) Bernstein, M. P.; Dworkin, J. P.; Sandford, S. A.; Cooper, G. W.; Allamandola, L. J. *Nature* **2002**, *416*, 401.

(7) Chyba, C.; Sagan, D. *Nature* **1992**, *335*, 125.

(8) Cooper, G. W.; Onwo, W. M.; Cronin, J. R. *Geochim. Cosmochim. Acta* **1992**, *57*, 4109.

(9) Hollis, J. M.; Snyder, L. E.; Suenram, R. D.; Lovas, F. J. *Astrophys. J.* **1980**, *241*, 1001.

(10) Snyder, L. E.; Hollis, J. M.; Suenram, R. D.; Lovas, F. J.; Brown, L. W.; Buhl, D. *Astrophys. J.* **1983**, *268*, 123.

(11) Ceccarelli, C.; Loinard, L.; Castets, A.; Faure, A.; Lefloch, B. *Astron. Astrophys.* **2000**, *362*, 1122.

(12) Hollis, J. M.; Pedelty, J. A.; Boboltz, D. A.; Liu, S.-Y.; Snyder, L. E.; Palmer, P.; Lovas, F. J.; Jewell, P. R. *Astrophys. J. Lett.* **2003**, *596*, L235.

(13) Hollis, J. M.; Pedelty, J. A.; Snyder, L. E.; Jewell, P. R.; Lovas, F. J.; Palmer, P.; Liu, S.-Y *Astrophys. J.* **2003**, *588*, 353.

(14) Kuan, J.-Y.; Charnley, S. B.; Huang, H.-C.; Tseng, W.-L.; Kisiel, Z. *Astrophys. J.* **2003**, *593*, 848.

(15) Snyder, L. E.; Lovas, F. J.; Hollis, J. M.; Friedel, D. N.; Jewell, P. R.; Remijan, A.; Ilyushin, V. V.; Alekseev, E. A.; Dyubko, S. F. *Astrophys. J.* **2005**, *619*, 914.

(16) Irvine, W. *Origins Life Evol. Biosphere* **1998**, *28*, 365.

(17) Ehrenfreund, P.; Bernstein, M.; Dworkin, J. P.; Sandford, S. A.; Allamandola, L. *Astrophys. J.* **2001**, *550*, L95.

(18) Belloche, A.; Menten, K. M.; Comito, C.; Müller, H. S. P.; Schilke, P.; Ott, J.; Thorwirth, S.; Hieret, C. *Astron. Astrophys.* **2008**, *482*, 179.

(19) Arnaud, R.; Adamo, C.; Cossi, M.; Milet, A.; Vallee, Y.; Barone, V. *J. Am. Chem. Soc.* **2000**, *122*, 324.

(20) Basiuk, V. A. *J. Phys. Chem. A* **2001**, *105*, 4252.

(21) Koch, D. M.; Toubin, C.; Peslherbe, G. H.; Hynes, J. T. *J. Phys. Chem. C* **2008**, *112*, 2972.

(22) Maeda, S.; Ohno, K. *Chem. Phys. Lett.* **2004**, *398*, 240.

(23) Maeda, S.; Ohno, K. *Chem. Lett.* **2004**, *33*, 1372.

(24) Maeda, S.; Ohno, K. *Astrophys. J.* **2006**, *640*, 823.

(25) Herbst, E. *Chem. Soc. Rev.* **2001**, *30*, 168.

(26) Blagojevic, V.; Petrie, S.; Bohme, D. K. *Mon. Not. R. Astron. Soc.* **2003**, *339*, L7.

(27) Snow, J. L.; Orlova, G.; Blagojevic, V.; Bohme, D. K. *J. Am. Chem. Soc.* **2007**, *129*, 9910.

(28) Largo, A.; Redondo, P.; Barrientos, C. *Int. J. Quantum Chem.* **2004**, *98*, 355.

(29) van Dishoeck, E. F.; Jansen, D. J.; Schilke, P.; Phillips, T. G. *Astrophys. J.* **1993**, *416*, L83.

(30) Mehringer, D. M.; Snyder, L. E.; Miao, Y.; Lovas, F. *Astrophys. J.* **1997**, *480*, L71.

(31) Jackson, D. M.; Stibrich, N. J.; McLain, J. L.; Fondren, L. D.; Adams, N. G.; Babcock, L. M. *Int. Mass. Spectrom.* **2005**, *247*, 55.

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

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

(34) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

(35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millan, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.;

Gas-Phase Reaction of NH$_2^+$ with Acetic Acid

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2093**

Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelly, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challa-combe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Gaussian Inc.: Pittsburgh, PA, 1998.

(36) Lattelais, M.; Ellinger, Y.; Zanda, B. *Int. J. Astrobiol.* **2007**, *6*, 37.

JCTC Journal of Chemical Theory and Computation

Prediction of Vibronic Coupling and Absorption Spectra of Dimers from Time-Dependent Density Functional Theory: The Case of a Stacked Streptocyanine

Julien Guthmuller,[*,†] Freddy Zutterman,[‡] and Benoît Champagne[†]

Laboratoire de Chimie Théorique Appliquée, Facultés Universitaires Notre-Dame de la Paix, Rue de Bruxelles 61, B-5000 Namur, Belgium, and Agfa-Gevaert N.V., Septestraat 27, B-2640 Mortsel, Belgium

Received July 29, 2008

Abstract: Methods based on density functional theory calculations have been used to simulate the absorption spectra of a streptocyanine and of its covalently bonded dimer. Two approaches, based on multimode Franck−Condon overlap integrals, have been employed. In the first approach the monomer and the dimer are treated as single molecules, and the Franck−Condon factors are determined for both systems. The second approach is based on the diagonalization of the dimer Hamiltonian which is constructed from the monomer Franck−Condon overlap integrals and quantities describing the intermonomer electronic coupling. Both approaches succeed in reproducing the hypsochromic shift of the maximum of absorption occurring upon dimerization with an accuracy of 0.05 eV. The vibronic structure of the monomer is also in good agreement with experiment and depends little on the inclusion of Duschinsky rotation effects. The shape and relative intensity of the dimer spectrum is qualitatively reproduced by the two methods, each of them being able to describe most of the vibronic features. Moreover, accounting for the solvent effects in the calculation of the intermonomer electronic coupling improves the agreement with experiment by reducing the intensity of the maximum and by enlarging the spectrum at longer wavelengths.

1. Introduction

The description of the vibronic coupling in molecular aggregates and of the corresponding absorption spectrum have been the subject of many studies in the recent literature. For example, Eisfeld et al.[1] have employed the so-called Coherent Exciton Scattering (CES) approximation[2,3] to rationalize the J and H bands of organic dye aggregates. Eisfeld[4] has also recently proposed a simple method based on dipole−dipole coupled dimers to connect the absorption properties of the dimer to its conformation. The vibronic coupling in dimers has also been simulated from the diagonalization of the dimer Hamiltonian in a basis of monomer eigenstates.[5−8] This approach allows the investigation of the weak and strong electronic coupling limits and

enables the introduction of different approximations by eliminating some types of eigenstates from the basis set.[6] However, in the above-mentioned approaches[5−8] the absorption spectrum is usually described in terms of a single effective vibrational normal mode coordinate, while the electronic coupling parameters - representing the intermonomer interaction - are adjusted to reproduce the experimental spectrum. Moreover, a multimode vibronic treatment of coupled molecular dimers has been presented by Myers Kelley[9] and has been applied to the absorption, resonance Raman scattering, and hyper-Rayleigh scattering of molecular dimers.[9,10] In this approach the multimode vibronic properties are deduced from the analysis of the experimental resonance Raman intensities, and the strength of the intermonomer coupling is treated as an empirical parameter chosen to reproduce the experimental absorption spectrum. Finally, quantum chemical calculations describing the dimer as a supermolecule have been reported.[11,12] These methods allow

* Corresponding author e-mail: julien.guthmuller@fundp.ac.be.
† Facultés Universitaires Notre-Dame de la Paix.
‡ Agfa-Gevaert N.V.

DFT Vibronic Coupling: Stacked Streptocyanine

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2095**

Figure 1. Structure of the streptocyanine monomer and its covalently bonded dimer.

a direct evaluation of the intermonomer interaction, but their application has so far been limited to strongly coupled dimers in which the absorption spectrum can be represented as a single allowed electronic transition.

The aim of the present study is to investigate the ability of first principles theoretical schemes to predict the vibronic coupling and absorption spectrum of dimers. Two approaches based on time-dependent density functional theory (TDDFT) calculations are considered. In the first approach, similarly to our previous studies,[13−15] the absorption spectra of the monomer and of the dimer are simulated using DFT/TDDFT calculations treating the monomer and the dimer as simple molecules. In this approach, the geometrical relaxation and harmonic frequencies in the excited-state are computed for the state with the largest oscillator strength. Then, the absorption spectrum is constructed from the Franck−Condon (FC) factors, which are determined using a recursive scheme. In a second approach, the multimode vibronic treatment of coupled molecular dimers[9] is employed. In this method, the absorption spectrum of the dimer is obtained from the vibronic structure of the monomer and from intermonomer electronic coupling parameters. Furthermore, the amplitude of the coupling is evaluated from TDDFT calculations, which are performed at the dimer geometry. The two approaches are applied to the absorption properties of a streptocyanine dye and of its covalently bonded stacked dimer (Figure 1). These systems have been previously studied experimentally in methanol solution and theoretically at the INDO/S−CI level by Katoh et al.[16] The covalently bonded stacked dimer is particularly suitable because its geometry is well defined and does not require assuming the relative position of its chromophoric moieties which can be the case for many aggregates in solution.

2. Theoretical Method

2.1. Vibronic Coupling and Franck−Condon Factors for a Single Molecule. Quantum chemical calculations were carried out using the TURBOMOLE package.[17] The geometries and harmonic frequencies of vibration in the ground-state were calculated at the density functional theory (DFT) level by employing the B3LYP exchange-correlation (XC) functional and the TZVP basis set. The vertical transition energies $E_{vert} = E_e - E_g$ were obtained from TDDFT with the same XC functional and basis set. In order to investigate the absorption properties, the geometry of the excited-state with the largest oscillator strength was optimized. The simulation of the absorption spectrum requires the determination

of multidimensional FC overlap integrals $\langle \theta_{g0}|\theta_{ev}\rangle$. In this notation $|\theta_{g0}\rangle$ is the vibrational wave function of the electronic ground state (g) in its vibrational ground state (0), and $|\theta_{ev}\rangle$ is the vibrational wave function of the electronic excited state (e), where ν is a multi-index representing the harmonic quantum numbers of the 3N-6 vibrational normal modes. The FC overlap integrals were calculated with a homemade program (MultiFC), which makes use of the recursive relations given by P. T. Ruhoff.[18] The FC integrals were first calculated within the independent mode displaced harmonic oscillator model (IMDHOM), which only takes into account the geometrical differences between the ground and excited states. The absorption spectrum at 0 K is then obtained from eq 1:

$$A(\omega) \propto \omega \sum_{\nu} \frac{\Gamma}{\pi} \frac{\mu_{ge}^2 \langle \theta_{g0}|\theta_{ev}\rangle^2}{(\omega - E_{ev})^2 + \Gamma^2} \qquad (1)$$

The summation is taken over all vibronic levels in the excited-state and the FC amplitudes are broadened by a Lorentzian function describing the homogeneous broadening with a full width at half-maximum (fwhm) equal to 2Γ. μ_{ge} is the dipole transition moment, and E_{ev} is the excitation energy accounting for both the electronic and vibrational levels. Typically, several thousands of FC overlap integrals are considered in order to get converged spectra. This method will be referred to in the following as the single molecule approach. Within this single molecule approach, the impact of the Duschinsky[19] rotations was investigated for the monomer. In that case, the determination of the FC overlap integrals requires the evaluation of the excited-state normal coordinates of vibration.

2.2. Multimode Vibronic Theory of Coupled Monomers. The absorption spectrum of the dimer was also simulated using a second approach based on the multimode vibronic theory of coupled identical monomers.[9] This approach allows the calculation of the dimer absorption spectrum only from the knowledge of the monomer properties and from the electronic coupling between both monomers. In addition to the ground state (g), only the single excited electronic state (e) - with the largest oscillator strength - is taken into account, and the two monomers (A and B) are coupled by an intermonomer electrostatic interaction. In the present treatment the dimer is initially in its ground state $|(g0)_A(g0)_B\rangle$, and the final excited states are sought as linear combinations of the basis vectors $(|(g0)_A(e\nu)_B\rangle, |(e\nu)_A(g0)_B\rangle)$, in which ν is again a multi-index representing the vibrational quantum numbers. For example, in this notation $(g0)_A$ denotes the ground electronic and vibrational state of monomer A, while $(e\nu)_B$ is a vibronic state of monomer B corresponding to the vibrational excitation ν of the electronic excited state (e). Then, the dimer vectors $|(g0)_A(e\nu)_B\rangle$ are constructed as products of the monomer states $|(g0)_A\rangle$ and $|(e\nu)_B\rangle$. It is assumed in this basis that the vibrational excitation always strictly accompanies the electronic excitation. This approach has been referred to by Andrzejak and Petelenz[6] as the modified strong coupling (MSC) approach and will be referred to as so in the following. It is convenient to introduce the symmetric and antisymmetric dimer states

$$|g0, ev\rangle_{\pm} = \frac{1}{\sqrt{2}}(|(g0)_A(ev)_B\rangle \pm |(ev)_A(g0)_B\rangle) \quad (2)$$

In this basis the dimer Hamiltonian $H = H_0 + V$ is block diagonal and is given by

$$_{\pm}\langle g0, ev|H|g0, ev'\rangle_{\pm} = E_{ev}\delta_{vv'} +$$
$$\Delta\delta_{vv'} \pm E_{exc}\langle\theta_{g0}|\theta_{ev'}\rangle\langle\theta_{ev}|\theta_{g0}\rangle \quad (3)$$

where E_{ev} are the monomer excited-state energies, E_{exc} is the exciton shift responsible for the splitting of the zero-order excited states, Δ is a dimerization shift[4] describing both the modification of the initial and final state energies due to the intermonomer interaction, and $\langle\theta_{g0}|\theta_{ev}\rangle$ are the monomer FC overlap integrals evaluated as discussed in section 2.1. The two blocks of eq 3 can be diagonalized separately and lead to the true eigenvalues $E_{s\pm}$ and eigenvectors of the dimer excited states[9]

$$|\psi_{s\pm}\rangle = \sum_v C_{s\pm;0,v}|g0, ev\rangle_{\pm} \quad (4)$$

The calculation of the absorption spectrum of the dimer is then straightforward and can be performed by calculating the matrix elements of the dipole operator $\vec{\mu}^A + \vec{\mu}^B$ between the ground-state and the excited vibronic states. For parallel H-dimers ($\vec{\mu}_{ge}^A \cdot \vec{\mu}_{ge}^B = -\mu_{ge}^2$), these matrix elements are only nonzero for the (−) block. In this case, the absorption spectrum is given by[9]

$$A(\omega) \propto \omega \sum_s \frac{\Gamma}{\pi} \frac{2\mu_{ge}^2(\sum_v C_{s-;0,v}\langle\theta_{g0}|\theta_{ev}\rangle)^2}{(\omega - E_{s-})^2 + \Gamma^2} \quad (5)$$

Adopting this MSC approach requires the knowledge of E_{exc} and Δ. These quantities have often been used as parameters[1−3,5−8] tuned to fit to experiment. In this work they were evaluated from the vertical excitation energies of the monomer (E_{vert}^M) and dimer (E_{vert}^D) following eqs 6 and 7.

$$E_{exc} = \frac{E_{vert}^D(S_2) - E_{vert}^D(S_4)}{2} \quad (6)$$

$$\Delta = \frac{E_{vert}^D(S_4) + E_{vert}^D(S_2)}{2} - E_{vert}^M(S_1) \quad (7)$$

The electronic excited states S_2 and S_4 are the two excitonic components of the dimer, while S_1 is the monomer excited-state with the largest oscillator strength (see section 3.1).

2.3. Other Theoretical and Computational Aspects. The effect of the solvent (methanol) on E_{exc} and Δ was determined by using the program GAUSSIAN 03[20] (G03) and the integral equation formalism of the polarizable continuum model[21] (IEFPCM). The ground-state geometries and vertical transition energies of the monomer and dimer were calculated at the DFT/TDDFT/B3LYP/6−311G* level of approximation. The calculations were also performed in vacuum and showed similar results to the one obtained with TURBO-MOLE at the B3LYP/TZVP level (E_{exc} and Δ differ at most by 0.01 eV between the two approaches).

3. Results

3.1. Excited States. Table 1 presents the vertical transition energies and oscillator strengths of the monomer and dimer

Table 1. Vertical Excitation Energies, Oscillator Strengths, and Singly-Excited Configurations for the Lowest Energy Singlet Excited States of the Monomer and the Stacked Dimer[a]

	state	transition	wt (%)	B3LYP/TZVP (vacuum) E_{vert} (eV)	f	B3LYP/6−311G* (methanol) E_{vert} (eV)	f
monomer	S_1	H→L	99	3.77	1.410	3.65	1.487
dimer	S_1	H→L	98	2.41	$<5.10^{-4}$	2.46	$<5.10^{-4}$
	S_2	H→L+1	55	2.91	0.006	2.97	0.011
		H−1→L	45				
	S_3	H−1→L+1	97	3.74	$<5.10^{-4}$	3.77	$<5.10^{-4}$
	S_4	H−1→L	54	4.03	2.827	3.86	3.115
		H→L+1	44				

[a] The results in vacuum are calculated with TURBOMOLE at the TDDFT/B3LYP/TZVP level of approximation, and the effect of the solvent (methanol) is described using G03 with the IEFPCM approach at the TDDFT/B3LYP/6−311G* level of approximation. H and L correspond to the highest occupied and lowest unoccupied molecular orbitals, respectively.
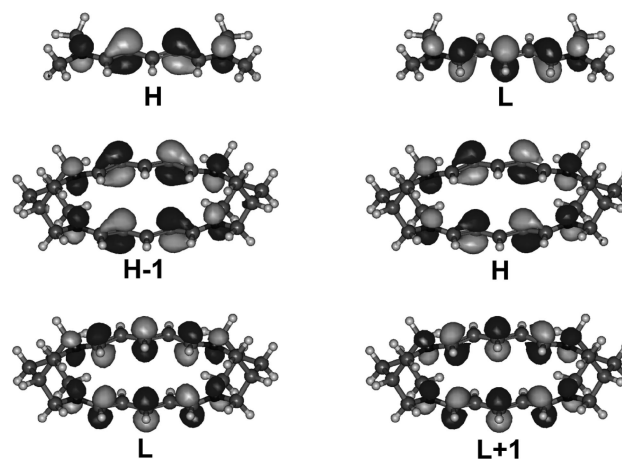
Figure 2. Molecular orbitals involved in the dominant configurations of the monomer and dimer low-energy excited states. H and L correspond to the highest occupied and lowest unoccupied molecular orbitals, respectively.

calculated in vacuum and methanol solution. The orbitals involved in the transitions are depicted in Figure 2 and have similar shapes in vacuum and methanol. The first excited-state of the monomer S_1 is associated with a H→L configuration and presents a large oscillator strength of 1.410. Therefore, this state is responsible for the absorption properties of the molecule in this energy domain. Indeed, the next excited-state of the monomer, S_2, is calculated with a higher energy of 5.44 eV and a much smaller oscillator strength of 0.023. For the dimer, state S_4 possesses the largest oscillator strength (2.827). This value is exactly twice larger than the S_1 oscillator strength of the monomer. Therefore, state S_4 corresponds to the allowed (ψ_{s-}) component of the molecular exciton theory. Going from the monomer to the stacked dimer, the transition energy to this strongly dipole-allowed state shows a hypsochromic shift of 0.26 eV, which is characteristic of H-dimers. State S_2 presents a small oscillator strength of 0.006 and is composed of comparable configura-

DFT Vibronic Coupling: Stacked Streptocyanine

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2097**

tions, H→L+1 and H-1→L, to S_4. Therefore, S_2 can be associated to the forbidden (ψ_{s+}) component of the molecular exciton theory. Two additional states, S_1 and S_3, are also calculated, but they do not contribute to the absorption spectrum owing to their negligible oscillator strengths. Finally, note that the TDDFT results are in agreement with the molecular exciton theory, showing that only one electronic state is responsible for the absorption properties of the stacked H-dimer.

The inclusion of the solvent effects within the IEFPCM approach leads to non-negligible modifications of the vertical transition energies. On the one hand, a bathochromic shift is found for the allowed states S_1 and S_4 of the monomer and dimer, respectively. On the other hand, the S_2 state of the dimer is hypsochromically shifted in methanol solution.

The exciton and dimerization shifts can then be evaluated in both environments from the energies of these states using eqs 6 and 7. In vacuum, E_{exc} and Δ amount to -0.56 eV and -0.30 eV and to -0.45 eV and -0.24 eV in methanol, respectively. Thus, accounting for the effects of the solvent leads to a reduction of the absolute values of both E_{exc} and Δ, which will be reflected in the simulated absorption spectra.

When comparison is performed with the experimental absorption maxima in methanol,[16] it appears that the calculated vertical transition energies of the monomer and dimer are overestimated. Indeed, the theoretical value of 3.77 eV for the monomer is 0.75 eV above the experimental maximum of 3.02 eV. Similarly, the calculated value of 4.03 eV for the dimer is 0.68 eV higher than the experimental value of 3.35 eV. In the case of the monomer, the inclusion of the geometry relaxation in the excited-state and of the vibrational structure leads to a theoretical maximum of 3.55 eV, which is still 0.53 eV above the experimental value. Additional corrections can be included by taking into account the interactions with the solvent. The calculations performed with the IEFPCM model and the program G03 for the vertical transition energies suggest that the solvent effect leads to a bathochromic shift of about 0.15 eV. Therefore, a value of 3.40 eV is our best theoretical estimate for the absorption maximum of the monomer in methanol. The remaining deviation of 0.38 eV with respect to experiment can be attributed to the limitations of the theoretical method. These limitations arise from the approximate XC functional (B3LYP) and from the treatment of the solvent. In the later, specific interactions with solvent molecules and, in the present case, counterions are not taken into account. In the former, one can distinguish between the limitations of both the exchange and correlation parts of the functional. For instance, Peach et al.[22] related the excitation energy underestimations to the charge transfer character of the transition. The situation is nevertheless more complicated here and involves correlation effects since the excitation energy is overestimated. However, in previous studies on structurally related cyanine dyes[13,14] it has been shown that these errors are rather systematic and positive. Consecutively, it has been shown that the overestimation of the excitation energies for compounds belonging to the same family can be corrected by applying a linear scaling procedure, which reduces the deviations within 0.1 eV.
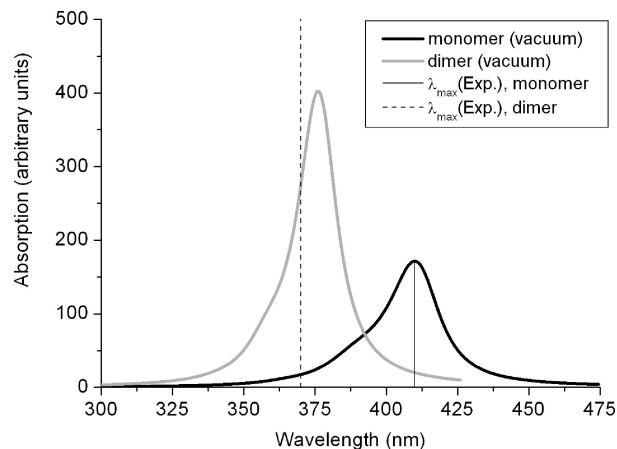
Figure 3. Absorption spectra of the monomer and stacked dimer within the IMDHOM. The results are obtained from a single molecule calculation at the TDDFT/B3LYP/TZVP level of theory in vacuum. The theoretical $\omega_{e0,g0} = E_{e0} - E_{g0}$ origins have been shifted so that the experimental and theoretical absorption maxima of the monomer coincide. The broadening is described by a fwhm set to 1000 cm^{-1}. The experimental maxima in methanol[16] (410 nm and 370 nm) are represented by vertical lines.

3.2. UV/Visible Absorption Spectra. Figure 3 presents the absorption spectra of the monomer and stacked dimer obtained from the single molecule approach within the IMDHOM. In order to correct the spectra for the systematic excitation energy overestimation, they were shifted so that the experimental and theoretical absorption maxima of the monomer coincide. Each vibronic transition was broadened by a Lorentzian function (see eq 1) with a fwhm of 1000 cm^{-1}. The value of the fwhm was chosen in order to reproduce the experimental broadening of the spectra.[16] Due to the rather low resolution of the experimental spectra, only homogeneous broadening was assumed and appears adequate to simulate the observed broadening of the spectra. The spectrum of the monomer shows a vibronic shoulder between 375 nm and 400 nm. The spectrum of the dimer displays a narrower shape with a rather weak vibronic shoulder at shorter wavelengths. The ratio between the dimer and the monomer intensities at the maximum is equal to 2.35. The position of the dimer maximum with respect to the monomer is accurately predicted by the theory. An underestimation of 0.05 eV is found in comparison to experiment. The theoretical spectra also show a general good agreement with the experimental spectra reported by Katoh et al.[16] for both the intensities and the shapes. Indeed, for the monomer and dimer the shoulders at shorter wavelengths are properly reproduced by the calculation. Nevertheless, the experimental shoulder at larger wavelength for the dimer is not present in the calculated spectrum.

Figure 4 compares the monomer spectrum with the absorption spectra of the model dimer obtained from the diagonalization approach in the MSC basis set, where the (monomer) FC overlap integrals are evaluated within the IMDHOM. The spectra in vacuum and methanol were simulated using the exciton and dimerization shifts given in section 3.1. It should be mentioned that the spectrum in methanol was calculated using the monomer overlap integrals
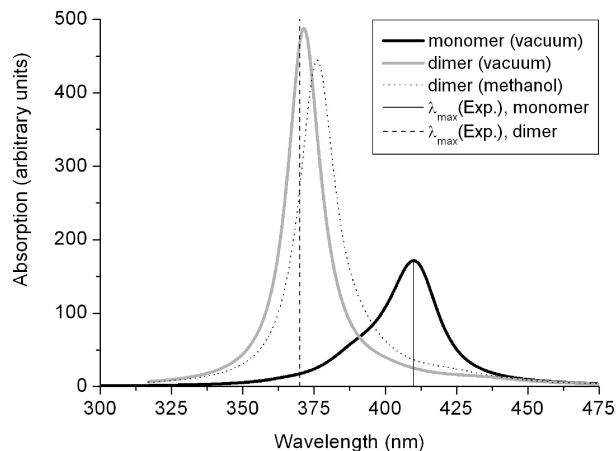
Figure 4. Absorption spectra of the monomer and dimer. The monomer spectrum is obtained within the IMDHOM while the dimer spectra are obtained from eq 5 after diagonalization of eq 3, of which the different quantities are evaluated at the TDDFT/B3LYP level of theory. The theoretical $\omega_{e0,g0} = E_{e0} - E_{g0}$ origins have been shifted so that the experimental and theoretical absorption maxima of the monomer coincide. The broadening is described by a fwhm set to 1000 cm^{-1}. The experimental maxima in methanol[16] (410 nm and 370 nm) are represented by vertical lines.

Figure 5. Absorption spectra of the monomer and dimer. The monomer spectrum is obtained including Duschinsky rotation effects, while the dimer spectra are obtained from eq 5 after diagonalization of eq 3, of which the different quantities are evaluated at the TDDFT/B3LYP level of theory. The theoretical $\omega_{e0,g0} = E_{e0} - E_{g0}$ origins have been shifted so that the experimental and theoretical absorption maxima of the monomer coincide. The broadening is described by a fwhm set to 1000 cm^{-1}. The experimental maxima in methanol[16] (410 nm and 370 nm) are represented by vertical lines.

in vacuum and therefore that the effects of the solvent are only included in the exciton and dimerization shifts. This is however a good approximation since the shape of the theoretical monomer spectrum in vacuum is in good agreement with the experimental spectrum in methanol.[16] The positions of the absorption maxima for the dimer spectra are in very good agreement with experiment. The calculated values are underestimated by 0.01 eV and 0.05 eV in vacuum and methanol, respectively. The ratios between the dimer and monomer intensities are larger with values of 2.85 and 2.59 in vacuum and methanol, respectively. The dimer spectrum in vacuum shows a rather single band shape, whereas the spectrum in methanol displays a weak shoulder at larger wavelengths. The presence of this shoulder is connected to the smaller value of the exciton shift in methanol and is in agreement with the experimental spectrum reported by Katoh et al.[16]

Figure 5 shows similar spectra to Figure 4 obtained by including the Duschinsky rotation effect. On the one hand, the monomer spectrum displays a larger vibronic shoulder and a decreased intensity at the maximum in comparison to the IMDHOM scheme (Figures 3 and 4), which neglects Duschinsky rotations. On the other hand, the intensities of the dimer spectra are almost similar to those obtained from the monomer FC overlap integrals calculated within the IMDHOM approximation. This leads to larger dimer/monomer intensity ratios of 3.28 and 2.90 in vacuum and methanol, respectively. Moreover, the absorption maximum in vacuum is overestimated by 0.005 eV, whereas it is underestimated by 0.04 eV in methanol. In general, one can conclude that for this compound, the Duschinsky effect has a small impact on the shape and position of the dimer spectra. In agreement with previous studies dealing with absorption and resonance Raman spectra,[23,24] this shows that the

IMDHOM can be considered as a good first approximation for the determination of the vibronic structure of excited states.

3.3. Further Discussion. As presented in the previous section, general agreement is found between both approaches and the experimentally reported spectra: i) the position of the dimer maximum with respect to the monomer is predicted with an accuracy of about 0.05 eV and ii) the global vibronic structure and shapes of the spectra are reproduced. Nevertheless, some remarks can be drawn concerning the differences and limit of applicability of both approaches. The comparison between the single molecule approach on the stacked dimer and the MSC results on the model dimer reveals some differences. The single molecule calculation shows a shoulder at shorter wavelengths, whereas the MSC method in methanol produces a shoulder at larger wavelengths. Subsequently, the second approach leads to a further increase of the intensity of the absorption maximum.

On the one hand, the single molecule calculation on the stacked dimer is based on the assumption that the Born–Oppenheimer approximation is valid for the dimer as a whole. As a consequence, the dimer displays a spectrum comparable to the monomer with a regular FC progression at shorter wavelengths. On the other hand, as discussed recently by Andrzejak and Petelenz,[6] the MSC approach is truly valid for weak electronic coupling (exciton shift small with respect to the monomer bandwidth) and can be considered as a good approximation for intermediate electronic coupling (exciton shift comparable to the monomer bandwidth). This approach is able to describe the complex interactions between the vibronic states in the intermediate coupling regime, which lead to a vibronic shoulder at longer wavelengths. However, when the exciton shift becomes large the intensity is mainly transferred to a single band, and the

DFT Vibronic Coupling: Stacked Streptocyanine

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2099**

method is unable to reconstruct the regular FC progression at shorter wavelengths. For the dimer investigated in this study the ratio between the exciton shift and the monomer bandwidth indicates a coupling lying between the intermediate and strong coupling limits. This is also suggested by the rather short optimized intermonomer spacing of 4.07 Å. Anyway, each method appears to be able to reproduce most features of the dimer absorption spectrum, partly due to the rather low resolution of the spectra. Improvements can however be achieved within the second approach by including in the dimer basis set vectors where both monomers are vibrationally excited. In this situation, it has been shown[5,6] using the monomode dimer model that the inclusion of these states strongly improves the description of the absorption spectrum for large exciton shifts. In particular, the regular FC progression is reproduced in the strong electronic coupling limit. The inclusion of such effects in a multimode approach has been presented by Myers Kelley[9] by employing experimentally determined quantities. In the future, an efficient implementation of such a scheme based on theoretical quantum chemistry approaches is expected to lead to more accurate predictions of the dimers and aggregates absorption spectra. Nevertheless, the use of the MSC approximation is still meaningful in situations like here and can lead to quantitative predictions for the apparent energy shift arising upon dimerization.

4. Conclusions

DFT and TDDFT schemes have been used to simulate the absorption spectra of a streptocyanine and of its covalently bonded dimer. Two approaches, based on multimode FC overlap integrals, have been employed. The first approach corresponds to a single molecule calculation where the FC factors are determined for the monomer and the stacked dimer separately. The second approach is based on the diagonalization of the dimer Hamiltonian, which is constructed from the monomer FC overlap integrals and quantities (E_{exc} and Δ) describing the intermonomer electronic coupling. Both approaches succeed in reproducing the hypsochromic shift of the maximum of absorption occurring upon dimerization with an accuracy of 0.05 eV. The vibronic structure of the monomer is also in good agreement with experiment and depends little on the inclusion of Duschinsky rotation effects. The shape and relative intensity of the dimer spectrum is qualitatively reproduced by the two methods, each of them being able to describe most of the vibronic features. Moreover, accounting for the solvent effects in the calculation of the exciton and dimerization shifts improves the agreement with experiment by reducing the intensity of the maximum and by enlarging the spectrum at longer wavelengths. Finally, this study illustrates the efficiency of quantum chemical calculations for the prediction of dimer absorption spectra and can be seen as a first step toward first principles description of vibronic coupling in molecular aggregates.

References

(1) Eisfeld, A.; Briggs, J. S. *Chem. Phys.* **2006**, *324*, 376.

(2) Eisfeld, A.; Briggs, J. S. *Chem. Phys.* **2002**, *281*, 61.

(3) Eisfeld, A.; Briggs, J. S. *Phys. Rev. Lett.* **2006**, *96*, 113003.

(4) Eisfeld, A. *Chem. Phys. Lett.* **2007**, *445*, 321.

(5) Eisfeld, A.; Braun, L.; Strunz, W. T.; Briggs, J. S.; Beck, J.; Engel, V. *J. Chem. Phys.* **2005**, *122*, 134103.

(6) Andrzejak, M.; Petelenz, P. *Chem. Phys.* **2007**, *335*, 155.

(7) Fulton, R. L.; Gouterman, M. *J. Chem. Phys.* **1964**, *41*, 2280.

(8) Kopainsky, B.; Hallermeier, J. K.; Kaiser, W. *Chem. Phys. Lett.* **1981**, *83*, 498.

(9) Kelley, A. M. *J. Chem. Phys.* **2003**, *119*, 3320.

(10) Leng, W.; Würthner, F.; Kelley, A. M. *J. Phys. Chem. B* **2004**, *108*, 10284.

(11) Gavrilenko, V. I.; Noginov, M. A. *J. Chem. Phys.* **2006**, *124*, 044301.

(12) Clark, A. E.; Qin, C.; Li, A. D. Q. *J. Am. Chem. Soc.* **2007**, *129*, 7586.

(13) Guillaume, M.; Champagne, B.; Zutterman, F. *J. Phys. Chem. A* **2006**, *110*, 13007.

(14) Champagne, B.; Guillaume, M.; Zutterman, F. *Chem. Phys. Lett.* **2006**, *425*, 105.

(15) Guillaume, M.; Liégeois, V.; Champagne, B.; Zutterman, F. *Chem. Phys. Lett.* **2007**, *446*, 165.

(16) Katoh, T.; Inagaki, Y.; Okazaki, R. *Bull. Chem. Soc. Jpn.* **1997**, *70*, 2279.

(17) *TURBOMOLE, Program Package for ab initio Electronic Structure Calculations*; University of Karlsruhe: Germany, Ver. 5.71.

(18) Ruhoff, P. T. *Chem. Phys.* **1994**, *186*, 355.

(19) Duschinsky, F. *Acta Physicochim. URSS* **1937**, *7*, 551.

(20) *Gaussian 03, Revision C.02*; 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.; Raghava-

chari, 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, Inc.: Wallingford CT, 2004.

(21) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.

(22) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. *J. Chem. Phys.* **2008**, *128*, 044118.

(23) Guthmuller, J.; Champagne, B. *J. Chem. Phys.* **2007**, *127*, 164507.

(24) Guthmuller, J.; Champagne, B. *J. Phys. Chem. A* **2008**, *112*, 3215.

JCTC Journal of Chemical Theory and Computation

Covalent Excited States of Polyenes $C_{2n}H_{2n+2}$ ($n = 2-8$) and Polyenyl Radicals $C_{2n-1}H_{2n+1}$ ($n = 2-8$): An Ab Initio Valence Bond Study

Junjing Gu,[†] Yonghui Lin,[†] Ben Ma,[†] Wei Wu,*,[†] and Sason Shaik*,[‡]

State Key Laboratory of Physical Chemistry of Solid Surfaces and Department of Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China, and The Institute of Chemistry and the Lise Meitner-Minerva Center for Computational Quantum Chemistry, The Hebrew University, Jerusalem 91904, Israel

Abstract: The ab initio valence bond (VB) methods, VBSCF and VBCI, are applied to the ground states and the covalent excited states of polyenes $C_{2n}H_{2n+2}$ ($n = 2-8$) and polyenyl radicals $C_{2n-1}H_{2n+1}$ ($n = 2-8$). The excitation energy gap was computed at the ab initio VB level, which is in good agreement with the semiempirical VB method, VBDFT(s), and the experimental values as well as with the molecular orbital theory based methods, CASPT3 and MRCI. The ab initio VB wave functions of systems are also in very good agreement with those of the VBDFT(s) method, even though the former is based on the ab initio VB scheme while the latter is a semiempirical Hückel type method, in which no orbital optimization procedure is performed. The computational results show that the ab initio VB method is capable now of providing numerical accuracy not only for bond forming and breaking processes, as shown in the past, but also for excitation energies, as shown here. In addition, the computational results validate the efficiency of the VBDFT(s) method, which is a simple VB model with less computational effort but which provides intuitive insights into the excited states of conjugated molecules.

Introduction

Polyenes are versatile molecules with great appeal for the chemical, biochemical, and physical communities. The excited states of these molecules are especially important in several processes in nature, such as vision and light harvesting. It is no surprise therefore that this molecular family has attracted much interest both experimentally and theoretically. There is special interest in the properties of the longer chain members of the polyene series because of their potential use in nanodevices. As such, the present paper uses the modern valence bond (VB) theory to calculate the excitation energies and elucidate the properties of the covalent states of the even-

and odd-membered polyene series and their asymptotic behavior as the polyene grows.

The ground-state of polyenes with an even number of carbon atoms is 1^1A_g. One of the key excited states is the so-called dark state 2^1A_g, which has drawn much attention since the discovery by Hudson and Kohler.[1,2] The pioneering study of Karplus and Schulten[3] has shown that 2^1A_g is the first adiabatic excited-state for small polyenes. After their study, more sophisticated calculations established the same trend for larger members, up to $C_{10}H_{12}$.[4−12]

The odd carbon series, the so-called polyenyl radicals, share an unpaired electron. The smallest polyenyl radical is allyl, which is of particular interest due to its doublet instability observed in molecular orbital (MO)-based calculations.[13−18] Higher polyenyl radicals have also been subjects of interest as models for electron transport in polyacetylene, in terms of migrations of neutral solitons.[19−21] The covalent

* Corresponding author e-mail: weiwu@xmu.edu.cn (W.W.), sason@yfaat.ch.huji.ac.il (S.S.).

† Xiamen University.

‡ The Hebrew University.

2102 *J. Chem. Theory Comput., Vol. 4, No. 12, 2008*

Gu et al.

excited states in this series have state symmetries A_2 and B_1 that alternate along the series, starting from the smallest member.

VB theory provides a powerfully intuitive tool for understanding the electronic structures and reactivity patterns of excited states. Indeed, quite a few VB approaches have been designed and utilized for theoretical studies of the covalent excited states of polyenes.[11−14,22−31] Recently, we have developed a semiempirical VB method,[32−35] called VBDFT(s), which is scaled with a single parameter to give the energy of the density functional theory (DFT). We used the method to discuss the electronic structures of covalent excited states in terms of the chemical VB structures, called also the Rumer structures. In the previous studies, we derived the ground rules needed to understand the results and used these rules to discuss the asymptotic behavior of these molecules as the number of carbon atoms goes to infinity. Furthermore, the spin density distribution was discussed for radical series. These studies showed that the VB theory provides a coherent and lucid understanding of the properties of the covalent excited state, such as the makeup of the various states, their energies and geometries, the opposite bond alternation properties of the ground and covalent excited state, isomerization patterns, soliton characteristics, and so on.

Although the VBDFT(s) excitation energies were in good accord with sophisticated MO-based methods, such as CASPT2 and MRCI, there is still a challenging question; whether ab initio VB approaches are also capable of reproducing quantitatively correct results and offering at the same time lucid insight native to the VB method. Ab initio VB applications to chemical problems are still lagging behind the MO theory, due to their heavier computational cost, associated with the use of nonorthogonal atomic orbitals in VB theory. However, thanks to the rapid development in computational science, VB theory has been enjoying some revival for the last two-three decades.[36−38] Very recently, we have developed a VB method,[39] called VBCI, which incorporates configuration interaction into the VB method, while still keeping the minimal set of VB structures used in the simple VBSCF method. Several applications[40] show that the computational results of the VBCI method are in good agreement with those of MO-based sophisticated methods. We have therefore decided to apply in this paper two VB methods, VBSCF and VBCI, to explore the ground and covalent excited states of polyenes and polyenyl radicals and to compare the results with CASPT3 and MRCI calculations.

Computational Methods

A Brief Summary of VB Methods. A many-electron VB wave function is expressed in terms of VB structures Φ_K

$$\Psi = \sum_K C_K \Phi_K \tag{1}$$

where Φ_K may be a spin-coupled VB function or a spin-free form of a VB function.[41−43] The coefficients C_K in eq 1 are subsequently determined by solving the usual secular equation. Since VB structures are not mutually orthogonal, normalized structure weights are defined as[44]

$$W_K = \sum_L C_K M_{KL} C_L \tag{2}$$

where M_{KL} is the overlap matrix element between structures K and L.

One of the significant features of modern VB methods is that the VB orbitals are optimized during the calculation. In the VBSCF method,[45] both the VB orbitals and structural coefficients C_K are optimized simultaneously to minimize the total energy. The VBSCF method takes care of the static electron correlation but lacks dynamic correlation,[46] which is an absolutely essential ingredient for the goal of quantitative accuracy. Recently, an improved VB method, called VBCI,[39,40] was introduced, which uses a configuration interaction technique to describe electron correlation, both static and dynamic ones. Taking optimized VB orbitals from the VBSCF calculation, a subsequent VBCI calculation includes, in addition to the entire set of the VBSCF space, also those excited VB structures, which are generated by replacing occupied orbitals with virtual orbitals. In order to keep the one-to-one correspondence of VB functions to classical VB structures, the virtual orbitals in the VBCI method are strictly localized, and during the CI procedure the occupied orbitals are replaced only by those virtual orbitals that are localized on the same atoms. In this manner the VBCI wave function can be written as a linear combination of the same set of VB structures as in the VBSCF method. Two levels of VBCI methods, VBCIS and VBCISD, are applied in this paper, where VBCIS involves only single excitations, while VBCISD involves also double excitations.

The Rumer Structure Set for Covalent States. For a system of electron number N and spin S, the total number of Rumer covalent structures is

$$d_{full} = \binom{N}{N/2 - S} - \binom{N}{N/2 - S - 1} \tag{3}$$

For polyenes ($N = 2n$), with one π electron per π atomic orbital, each VB structure contains n covalent π-spin pairs, while for polyenyl radicals ($N = 2n$-1), each structure possesses n-1 covalent π-pairs and one unpaired electron. For linear polyenes and polyenyl radicals, there are spin-pairs over adjacent carbon atoms, which are actual π-bond and there are spin-pairs over longer distances; the longest being the one that links the two terminal carbon atoms. In this paper, a π-bond of two neighboring atoms is called a short bond, and any more distant spin-pair is called a long π-bond. Although the energies of Rumer structures are computed at the ab initio level in this paper, the spectrum of the Rumer structures may be discussed qualitatively by counting the number of short π-bonds.[32−35]

Scheme 1 shows the spectrums in terms of blocks that differ in the number of short-bonds, where parts a and b of Scheme 1 are for polyenes and polyenyl radicals, respectively. For polyenes ($N = 2n$), the lowest energy structure is the fundamental Rumer structure which involves n short π-bonds, denoted as R(0). The uppermost Rumer structure possesses only a single π-bond in the center of the molecule, while the rest of the electrons are paired into n-1 long bonds. A Rumer structure belonging to a general block is indicated by R(k), where k is the excitation rank and is simply the

Covalent Excited States of Polyenes and Polyenyl Radicals

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2103**
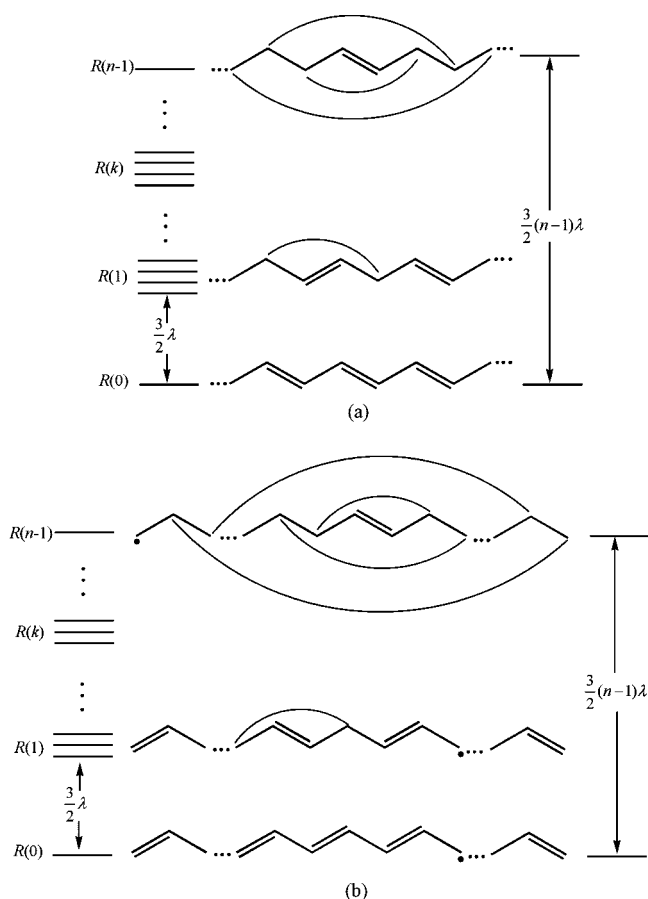
number of long bonds possessed by all the Rumer structures of the k block.

For polyenyl radicals ($N = 2n\text{-}1$), the lowest energy block contains n fundamental Rumer structures, which involve $n\text{-}1$ short bonds and an unpaired electron located on an odd-numbered carbon atom. The uppermost block consists of only one pair of twin Rumer structures, each of which possesses only one short π-bond near the center. In the VBDFT(s) method,[32–35] for uniform C–C bond lengths, all Rumer structures in the same block share the same energy, and the energy difference between the two neighboring blocks is 1.5λ, where λ is a parameter for the energy of a short π-bond. This is because the excitation breaks one π-bond and adds an additional nonbonded repulsion, which costs 0.5λ.

It can be seen from eq 3 that the number of total Rumer structures increases exponentially with the number of carbon atoms. In practice, it is impossible to include the full set of Rumer structures in the calculations of long polyenes, especially for ab initio VB treatments. The VBDFT(s) studies[33–35] showed that the ground-state properties were well reproduced by using only the fundamental, the singly and doubly excited Rumer structure blocks, while the excited states naturally require to include also the triply excited Rumer structures. In this paper, this truncation technique is applied for the long polyenes for which the full Rumer set is much too extensive. Therefore, two types of truncations are employed in this paper and are symbolized as VB-

SCF(S,D), VBCIS(SD), etc. One type reduces the VBSCF space by including only parts of the Rumer structures. This information is designated with "S" or "S, D" in parentheses, for example, VBSCF(S,D) means that one utilizes only those Rumer structures belonging to the fundamental, singly excited and doubly excited blocks of Scheme 1 for the VBSCF structure space. The other type of truncation information is for VBCI space. This is denoted as VBCIS and VBCISD, where the S and SD refer now to the excitation levels from occupied to virtual orbitals. Thus, for example, using a label such as VBCIS(S,D) would mean that the VBSCF space is (S,D), and the CI space of the VBCI involves only singles.

State Symmetries. In our previous papers,[33,35] the symmetries of states have been didactically and intuitively illustrated by means of the VB theory. In short, for polyenes, both of the ground-state and the dark state are A_g symmetry, designated as 1^1A_g and 2^1A_g, respectively. For polyenyl radicals, the ground-state is A_2 symmetry for even-membered n and B_1 symmetry for odd-membered n, respectively. Conversely, the symmetry for the first covalent excited-state is $B_1(A_2)$ for even(odd)-membered n.

Software. All VBSCF and VBCI calculations were carried out using the Xiamen Valence Bond (XMVB) package.[47] CASPT3 and MRCI calculations were performed using the MOLPRO package,[48] in which all π-electrons are included and N orbitals are taken as active space for C_NH_{N+2} molecule ($N = 2n$ for polyenes or $2n\text{-}1$ for polyenyl radicals). The MRCI(SD) level is used for MRCI calculations.

Basis Sets and Geometries. In order to produce good accuracy for the studies of excited states, a large basis set is usually required. However, our goal in this paper is to reach the chains of polyenes and polyenyl radicals as long as possible, and therefore large basis sets are prohibitive. Instead we use a D95V basis set, which is of a double-ζ quality and is used for all calculations, including VB-based VBSCF and VBCI methods and the MO-based CASPT3 and MRCI methods. The 6–31G(d) is applied for $C_{11}H_{13}$ and $C_{12}H_{14}$ to test basis set dependence. Optimized geometries for the ground states of polyenes and polyenyl radicals, C_3H_5–$C_{16}H_{18}$, were determined using B3LYP/D95V calculations implemented in the Gaussian 03 package.[49] The calculations were carried out in C_{2v} symmetry for polyenes and C_{2h} symmetry for polyenyl radicals. The B3LYP/D95V geometries are in good agreement with those obtained[17e] using the CASSCF method.

Results and Discussion

The Effect of Truncating the Rumer Basis. To check the errors that arise from the truncation of Rumer structures, the polyenyl radical $C_{11}H_{13}$ and the polyene $C_{12}H_{14}$ were tested at different truncation levels. Table 1 shows the effect of truncating the Rumer basis on the excitation energies with various VB methods. It can be seen that for $C_{11}H_{13}$ the excitation energies with (S) and (S,D) structure truncation are in good agreement with their corresponding values for all three methods, VBSCF, VBCIS, and VBCISD, in which the full Rumer set is used in the VBSCF calculation. The

Table 1. Effect of Truncation of the Rumer Space on the Vertical Excitation Energy (eV) for $C_{11}H_{13}$ and $C_{12}H_{14}$ Computed with the D95V Basis Set

molecule	space of VB structures[a]	no. of Rumer structures	VBSCF	VBCIS	VBCISD
$C_{11}H_{13}$	full	132	1.534	1.527	1.586
	(S)	46	1.541	1.534	1.563
	(S,D)	106	1.535	1.528	1.551
$C_{12}H_{14}$	full	132	3.562	3.548	3.497
	(S)	16	4.966	4.955	5.085
	(S,D)	66	3.564	3.551	3.602

[a] The space of the structure set which is used for the VB calculation, where "full" denotes the full Rumer set, while (S) means that those Rumer structures belonging to the fundamental and singly excited blocks are used, and (S, D) adds also for doubly excitations.

maximal deviation is 0.035 eV for VBCISD calculation. For $C_{12}H_{14}$, the difference in excitation energy of singly truncation (S) from the full Rumer set is pretty large, ca. 1.5 eV, while the deviation of doubly excitation truncation (S,D) is quite small, 0.002, 0.003, and 0.105 eV for VBSCF, VBCIS, and VBCISD, respectively. Therefore, these tests show that for polyenyl radicals, reliable results can be obtained when the Rumer set is truncated at the single excitation, while doubly excitation should be included in calculations for the polyene series.

VB Excitation Energies and the Convergence of the Excitation Gap. Table 2 shows the VB vertical excitation energies of polyenes, alongside the results of the semiempirical VB method, VBDFT(s), and sophisticated ab initio methods such as CASPT3 and MRCI. Full Rumer set calculations were performed for $C_4H_6-C_{12}H_{14}$ except for the VBCISD of $C_{12}H_{14}$, which is calculated at the VBCISD(S,D) level. For $C_{14}H_{16}$, there are 429 Rumer structures, which are too many to perform a VBSCF calculation with complete orbital optimization. A more effective way is to use only the singly excited structures (S) during the orbital optimization procedure, followed by VBSCF(full) and VBCIS(S) calculations with the so obtained optimized orbitals. For $C_{16}H_{18}$, only the fundamental and eight singly excited Rumer structures were included in orbital optimization, and then VBSCF(S,D) was performed. As can be seen from Table 2, the excitation energies of the all three levels of ab initio VB methods match one another very well. The deviation among the different methods is within only 0.09 eV. This indicates that VBCI accounts well and on equal footing for the dynamic correlation energies for both the ground-state and the covalent excited states. The excitation energies of ab initio VB methods are slightly higher than their corresponding values of VBDFT(s) by 0.3−0.4 eV. The results of the two sophisticated MO-based method, CASPT3 and MRCI, are slightly different, where the excited energies of the MRCI method are a little lower than those of CASPT3 by 0.2−0.4 eV. Note that the results of the ab initio VB methods are in good agreement with those of the MRCI method; especially good is the VBCISD method. The deviation of VBCISD from the MRCI method is in a range of 0.03−0.08 eV. Compared to experimental values, both of the ab initio VB and MO-based methods predict higher excitation energies, while VBDFT(s) values are in better agreement with experiments. It should be noted that the theoretical vertical

excitation energy is generally higher than experimental values, and thus the agreement between VBDFT(s) and experimental values does not show that the VBDFT results are better than those of ab initio VB and MRCI results.

Table 3 shows the VB vertical excitation energies for polyenyl radicals. Similar to the polyene series, a full Rumer set results are tabulated for $C_3H_5-C_{11}H_{13}$, while for $C_{13}H_{15}$ only singles (S) are involved in orbital optimization procedure, followed by VBSCF(full), VBCIS(S), and VBCISD(S) calculations, with no further orbital optimization. Similarly, for $C_{15}H_{17}$, only eight fundamental Rumer structures of the lowest energy were included in the orbital optimization procedure, followed by VBSCF(S) and VBCIS(S) calculations. As can be seen from Table 3, all the results of the three ab initio VB levels, VBSCF, VBCIS, and VBCISD, are in very good agreement mutually. The maximum of deviation is 0.06 eV. At the same time, all three ab initio VB results are seen to be in very good agreement with the corresponding semiempirical VBDFT(s) data. This latter match is very gratifying considering the fact that VBDFT(s) is just a Hückel-type VB method. Different from the polyene series, here the MO-based methods, CASPT3 and MRCI, match each other very well. The match between ab initio VB results with those of CASPT3 and MRCI is also good; the difference between VBCISD and MRCI methods are in a range of 0.03−0.06 eV.

In order to check the basis set dependence, Tables 4 and 5 collect the excitation energies for polyenes and polyenyl radicals with the 6−31G(d) basis set, respectively. It can be seen from Tables 2 and 4 that the VBSCF values of 6−31G(d) are higher than those obtained by using D95V by 0.12−0.30 eV for the polyene series, while for the VBCI levels, the values of 6−31G(d) are lower than those of D95V. Different from polyenes, for polyenyl radicals the excitation energies of ab initio VB methods are almost identical and are in very good agreement with those of 6−31G(d). Furthermore, the excitation energies of ab initio VBSCF methods for 6−31G(d) are in very good agreement with those of MRCI and CASPT3.

Figure 1 shows the excitation energies of various methods plotted against the number of carbons in the polyene chain, while Figure 2 exhibits the same information for the radical series. For the sake of economy, only VBSCF and VBDFT(s) as well as MRCI curves are shown in the same plot. As can be seen from the figures, all the three curves match very well. The VBSCF curve is close to that of MRCI, compared to that of semiempirical VBDFT(s) curve. The plots show the falloff of the excitation energy as the chain gets longer for both polyene and radical series. In a similar fashion to the VBDFT(s) studies,[33,35] an exponential fit to the VBSCF curve in Figure 1 leads to eq 4

$$\Delta E = 2.63 + 7.848 e^{-0.179(2n)}, R^2 = 0.99987; n = 2, 3, 4, ... \quad (4)$$

The intercept value of 2.63 eV at $n \rightarrow \infty$ is larger than that of VBDFT(s), 2.05 eV. This means that long polyenes will have a residual energy gap, presumably due to bond alternation.

Table 2. VB Vertical Excitation Energies (eV) for Polyenes $C_{2n}H_{2n+2}$ ($n = 2-8$) Computed with the D95V Basis Set

n	no. of Rumer structures	VBSCF	VBCIS	VBCISD	VBDFT(s)[a]	CASPT3	MRCI	exp.
2	2	6.47	6.47	6.56	6.28	6.75	6.53	
3	5	5.29	5.28	5.36	5.02	5.57	5.42	5.21[b]
4	14	4.49	4.49	4.55	4.19	4.79	4.60	4.41[c]
5	42	3.95	3.94	4.00	3.63	4.48	4.07	3.48[d]
6	132	3.56	3.55	3.60[e]	3.32	3.90	3.68	>2.68[f]
7	429	3.27	3.26[e]		2.93			>2.34[f]
8	225[e]	3.06[e]			2.70			

[a] The semiempirical method of ref 33. [b] Reference 8. [c] References 9 and 50. [d] Reference 51. [e] VB(S, D) type. [f] Reference 52.

Table 3. VB Vertical Excitation Energies (eV) for Polyenyl Radicals $C_{2n-1}H_{2n+1}$ ($n = 2-8$) Computed with the D95V Basis Set

n	no. of Rumer structures	VBSCF	VBCIS	VBCISD	VBDFT(s)[a]	CASPT3	MRCI
2	2	3.21	3.21	3.26	3.26	3.29	3.29
3	5	2.54	2.53	2.57	2.49	2.63	2.62
4	14	2.07	2.07	2.10	2.01	2.16	2.16
5	42	1.76	1.75	1.78	1.68	1.85	1.84
6	132	1.53	1.53	1.59	1.45	1.63	1.61
7	429	1.40	1.41[b]	1.44[b]	1.29		
8	120[c]	1.27[b]	1.19[b]		1.16		

[a] Reference 35. [b] VB(S) type.

**Table 4.** VB Vertical Excitation Energies (eV) for Polyenes $C_{2n}H_{2n+2}$ ($n = 2-5$) Computed with 6-31G(d)

n	structure	VBSCF	VBCIS	VBCISD	CASPT3	MRCI
2	2	6.59	6.30	6.36	6.67	6.65
3	5	5.49	5.19	5.24	5.57	5.42
4	14	4.75	4.45	4.49	4.72	4.70
5	42	4.25	3.94	3.85	4.17	4.15

Table 5. VB Vertical Excitation Energies (eV) for Polyenyl Radicals $C_{2n-1}H_{2n+1}$ ($n = 2-5$) Computed with 6-31G(d)

n	structure	VBSCF	VBCIS	VBCISD	CASPT3	MRCI
2	2	3.13	3.08	3.11	3.33	3.33
3	5	2.53	2.45	2.47	2.66	2.66
4	14	2.09	2.01	2.03	2.19	2.19
5	42	1.80	1.72	1.73	1.87	1.86

For radical series, fitting the VBSCF curve in Figure 2 as a cosine function leads to eq 5

$$\Delta E = 5.037\cos\frac{n-1}{2n}\pi, R^2 = 0.86629; n = 2, 3, 4, \ldots \quad (5)$$
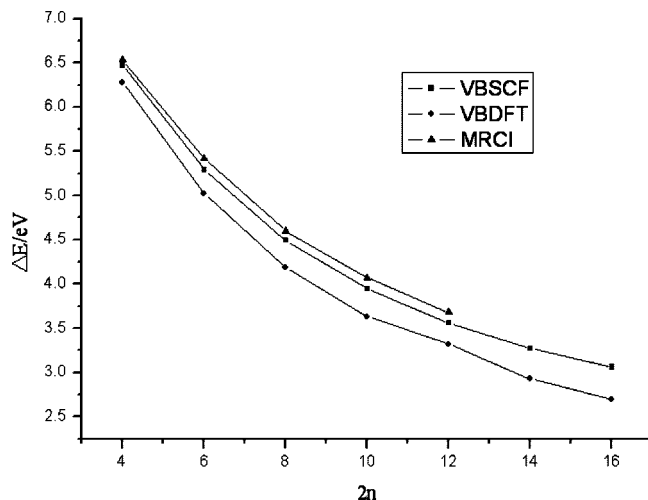
Figure 1. Dependence of the $1^1A_g \rightarrow 2^1A_g$ vertical excitation energy on the polyene length for $C_{2n}H_{2n+2}$.
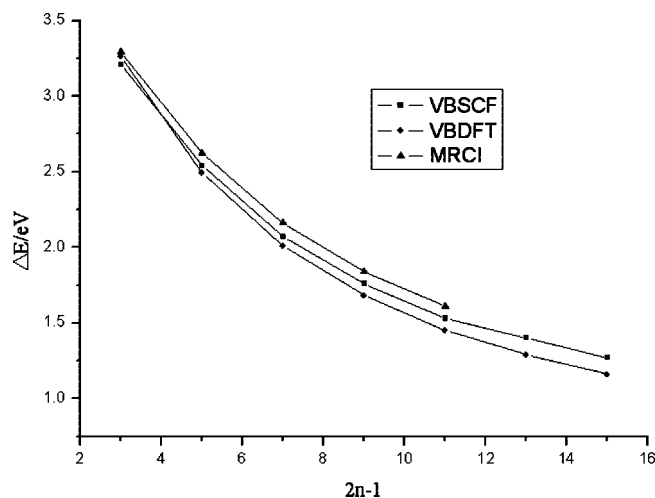
Figure 2. Dependence of the vertical excitation energy on the polyenyl radical chain length for $C_{2n-1}H_{2n+1}$.

The value of amplitude 5.04 eV in eq 4 is virtually identical to that of VBDFT(s), 4.96 eV. As discussed in the previous paper,[35] since the plot of radicals is done for ground-state geometries, this residual gap results from the nonuniform C−C bond-length of the chain and represents thereby the "soliton binding energy" *(stabilization of the radical by geometric distortion)*. For *an infinite chain* with uniform C−C bond length, the state symmetry is independent of the parity of *n*. As such, the symmetries of ground and covalent excited states, A_2 and/or B_1, collapse into a degenerate state of E symmetry. Therefore, the gap between the ground and covalent excited states for uniform C−C bond length should converge to zero.

Conclusion

This paper utilizes the ab initio VB methods, VBSCF and VBCI, to compute the relative energies and electronic structures of the ground and covalent excited states of

polyenes $C_{2n}H_{2n+2}$ ($n = 2-8$) and polyenyl radicals $C_{2n-1}H_{2n+1}$ ($n = 2-8$). The ab initio VB excitation energies are in good agreement with MRCI and CASPT3, with available experimental data, and with the formerly developed semiempirical VB method, VBDFT(s). VBDFT(s) performs extremely well compared with all the ab initio methods. Although not shown in the paper, the match is apparent also for the VB wave functions, for which VBDFT(s) and VBSCF have very similar results, despite the fact that VBSCF performs orbital optimization, while VBDFT(s) does not. As such, the computational results validate the efficiency of VBDFT(s) method, which is a simple nonexpensive VB model. Accordingly, VBDFT(s) can be used to obtain intuitive insight into the covalent excited states of conjugated molecules, as demonstrated in the previous publications.[32-35]

From the perspective of method development, the computational results show that the ab initio VB theory is able to provide numerical accuracy for the physical properties of excited states. We believe that this ability to deal with excited states is one more step in the revival of the VB theory.

Finally, while the present paper focuses on covalent excited states, the ab initio VB method is also capable of exploring the ionic excited states of polyenes, which are important in excited-state chemistry and photophysics. For the ionic excited states, there are many more Rumer structures than those of covalent excited states. We are currently studying the levels of truncation of the Rumer set that will provide good accuracy and compact wave functions.

Supporting Information Available: Two tables and the full reference for ref 49. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hudson, B. S.; Kohler, B. E. *Chem. Phys. Lett.* **1972**, *14*, 299–304.

(2) Hudson, B. S.; Kohler, B. E. *J. Chem. Phys.* **1973**, *59*, 4984–5002.

(3) Schulten, K.; Karplus, M. *Chem. Phys. Lett.* **1972**, *14*, 305–309.

(4) Hudson, B. S.; Kohler, B. E.; Schulten, K. *Excited States*; Lim, E. C., Ed.; Academic Press: New York, 1982; *Vol. 6*, pp 1–95.

(5) Kohler, B. E. *Chem. Rev.* **1993**, *93*, 41–54.

(6) Orlandi, G.; Zerbetto, F.; Zgierski, M. Z. *Chem. Rev.* **1991**, *91*, 867–891.

(7) Branchadell, V.; Sodupe, M.; Oliva, A.; Bertrán, J. Contribution of Quantum Chemistry to the Study of Dienes and Polyenes. In *The Chemistry of Dienes and Polyenes*;

Rappoport, Z., Ed.; John Wiley & Sons: Chichester, 1997; *Vol. 1*, pp 1–23.

(8) Serrano-Andrés, L.; Merchán, M.; Nebot-Gil, I.; Lindh, R.; Roos, B. O. *J. Chem. Phys.* **1993**, *98*, 3151–3162.

(9) Serrano-Andrés, L.; Lindh, R.; Roos, B. O.; Merchán, M. *J. Phys. Chem.* **1993**, *97*, 9360–9368.

(10) Serrano-Andrés, L.; Roos, B. O.; Merchán, M. *Theor. Chim. Acta* **1994**, *87*, 387–402.

(11) Nakayama, K.; Nakano, H.; Hirao, K. *Int. J. Quantum Chem.* **1998**, *66*, 157–175.

(12) Kawashima, Y.; Nakayama, K.; Nakano, H.; Hirao, K. *Chem. Phys. Lett.* **1997**, *267*, 82–90.

(13) (a) Karadakov, P. B.; Gerratt, J.; Raos, G.; Cooper, D. L.; Raimondi, M. *J. Am. Chem. Soc.* **1994**, *116*, 2075–2084. (b) Oliva, J. M.; Gerratt, J.; Cooper, D. L.; Karadakov, P. B.; Raimondi, M. *J. Chem. Phys.* **1997**, *106*, 3663–3672.

(14) (a) Levin, G.; Goddard, W. A., III *J. Am. Chem. Soc.* **1975**, *97*, 1649–1656. (b) Voter, A. F.; Goddard, W. A., III *Chem. Phys.* **1981**, *57*, 253–259.

(15) (a) Wiberg, K. B.; Breneman, C. M.; Laidig, K. E.; Rosenberg, R. E. *Pure Appl. Chem.* **1989**, *61*, 635–642. (b) Wiberg, K. B.; Breneman, C. M.; LePage, T. J. *J. Am. Chem. Soc.* **1990**, *112*, 61–72.

(16) Gobbi, A.; Frenking, G. *J. Am. Chem. Soc.* **1994**, *116*, 9275–9286.

(17) (a) Paldus, J.; ČížEk, J. *J. Chem. Phys.* **1971**, *54*, 2293–2303. (b) Paldus, J.; Veilard, A. *Mol. Phys.* **1978**, *35*, 445. (c) Kikuchi, O. *Chem. Phys. Lett.* **1980**, *72*, 487–488. (d) Takada, T.; Dupuis, M. *J. Am. Chem. Soc.* **1983**, *105*, 1713–1716. (e) Sim, F.; Salahub, D. R.; Chin, S.; Dupuis, M. *J. Chem. Phys.* **1991**, *95*, 4317–4326. (f) Feler, D.; Davidson, E. R.; Borden, W. T. *J. Am. Chem. Soc.* **1984**, *106*, 2513–2519.

(18) McConnell, H. M.; Chesnut, D. B. *J. Chem. Phys.* **1958**, *28*, 107–117.

(19) Etemad, S.; Heeger, A. J. *Ann. Rev. Phys. Chem., Neue Folge* **1981**, *126*, 205.

(20) Brédas, J. L.; Street, G. B. *Acc. Chem. Res.* **1985**, *18*, 309–315.

(21) Bally, T.; Hrovat, D. A.; Borden, W. T. *Phys. Chem. Chem. Phys.* **2000**, *2*, 3363–3371.

(22) Ohmine, I.; Karplus, M.; Schulten, K. *J. Chem. Phys.* **1978**, *68*, 2298–2318.

(23) Simonetta, M.; Gianinetti, E.; Vandoni, I. *J. Chem. Phys.* **1968**, *48*, 1579–1594.

(24) Dunning, T. H., Jr.; Hosteny, R. P.; Shavitt, I. *J. Am. Chem. Soc.* **1973**, *95*, 5067–5068.

(25) (a) Said, M.; Maynau, D.; Malrieu, J.-P.; Garcia Bach, M.-A. *J. Am. Chem. Soc.* **1984**, *106*, 571–579. (b) Said, M.; Maynau, D.; Malrieu, J.-P. *J. Am. Chem. Soc.* **1984**, *106*, 580–587. (c) Guihery, N.; Amor, N. B.; Maynau, D.; Malrieu, J. P. *J. Chem. Phys.* **1996**, *104*, 3701–3708.

(26) (a) Klein, D. J.; García-Bach, M. A. *Phys. Rev. B* **1979**, *19*, 877–886. (b) Klein, D. J.; Schmalz, T. G.; Seitz, W. A.; Hite, G. E. *Int. J. Quantum Chem. Symp.* **1986**, *19*, 707–718.

(27) Li, X.; Paldus, J. *Int. J. Quantum Chem.* **1999**, *74*, 177–192.

(28) Hirao, K.; Nakano, H.; Nakayama, K.; Dupuis, M. *J. Chem. Phys.* **1996**, *105*, 9227–9239.

Covalent Excited States of Polyenes and Polyenyl Radicals

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2107**

(29) Wu, J.; Copper, D. L. *Phys. Chem. Chem. Phys.* **2001**, *3*, 2419–2423.

(30) Schmalz, T. G. Valence Bond Theory: A Valence Bond View of Fullerenes. In *Theoretical and Computational Chemistry*; Cooper, D. L., Ed.; Elsevier: Amsterdam, 2002; *Vol. 10*, pp 535.

(31) Klein, D. J. Valence Bond Theory: Resonating Valence-Bond Theories for Carbon p-Networks and Classical/Quantum Connections. In *Theoretical and Computational Chemistry*; Cooper, D. L., Ed.; Elsevier: Amsterdam, 2002; *Vol. 10*, p 447.

(32) Wu, W.; Zhong, S. J.; Shaik, S. *Chem. Phys. Lett.* **1998**, *292*, 7–14.

(33) Wu, W.; Danovich, D.; Shurki, A.; Shaik, S. *J. Phys. Chem. A* **2000**, *104*, 8744–8758.

(34) Wu, W.; Luo, Y.; Song, L.; Shaik, S. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5459–5465.

(35) Luo, Y.; Song, L.; Wu, W.; Danovich, D.; Shaik, S. *Chem. Phys. Chem.* **2004**, *5*, 515–528.

(36) Cooper, D. L. *Valence Bond Theory*; Elsevier: Amsterdam, The Netherlands, 2002.

(37) Shaik, S.; Hiberty, P. C. In *Reviews in Computational Chemistry*; 2004; pp 1−100.

(38) A special issue on 90 Years of the Lewis concept: *J. Comput. Chem.* **2007**, 28.

(39) (a) Wu, W.; Song, L.; Cao, Z.; Zhang, Q.; Shaik, S. *J. Phys. Chem. A* **2002**, *106*, 2721–2726. (b) Song, L.; Wu, W.; Zhang, Q.; Shaik, S. *J. Comput. Chem.* **2004**, *25*, 472–478.

(40) (a) Song, L.; Wu, W.; Hiberty, P. C.; Danovich, D.; Shaik, S. *Chem. Eur. J.* **2003**, *9*, 4540–4547. (b) Su, P.; Song, L.; Wu, W.; Hiberty, P. C.; Shaik, S. *J. Am. Chem. Soc.* **2004**, *126*, 13539–13549. (c) Su, P.; Song, L.; Wu, W.; Hiberty, P. C.; Shaik, S. *J. Comput. Chem.* **2007**, *28*, 185–197. (d) Su, P.; Wu, W.; Shaik, S.; Hiberty, P. C. *Phys. Chem. Phys.* **2008**, *9*, 1442–1452.

(41) Wu, W.; Mo, Y.; Cao, Z.; Zhang, Q. A Spin-Free Approach for Valence Bond Theory and Its Application. In *Valence Bond Theory*; Cooper, D. L., Ed.; Elsevier: Amsterdam, 2002; pp 143−185.

(42) Li, X.; Zhang, Q. *Int. J. Quantum Chem.* **1989**, *36*, 599–632.

(43) Mcweeny, R. *Int. J. Quantum Chem.* **1990**, *24*, 733–752.

(44) Chirgwin, H. B.; Coulson, C. A. *Proc. R. Soc. London* **1950**, *A2*, 196–209.

(45) (a) van Lenthe, J. H.; Balint-Kurti, G. G. *Chem. Phys. Lett.* **1980**, *76*, 138–142. (b) van Lenthe, J. H.; Balint-Kurti, G. G. *J. Chem. Phys.* **1983**, *78*, 5699–5713. (c) Verbeek, J.; van Lenthe, J. H. *J. Mol. Struct. (THEOCHEM)* **1991**, *75*, 115–137.

(46) Hiberty, P. C.; Shaik, S. *Theor. Chem. Acc.* **2002**, *108*, 255–272.

(47) (a) Song, L.; Wu, W.; Mo, Y.; Zhang, Q. *XMVB: An Ab Initio Nonorthogonal Valence Bond Program*; Xiamen University: Xiamen, 1999. (b) Song, L.; Mo, Y.; Zhang, Q.; Wu, W. *J. Comput. Chem.* **2005**, *26*, 514.

(48) MOLPRO (version 98.1) is a package of ab initio programs written by Werner, H.-J. and Knowles, P. J. with contributions from Almlöf, J.; Amos, R. D.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Elbert, S. T.; Hampel, C.; Lindh, R.; Lloyd, A. W.; Meyer, W.; Nicklass, A.; Peterson, K.; Pitzer, R.; Stone, A. J.; Taylor, P. R.; Mura, M. E.; Pulay, P.; Schü tz, M.; Stoll, H.; Thorsteinsson, T.

(49) Frisch, M. J. et al. *Gaussian 03 Revision D.01*; Gaussian, Inc.: Pittsburgh, PA, 2004.

(50) (a) Heimbrook, L. A.; Kenny, J. E.; Kohler, B. E.; Scott, G. W. *J. Chem. Phys.* **1981**, *75*, 4338–4342. (b) Heimbrook, L. A.; Kohler, B. E.; Levy, I. J. *J. Chem. Phys.* **1984**, *81*, 1592–1597. (c) Leopoid, D. G.; Vanda, V.; Granville, M. F. *J. Chem. Phys.* **1984**, *81*, 4210–4217. (d) Alian, M.; Neuhaus, L.; Haselbach, E. *Helv. Chim. Acta* **1984**, *67*, 1776–1782. (e) Bouwman, W. G.; Jones, A. C.; Philips, D.; Thibodeau, P.; Friel, C.; Christensen, R. L. *J. Phys. Chem.* **1990**, *94*, 7429–7434. (f) Petek, H.; Bell, A. J.; Choi, Y. S.; Yoshihara, K.; Tounge, B. A.; Christensen, R. L. *J. Chem. Phys.* **1993**, *98*, 3777–3794.

(51) D'Amico, K. L.; Manos, C.; Christensen, R. L. *J. Am. Chem. Soc.* **1980**, *102*, 1777–1782.

(52) Kohler, B. E.; Spangler, C.; Westerfield, C. *J. Chem. Phys.* **1988**, *89*, 5422–5428.

JCTC Journal of Chemical Theory and Computation

Explaining Asymmetric Solvation of Pt(II) versus Pd(II) in Aqueous Solution Revealed by Ab Initio Molecular Dynamics Simulations

Elizabeth C. Beret,[†] José M. Martínez,[†] Rafael R. Pappalardo,[†]
Enrique Sánchez Marcos,*,[†] Nikos L. Doltsinis,[‡,§] and Dominik Marx*,[‡]

*Departamento de Quimica Fisica, Universidad de Sevilla, 41012 Sevilla, Spain, and
Lehrstuhl für Theoretische Chemie, Ruhr—Universität Bochum, 44780 Bochum, Germany*

Abstract: The solvation behavior of Pt(II) versus Pd(II) has been studied in ambient water using ab initio molecular dynamics. Beyond the well-defined square-planar first solvation shell encompassing four tightly bonded water molecules as predicted by ligand field theory, a second coordination shell containing about 10 H_2O is found in the equatorial region. Additional solvation in the axial regions is observed for both metals which is demonstrated to be induced by the condensed phase. For the Pt(II) aqua complex, however, this water molecule is bonded with one of its hydrogen atoms toward the cation, thus establishing a typical anionic solvation pattern, which is traced back to the electronic structure of Pt^{2+} versus Pd^{2+} cations, in particular to the anisotropic polarizability of their tetrahydrates. Systematic model calculations based on suitable aqua complex fragments embedded in a polarizable continuum solvent support the idea that anionic hydration is facilitated by the liquid. Furthermore, transient protolysis of water molecules in the first shell is observed for both divalent transition metal cations, being more pronounced for Pt(II) versus Pd(II). The relevance of these solvation features is discussed with respect to the different acidity of Pt^{2+} versus Pd^{2+} aqua ions in water, their different water ligand exchange rates, and force field modeling approaches.

1. Introduction

Metal-ion water solutions have received a large amount of attention from the scientific community during the last three decades. However, there is only scarce information available for Pd(II) and Pt(II) from experiment[1−5] and theory.[6−10] Nevertheless, their aqua ions are both well-known to form square-planar tetrahydrates,[11] which introduces a strong asymmetry in their hydration patterns. As a result, the "concentric shell model" of Frank and Evans[12] is no longer valid and thus cannot be used to understand the solvation structure in the nonequatorial regions. In this sense, although

overall consensus between different experimental and theoretical investigations is achieved on the gross square-planar structure of these aqua ions, describing the axial regions above and below the metal-oxygen plane remains controversial. Previous EXAFS[3,4] and LAXS[3] studies on Pt(II) solutions depict a completely unpopulated axial region, while classical molecular dynamics (MD) simulations, employing interaction potentials obtained from first principles,[7,8] find a so-called "meso-shell". It consists of two labile water molecules axially coordinated to the metal cation, M^{2+}, at a M-O distance of about 2.67 and 2.95 Å for Pd(II) and Pt(II), respectively. A similar XAFS study on Pd(II) supports this picture,[5] while only one water molecule in the axial region is predicted from a QM/MM MD simulation of Pd(II) in ref 9. A more recent combined theoretical (QM/MM MD) and experimental (XAFS) study[10] reports, instead, the finding of two water molecules in the axial region of Pd(II), at

---

* Corresponding author e-mail: sanchez@us.es (E.S.M.) and dominik.marx@theochem.ruhr-unibochum.de (D.M.).

[†] Universidad de Sevilla.

[‡] Ruhr—Universität Bochum.

[§] Present address: Department of Physics, King's College London, Strand, London WC2R 2LS, United Kingdom.

Asymmetric Solvation of Pt(II) vs Pd(II)

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2109**

2.7−2.8 Å from the metal, in agreement with the previous meso-shell picture.[7]

Previous analysis of X-ray scattering data on $PdCl_4^{2-}$ and $PtCl_4^{2-}$ showed two bonded water molecules at the apical sites of these square-planar ions, at a distance of 2.77 Å to the metal,[13] indicating a similar axial hydration pattern as that of the meso-shell obtained from classical MD.[7,8] However, in Monte Carlo[14] and Car-Parrinello MD[15] simulations of a solution of cisplatin (*cis*-Pt(NH$_3$)$_2$(Cl)$_2$) in water and in finite cluster solvation studies[16] no axial hydration structure such as the meso-shell was found at equilibrium, but apparently no further detailed analysis of coordination in the axial region was performed in those investigations. An analysis of the interaction energy between the [Pt-(NH$_3$)$_4$]$^{2+}$ or *trans*-[Pt(OH)$_2$(NH$_3$)$_2$] complex and a water molecule approaching the metal along the axial region in vacuum showed that a linear HO-H· · ·Pt arrangement, called "inverse hydration", is stabilized by electron dispersion contributions to bonding.[17,18] In particular, in the case of the cationic compound the interaction occurs preferentially via coordination through the oxygen atom, although "inverse hydration" is observed in terms of a metastable local minimum, while for the neutral compound an approach through the hydrogen atom is clearly favored.

Elucidation of the axial coordination pattern of Pd(II) and Pt(II) in aqueous solution not only is interesting from this fundamental physico-chemical point of view but also impacts onto biochemistry. Rate constants for water exchange at 298 K in these aqua ions are about 6 orders of magnitude higher for Pd(II) than for Pt(II), the difference being attributed to their corresponding activation enthalpies.[11] The exchange reaction is believed to follow an associative interchange mechanism[6] in which the transition state defines a trigonal bipyramid for a 5-fold hydrated cation. A similar behavior is observed for other Pd(II) and Pt(II) square-planar complexes including ligands such as H$_2$O, NH$_3$, or Cl$^-$ when exchange of a water molecule takes place: the mechanism is associatively activated, and the reaction is faster for the Pd(II) case.[19,20] Some of these square-planar derivatives of Pt(II) such as cisplatin are reactive as anticancer drugs when they exchange their first-shell ligands for water molecules, while analogous coordination compounds of Pd(II) are not. The basis for this different behavior is thought to be the rate of ligand exchange, which happens too fast in the Pd(II) case to allow for the complex to reach its pharmacological target.[21] Thus, according to the water exchange mechanism, the axial hydration structure may play an important and discriminating role between Pd(II) and Pt(II) with respect to their pharmacological activities.

In the present study, Car-Parrinello molecular dynamics[22,23] (CP-MD) simulations for Pd(II) and Pt(II) in water at ambient conditions are presented, thus introducing dynamical nuclear and electronic degrees of freedom not considered in previous, classical MD simulations published by some of the authors.[7,8] The focus is on the hydration pattern in the axial versus equatorial regions beyond the first, square-planar hydration shell and, in particular, on the interplay of interactions defining them, thus going far beyond our preliminary short communication.[24] These dynamical condensed phase ab initio simulations in fully explicit solvent are complemented by ab initio calculations of finite, solvated tetrahydrate complexes within the Polarizable Continuum Model (PCM)[25−27] in order to disentangle the solvent effects on a qualitative level.

Furthermore, particular attention is paid to important changes in polarization and many-body effects that lead to transient protolysis of water molecules in the first coordination shell of the Pd(II) and Pt(II) tetrahydrates. Although proton transfer, autoprotolysis, and Grotthuss structural diffusion have been studied extensively in liquid water using the ab initio MD framework[28−32] as reviewed recently,[33] reports on ion-induced dissociation of water molecules as a source of proton transfer remain scarce. Temporary proton transfer events in aqueous solutions of *trivalent* Al(III) have been previously reported,[34,35] and even the 5-fold coordinated complex [Al(H$_2$O)$_4$OH]$^{2+}$ has been advocated as the predominant form of Al(III) in water under ambient conditions.[36] In addition, the mechanism and free energies of water dissociation induced by the *trivalent* transition metal cations Cr^{3+} and Fe^{3+} in aqueous solution has been revealed recently[37] using CP-MD. Since Al^{3+} features a distinctly different behavior, it has been concluded that "electrostatic effects alone do not fully account for water dissociation".[35] Here, we analyze quantitatively the electronic structure of the *divalent* transition metal Pd(II) and Pt(II) aqua complexes, and we gain additional insights into transient protolysis which is qualitatively related to experimental pK_a measurements.[1,2]

2. Methods

Ab initio molecular dynamics simulations within the Car-Parrinello approach,[22,23] CP-MD, have been performed of Pd(II) and Pt(II) cations in aqueous solution using the CPMD program package.[23,38] The PBE exchange and correlation GGA functional[39,40] was employed together with ultrasoft pseudopotentials[41] in conjunction with a plane wave basis set truncated at 30 Ry. Each system was set up to contain one metal dication and 70 water molecules in a periodically repeated cubic simulation box of length 13.140 Å and 13.418 Å for Pd and Pt, respectively, in order to reproduce the density of water at ambient conditions. Nose-Hoover chain thermostats were applied to both nuclei and electrons. The temperature was kept at 350 K in order to approximately correct for the underestimation of temperature using GGA-type functionals[42] thereby emulating room temperature, ≈300 K. The equations of motion were integrated using a time step of 4 au together with a fictitious electron mass parameter of 400 au and the hydrogen mass for H.

The starting configurations were taken from previous classical MD runs performed with Pd and Pt-water interaction potentials based on ab initio parametrization.[7,8] A 2 ps thermalization plus a 2 ps system relaxation period preceded each production run of over 11 ps length. The system configuration was saved every 10 MD steps (≈ 0.97 fs) for further analysis. For electronic structure analysis norm-conserving dual-space Gaussian pseudopotentials[43] and a plane wave cutoff of 80 Ry were employed.

The dynamical liquid phase simulations were supplemented by static gas phase electronic structure calculations

using finite solvation complexes subject to both vacuum conditions and the solvent effects induced by a continuum solvation model. Plane wave PBE gas phase calculations with the CPMD code were performed by applying cluster boundary conditions[23] to avoid spurious electrostatic coupling of periodic images of the charged clusters. In addition, PBE and MP2 gas phase calculations using Gaussian basis sets were carried out. For these calculations energy-consistent pseudopotentials from the Stuttgart/Köln group (ECP28MWB for Pd and ECP60MWB for Pt) and their corresponding optimized valence basis sets (cc-pVDZ type) for Pd and Pt[44] were used within the Gaussian03 package,[45] together with aug-cc-pVDZ basis sets[46] for O and H. The continuum calculations were used to study solvation complexes systematically as a function of a reaction coordinate describing hydration in the axial region and were carried out with the Polarizable Continuum Model (PCM)[25,27] in its Integral Equation Formalism (IEFPCM)[47−50] as implemented in the Gaussian03 package.[47] The dielectric continuum was chosen to represent liquid bulk water at 298 K, i.e. $\varepsilon = 78.39$. The cavities employed here were built by using atomic radii from the Universal Force Field[51] scaled by an alpha factor of 1.2 (r_{Pd}: 1.450 Å, r_H: 1.443 Å), and all hydrogen atoms were explicitly considered in the cavity definition.

The polarizabilities of the tetrahydrates in the gas phase were computed using the Gaussian03 package[45] with the PBE exchange and correlation functional. The tetrahydrate geometry was first optimized using the chosen electronic structure method. It has been reported that good values of polarizabilities of metal cations can be computed within the density functional framework if triple-zeta quality basis sets extended with diffuse and polarization functions are employed.[51] For the $[Pd(H_2O)_4]^{2+}$ case, two different pseudopotentials from the Stuttgart/Köln group, together with the corresponding optimized valence basis sets, were tested. The ECP28MWB pseudopotential with cc-pVDZ quality basis sets for Pd(II)[44] together with aug-cc-pVDZ basis sets[46] for O and H yields virtually identical results as the ECP28MDF pseudopotential with aug-cc-pVTZ-PP basis sets for Pd(II)[53] and aug-cc-pVTZ basis sets[46] for O and H. For Pt(II) no triple-zeta quality basis set from the Stuttgart/Köln group is available, so the ECP60MWB pseudopotential with cc-pVDZ quality basis set[44] was employed for this metal cation, together with aug-cc-pVDZ basis sets[43] for O and H.

3. Results

3.1. General Hydration Pattern. *3.1.1. Radial Distribution Functions.* The spherically averaged solvation structure around the Pd and Pt cations is first analyzed in terms of the usual metal-oxygen and metal-hydrogen radial distribution functions (RDFs) shown in Figure 1. The positions of the peak maxima r_M and the coordination numbers n are gathered in Table 1. The first solvation shell is sharply defined and integrates up to exactly four water molecules in both cases. Zero density in the RDFs after this M-O peak means that these four H_2O molecules remain in the first solvation shell during the whole simulation, as expected from previous data on mean residence times of H_2O in the closest environment
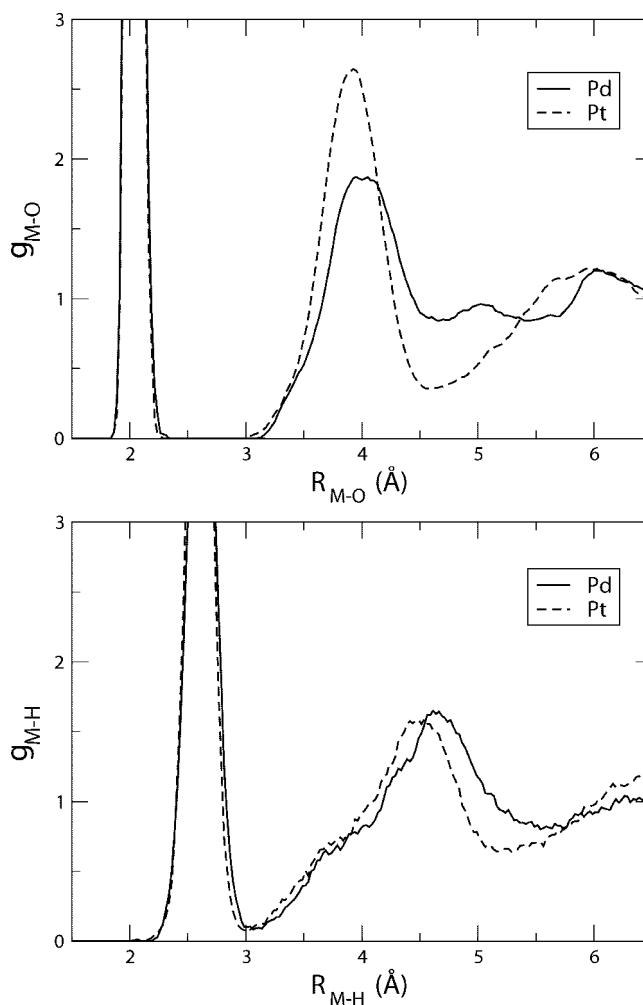
Figure 1. M-O and M-H radial distribution functions for aqueous Pd(II) (solid line) and Pt(II) (dashed line) solutions.

Table 1. Peak Maxima (r_M in Å), Coordination Numbers (n), and Underlying Integration Ranges (r_L in Å) from M-O and M-H Radial Distribution Functions of Aqueous Pd(II) and Pt(II) Solutions

	1st shell			2nd shell		
	r_M	n	r_L	r_M	n	r_L
Pd−O	2.04	4.0	0−2.31	4.02	10.3	2.31−4.64
Pd−H	2.61	8.0	0−3.10	4.64	37.1	3.10−5.56
Pt−O	2.04	4.0	0−2.26	3.91	9.9	2.26−4.56
Pt−H	2.61	8.6	0−3.10	4.50	26.2	3.10−5.19

of these cations.[19] The first M-O maximum appears at 2.04 Å for both Pd and Pt which is consistent with previous experimental and theoretically obtained values ranging from 2.00 to 2.07 Å, see refs 3−5 and 7−9. The corresponding M-H maxima are found at 2.61 Å which can be compared to a value of ≈2.7 Å found previously using classical MD simulations employing interaction potentials[7,8] based on MP2 calculations and also by QM/MM MD simulations at the Hartree-Fock level.[9,10]

The second hydration shell peaks are also clearly defined but quite different in Pd and Pt as the peak is higher and narrower in the case of Pt. In addition, a clear minimum appears after the second Pt−O peak, whereas this is absent

Asymmetric Solvation of Pt(II) vs Pd(II)

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2111**
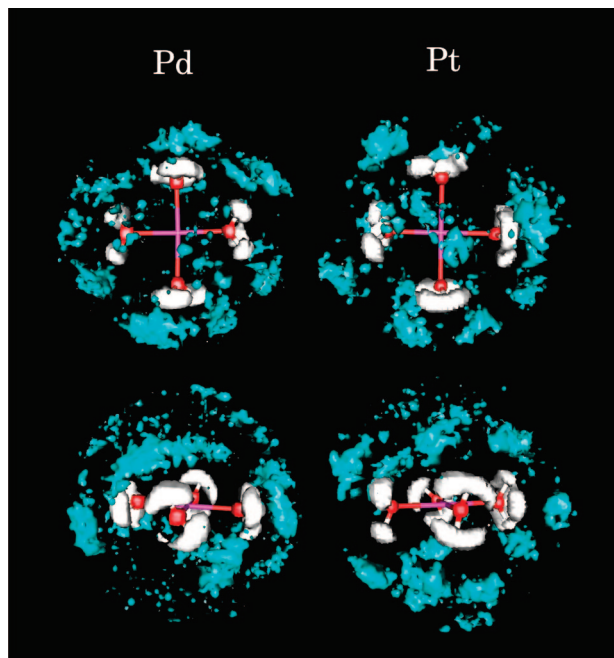
Figure 2. Spatial distribution functions for hydrogen atoms in the first solvation shell (white surfaces) and oxygen atoms in the first and in the second solvation shells (red and cyan surfaces, respectively) of Pd(II) and Pt(II), from apical (up) and equatorial (below) points of view. The average square-planar geometry of the tetrahydrates is represented as pink, red, and white bars.

in the Pd−O RDF. Thus the second solvation shell is more pronounced in the case of Pt compared to Pd. The second M-O peaks have maxima at 4.02 and 3.91 Å for Pd and Pt, respectively. These distances are shorter than those obtained from previous, classical MD simulations[7,8] (4.08 Å for both cations) and QM/MM MD simulations of Pd(II) (4.50 Å in ref 9 and 4.40 Å in ref 10) and integrate up to about 10 H_2O molecules.

3.1.2. Spatial Distribution Functions. The M-O and M-H RDFs merely provide spherically averaged information about the distribution of water molecules in the solution around the metal cation. Even though this purely radial information is of great value in itself, it is not sufficient when studying systems whose symmetry is clearly not spherical as in the present case. Here, spatial distribution functions (SDFs) presented in Figure 2 offer three-dimensional views of the solvation neighborhood around the metal cations.

The spatial picture obtained for Pd(II) and Pt(II) solutions is that of a well-defined square-planar first solvation shell, see Figure 2, allowing for the definition of an average plane containing the metal cation and those four oxygen atoms of H_2O that directly coordinate to the metal. The average orientation of the hydrogen atoms in these first-shell water molecules is distinctly different in Pd(II) and Pt(II) solution. In the Pd(II) case, the molecular plane of two H_2O are co-planar with the metal-oxygen plane, while the other two H_2O are perpendicular to it. For the Pt(II) tetrahydrate on the other hand none of the eight hydrogen atoms are contained in the metal-oxygen plane.

Most interesting is the behavior beyond the first solvation shell where a dense toroidal-shaped second hydration shell

is observed to enclose the tetrahydrate cores. In stark contrast to the equatorial region, the axial regions remain sparsely populated. Still, some H_2O density is observed there at second-shell distances. However, these H_2O molecules are not strictly second-shell molecules, as they are not bonded to H_2O in the first hydration shell. Based on this qualitative knowledge a more quantitative analysis can be performed by using angle-resolved RDFs that focus onto the axial and equatorial solvation regimes.

3.1.3. Angular Decomposition of RDFs. The solvation space around the metal cations is split into different regions, which are defined by an azimuthal angle θ relative to the axis perpendicular to the average molecular plane of the tetrahydrate around the metal center. Then the RDFs for the molecules included in each of these regions can be obtained separately. Let us consider two hypothetical compounds, one with an octahedral arrangement of the oxygen atoms, O, around the metal center, M, and the other one with an hexagonal-planar structure according to the sketch in Figure 3. If all M-O average distances oscillate around one R_{MO} equilibrium value, both compounds will yield identical M-O *global* RDFs, their corresponding peaks centered at R_{MO} with a half-width of σ_{M-O} due to molecular vibrations. In order to distinguish between the two qualitatively different coordination structures, one can define a molecular plane and formally split the two hemispheres above and below into regions defined by the solid angle with respect to the axis normal to the molecular plane. The revolution of the generatrix line around this axis defines symmetric conical and semi-crown-shaped regions above and below the molecular plane. Then angle-resolved RDFs can be computed taking into account those molecules that are contained in each of those regions. To this aim, a reference *homogeneous* density ρ_{ref} is defined as the total number of O atoms in the simulation box divided by the total box volume V, as in a standard RDF computation. For each angular region the partial RDF, $g_p(r)$, is obtained by dividing the current oxygen density number (number of O atoms in a fraction of volume centered in r within the given angular region) by the reference density, ρ_{ref}. An application of this definition is shown in Figure 3, for the model example of two hexaco-ordinated metal compounds, MO_6, with the same average distance, R_{M-O}, but two different spatial distributions of O ligands around the metal M (octahedral and hexagonal-planar). Let us define two regions by means of the azimuthal angle: $\theta \in [0, 60]°$ and $\theta \in [60, 90]°$ where the two hemispheres, above and below the molecular plane, are considered to be equivalent due to symmetry. Each of these regions has a volume equal to half the total volume, $V_{[0, 60]°} = V_{[60, 90]°} = V/2$. The global RDFs show peaks of intensity $(6)/(\rho_{ref}V)$, and thus no distinction is observed between the octahedral and hexagonal-planar MO_6 compounds. However, when we compute angle-resolved RDFs, in the octahedral case peaks of intensities $(4)/(\rho_{ref}V)$ and $(8)/(\rho_{ref}V)$ are observed for the $[0, 60]°$ and $[60, 90]°$ regions, respectively. On the other hand, for the planar compound no signal is observed in the $[0, 60]°$ region, while a peak of intensity $(12)/(\rho_{ref}V)$ appears in the $[60, 90]°$ region. Thus, this type of decomposition of global RDFs into angle-resolved RDFs
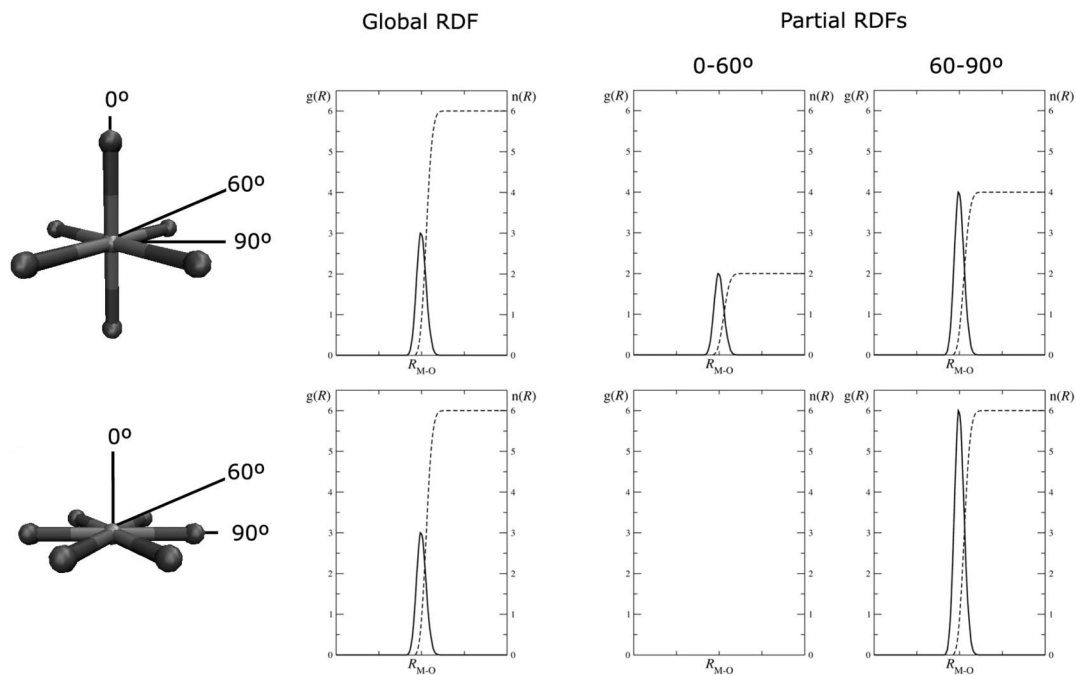
Figure 3. Decomposition of M-O global RDFs (left) into contributions coming from different angular regions for an octahedral (top) and a hexagonal-planar (bottom) MO_6 compound, see text for further details. RDFs and running integration numbers are represented by solid and dashed lines, respectively.
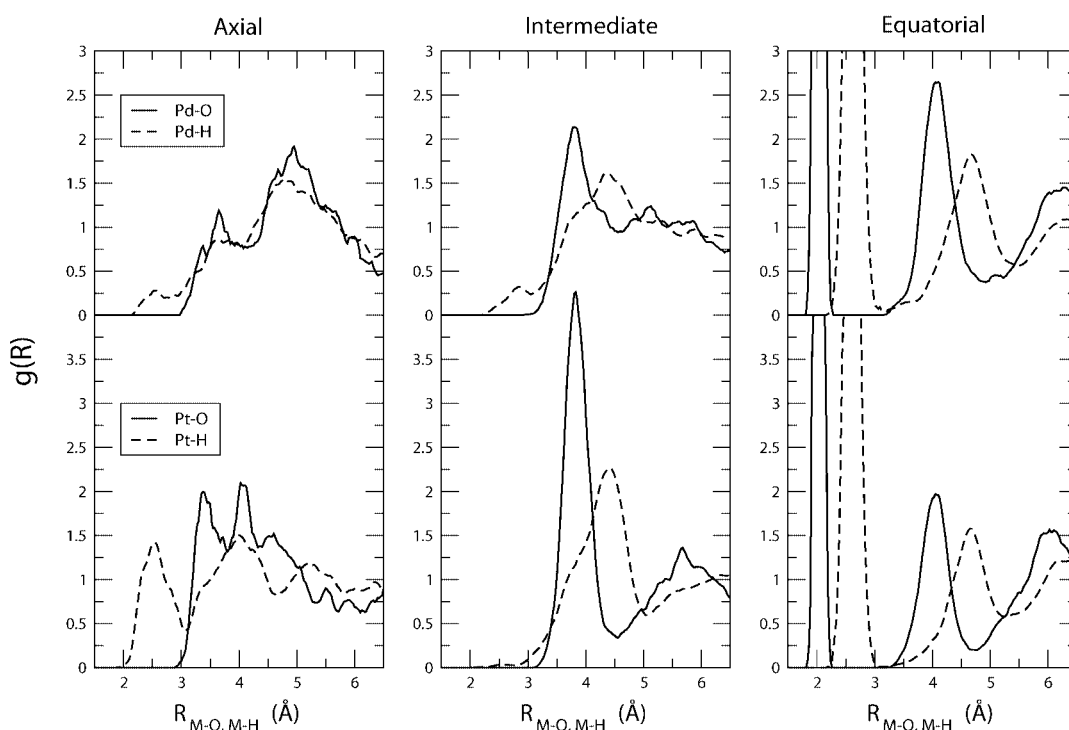
Figure 4. Decomposition of M-O (solid line) and M-H (dashed line) global RDFs for aqueous Pd(II) (top) and Pt(II) (bottom) solutions into three angular regions defined by the solid angle θ defined in Figure 3: axial region: $\theta \in [0, 30]°$, intermediate region: $\theta \in [30, 60]°$, and equatorial region: $\theta \in [60, 90]°$.

helps to extract a quantitative understanding from the distribution of ligands around a given metal center.

In the case of the Pd(II) and Pt(II) tetrahydrates with a square-planar first solvation shell, three angular regions were defined within the angular ranges $\theta \in [0, 30]°$ (termed *axial* region), $\theta \in [30, 60]°$ (*intermediate* region), and $\theta \in [60, 90]°$ (*equatorial* region). For each configu-

ration snapshot selected for this analysis all water molecules were uniquely assigned to one of the three regions. Figure 4 depicts the resulting angular decomposition of the global M-O and M-H RDFs for the Pd(II) and Pt(II) aqueous solutions. In these angle-resolved RDFs one can clearly observe that the closest water molecules appear in the equatorial region, which is quite similar for both

Asymmetric Solvation of Pt(II) vs Pd(II)

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2113**

Table 2. Reconstruction of the Second Hydration Shell in Aqueous Pd(II) and Pt(II) Solutions Based on the Different Contributions in Terms of Angle-Resolved RDFs Introduced in Figure 4[a]

	r_L	[0,30]°	[30,60]°	[60,90]°	n
Pd−O	2.31−4.64	1.0	4.0	5.3	10.3
Pt−O	2.26−4.56	1.7	4.6	3.6	9.9

[a] [0, 30]°, [30, 60]°, and [60, 90]° label the integration numbers obtained for the angle-resolved RDFs within the reported r_L integration limits (in Å) corresponding to axial, intermediate, and equatorial regions, respectively, and n is the sum of the three angle-resolved contributions.

Pd and Pt solutions. It includes peaks from the first hydration shell, in accordance with the average square-planar geometry of the tetrahydrates as well as part of the second hydration shell. Furthermore, the Pt−H RDF also features a pronounced peak at short distances which is not observed for the Pd case.

Remarkable differences between the two systems can be found in the intermediate and axial regions. The second solvation shell is mainly distributed among the [30, 60]° and [60, 90]° regions. The differences between Pd(II) and Pt(II) second shells are more clearly observed in these partial RDFs than in the global RDFs. In the case of Pt(II), a strong M-O peak appears in the [30, 60]° region, and smaller contributions are found in the axial (peak centered at ≈4 Å) and equatorial regions. For Pd(II) the second hydration shell is more evenly distributed among the [30, 60]° and [60, 90]° regions. These different distributions of the second solvation shell can be understood taking into account the average orientations of the hydrogen atoms in the first solvation shell observed in the SDFs (see Figure 2). In the Pt(II) case, most hydrogen atoms are located above and below the metal−oxygen plane, therefore forcing the second-shell H_2O molecules, which are directly bonded to them, to be further away from the metal-oxygen plane than in the Pd(II) case.

In the axial region a maximum at 2.55 Å can be found in the Pt−H RDF which integrates to 0.6 hydrogen atoms. A corresponding Pt−O peak can be found at around 3.50 Å integrating also to 0.6 oxygen atoms. These two peaks are evidence for a water molecule that is weakly bonded to the metal cation through a hydrogen atom, i.e. this water molecule is oriented as if it was coordinated to an anion. Therefore this coordination pattern may be termed *anionic hydration*. A second Pt−O peak at 4 Å is observed, integrating up to 0.8 oxygen atoms, which represents a second H_2O molecule in the axial region but at a second-shell distance and bonded through the oxygen atom. In the Pd(II) case the Pd−H peak can also be found, but it is much less defined. There is only one Pd−O peak around 3.5 Å (close to second-shell distances) which integrates up to 1.1 oxygen atoms. For Pd the axial water molecule is poorly oriented, its interaction with the Pd(II) cation being much weaker than in the Pt(II) case and being established through the oxygen atom. It is worth noticing that particular structural features hardly visible in the global RDFs are clearly observed by means of constructing these partial RDFs.

In Table 2 the number of water molecules contributing to the second peaks found in the global RDFs are reconstructed based on the contributions from the angle-resolved RDFs. Putting together all the structural information so far obtained, the solvation pattern of Pd(II) and Pt(II) can be described as follows. For both cations the first hydration shell is well defined and composed of four water molecules arranged in an average plane defined by the metal cation and the four first-shell oxygen atoms. At typical second hydration shell distances about 10 H_2O are found, which are mainly located in the *equatorial region* thereby enclosing the square-planar aqua ions in a crownlike arrangement. Eight of them are directly bonded to the four H_2O molecules in the first hydration shell, thus constituting a proper second hydration shell. In the Pt(II) case, the two additional H_2O molecules are disposed in the axial region, their orientations being anionic for one of them and through the oxygen atom for the other one as shown in Figure 5 using a representative configuration. However, their residence time in the axial region is rather short, indicating a weak interaction with the metal cation. For Pd(II), only one H_2O molecule enters the axial region, interacting with the metal through the oxygen atom and residing for a shorter time than in the Pt(II) case, whereas the other water molecule is a *bridge* H_2O, which helps to strengthen the structure within the second hydration shell (see Figure 5).

Although the overall solvation scenario in terms of a rigid square-planar first solvation shell and a less defined second shell is in agreement with other theoretical[6−10] and experimental[1−5] investigations, the previously found axial hydration structures for Pd(II) and Pt(II) were quite different. In classical MD simulations[7,8] a pronounced meso-shell consisting of two H_2O molecules bonded to the metal cation through their oxygen atoms was found for both Pd(II) and Pt(II). These axial H_2O showed a much higher persistence than in ab initio MD and were located at a distance of 2.67 and 2.95 Å from the metal for Pd(II) and Pt(II), respectively, in agreement with XAFS experimental results.[5] A similar orientation was found within the QM/MM simulation framework[9,10] for one or two tighter H_2O molecules in the axial region of Pd at a distance of 2.5−2.8 Å from the metal cation. However, the possibility of square-planar compounds of Pt(II) accepting coordination from an axial water molecule through its hydrogen atom (referred to as "inverse hydration") was already suggested for the case of the neutral microsolvated *trans*-[Pt(OH)$_2$(NH$_3$)$_2$]. H_2O complex in vacuume.[17,18]

When dealing with cations in aqueous solution, a perfect, collinear ion-dipole orientation of solvation water with respect to the metal cation would be characterized by a tilt angle of 180° (being the angle between the geometric dipole moment vector of the bonded H_2O molecule and the vector between the cation and the oxygen atom of this molecule). However, already the pioneering classical MD study of an electrolyte solution[54] observed a significant tilt which is consistent with an arrangement where one of the lone pairs rather than the dipole vector points toward the cationic center. This picture has been confirmed by subsequent simulations as well as by diffraction experiments.[55] In the current CP-MD simulations the first shell water molecules show an average tilt angle of ca. 135° and 132° for Pd(II) and Pt(II), respectively, to be compared with ≈165° for both Pd(II) and
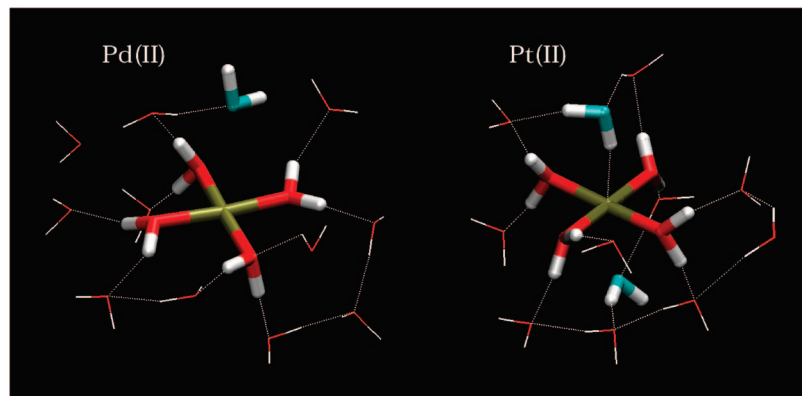
Figure 5. Snapshots sampled from the simulations of aqueous Pd(II) and Pt(II) solutions showing the tetrahydrate cores (licorice representation with red oxygen atoms), water molecules in the axial region (licorice representation with blue oxygen atoms), and some second shell and bulk water molecules; hydrogen bonds are sketched as dotted white lines.

Pt(II) obtained from classical MD and ≈175° in the above-mentioned Pd(II) QM/MM simulations. When the axial H_2O molecule comes close to the cation, its average tilt angle decreases to ≈100° for Pd(II) in the present CP-MD simulation to be compared to ≈100° and 90° in the above-cited classical MD and QM/MM MD approaches, respectively. An "anti ion-dipole" orientation with a tilt angle close to 0° would represent the other limiting behavior. On the other hand, the typical solvation of anions such as Cl^- as the prototype occurs preferentially via the hydrogen atoms of the first shell H_2O molecules yielding a tilt angle of typically $50 \pm 10°$, see refs 54 and 55. These angles should be compared to an average tilt angle of ≈40° observed in the CP-MD simulation for the axial water molecules around Pt(II). Clearly, the orientation of these axial water molecules with respect to the cation cannot be classified as a typical cationic solvation pattern, whereas it fits rather well the anionic hydration scenario. In the following, this peculiar solvation beyond the square-planar first shell will be dissected.

3.2. Axial Hydration. At first glance, the hydration of Pd(II) and Pt(II) aqua ions in the axial regions above or below the square plane, although dynamical and scarce, seems striking for various reasons. First of all, standard ligand field theory predicts a square-planar 4-fold coordination to be the most stable situation for these aqua ions leaving the axial regions as completely void spaces between the tetrahydrate and the bulk. Second, the hydration pattern in the axial region is very different between these two tetrahydrates, despite the similarities found in the intermediate and equatorial regions. In the Pd(II) case, one H_2O slightly escapes the second solvation shell, entering the axial region oriented toward the metal cation through the oxygen atom. For Pt(II), in contrast, one of the axial H_2O approaches the metal cation not through the oxygen atom, as expected, but through the hydrogen atom like in an anionic solvation shell. In order to gain further insights into these surprising results two different procedures have been employed: analysis of the electronic structure and description of interaction energies in gas phase and in solution.

3.2.1. Electronic Structure Analysis. Random snapshots were sampled from the simulation trajectory of Pt^{2+} in aqueous solution. The system was conceptually separated

into the following components: the tetrahydrate $[Pt(H_2O)_4]^{2+}$, the axial H_2O molecule with anionic orientation, a second-shell H_2O molecule, and the remaining solvent. The electronic (valence) density and the total electrostatic potential were computed for each of these entities which were subtracted from the functions corresponding to the whole system. This strategy allows for a comparison of the axial interaction with a standard hydrogen bond between a water molecule in the first solvation shell of Pt and a second-shell H_2O molecule.

This analysis shows that there is a certain charge transfer from the axial water molecule and the metal cation toward the region between them, as seen by a net displacement of the electronic density toward that region (white lobe) in the left panel of Figure 6a. In addition, the relative electrostatic potential is observed to become more negative, i.e. less electron attracting, in that region depicted by the red area in the right panel of Figure 6a. This pattern is qualitatively similar to the hydrogen bond between the first and second solvation shells observed for the same snapshot (see Figure 6b) The main effect of the interaction is the increase of electronic density (or negative charge character) in the region between the atoms involved, revealed by a pronounced white lobe in that region of space concurrent to a more negative relative electrostatic potential (red area). Therefore this anionic interaction behaves like a weak hydrogen bond formed between the hydrogen atom of the axial H_2O molecule and the polarized Pt(II) cation. This interpretation is reinforced by an additional analysis in terms of maximally localized Wannier functions.[56] When anionic axial hydration is established, the Wannier center located along the Pt−H direction moves slightly away from the Pt ion toward the hydrogen atom of the axial H_2O molecule, enlarging the average metal to Wannier center distance of 0.38 Å to 0.44 Å signaling polarization. In addition, we have computed Kohn-Sham canonical orbitals for a snapshot extracted from the Pt(II) simulation in which anionic hydration takes place. The resulting orbitals indicate an interaction between the $5d_{z^2}$, $5d_{xz}$, and $5d_{yz}$ orbitals of the metal and orbitals of the axial water molecule, the contribution of the $5d_{z^2}$ being more important than that of the $5d_{xz}$ and $5d_{yz}$ orbitals.

In solution, the more pronounced localization of the solvent electronic density in the equatorial region induces a

Asymmetric Solvation of Pt(II) vs Pd(II)

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2115**
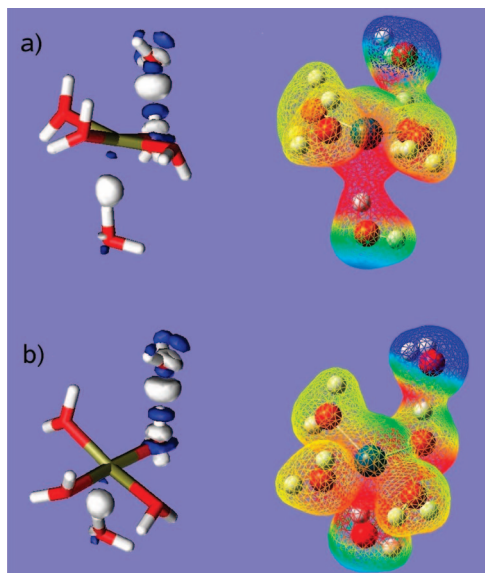
Figure 6. Electronic structure analysis of the anionic hydration of the solvated Pt(II) tetrahydrate compared to a proper hydrogen bond between a first-shell and a second-shell water molecule. A representative snapshot is shown in two orientations revealing the axial (a) and equatorial (b) interactions. Left: Electronic (valence) density difference (see text) shown by isosurfaces at +0.005|e| (white lobe) and −0.005|e| (blue lobe). Right: Electronic (valence) density surface (isosurface at +0.02|e|) onto which the change in electrostatic potential (see text) is mapped ranging from −0.02 au (red) via yellow and green to +0.02 au (blue).

Table 3. Isotropic Polarizability, α_{iso}, and Anisotropic Components, $\alpha_{xx} = \alpha_{yy}$ and α_{zz}, in au^3 Computed for the Pd(II) and Pt(II) Tetrahydrates in Vacuum[a]

complex	basis set	α_{iso}	$\alpha_{xx} = \alpha_{yy}$	α_{zz}
[Pd(H$_2$O)$_4$]$^{2+}$	ECP28MWB	52.2	58.7	39.2
	ECP28MDF	52.3	58.7	39.6
[Pt(H$_2$O)$_4$]$^{2+}$	ECP60MWB	53.7	58.8	43.6

[a] The square-planar tetrahydrate lies in the *xy* plane and *z* is the axial direction. The basis sets are labeled according to the pseudopotential chosen for the metal cation (see section 2 Methods).

slight electronic density redistribution of the metal, where the more diffuse character of the d_{z^2} orbital occupying the axial region stabilizes the system. In the case of the Pt(II) tetrahydrate this delocalization is sufficient for a dynamic anionic arrangement with a single H$_2$O molecule to take place. In the Pd(II) species the $4d_{z^2}$ orbital is less diffuse than the $5d_{z^2}$ in Pt(II), thus not being able to achieve an anionic coordination with an axial H$_2$O.

This conclusion can be quantified by considering the polarizability tensors of the two tetrahydrate cores, i.e. of [M(H$_2$O)$_4$]$^{2+}$ for M = Pt^{2+} and Pd^{2+}, reported in Table 3. Although the isotropic polarizability α_{iso} of this complex is virtually identical for Pt^{2+} and Pd^{2+}, there are significant differences when it comes to the anisotropy of the polarizability tensor. Within the square plane of the tetrahydrates the $\alpha_{xx} = \alpha_{yy}$ values are, again, very similar for both Pt^{2+} and Pd^{2+}, but the polarizability α_{zz} along the *z*-direction, i.e. perpendicular to this plane, is significantly larger for Pt^{2+}. Thus, the [Pt(H$_2$O)$_4$]$^{2+}$ tetrahydrate is expected to be more

polarizable in the axial regions in comparison to the respective Pd^{2+} complex, which is exactly the region where anionic solvation is observed for Pt(II) in aqueous solution.

Overall, these four different ways to analyze the electronic structure yield the same qualitative picture when it comes to their solvation behavior beyond the first, tightly bonded, square-planar shell. Thus, the anionic hydration as such and also the qualitative differences between the behavior of Pt^{2+}(aq) and Pd^{2+}(aq) can be traced back in the framework of ab initio MD simulations to the electronic structure properties of the particular transition metal.

*3.2.2. Metal-Water Interaction Energies in the Gas Phase.* One could think that the unexpected anionic orientation found in the Pt(II) case is originated as an artifact of the electronic structure method underlying the CP-MD simulations. In order to make sure that anionic hydration is not derived from the use of a DFT method with plane waves, a number of calculations for finite solvation complexes in the gas phase have been performed using PBE/*plane waves* (*p.w.*), PBE/*Gaussian functions* (*g.f.*), and MP2/*g.f.* The *g.f.* used were of the aug-cc-pVDZ type for O and H, whereas for Pd and Pt the pseudopotentials and optimized valence basis sets (MWB pseudopotentials and cc-pVDZ basis sets) of the Stuttgart/Köln group were applied. The so-called interaction energy between the metal tetrahydrate solvation core, [M(H$_2$O)$_4$]$^{2+}$, and an axial H$_2$O molecule approaching the metal center in anionic and ion-dipole orientations, i.e.

$$\Delta E(d) = E([M(H_2O)_4]^{2+} \bullet H_2O) - E([M[H_2O]_4]^{2+}) - E(H_2O) \quad (1)$$

was computed in vacuum as a function of the metal-oxygen or metal-hydrogen distances, i.e. $d = d_{M-O}$ and $d_{M\text{-}H}$, using the PBE and MP2 methods. The structures of the fragments [M(H$_2$O)$_4$]$^{2+}$ and H$_2$O were kept frozen along the reaction path d. The counterpoise (CP) correction[57] to the basis set superposition error (BSSE) was applied to the raw data obtained with atom-centered gaussians, whereas the plane wave based PBE/*p.w.* calculations do not suffer from BSSE.

Figure 7 shows that both methods, PBE and MP2, yield a strong preference of the ion-dipole orientation over the repulsive anionic arrangement for both metal cations in gas phase (note the difference in energy scales of the left and right panels). In particular, both methods predict for the ion-dipole orientation well defined minima at consistent positions and attractive interaction energies for both Pd(II) and Pt(II). For the Pd(II) cation in the anionic orientation, the three methods yield curves with no minima, albeit at long distances the agreement is not good, since the PBE/*p.w.* curve shows a stronger dissociative behavior than the PBE/*g.f.* and MP2/*g.f.* curves.

In the anionic solvation case of Pt(II), shown in the lower left panel of Figure 7, there is seemingly a qualitative disagreement between the MP2 and PBE behaviors: the two PBE curves yield a very shallow minimum of about half a kcal/mol only, whereas the MP2 curve is steeply repulsive upon close approach, flat at intermediate distances, and asymptotically dissociative without displaying any metastable state. However, when it comes to assessing MP2 interaction energies, it is well documented that the full (100%) coun-
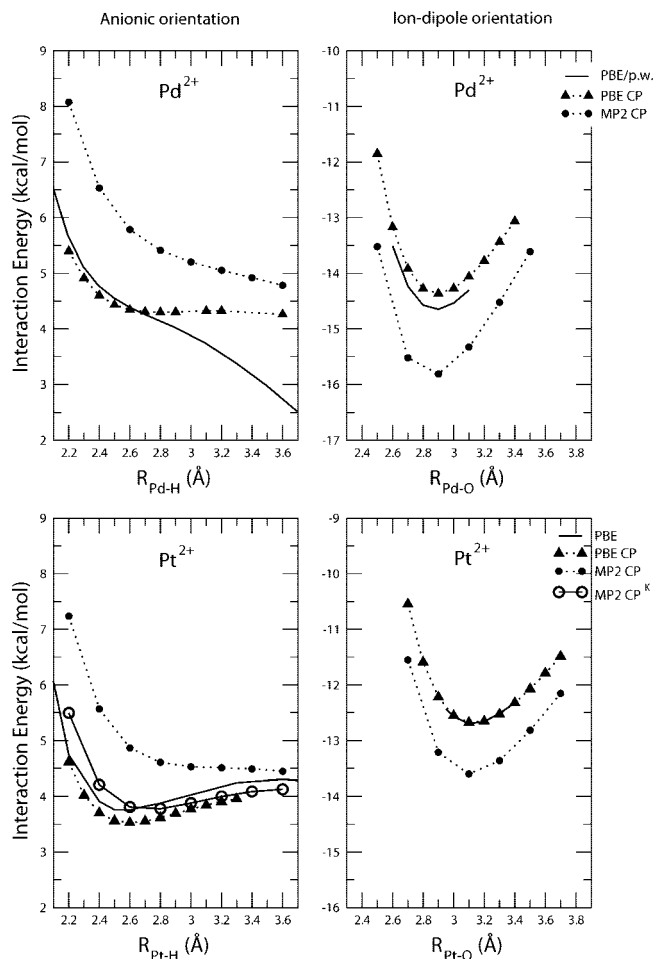
Figure 7. Interaction energies according to eq 1 between the Pd(II) (upper panels) and Pt(II) (lower panels) tetrahydrates and an additional water molecule in an axial position with anionic (left) and ion-dipole (right) orientations as a function of metal-hydrogen atom and metal-oxygen atom distances with respect to the axial H_2O molecule, respectively. The zero of energy corresponds to the isolated tetrahydrate and H_2O fragments. PBE/*g.f.* energies have been corrected for the BSSE using the (full, 100%) CP-correction,[57] whereas the PBE/*p.w.* calculations are free of BSSE. For MP2/*g.f.* energies two correction schemes have been applied: full (100%) CP-correction and, in the anionic coordination case of Pt, half (50%) CP-correction ("MP2/*g.f.* CP^K") according to ref 59.

terpoise correction overestimates the BSSE.[58] Therefore, we apply only half (50%) of the CP-correction as suggested by Kim and co-workers,[59] CP^K. It should be noted that the CP^K BSSE correction scheme is considered to be much more realiable for this type of interaction than the full CP-correction as can be judged based on previous experience on different systems, including hydration of halide anions.[59,60] The CP^K scheme yields a curve which does show a shallow minimum like the BSSE-free PBE data, whereas the full CP-correction leads to a purely repulsive interaction potential with respect to the axial water molecule in anionic orientation.

Considering all this evidence derived from PBE and MP2 calculations, the present analysis strongly suggests that the anionic orientation of the axial water molecule appears as a shallow metastable minimum for Pt(II) in the gas phase. This result is in agreement with an analogous analysis of the

interaction energy between the $[Pt(NH_3)_4]^{2+}$ complex in vacuum approached by a water molecule in axial position.[17,18] Although this minimum does not account for the hydrogen pattern observed in the CP-MD simulation, in the next subsection it will be demonstrated that bulk solvation effects finally stabilize this peculiar axial water molecule in aqueous Pt(II) solutions.

3.2.3. Ion-Dipole vs Anionic Orientation in Aqueous Solution and Ligand Exchange. Once the electronic structure method employed in the CP-MD simulations has been scrutinized for the quantum mechanical interaction energy, now we study how condensed phase effects affect the two types of relative orientations of axial water molecules. The energy profile of a water molecule approaching the Pt(II) tetrahydrate along the axial direction, which is perpendicular to the metal−oxygen molecular plane of the tetrahydrate, has been built taking as geometrical coordinate the metal-water distance along that direction. In each curve, the region close to the minimum has been scanned. The reference geometries selected for these calculations (Figure 8) were extracted from a representative snapshot for the two axial coordination modes in the Pt(II) CP-MD trajectory. Specific solvent effects have been included explicitly by considering the set of three water molecules that surround the axial water molecule in the reference snapshot. Two of these additional H_2O are directly bonded to first-shell water molecules and therefore are kept fixed along the energy profile. The third H_2O is displaced with the axial water molecule, partially representing its interaction with the water network of the bulk.

Average bulk solvent effects have been introduced by means of the polarizable continuum model (PCM)[25,27] where the solvent is represented by a dielectric continuum characterized by its dielectric permittivity, ε, thus including electrostatic components of the solvation in a mean-field manner. In the case studied here, the dielectric continuum surrounds the molecular complex formed by the Pt(II) tetrahydrate and the water molecule cluster. Then, formally a semicontinuum solvation model has been adopted. This approach allows, on one hand, the definition of a part of the solvation at a molecular level for a reduced but significant number of solvent molecules and, on the other hand, satisfies the condensed medium effects by including a dielectric continuum. It is worth mentioning that although the number of explicitly included water molecules is small, the use of cavities adapted to the molecular shape allows for inclusion of certain ingredients of specific solvation.[26] In this spirit, the energies obtained with the PCM method can be used to study qualitatively the hydration of the Pt(II) tetrahydrate along a defined region, such as the axial one, the information being derived from a source independent from the CP-MD simulations.

Figure 9 shows the total PBE/*g.f.* energy as a function of the Pt−O or Pt−H distances for the two types of relative orientation of the axial water molecule illustrated in Figure 8. The energy gap between the ion-dipole and anionic orientation minima in the microsolvated representation (see Figure 9a) is 22.2 kcal/mol. When this energy gap is computed using PBE/*p.w.* the resulting value is 22.8 kcal/
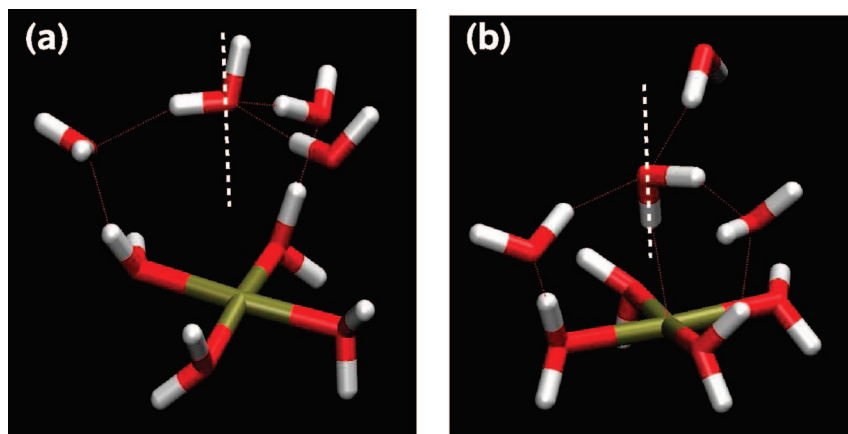
Figure 8. Structures for the water clusters, $(H_2O)[H_2O]_3$, approaching the Pt(II) tetrahydrate following a ion-dipole (left panel **a**) or anionic orientation, (right panel **b**).
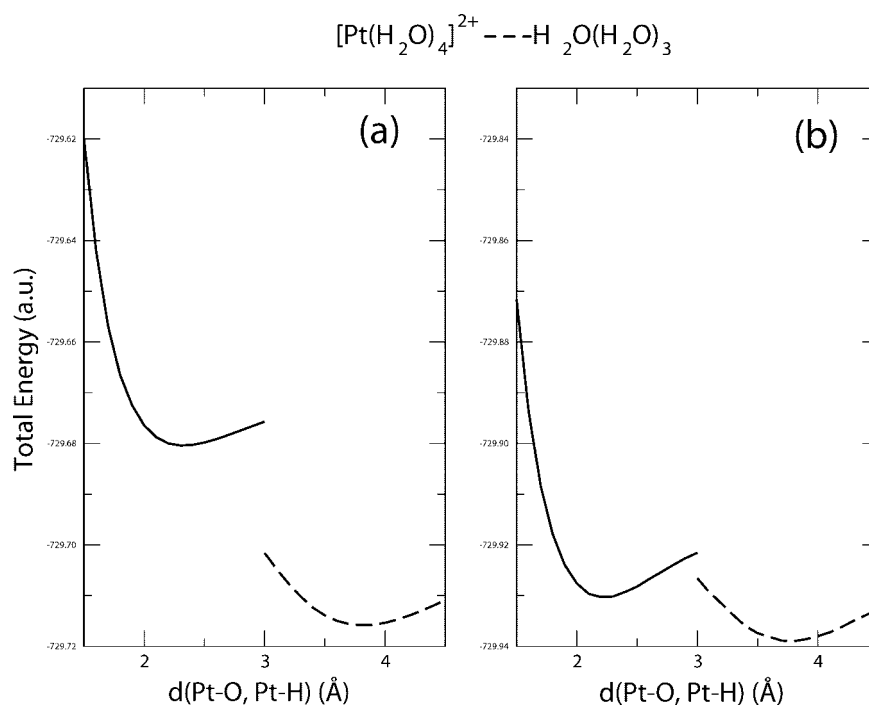
Figure 9. Interaction energies according to eq 1 between the Pt(II) tetrahydrates and an additional water molecule solvated with three water molecules in an axial position $(H_2O)[H_2O]_3$ with anionic (solid line) and ion-dipole (dashed line) orientations as a function of metal-hydrogen atom and metal-oxygen atom distances with respect to the axial H_2O molecule using the PBE/*g.f.* method. Gas phase data are reported in the left panel (a), whereas the liquid phase PCM continuum data are shown in the right panel (b).

mol. Then we can be confident that the analysis derived from these curves is not affected by different quantum mechanical ingredients than those of the CP-MD simulations. The minimum for the anionic orientation curve is defined at ca. 2.3 Å (Pt−H distance), whereas the minimum for the ion-dipole orientation is close to 3.8 Å (Pt−O distance). Interestingly, the interaction energy curve for the ion-dipole approach of one water molecule to the Pt(II) tetrahydrate (see Figure 7) shows a minimum to a much shorter Pt−O distance, ca. 3.1 Å. This indicates that the second-shell water molecules, which are strongly coordinated to the equatorial first shell, prevent a closer approach of the axial water molecule to the Pt cation. Bearing in mind that the axial partial RDFs for Pt−O show a peak at ca. 4.0 Å (see left panel of Figure 4) which corresponds to this ion-dipole

orientation, it may be concluded that the main responsible of this long distance is the strong water network formed by the first and second equatorial hydration shells. The inclusion of bulk solvent effects by means of the PCM solvation model does not change significantly the position of the minima but reduces the energy gap among the minima corresponding to the two water orientations. This means that average electrostatic solute−solvent interactions preferentially favor the anionic orientation versus the ion-dipole one.

Together with intricate hydrogen bonding effects of the axial water molecule in contact with the bulk solvent as the embedding medium, which cannot be reliably modeled by any continuum approach, the anionic coordination is finally stabilized as observed in the Pt(II) solution by ab initio MD using explicit water molecules. Thus the presence of such

dynamical axial water molecules is clearly a condensed phase effect which is not expected to be observed in the gas phase. When it comes to the meso-shell found in classical MD, a similar situation has been described:[7,8] the axial hydration structure appears only when the metal cation is immersed in the solution, whereas no minimum is found for the interaction energy in the axial region in gas phase. In the present work, however, this subtle condensed phase effect is clearly worked out by comparing dynamical ab initio MD in explicit water to static ab initio approaching path calculations in both the gas phase and in implicit continuum solvent. The view supplied by the consideration of specific interactions between the axial water molecule and a reduced representation of the water network shows a significant role in the definition of the final structural arrangement by restricting particularly the closest distance approach of the ion-dipole orientation.

Pd(II) and Pt(II) square-planar coordination complexes show very different rate constants for the ligand exchange process in water. In particular, this value is about four to six *orders of magnitude* larger for Pd(II) than for Pt(II) according to refs 2, 19, and 20. Given that the exchange mechanism is thought to be associatively activated, i.e. first the incoming ligand has to attach itself to the cation before a H_2O molecule in the first shell can leave this transient complex, this different behavior can be rationalized by the presence of anionic hydration in Pt(II) and its absence in the Pd(II) case. On the one hand, the axial H_2O molecule, in spite of its short persistence, represents a sterical hindrance for the incoming ligand, which can only attack through one of the two hemispheres above and below the square plane. On the other hand, the axial H_2O polarizes the d_{z^2} electronic density of the metal toward the hemisphere it is occupying, i.e. it stabilizes the solvated tetrahydrate and thus competes with the incoming ligand for the electronic density. Furthermore, when the incoming ligand is a water molecule, its preferred orientation toward the metal cation will be of the anionic type. Therefore, in order to achieve the proposed transition structure,[6,16] i.e. a trigonal-bipyramidal coordination around the metal cation, the axial H_2O has to reorient itself in order to solvate the metal properly in a cationic fashion. Since this water molecule, itself, is engaged in hydrogen bonding (see Figure 5) this implies a restructuring of its hydrogen bond network with respect to the bulk and thus an increase in the total energetic cost of the ligand exchange reaction. In the Pd(II) case the necessary rearrangement is much easier as the incoming H_2O molecule would initially coordinate through the oxygen atom, i.e. with the proper cationic orientation, implying that the exchange process is energetically less costly. Taken together, the peculiar anionic axial solvation pattern of Pt(II) versus Pd(II) would structurally, electronically, and energetically be consistent with a much less efficient ligand exchange for Pt(II) compared to Pd(II).

3.3. Transient Protolysis of Water Molecules. Based on the information extracted from the M-O RDFs, water molecules in the solution can be divided into three regions: first and second solvation shells and bulk. For each of these regions statistical distributions of O−H distances and

Table 4. Average, Most Probable and Standard Deviation of O-H Distances ($\langle d \rangle$, d_{max}, and σ_d all in Å) and H-O-H Angles ($\langle \theta \rangle$, θ_{max}, and σ_θ all in deg) for Water Molecules in the First Hydration Shell (H_2O) and in the Bulk (H_2O) of Aqueous Pd(II) and Pt(II) Solutions

		OH distance			HOH angle		
		$\langle d \rangle$	d_{max}	σ_d	$\langle \theta \rangle$	θ_{max}	σ_θ
Pd(II):	$(H_2O)_I$	1.029	1.015	0.049	107.8	107.1	6.6
	$(H_2O)_{Bulk}$	0.999	0.993	0.031	106.0	105.7	5.7
Pt(II):	$(H_2O)_I$	1.028	1.017	0.043	109.4	109.0	6.5
	$(H_2O)_{Bulk}$	0.996	0.992	0.029	105.7	106.4	5.8

H−O−H angles can be computed. A close examination of these parameters indicates that the structure of water molecules in the second solvation shell is already similar to that of H_2O in the bulk region so they are both considered together; details of the statistical distributions thus obtained can be found in Table 4.

Water molecules in the first solvation shell are more strongly distorted than bulk molecules. These distortions, whose origin can be assigned to electronic charge polarization due to coordination to metal cations, are very pronounced for the O−H distances. Elongations beyond 1.2 Å suggest the onset of transient proton transfer events taking place between H_2O molecules in the first and second solvation shells. As a result of this observation the trajectories of the Pd and Pt simulations were analyzed in more detail to extract deeper insights into this phenomenon. To detect any hydronium ions, H_3O^+, the number of hydrogen atoms, n_H, within a radius of 1.20 Å of a given oxygen atom was counted.[28,29] The average lifetime of a transient hydronium ion was measured as the average period of time over the entire trajectory during which $n_H = 3$ continuously. To complete the structural characterization of these H_3O^+ ions, also the degree of pyramidalization[61] was computed along the trajectory.

In the Pd case, 23 events consisting of first shell H_2O molecules transferring a proton to a second shell H_2O were found, whereas for Pt only 9 such events were observed. The pyramidalization degree of the H_3O^+ ion fluctuates between about 5 and 25° in both simulations. Effective protolysis of H_2O in the first hydration shell would imply translocation of the dissociated proton further away from the second solvation shell[30,37] and its subsequent diffusion through the bulk.[33] In all cases observed, however, the proton bonded back to the first-shell water molecule, i.e. the nascent hydronium did not detach from the solvation complex and thus did not undergo Grotthuss structural diffusion[28,31] so that no proton transport was observed.[31,32] Thus, we observed *spontaneous cation-induced transient protolysis* in these aqueous Pt(II) and Pd(II) solutions. Although these results clearly do not allow for a quantitative prediction of the pK_a values for Pd(II) and Pt(II) tetrahydrates, they do agree qualitatively with experimental measures of the acidic character of these two aqua ions the Pd(II) hydrate being slightly more acidic than the Pt(II) one (with measured pK_a values of 2.3 for Pd(II)[1] and 2.5 for Pt(II)[2]). It is mentioned in passing that no correlation between transient protolysis events and anionic axial hydration events was found.

What are the implications of this phenomenon on the modeling of such transition metal solutions with the aim to extend the length and time scales beyond what is accessible by ab initio MD? The average lifetime of H_3O^+ in the second solvation shell was 7.4 fs in the Pd solution and 4.9 fs in the Pt case. Thus, the tetrahydrated cations $[M(H_2O)_4]^{2+}$ transformed spontaneously into the $[M(H_2O)_3(OH)]^+$ aquo complexes during about 1.5% and 0.4% of the total simulation time for Pd and Pt, respectively. Water molecules directly coordinated to the metal cation suffer from polarization and partial electron transfer toward the metal and thus exhibit a stronger acid behavior than in the bulk. Thus, the tetrahydrate complexes experience a certain number of hydrolysis events, during which the metal bears a lower effective charge. In order to achieve such a fine description of the system, which might be necessary in order to deal with ligand exchange e.g. in cisplatin-like compounds, the solvated aqua ion normally represented by $[M(H_2O)_4]^{2+}$ should be considered as the more complex *dynamical entity* $[M(H_2O)_4]^{2+} \cdot (H_2O)_m \leftrightarrow [M(H_2O)_3(OH)]^+ \cdot [(H_3O)(H_2O)_{m-1}]^+$. This level of modeling might be relevant in order to build more reliable intermolecular potentials to be used in classical MD simulations of hydrates of transition metal cations, possibly following similar strategies as those employed previously for neat liquid water.[62−64]

4. Concluding Remarks and Outlook

The structure, energetics, and electronic structure of the hydration pattern of Pd(II) and Pt(II) cations in bulk water has been studied using Car-Parrinello ab initio MD at ambient conditions. In addition to the well-established general feature of a rather rigid square-planar first solvation shell, $[M(H_2O)_4]^{2+}$, two important special features of these solvated transition metal complexes have been analyzed in detail: "anionic solvation" in the axial regions and "transient protolysis" in the first solvation shell. In the case of the aqueous Pt(II) solution, an additional fifth water molecule could be identified in the *axial regions* being weakly coordinated to the metal cation via its *hydrogen atom*, which is the generic orientation of first-shell solvent water around *anions*. These features could only be unravelled using angle-resolved radial distribution functions; they cannot be resolved by standard spherically-averaged radial distribution functions. Electronic structure analyses suggest that this contact can be viewed as a weak hydrogen bonding interaction established between the d_{z^2} orbital of the metal and one hydrogen atom of the water molecule as a result of polarization effects. In contrast, axial water molecules at a second-shell distance from Pd^{2+} are characterized by a normal cationic orientation dictated by the dipole moment vector of H_2O. The qualitative difference in axial hydration between Pd(II) and Pt(II) may well be connected to the lower polarizability of the Pd(II) compared to the Pt(II) tetrahydrate along the *axial* direction. Detailed analyses uncover that the stabilization of a fifth water molecule in an axial position is a condensed phase effect that is absent in microsolvated complexes in the gas phase. Additionally, static quantum-chemical calculations of interaction energies as a function of distance between the square-planar tetrahydrates and a fifth water molecule using the Polarizable Continuum Model approach support the idea

that anionic hydration is indeed a direct consequence of the presence of the solvent. Axial solvation refines the previously proposed meso-shell concept[7,8] uncovered in classical MD simulations of these transition metals in solution, where the solvation is symmetric with respect to the square plane, better defined, and cationic for both species.

This peculiar axial solvation pattern also impacts the first-shell ligand exchange of square-planar Pd(II) and Pt(II) complexes in water, which is several orders of magnitude faster for Pd(II) compared to Pt(II). The accepted mechanism for these water exchange reactions is associative and includes a transition state with a trigonal-bipyramidal structure with all oxygens pointing toward the metal, which involves a relevant participation of the regions above and below the metal-oxygen plane before its formation. The main observable structural difference between Pd(II) and Pt(II) aqua ions is an anionic axial solvation in the case of Pt, whereas a less pronounced and cationic solvation is seen for Pd above and below the metal-oxygen plane. It is thus conceivable that the *anionic* arrangement in the case of Pt is unfavorable for associative ligand exchange via a 5-fold cationically coordinated intermediate, whereas the *cationic* orientation in Pd favors that process. This might have further ramifications for ligand exchange rates of novel cisplatin-based drugs, currently investigated within our collaboration.

Water molecules in the first hydration shell of Pd(II) and Pt(II) show also an enhanced "acidity" compared to H_2O in the bulk due to polarization and charge transfer effects with the metal cation favoring cation-induced transient protolysis in the first shell. The underlying proton transfer from first-shell water molecules into the second shell has been analyzed and found to be more pronounced for Pt in qualitative agreement with experimental pK_a differences. Although hydrolysis phenomena have been characterized previously for several *trivalent* cations, such as $Al(III)^{[34−37]}$ as well as Cr(III) and Fe(III),[37] these aqua ions are generally well coordinated, whereas in the case of divalent Pd(II) and Pt(II) only four H_2O in a very open square-planar coordination stabilize the +2 charge at the center. This seems to allow for protolysis in the neighborhood of these divalent cations, although to a much lesser extent compared to trivalent cations. To take into account this feature together with proper axial solvation is a challenge to future force field based modeling of Pt(II) metals in aqueous solution and, in particular, of cisplatin-based anticancer drugs.

References

(1) Nabivanets, B. I.; Kalabina, L. V. *Russ. J. Inorg. Chem.* **1970**, *15*, 818.

(2) Elding, L. I. *Inorg. Chim. Acta* **1976**, *20*, 65.

(3) Hellquist, B.; Bengtsson, L. A.; Holmberg, B.; Hedman, B.; Persson, I.; Elding, L. I. *Acta Chem. Scand.* **1991**, *45*, 449–455.

(4) Ayala, R.; Sánchez Marcos, E.; Díaz-Moreno, S.; Armando Solé, V.; Muñoz-Páez, A. *J. Phys. Chem. B* **2001**, *105*, 7588–7593.

(5) Purans, J.; Fourest, B.; Cannes, C.; Sladkov, V.; David, F.; Venault, L.; Lecomte, M. *J. Phys. Chem. B* **2005**, *109*, 11074–11082.

(6) Deeth, R. J.; Elding, L. I. *Inorg. Chem.* **1996**, *35*, 5019–5026.

(7) Martínez, J. M.; Torrico, F.; Pappalardo, R. R.; Sánchez Marcos, E. *J. Phys. Chem. B* **2004**, *108*, 15851–15855.

(8) Torrico, F.; Pappalardo, R. R.; Marcos, E. S.; Martínez, J. M. *Theor. Chem. Acc.* **2006**, *115*, 196–203.

(9) Shah, S.; Hofer, T. S.; Fatmi, M. Q.; Randolf, B. R.; Rode, B. M. *Chem. Phys. Lett.* **2006**, *426*, 301–305.

(10) Hofer, T. S.; Randolf, B. R.; Shah, S.; Rode, B. M.; Persson, I. *Chem. Phys. Lett.* **2007**, *445*, 193–197.

(11) Richens, D. T. *The Chemistry of Aqua Ions*; John Wiley and Sons: 1977.

(12) Frank, H. S.; Evans, M. W. *J. Chem. Phys.* **1945**, *13*, 507–532.

(13) Caminiti, R.; Carbone, M.; Sadun, C. *J. Mol. Liq.* **1998**, *75*, 149–158.

(14) Lopes, J. F.; de A. Menezes, V. S.; Duarte, H. A.; Rocha, W. R.; de Almeida, W. B.; Dos Santos, H. F. *J. Phys. Chem. B* **2006**, *110*, 12047–12054.

(15) Carloni, P.; Sprik, M.; Andreoni, W. *J. Phys. Chem. B* **2000**, *104*, 823–835.

(16) Song, T.; Hu, P. *J. Chem. Phys.* **2006**, *125*, 091101-1-3.

(17) Kozelka, J.; Bergès, J.; Attias, R.; Fraitag, J. *Angew. Chem., Int. Ed.* **2000**, *39*, 198–201.

(18) Bergès, J.; Caillet, J.; Langlet, J.; Kozelka, J. *Chem. Phys. Lett.* **2001**, *344*, 573–577.

(19) Helm, L.; Merbach, A. E. *Chem. Rev.* **2005**, *105*, 1923–1959.

(20) Rotzinger, F. P. *Chem. Rev.* **2005**, *105*, 2003–2037.

(21) Gill, D. S. In *Platinum Coordination Complexes in Cancer Chemotherapy*; Hacker et al. Eds.; Martinus Nijhoff Publishing: 1984; pp 267–278.

(22) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.

(23) Marx, D.; Hutter, J. Ab Initio Molecular Dynamics: Theory and ImplementationIn *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; NIC, FZ Jülich: 2000. See www.theochem.rub.de/go/cprev.html (accessed October 2008).

(24) Beret, E. C.; Pappalardo, R. R.; Doltsinis, N. L.; Marx, D.; Sánchez Marcos, E. *ChemPhysChem* **2008**, *9*, 237–240.

(25) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.

(26) Martínez, J. M.; Pappalardo, R. R.; Sánchez Marcos, E.; Mennucci, B.; Tomasi, J. *J. Phys. Chem. B* **2002**, *106*, 1118–1123.

(27) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3094.

(28) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **1995**, *103*, 150–161.

(29) Trout, B. L.; Parrinello, M. *J. Phys. Chem. B* **1999**, *103*, 7340–7345.

(30) Geissler, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. *Science* **2001**, *291*, 2121.

(31) Tuckerman, M. E.; Chandra, A.; Marx, D. *Acc. Chem. Res.* **2006**, *39*, 151–158.

(32) Chandra, A.; Tuckerman, M. E.; Marx, D. *Phys. Rev. Lett.* **2007**, *99*, 145901-1-4.

(33) Marx, D. *ChemPhysChem* **2006**, *7*, 1848–1870. See addendum: *ChemPhysChem* **2007**, *8*, 209-210.

(34) Sillanpää, A. J.; Päivärinta, J. T.; Hotokka, M. J.; Rosenholm, J. B.; Laasonen, K. E. *J. Phys. Chem. A* **2001**, *105*, 10111–10122.

(35) Ikeda, T.; Hirata, M.; Kimura, T. *J. Chem. Phys.* **2003**, *119*, 12386–12392.

(36) Swaddle, T. W.; Rosenqvist, J.; Yu, P.; Bylaska, E.; Phillips, B. L.; Casey, W. H. *Science* **2005**, *308*, 1450–1453.

(37) Coskuner, O.; Jarvis, E. A. A.; Allison, T. C. *Angew. Chem., Int. Ed* **2007**, *46*, 7853–7855.

(38) Hutter, J. et al. *CPMD V.3.11*; Copyright IBM Corp. 1990−2006, Copyright MPI für Festkörperforschung Stuttgart 1997−2001.

(39) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(40) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396.

(41) Vanderbilt, D. *Phys. Rev. B* **1985**, *32*, 8412–8415.

(42) Fernández-Serra, M. V.; Artacho, E. *J. Chem. Phys.* **2004**, *121*, 11136–11144.

(43) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B: Condens. Matter* **1996**, *54*, 1703–1710.

(44) Andrae, D.; Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123–141.

(45) 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.

(46) Dunning, T. H. J. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(47) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151.

(48) Cancès, M. T.; Menucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

(49) Mennucci, B.; Cances, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506.

(50) Tomasi, J.; Mennucci, B.; Cances, E. *J. Mol. Struct. (Theochem)* **1999**, *464*, 211.

(51) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

(52) Serr, A.; Netz, R. R. *Int. J. Quantum Chem.* **2006**, *106*, 2960–2974.

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

(54) Heinzinger, K.; Vogel, P. C. *Z. Naturforsch. A* **1974**, *29*, 1164–1171.

(55) Ohtaki, H.; Radnai, T. *Chem. Rev.* **1993**, *93*, 1157–1204.

(56) Marzari, N.; Vanderbilt, D. *Phys. Rev. B* **1997**, *56*, 12847.

(57) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(58) Rayon, V. M.; Sordo, J. A. *J. Phys. Chem. A* **1997**, *101*, 7414–7419.

(59) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. *Chem. Rev.* **2000**, *100*, 4145–4185.

(60) Singh, N. J.; Olleta, A. C.; Kumar, A.; Park, M.; Yi, H.; Bandyopadhyay, I.; Lee, H. M.; Tarakeshwar, P.; Kim, K. S. *Theor. Chem. Acc.* **2006**, *115*, 127–135.

(61) Haddon, R. C. *J. Am. Chem. Soc.* **1990**, *112*, 3385–3389.

(62) Halley, J. W.; Rustad, J. R.; Rahman, A. *J. Chem. Phys.* **1993**, *98*, 4110–4119.

(63) Mahadevan, T. S.; Garofalini, S. H. *J. Phys. Chem. B* **2007**, *111*, 8919–8927.

(64) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896–10913.

JCTC Journal of Chemical Theory and Computation

Influence of Structure on the Polarizability of Hydrated Methane Sulfonic Acid Clusters

Alisa Krishtal,[†] Patrick Senet,[‡] and Christian Van Alsenoy*,[†]

Chemistry Department, University of Antwerp, Universiteitsplein 1, B2610 Antwerp, Belgium, and Institut Carnot de Bourgogne, UMR 5209 CNRS, Université de Bourgogne, 9 Avenue Alain Savary BP 47870, F-21078 Dijon Cedex, France

Received July 24, 2008

Abstract: The relationship between polarizability and structure is investigated in methane sulfonic acid (MSA) and in 36 hydrated MSA clusters. The polarizabilities are calculated at B3LYP and MP2 level and further partitioned into molecular contributions using classic and iterative Hirshfeld methods. The differences in the two approaches for partitioning of polarizabilities are thoroughly analyzed. The polarizabilities of the molecules are found to be influenced in a systematic way by the hydrogen bond network in the clusters, proton transfer between MSA and water molecules, and weak interactions between water molecules and the methyl group of MSA.

1. Introduction

Methane sulfonic acid (MSA) has received some attention in the literature recently due to its presence in aerosol. The mystery of its formation has been the main topic of the research around this compound,[1−4] as well as its concentration,[6,8−13] but its role in the growth of particle within the aerosol[7,14] and its role as catalyst in oxidation processes[5] have also been investigated. Wang examined the ability of MSA to act as a center of nucleation with water molecules by studying the stability of several hydrated MSA clusters.[21] He found that for the smaller clusters the hydrogen bonds network between MSA and the surrounding water molecules are the main stabilizing factor, whereas for clusters containing three water molecules or more a proton transfer occurs from MSA to a neighboring water molecule, thus creating an ion−pair complex. As the aggregation number increases the probability for a proton transfer increases, as well as the corresponding stabilization.

One of the factors that can influence the interaction of those clusters with the environment is their polarizability because of the electrostatic polarization and the proportional-ity of the van der Waals interactions to the polarizabilities of the interacting molecular moieties. The global polarizability of the cluster as well as the individual polarizability of the MSA and the different water molecules in the cluster are of great interest. One can expect the polarizabilities to be dependent on the structure of the clusters in a systematic fashion. The number and strength of hydrogen bonds between the molecules, the eventual proton transfer and the steric hindrance are some of the factors that can influence the values of polarizability. A thorough analysis of such relation can lead to a deeper understanding of this property.

In a previous work, a method for partitioning of cluster polarizabilities into molecular contributions, based on the Hirshfeld scheme,[15,16] was introduced and applied to water clusters.[17] This method was further extended to the iterative Hirshfeld scheme, recently introduced by Bultinck et al.[19] The iterative Hirshfeld scheme brought a number of fundamental improvements to the classic Hirshfeld weight function and the resulting atomic charges, such as the elimination of the random nature of the weight function and the extension of applicability to charged systems. In this work, the effect of the iterative nature of the new scheme on the values of polarizability will be examined by comparing the values obtained by means of both schemes.

In previous works,[17,18] the polarizabilities of the clusters were studied only at DFT level because of the extensive size

* To whom correspondence should be addressed. E-mail: kris.vanalsenoy@ua.ac.be. Phone: +32 3 8202366. Fax: +32 3 8202310.
† University of Antwerp.
‡ Université de Bourgogne.

Hydrated Methane Sulfonic Acid Clusters

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2123**

of the systems. The quality of the values obtained at DFT level will be challenged here by comparison with MP2 method.

2. Method

The Hirshfeld method[15,16,19] allows us to write the *ij*-element of the polarizability tensor of a cluster as a sum of atomic contributions

$$\alpha_{ij} = \sum_A \alpha_{ij}^A + q_A^{(i)} j_A \tag{1}$$

where the *ij*-element of the intrinsic polarizability tensor of atom *A* and its first-order perturbed atomic charge in the Cartesian direction $j = x, y,$ or z are defined in eqs 2 and 3, respectively[17]

$$\alpha_{ij}^A = -\int i_A \omega_A(\mathbf{r}) \rho^{(j)}(\mathbf{r}) d\mathbf{r} \tag{2}$$

$$q_A^i = -\int \omega_A(\mathbf{r}) \rho^{(j)}(\mathbf{r}) d\mathbf{r} \tag{3}$$

The three first-order perturbed electronic densities $\rho^{(i)}$, $i = x, y, z$, can be obtained analytically by means of the CPHF method or numerically by means of the Finite Field method.

The atomic Hirshfeld weight function $\omega_A(r)$ can be defined in two ways:

• The classic Hirshfeld method,[15,16] where the weight function is constructed from the electronic densities of the spherical atoms

$$\omega_A(\mathbf{r}) = \frac{\rho_A^{(0)}(\mathbf{r})}{\sum_B \rho_B^{(0)}(\mathbf{r})} \tag{4}$$

• The iterative Hirshfeld method,[19] where the weight function is repeatedly constructed in each iteration from the atomic electronic densities obtained in the previous iteration, until conversion is achieved

$$\omega_A^n(\mathbf{r}) = \frac{\rho_A^{(n-1)}(\mathbf{r})}{\sum_B \rho_B^{(n-1)}(\mathbf{r})} \tag{5}$$

The electronic densities of the spherical atoms are used as an initial guess.

The intrinsic molecular polarizability of a molecule in a cluster is given by the sum of the intrinsic polarizabilities of the atoms in the molecule,

$$\alpha_{ij}^{int,mol} = \sum_{A(mol)} \alpha_{ij}^A \tag{6}$$

whereas the total molecular polarizability also includes the intramolecular charge delocalization contribution

$$\alpha_{ij}^{tot,mol} = \sum_{A(mol)} \alpha_{ij}^A + q_A^{(i)}(j_A - j_{mol}) \tag{7}$$

where j_{mol} represents the j coordinate of the geometrical center of the molecule.

Because the separate elements of the polarizability tensors are not rotationally invariable, only the isotropic values of the polarizabilities, given by the trace of the tensor, will be discussed in this work

$$\alpha = \frac{\alpha_{xx} + \alpha_{yy} + \alpha_{zz}}{3} \tag{8}$$

3. Computational Details

The polarizabilities of MSA and 36 hydrated MSA clusters $CH_3SO_3H \cdot (H_2O)_n$, $n = 1, 5$, were calculated at the DFT level and the B3LYP functional and at the MP2 frozen-core level, using the 6−311++G(2df,p) basis set. The polarizabilities were subsequently partitioned using the classic and iterative Hirshfeld methods. The structures of the clusters were previously optimized by L. Wang[21] using the B3LYP/6−311++G(2df,p) method. The methodology for obtaining and partitioning the polarizabilities at the two levels of theory is as follows. In the first step, the density matrices in the presence of an electric field applied at the positive and negative x, y, and z directions were obtained using the Gaussian03 program.[20] The density matrices were further used to obtain the first-order perturbed density matrices by means of the Finite Field Theory using the Brabo program.[22] At last, the Stock program[16] was used to obtain the atomic charges and the atomic and molecular polarizabilities (eqs 2, 6, and 7).

4. Results and Discussion

4.1. Nomenclature. In the following discussion, the polarizabilities of the clusters will be related to their structures, in particular, the hydrogen bond (HB) network between the different molecules. The molecules will be classified according their HB pattern in the following way:

• A molecule that donates a hydrogen atom into a HB will be referred to as making a D-type HB.

• A molecule that accepts a HB through an oxygen atoms will be referred to as making an A-type HB.

• The water molecules in the clusters can be labeled by the number and type of HBs they make as A, D, AD, AAD, ADD, and AADD. The protonated ion H_3O^+ can be labeled as DDD.

• MSA can be labeled as A, AD, AAD, AAAD, and AAAAD according to the number and type of HBs it makes. The deprotonated ion SO_3^- can be labeled as AAA and AAAA.

4.2. Cluster Polarizabilities. Table 1 lists the polarizabilities of the free MSA and the hydrated MSA clusters, calculated at the B3LYP and MP2 levels and the 6−311++G-(2df,p) basis set. While B3LYP overestimates the polarizabilities by a mean 2.54%, the qualitative trends are identical for both levels of theory. It can be seen that the values depend on the number of water molecules that surround the MSA. Figure 1 depicts the strong linear relation between the number of water molecules n and the mean polarizability of the clusters containing n water molecules, for both levels of theory. The addition of a water molecule to a cluster causes a mean increase of the cluster polarizability by 9.40 or 9.08 au according to B3LYP or MP2 method, respectively.

Examination of the values for clusters with identical number of water molecules n reveals that the polarizabilities are influenced by the structure of the clusters. For example, within the clusters containing one water molecule, two groups

Table 1. Calculated Polarizabilities of MSA and Hydrated MSA Clusters at B3LYP and MP2 Levels with 6-311++G(2df,p) Basis Set[a]

	B3LYP	MP2		B3LYP	MP2		B3LYP	MP2
MSA	43.059	42.551	II-9	61.340	59.917	IV-2	81.545	79.494
I-1	52.207	51.217	II-10	61.348	59.907	IV-3	82.449	80.421
I-2	52.215	51.212	III-1	72.281	70.680	IV-4	80.370	78.031
I-3	51.787	50.821	III-2	70.894	69.002	IV-5	80.197	77.871
I-4	51.808	50.851	III-3	71.541	69.709	IV-6	80.317	77.972
II-1	62.168	60.743	III-4	71.027	69.153	IV-7	80.388	78.198
II-2	62.078	60.665	III-5	70.726	68.877	IV-8	80.573	78.225
II-3	62.242	60.792	III-6	70.978	69.052	V-1	91.891	89.401
II-4	62.122	60.708	III-7	70.784	68.818	V-2	90.976	88.576
II-5	61.090	59.617	III-8	70.905	68.925	V-3	90.268	87.745
II-6	61.102	59.636	III-9	71.166	69.126	V-4	90.318	87.774
II-7	61.065	59.610	III-10	71.170	69.133			
II-8	61.093	59.639	IV-1	82.061	80.011			

[a] All values are in au.
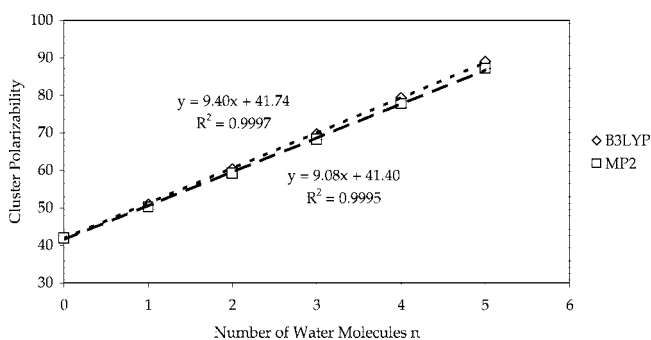
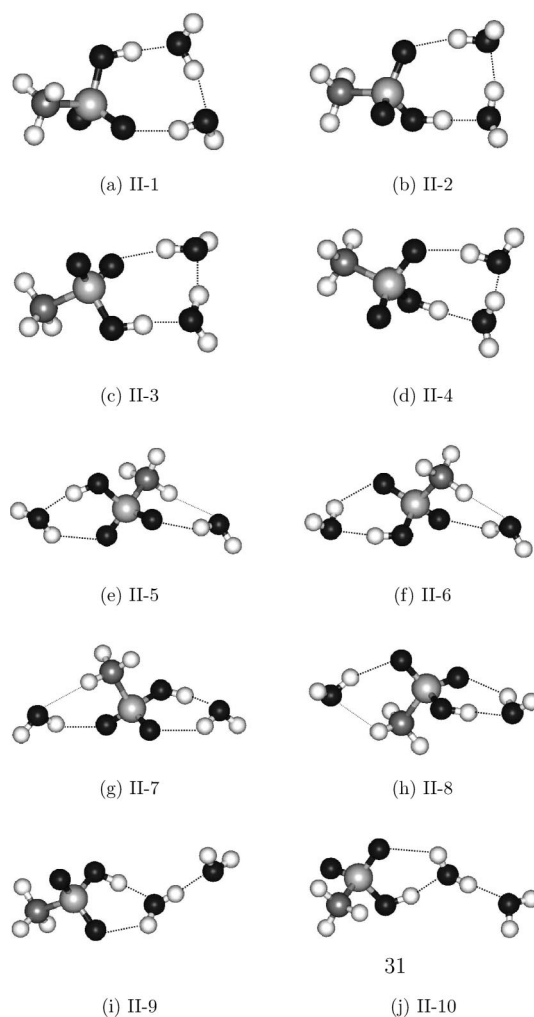***Figure 1.*** Mean polarizabilities of clusters containing identical number of water molecules n as function of n calculated at the B3LYP and MP2 levels with 6−311++G(2df,p) basis set.

(a) I-1   (b) I-2

(c) I-3   (d) I-4
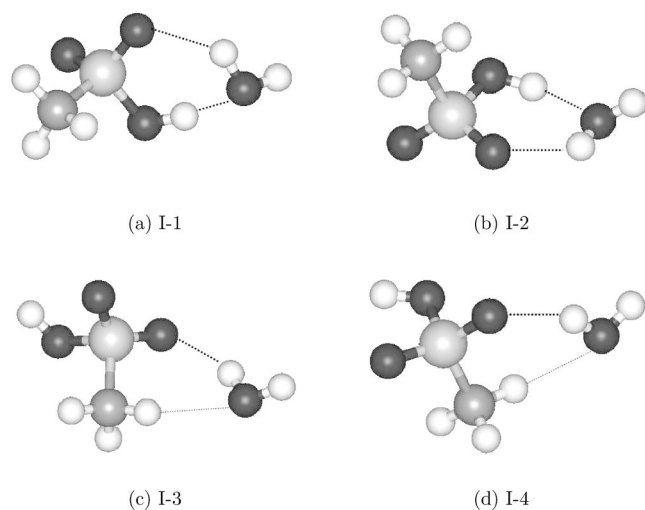
Figure 2. $CH_3SO_3H \cdot H_2O$ clusters.

(a) II-1   (b) II-2

(c) II-3   (d) II-4

(e) II-5   (f) II-6

(g) II-7   (h) II-8

(i) II-9   (j) II-10

Figure 3. $CH_3SO_3H \cdot (H_2O)_2$ clusters.

can be identified; clusters I-1 and I-2, with polarizability values around 52.2(51.2) au for the B3LYP(MP2) method, and clusters I-3 and I-4 with polarizability values around 51.8(50.8) au. The first two clusters contain an AD-type MSA and an AD-type water molecule, whereas the latter two clusters contain an A-type MSA and a D-type water molecule that additionally interacts through its oxygen atom with a hydrogen atom of the methyl group. A similar relation between structure and polarizability can be found in the clusters containing two water molecules, where the polar-

izability values can be separated into three groups. The first group contains the clusters II-1 to II-4, with values around 62.2(60.7) au, that consist of two AD-type water molecules and an AD-type MSA. The second group contains clusters II-5 to II-8 with values around 61.1(59.6), that consist of an AD-type water molecule, D-type water molecule that also interacts with the methyl group, and AAD-type MSA. The third group contains clusters II-9 and II-10, with values

Hydrated Methane Sulfonic Acid Clusters

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2125**

(a) III-1 (b) III-2 (c) III-3

(d) III-4 (e) III-5 (f) III-6

(g) III-7 (h) III-8 (i) III-9

(j) III-10
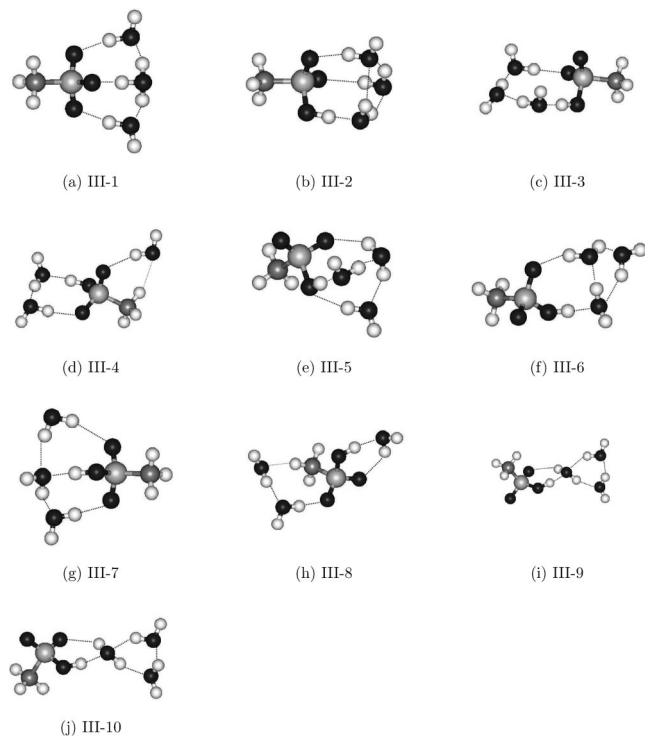
Figure 4. $CH_3SO_3H \cdot (H_2O)_3$ clusters.

around 61.3(59.9), that are formed out of an AAD-type water molecule, an A-type water molecule, and an AD-type sulfuric acid.

As the number of water molecules in the cluster increases, such straightforward connection between the polarizability values and structure becomes less evident, although one can observe that clusters that contain an ion−pair complex resulting from a proton transfer, such as III-1 and IV-1−IV-3, have slightly larger values than clusters with equivalent aggregation number without proton transfer.

4.3. Atomic and Molecular Charges. Table 2 lists the mean charges for the different entities in the cluster, calculated with the classic Hirshfeld and iterative Hirshfeld methods, both at the B3LYP and MP2 levels of theory. The values for the mean charges of the SO_3^- entity are similar in the classic and iterative Hirshfeld schemes, being around −0.46 au and −0.54 au for the B3LYP and MP2 methods, respectively. However, this is not the case for the values of the charge of the SO_3H entity, where the classic Hirshfeld method assigns a slightly negative charge of −0.19(−0.21) au for the B3LYP(MP2) method, whereas the iterative Hirshfeld assigns a positive charge of 0.11(0.17) au. This disagreement in the sign is also present in the values for the methyl group of the MSA: the classic Hirshfeld method assigns a positive charge of around 0.11(0.13) au to the methyl group, whereas the iterative Hirshfeld method assigns it a negative charge of −0.22(−0.21) au.

These differences can be explained by the charges of the atoms, which are essentially different in the two schemes. In the classic Hirshfeld scheme, the charges of the atoms are rather small, amounting to around 0.5 au for the S atoms, −0.3 au for the O atoms, −0.1 au for the C atoms, and 0.05 au for the H atoms of the methyl group. On the other hand, in the iterative Hirshfeld scheme the atomic charges are

(a) IV-1 (b) IV-2

(c) IV-3 (d) IV-4

(e) IV-5 (f) IV-6

(g) IV-7 (h) IV-8
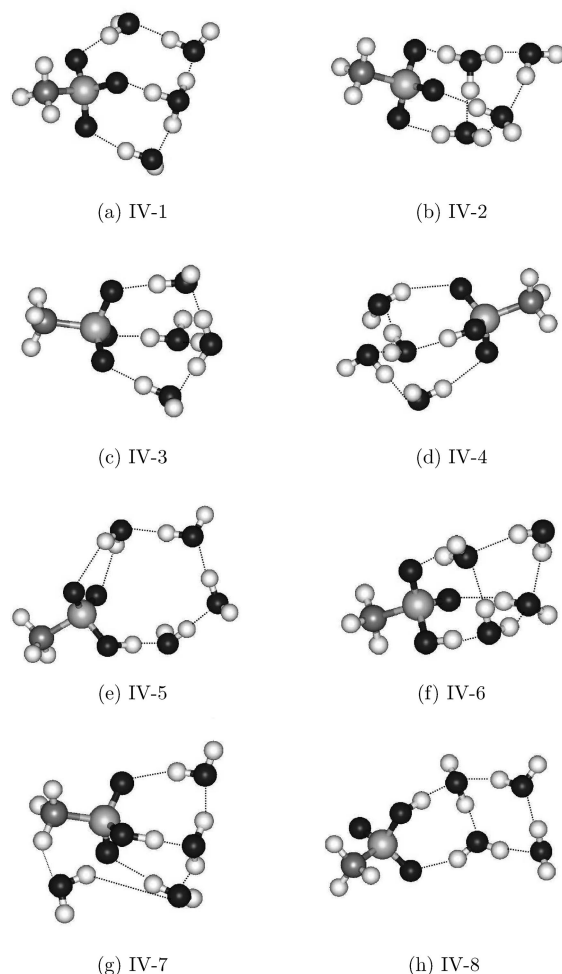
Figure 5. $CH_3SO_3H \bullet (H_2O)_4$ clusters.

(a) V-1 (b) V-2

(c) V-3 (d) V-4
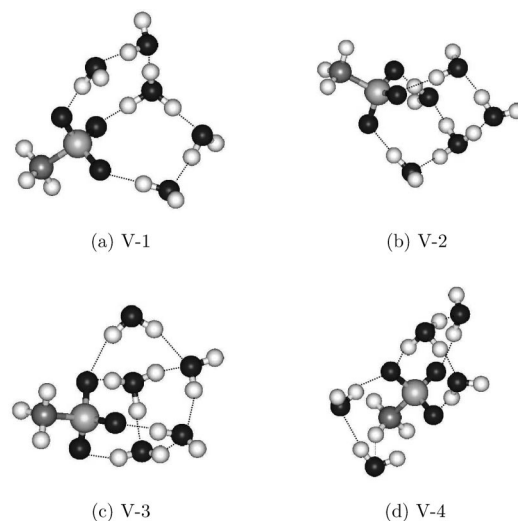
Figure 6. $CH_3SO_3H \cdot (H_2O)_5$ clusters.

considerably larger, amounting to around 2.2 au for the S atoms, −0.9 au for the O atoms, −0.9 au for the C atoms, and 0.2 au for the H atoms of the methyl group. As stated in the ref 19 the larger charges are typical for the iterative Hirshfeld scheme and may be seen as a consequence of the iterative nature of the scheme, where the boundaries of the atoms within the molecule are allowed to differ considerably from the free spherical atoms, thus being independent of the

Table 2. Mean Charges of the Different Entities in the Clusters Calculated at B3LYP and MP2 Levels with 6-311++G(2df,p) Basis Set and Partitioned with the Classic and Iterative Hirshfeld Schemes[a]

	B3LYP		MP2	
	classic	iterative	classic	iterative
SO_3^-	−0.47	−0.46	−0.54	−0.54
SO_3H	−0.19	0.17	−0.21	0.17
CH_3	0.11	−0.22	0.13	−0.21
A-H	0.12	0.05	0.11	0.03
D-H	−0.05	−0.01	−0.05	−0.01
AD-H	0.05	0.02	0.04	0.02
DD-H	−0.12	−0.04	−0.11	−0.02
AAD-H	0.07	0.02	0.06	0.01
ADD-H	0.02	0.03	0.02	0.02
AADD-H	0.06	0.03	0.05	0.01
DDD-H	0.26	0.63	0.31	0.72

[a] All values are in au.

initial choice of the promolecule. This effect is especially important for the atomic charge of the sulfur atom, which changes from 0.5 au to 2.2 au during the iterations, thus corresponding better with the high oxidation state of this atom in the molecule.

The values of the charges for the SO_3H and SO_3^- groups are not influenced by the number or type of HBs those entities form with the surrounding water molecules. However, the charges of the water molecules vary between positive and negative values according to the type of HBs they make. Generally, one can state that a D-type HB increases the electron density at the molecule, causing the D-type and DD-type of molecules to have a net negative charge, whereas an A-type HB decreases the electron density at the water molecule. Yet the relation between the charge of the water molecules and their type is not straightforward. Especially in the iterative Hirshfeld scheme, the net charges of the water molecules are smaller than in the classic Hirshfeld scheme, despite the larger values of atomic charges, and the charges are less dependent on the type of the molecules. For example, the ADD-type water molecule has a larger positive charge than the AAD-type water molecule in the iterative Hirshfeld scheme. This reduced charge transfer between molecules within a cluster implies that the molecules are better defined as separate entities within a cluster in the iterative Hirshfeld scheme.

4.4. Polarizability of MSA. MSA can be divided into two functional groups, the hydrophilic SO_3H or SO_3^- groups, that form HBs with the surrounding water molecules, and the hydrophobic methyl groups. Table 3 lists the mean intrinsic and total polarizabilities (eqs 6 and 7) of SO_3^- and SO_3H, calculated with the classical and iterative Hirshfeld schemes at the B3LYP and MP2 levels, as well as the correlation coefficient between the number of HBs and the polarizability values of the SO_3H entity. The polarizability values decrease with each additional HB, independent of its type.

The intrinsic polarizability values of SO_3H are strongly correlated with the number of HBs for both schemes at both levels of theory, as well as the total polarizability values for the classic Hirshfeld scheme at B3LYP and MP2 levels, having correlation coefficients above $R = 0.995$. However,

the correlation decreases to $R = 0.990$ for the total polarizabilities obtained with iterative scheme at both levels of theory. This can be attributed to the more drastic change in the definition of atoms in the cluster in the iterative scheme, resulting in a larger difference between the geometrical center of the group of atoms that is used as a reference point in eq 7 and the center of the electron density for those atoms. Another difference between the classic and iterative Hirshfeld scheme are the polarizability values of SO_3^-. In the classical scheme, the AAA- and AAAA-type SO_3^- have slightly lower values than the equivalent AAD- and AAAD-type SO_3H, while in the iterative Hirshfeld scheme the values for both entities are very similar for the same number of hydrogen bonds.

Although the values of the classic and iterative Hirshfeld schemes are quite similar, the contributions of the separate atoms are considerably different. For example, in the free MSA the sulfur atom has according to the classic Hirshfeld scheme an intrinsic polarizability (eq 2) of 3.7 au, while the oxygen atoms have an intrinsic polarizability of 3.2 au. The iterative Hirshfeld scheme assigns in the same molecule to the sulfur atom an intrinsic polarizability of 0.5 au and to the oxygen atoms an intrinsic polarizability of 4.1 au. This difference is in line with the differences in atomic charges between the two schemes: in the iterative Hirshfeld scheme, the sulfur atom has a significant positive charge and is therefore significantly less polarizable than the oxygen atoms.

Within the same type of SO_3H or SO_3^-, the values are also slightly influenced by the structure of the cluster. For example, Table 4 lists the polarizability values of the different AD-type SO_3Hs from different clusters. Those molecules can be classified into three different groups. The first group, that has the lowest values, contains the III-3 cluster, where SO_3H forms a ring with three water molecules. The second group, that has slightly larger values, contains the clusters II-1−II-4, III-6, and IV-8, where SO_3H forms a ring with two water molecules. Note that the fact that in the larger clusters those water molecules form also HBs with other water molecules outside the ring does not reflect on the polarizability values of the SO_3H. The third group, which has the largest values, contains the clusters I-1, I-2, II-9, II-10, III-9, and III-10, where the SO_3H forms a ring with only one water molecule. One can conclude that structures where the molecules form larger rings result in lower polarizabilities of the molecules involved. This can be explained by the weaker HBs that are formed in structures with smaller rings, because of ring strain, because it has been shown in previous studies[17,18] that weaker HBs lead to higher polarizability values.

The polarizabilities of the methyl groups in the clusters can be divided in two groups. The first group consists of the methyl groups that do not interact in any way with the water molecules and have values similar to the values of the methyl group in the free MSA. The second group, that has lower polarizability values, consists of the methyl groups that have one hydrogen atom interacting with an oxygen atom in a water molecule, thus forming a weak HB. Table 5 lists the mean values of the polarizabilites for both groups. An evident outlier in the second group is the methyl group in cluster III-8, that has polarizability values lower by 0.2 au than the

Hydrated Methane Sulfonic Acid Clusters

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2127**

Table 3. Mean Intrinsic and Total Polarizabilities of the SO$_3^-$ (AAA- and AAAA-types) and SO$_3$H (non-HB, A-, AD-, AAD-, and AAAD-types) Entities Calculated at B3LYP and MP2 Levels with 6-311++G(2df,p) Basis Set and Partitioned with the Classic Hirshfeld (H-C) and Iterative Hirshfeld (H-I) Schemes[a]

	intrinsic polarizability				total polarizability			
	B3LYP		MP2		B3LYP		MP2	
	H-C	H-I	H-C	H-I	H-C	H-I	H-C	H-I
				SO$_3^-$				
AAA	11.586	11.563	11.692	11.795	19.237	20.626	19.584	21.204
AAAA SO$_3$	10.819	10.919	10.948	11.169	18.136	20.174	18.515	20.866
				SO$_3$H				
non-HB	14.384	13.495	14.267	13.471	25.298	25.507	25.264	25.668
A	13.727	13.090	13.653	13.096	24.006	23.439	24.062	23.485
AD	13.083	12.257	13.038	12.310	22.581	22.848	22.700	23.118
AAD	12.278	11.706	12.279	11.798	21.217	21.364	21.415	21.653
AAAD	11.371	11.009	11.430	11.155	19.738	19.832	20.003	20.122
R	0.997	0.996	0.997	0.996	1.000	0.990	1.000	0.983

[a] *R* stands for the correlation coefficient between the values of SO$_3$H and the number of hydrogen bonds. All values are in au.

Table 4. Intrinsic and Total Polarizabilities of the AD-Type SO$_3$Hs in the Different Clusters Calculated at B3LYP and MP2 Levels with 6-311++G(2df,p) Basis Set and Partitioned with the Classic Hirshfeld (H-C) and Iterative Hirshfeld (H-I) Schemes[a]

	intrinsic polarizability				total polarizability			
	B3LYP		MP2		B3LYP		MP2	
	H-C	H-I	H-C	H-I	H-C	H-I	H-C	H-I
III-3	12.494	11.810	12.489	11.907	21.382	21.528	21.565	21.872
II-1	12.752	11.994	12.727	12.075	21.997	22.467	22.148	22.825
II-2	12.767	12.003	12.740	12.081	22.025	22.319	22.175	22.636
II-3	12.769	11.995	12.741	12.074	21.980	22.455	22.134	22.798
II-4	12.790	12.009	12.759	12.086	22.011	22.417	22.167	22.757
III-6	12.803	12.045	12.787	12.116	22.119	22.153	22.286	22.393
IV-8	12.807	12.042	12.789	12.116	22.115	22.133	22.280	22.379
I-1	13.386	12.499	13.315	12.530	23.285	23.560	23.353	23.795
I-2	13.423	12.541	13.350	12.568	23.339	23.672	23.404	23.915
III-9	13.443	12.560	13.381	12.582	23.258	23.221	23.343	23.384
III-10	13.444	12.552	13.382	12.575	23.269	23.523	23.354	23.730
II-9	13.597	12.638	13.512	12.650	23.409	23.676	23.467	23.888
II-10	13.603	12.657	13.523	12.674	23.367	23.894	23.429	24.162

[a] All values are in au.

Table 5. Mean Intrinsic and Total Polarizabilities of the HB CH$_3$ and non-HB CH$_3$ in the Different Clusters Calculated at B3LYP and MP2 levels with 6-311++G(2df,p) Basis Set and Partitioned with the Classic Hirshfeld (H-C) and Iterative Hirshfeld (H-I) Schemes[a]

	intrinsic polarizability				total polarizability			
	B3LYP		MP2		B3LYP		MP2	
	H-C	H-I	H-C	H-I	H-C	H-I	H-C	H-I
HB-CH$_3$	6.275	7.390	6.156	7.179	9.346	9.665	9.195	9.460
non-HB-CH$_3$	6.845	8.056	6.680	7.778	10.158	10.638	9.932	10.347

[a] All values are in au.

mean in Table 5. In this cluster, the methyl group forms a ring together with the MSA and two water molecules, as opposed to the rings with a single water molecule in the rest of clusters in the second group. Similar to the situation described above, the weak HB that the hydrogen atoms of the methyl group forms is shorter (2.21 au) than the mean HB length (2.46 au) between methyl groups and water molecules, resulting in lower polarizability values.

4.5. Polarizability of H$_2$O and H$_3$O$^+$. Table 6 lists the intrinsic and total polarizabilities of the different types of water molecules, calculated at the B3LYP and MP2 levels of theory and partitioned with the classic and iterative

Hirshfeld schemes. As a general trend the polarizabilities decrease with the number of HBs the molecule makes, although small differences can be distinguished between the different methods and schemes. The polarizability values of the DD-type water molecules are evident outliers because they consequently have the largest values, despite the fact that those molecules form two HBs. It must be noted that there are only three molecules of that type in the collection of clusters examined in this work, in clusters III-7, IV-6, and V-3. The reason for those high values will be discussed further below. The general order for the polarizability values of the different molecules calculated by the old Hirshfeld

Table 6. Mean Intrinsic and Total Polarizabilities of the Different Types of H_2O and the DDD-Type H_3O^+ Calculated at B3LYP and MP2 Levels with the 6-311++G(2df,p) Basis Set and Partitioned with the Classic Hirshfeld (H-C) and Iterative Hirshfeld (H-I) Schemes[a]

| | intrinsic polarizability | | | | total polarizability | | | |
| | B3LYP | | MP2 | | B3LYP | | MP2 | |
	H-C	H-I	H-C	H-I	H-C	H-I	H-C	H-I
				H_2O				
A	5.113	5.845	4.988	5.732	6.794	6.654	6.633	6.466
D	5.396	5.796	5.229	5.659	6.999	6.932	6.807	6.738
AD	4.784	5.332	4.682	5.268	6.149	6.219	6.034	6.103
DD	5.715	5.848	5.508	5.677	7.243	7.342	7.008	7.121
AAD	4.242	4.901	4.175	4.880	5.511	5.818	5.438	5.773
ADD	4.482	4.805	4.397	4.771	5.656	5.731	5.577	5.661
AADD	3.754	4.244	3.728	4.292	4.767	4.926	4.771	4.950
				H_3O^+				
DDD	4.060	3.905	3.941	3.846	5.064	4.756	4.957	4.676

[a] All values are in au.

scheme is D > A > AD > ADD > AAD > AADD, where the values decrease over a range of approximately 1.5 au or 30%. The order implies that A-type HBs have a larger influence on the polarizability than D-type HBs. This concept is not maintained in the intrinsic polarizabilities calculated by the iterative scheme: the D-type HBs have here larger influence because the values of A-type molecules are larger than the values of D-type molecules and the values of AAD-type molecules are larger than the values of ADD-type molecules. For the total polarizabilities obtained by the iterative Hirshfeld method the order is not altogether consequent, because the A-type molecules have larger values than the D-type molecules but the ADD-type moelcules have larger values than the AAD-type molecules. This lack of consequence can be explained by the drawback in the definition of total polarizabilities in the iterative scheme that has been mentioned above.

Table 7 lists the mean atomic polarizabilities of the oxygen, hydrogen-bonded hydrogen atoms (HB H), and non-hydrogen-bonded hydrogen atoms (non-HB H) in the different types of water molecules, calculated at the B3LYP and MP2 levels and partitioned with the classic and iterative Hirshfeld schemes. Although the values of the intrinsic polarizabilities of the molecules are quite similar for the classic and iterative Hirshfeld scheme, their distribution between the oxygen atom and the hydrogen atoms are very different. A similar situation was observed also for the SO_3H entity. The mean value encountered for the polarizabilities of oxygen atoms is 2.8 au in the classic Hirshfeld scheme and 4.3 au in the iterative Hirshfeld scheme. The HB Hs and non-HB Hs contribute 0.8 and 1.2 au in the classic scheme, whereas in the iterative scheme their values decrease to 0.3 and 0.6 au.

The DD-type water molecules appear to have polarizability values of oxygen atoms that are significantly higher than those in the rest of the water molecules, explaining the outlying values of intrinsic and total polarizabilities of those molecules in Table 6. This effect can be correlated to their somewhat remote position in the clusters, where the oxygen atoms experience less steric hindrance than the rest of water

Table 7. Mean Polarizabilities of the Oxygen Atoms, Hydrogen-Bonded Hydrogen Atoms (HB H), and Non-Hydrogen-Bonded Hydrogen Atoms (non-HB H) in the Different Types of Water Molecules Calculated at B3LYP and MP2 levels with 6-311++G(2df,p) Basis Set and Partitioned with the Classic Hirshfeld and Iterative Hirshfeld Schemes[a]

| | B3LYP/classic Hirshfeld | | | B3LYP/iterative Hirshfeld | | |
	O	HB H	non-HB H	O	HB H	non-HB H
A	2.805		1.154	4.582		0.631
D	3.149	0.929	1.319	4.704	0.404	0.688
AD	2.768	0.794	1.222	4.352	0.327	0.653
DD	3.542	1.087		4.952	0.448	
AAD	2.406	0.650	1.186	3.953	0.284	0.665
ADD	2.764	0.859		4.075	0.365	
AADD	2.292	0.731		3.564	0.340	

| | MP2/classic Hirshfeld | | | MP2/iterative Hirshfeld | | |
	O	HB H	non-HB H	O	HB H	non-HB H
A	2.778		1.105	4.575		0.578
D	3.096	0.885	1.248	4.673	0.361	0.625
AD	2.748	0.765	1.168	4.391	0.287	0.590
DD	3.455	1.027		4.873	0.402	
AAD	2.406	0.631	1.138	4.039	0.246	0.595
ADD	2.744	0.827		4.122	0.325	
AADD	2.302	0.713		3.688	0.302	

[a] All values are in au.

molecules in the cluster. As was shown in a previous study on methanol clusters,[18] steric hindrance tends to decrease the values of polarizabilities. Another possible effect is the difference in the local field those molecules experience compared to other water molecules in the clusters.

Examining the values in Table 7 closely allows us to recognize a certain cooperativity effect between the atoms. For example, the polarizability values of the oxygen atoms that make one A-type HB decrease with the number of HBs the hydrogen atoms make in the same molecule: A > AD > ADD. A similar effect is present in the values of the oxygen atoms that make two A-type HBs: AA > AAD > AADD. Note that this effect is larger in the iterative Hirshfeld scheme. In addition, the values for the HB Hs decrease in the order of D > AD > AAD and DD > ADD > AADD.

5. Conclusions

The polarizabilities of MSA and hydrated MSA clusters with up to five water molecules were calculated using the B3LYP and MP2 methods. The polarizabilities and the charges of the different entities in the clusters were partitioned using both classic and iterative Hirshfeld schemes. The quality of the values obtained with the B3LYP level in this study is overall high, with global polarizabilities varying by less than 2% from the MP2 values and the trends in the partitioned polarizability values and charges being very similar at both levels of theory.

The atomic and molecular charges obtained using classic and iterative Hirshfeld schemes tend to disagree. While according to classic Hirshfeld scheme the SO_3H group has a negative charge and the methyl group has a positive charge, the iterative Hirshfeld scheme claims the opposite. The atomic charges in the iterative Hirshfeld scheme are signifi-

Hydrated Methane Sulfonic Acid Clusters

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2129**

cantly larger, especially for the sulfur atom, where the classic Hirshfeld scheme assigns a charge of 0.2 au, whereas the iterative Hirshfeld scheme assigns a charge above 2 au, thus corresponding better with the high oxidation state of the sulfur atom in MSA.

The large differences in atomic charges results also in large differences in the atomic polarizability values, where for example in the iterative scheme the sulfur atom has a significantly smaller contribution to the polarizability of MSA because of its high charge. Nevertheless, the values for the molecular polarizabilities obtained by eqs 6 and 7 in the both schemes are mostly similar. Both schemes reveal a strong relation between the number, type, and strength of hydrogen bonds and the polarizabilities of the different entities in the cluster. A lack of consistency in the definition of the total molecular polarizability was detected for the iterative Hirshfeld scheme that most likely is connected to a larger shift of the center of the electron density of the molecule from its geometrical center, which is being used as a reference point for the intramolecular charge delocalization contribution.

References

(1) Asatryan, R.; Bozzelli, J. W. *Phys. Chem. Chem. Phys.* **2008**, *10*, 1769.

(2) Campolongo, F.; Saltelli, A.; Jensen, N. R.; Wilson, J.; Hjorth, J. *J. Atmos. Chem.* **1999**, *32*, 327.

(3) von Glasow, R.; Crutzen, P. J. *Atmos. Chem. Phys.* **2004**, *4*, 589.

(4) Barnies, I.; Hjorth, J.; Mihalopoulos, N. *Chem. Rev.* **2006**, *106*, 940.

(5) Key, J. M.; Paulk, N.; Johansen, A. M. *Environ. Sci. Technol.* **2008**, *42*, 133.

(6) Morganti, A.; Becagli, S.; Castellano, E.; Severi, M.; Traversi, R.; Udisti, R. *Anal. Chim. Acta* **2007**, *603*, 190.

(7) Hara, K.; Osada, K.; Kido, M.; Matsunaga, K.; Iwasaka, Y.; Hashida, G.; Yamanouchi, T. *Tellus B* **2005**, *57*, 230.

(8) Arimoto, R.; Hogan, A.; Grube, P.; Davis, D.; Webb, J.; Schloesslin, C.; Sage, S.; Raccah, F. *Atmos. Environ.* **2004**, *38*, 5473.

(9) Easter, R. C.; Ghan, S. J.; Zhang, Y.; Saylor, R. D.; Chapman, E. G.; Laulainen, N. S.; Abdul-Razzak, H.; Leung, L. R.; Bian, X. D.; Zaveri, R. A. *J. Geophys. Res., [Atmos.]* **2004**, *109*, D20210.

(10) Gondwe, M.; Krol, M.; Klaassen, W.; Gieskes, W.; de Baar, H. *Global Biogeochem. Cycles* **2004**, *18*, GB2006.

(11) Ooki, A.; Miura, K.; Uematsu, M. *J. Oceanogr.* **2003**, *59*, 799.

(12) Rankin, A. M.; Wolff, E. W. *J. Geophys. Res. [Atmos.]* **2003**, *208*, 4775.

(13) Berresheim, H.; Elste, T.; Tremmel, H. G.; Allen, A. G.; Hansson, H. C.; Rosman, K.; Dal Maso, M.; Makela, J. M.; Kulmala, M.; O'Dowd, C. D. *J. Geophys. Res. [Atmos.]* **2002**, *107*, 8100.

(14) Bigg, E. K.; Leck, C. *J. Geophys Res. [Atmos.]s* **2001**, *106*, 32101.

(15) Hirshfeld, F. L. *Theoret. Chim. Acta (Berlin)* **1977**, *44*, 129.

(16) Rousseau, B.; Peeters, A.; Van Alsenoy, C. *Chem. Phys. Lett.* **2000**, *324*, 189.

(17) Krishtal, A.; Senet, P.; Mingli, Y.; Van Alsenoy, C. *J. Chem. Phys.* **2006**, *125*, 034312.

(18) Krishtal, A.; Senet, P.; Van Alsenoy, C. *J. Chem. Theory Comput.* **2008**, *4*, 426.

(19) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbo-Dorca, R. *J. Chem. Phys.* **2007**, *126*, 144111.

(20) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.05; Gaussian, Inc.: Pittsburgh PA, 2003.

(21) Wang, L. *J. Chem. Phys. A* **2007**, *111*, 3642.

(22) Van Alsenoy, C.; Peeters, A. *J. Mol. Struct (THEOCHEM)* **1993**, *286*, 19.

JCTC Journal of Chemical Theory and Computation

Accurate First-Principle Prediction of ^{29}Si and ^{17}O NMR Parameters in SiO$_2$ Polymorphs: The Cases of Zeolites Sigma-2 and Ferrierite

Alfonso Pedone,[†] Michele Pavone,[‡,§] Maria Cristina Menziani,[†] and
Vincenzo Barone*,[‡,§,ll]

*Dipartimento di Chimica, Università di Modena e Reggio Emilia, Via G. Campi 183,
41100 Modena, Italy, Dipartimento di Chimica "Paolo Corradini" and CR-INSTM
"Village", Università di Napoli Federico II, Complesso Universitario Monte
Sant'Angelo Via Cintia I, 80126, Napoli, Italy, and Istituto per i Processi Chimico
Fisici CNR and CR-INSTM "Village", Area della Ricerca del CNR Via Moruzzi 1,
56124 Pisa, Italy*

Abstract: The magnetic shielding tensors of silica polymorphs have been investigated by means of quantum chemical calculations. Several levels of theory, from Hartree−Fock to the last generation of Density Functional Theory based approaches, have been tested on predicting ^{29}Si and ^{17}O isotropic and principal components of the chemical shift tensors together with ^{17}O quadrupolar coupling constants. The NMR parameters have been computed on all known silica systems, namely, α-quartz, α-cristobalite, coesite, Sigma-2, and ferrierite zeolites. Besides, cluster based approaches have been compared to a hybrid Quantum-Mechanics/Molecular-Mechanics (QM/MM) method, within the ONIOM scheme. The convergence of computed ^{17}O NMR parameters with respect to cluster size is found to be system-dependent. Excellent agreement between computed and experimental data has been found for ^{29}Si NMR parameters of the different Si sites of silica polymorphs and of Sigma-2 and ferrierite zeolites.

Introduction

Silicates are the major constituents of a large range of materials, from zeolites to clays, from minerals to glasses. These materials present very complex structural features: an accurate definition of their structure−property relationships is not always straightforward, although it is extremely important in both fields of technological industries and geosciences. For silicates, X-ray diffraction (XRD) and solid-state nuclear magnetic resonance (NMR) techniques provide the experimental framework to define the atomistic structural details of crystals (or powders) and the local chemical environment embedding specific sites, respectively. However, in some cases, when there are no chances to obtain single

crystals or the different magnetically active sites present very similar chemical surroundings, the rationalization of experimental data could strongly benefit from theoretical modeling techniques able to predict reliable structural properties and accurate NMR parameters. This last issue is the focus of the present contribution.

The dominant constituents of silicates are silicon and oxygen atoms, both the elements present isotopes, ^{29}Si and ^{17}O, amenable of NMR studies. The ^{29}Si atom (nuclear spin equal to 1/2) has a relatively high natural abundance (4.7%), which allowed the accumulation of several ^{29}Si NMR data over the years for natural minerals, glasses, and zeolites.[1,2] Some empirical correlations between ^{29}Si chemical shifts and a number of structural parameters have been proposed: in particular, the coordination of Si atoms, the connectivity of SiO$_4$ tetrahedra, the Si−O bond length, and the Si−O−Si angle.[3−10] The ^{17}O atom has a nuclear spin equal to 5/2, and it is possible to measure the chemical shifts and

* Corresponding author e-mail: baronev@unina.it.
† Università di Modena e Reggio Emilia.
‡ Università di Napoli Federico II.
§ CR-INSTM "Village".
ll IPCF-CNR.

^{29}Si and ^{17}O NMR Parameters in SiO$_2$ Polymorphs

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2131**

quadrupolar parameters. However, ^{17}O NMR technique is challenging for the low natural abundance (0.037%) of this oxygen isotope. Besides, being ^{17}O a quadrupolar nucleus, NMR peaks are broadened by electric quadrupolar interactions between the nuclear electric quadrupole and the electric field gradient (efg) at the nucleus. Thanks to enriched samples and to the development of high-resolution techniques, the amount of information from ^{17}O NMR spectra of silicates is rapidly increasing.[11–15] Nevertheless, the spectra are still very difficult to interpret, also because empirical correlations have not yet been proposed for enabling the assignment of chemical shifts to each site. Indeed, recent *ab initio* calculations by Xue et al.[16,17] on small silicate clusters showed that [17]O chemical shift is strongly affected by the size of the rings constituted by the Si–O–Si silicate network and by the puckering of the cluster where it resides. Moreover, it has been recently found that [17]O quadrupolar coupling constant (C_Q - *vide infra*) is strongly dependent on Si–O distance and Si–O–Si angle values, while the asymmetry parameter (η - *vide infra*) is dependent on Si–O–Si angle but not on Si–O distance.[18]

Under such circumstances, it is not hard to highlight how relevant is the accuracy of NMR *ab initio* calculations for allowing the assignment of each site in complex ^{29}Si- and ^{17}O NMR spectra of SiO$_2$-polymorphic systems. Most of the previous reports on this topic were restricted to calculation of isotropic shifts by using the Hartree–Fock (HF) level of theory on cluster models of silica polymorphs,[19] possible zeolite precursors,[20] and a number of different zeolites.[14,15,21] Brouwer et al.[22] first demonstrated that the principal components of zeolite [29]Si chemical shift (CS) tensors can be accurately calculated and used together with ultrahigh-field NMR experiments into an NMR crystallography structure refinement tool for zeolites.[23] The framework structure of Sigma-2 zeolite was solved from solid-state ^{29}Si NMR data and subsequently refined against the computed ^{29}Si CS tensors to give a NMR determined crystal structure that was in a very good agreement with the single-crystal XRD structure. The cluster model has also been used with success to calculate ^{1}H and ^{29}Si NMR chemical shifts of silane and silanol groups in silica[24,25] by employing the Density Functional Theory method with the hybrid B3LYP functional.[26,27] Moreover, periodic density functional theory (DFT-PBE) calculations on a number of silica polymorphs, including ferrierite and Faujasite zeolites, performed well[28] in the determination of the [29]Si and [17]O isotropic shifts and [17]O quadrupolar parameters. However, to our knowledge there is still the lack in recent literature of a systematic investigation able to verify the performance of different *ab initio* methods for computing ^{29}Si- and ^{17}O NMR parameters of silicates.

In the present paper, the accuracy of HF- and several DFT-based methods (PBE,[29,30] PBE0,[31] B3LYP,[26,27] CAM-B3LYP,[32] and M05–2X[33] - *vide infra*) for predicting magnetic parameters of α-quartz,[34] α-cristobalite,[35] coesite,[36] Sigma-2,[22] and ferrierite[37] all-silica zeolites has been tested. Besides, the effectiveness of a multiscale method, the ONIOM scheme,[38–40] which combines Quantum-Mechanics and Molecular Mechanics approaches (QM/MM) in order
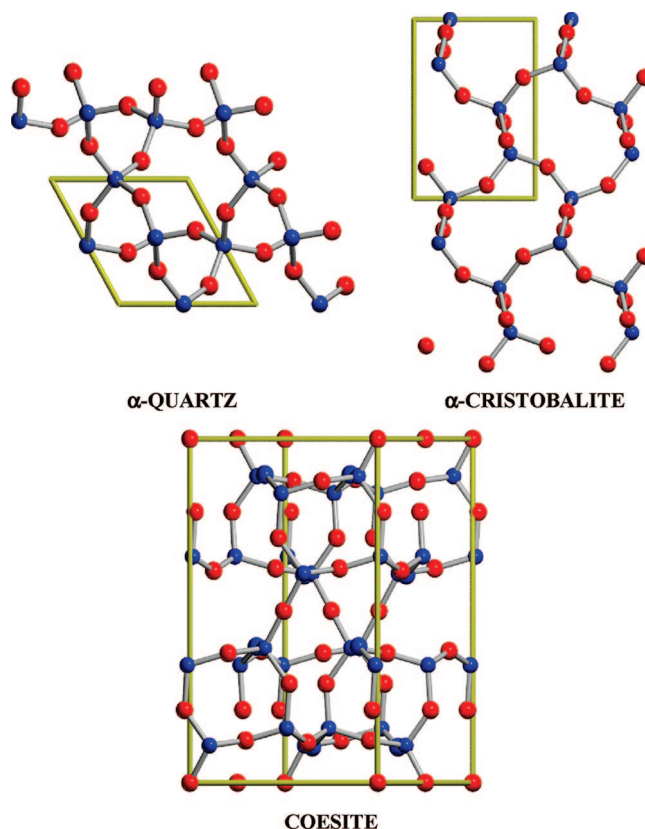

Figure 1. Unit cells of α-quartz,[34] α-cristobalite,[35] and coesite.[36] Light blue spheres represent Si and light red spheres represent O. Manipulation and visualization of structures have been dealt with the MOLDRAW program,[24] and molecular drawings have been rendered by the POVRAY program using input files prepared with MOLDRAW.

to account for long-range Coulombic interactions and to achieve convergence of computed properties with respect to cluster sizes has been investigated.

All the investigated silica polymorphs consist of three-dimensional SiO$_4$ tetrahedral networks but have different tetrahedral connectivity. In the structure of α-quartz and α-cristobalite, the SiO$_4$ tetrahedra form six-membered rings; each contains only a single Si and a single O site. In the structure of coesite, there are two Si and five O sites: O1 and O2 are part of large six- or eight-membered rings that cross-link the smaller four-membered rings made by O3, O4, and O5. Figure 1 reports the structure of α-quartz, α-cristobalite, and coesite. Sigma-2 and ferrierite are two well-known zeolites whose structures are reported in Figure 2.

Sigma-2 is known to be a clathrasil rather than a zeolite, i.e., it has cages but no channel systems. There are four Si and seven O atoms in the asymmetric unit, which are made up of two cages. The large cage is roughly spherical with a free diameter of 7.5 Å and $\bar{4}$ point symmetry. It can be visualized as a tennis ball with twelve five-rings forming the seam and rows of four six-rings filling in the spaces. There are four large cages and eight small ones per unit cell. The small cage consists of three four-rings and six five-rings. Although the structure of Sigma-2 is interesting in its own right, it is used here as a test case because both the structure and the [29]Si chemical shift tensor have been recently resolved with good accuracy.[22,23]
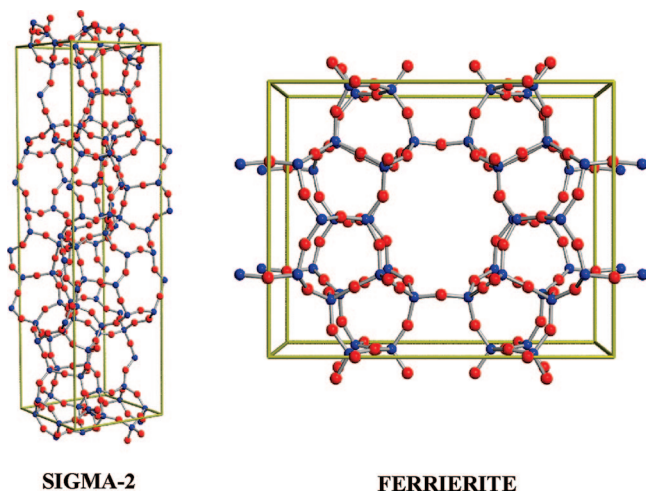
Figure 2. Unit cells of Sigma-2[22] and ferrierite[37] all-silica zeolites.

Ferrierite (FER)[41] is known to be a natural as well as a synthetic material with a framework structure of corner-sharing tetrahedral $TO_{4/2}$ units (T=Si[IV], Al[III]) that give a fully condensed 3D framework that contains a system of intersecting channels, circumscribed by 8 T atoms and 10 T atoms.[37,42,43] The aluminum-containing phase has shown potential as a deNOx catalyst for car exhaust systems[44] and is an excellent shape-selective catalyst for the isometization of n-butenes to isobutenes.[45] The latter is an important feedstock for the production of methyl tert butyl ether (MTBE), which is a commercial oxygenate additive in motor fuel.

Methods and Computational Details

Experimental NMR chemical shifts are related to the magnetic shielding tensors, which account for the local electronic environment surrounding the NMR-active nuclei. The nuclear magnetic shielding tensor is expressed as a mixed second derivative of the energy with respect to the external magnetic field, **B**, and the magnetic moment of nucleus X, $\boldsymbol{\mu}^X$

$$\sigma_{ij}^X = \left[\frac{\partial^2 E}{\partial \mu_i^X \partial B_j} \right] \quad (1)$$

where i and j are the components of the nuclear induced magnetic moment and the external magnetic field, respectively. The isotropic shielding constant (σ) is defined as one-third of the trace of the shielding tensor; the chemical shift δ is computed as

$$\delta_{sample} = \delta_{ref} + \sigma_{ref} - \sigma_{sample} \quad (2)$$

where δ_{sample} and σ_{sample} refer to the compound under study, δ_{ref} is the experimental chemical shift for a suitable reference system, and σ_{ref} is the corresponding computed isotropic shielding.

The calculations of [29]Si and [17]O shielding tensors were performed with the Gaussian code[46] via the gauge-including atomic orbital (GIAO) approach,[47] which uses basis functions that have and explicit field dependence, as first proposed by Ditchfield.[48] The electronic degrees of freedom were de-

scribed at different levels of theory: from Hartree−Fock (HF) to Density Functional Theory (DFT). In particular, we exploited the PBE, [29,30] PBE0,[31] B3LYP,[26,27] CAM-B3LYP,[32] and M05−2X[33] exchange-correlation density functional models. The 6−311+G(2df,p) basis set was adopted: this triple-ζ basis set augmented with diffuse and polarization functions has been extensively validated in recent literature.[16,17,19,22]

The systems under investigation were several silica polymorphs: α-quartz, α-cristobalite, coesite, Sigma-2, and ferrierite all-silica zeolites. For the purpose of comparing the aforementioned levels of theory, the ^{29}Si and ^{17}O NMR parameters were computed on experimental geometries, thus avoiding subtle indirect effects related to structural relaxation. In order to establish the most accurate and feasible computational procedure, two alternative approaches have been applied to compute the [29]Si and [17]O shielding constants of α-quartz and α-cristobalite, namely, the *cluster* approach and the hybrid Quantum-Mechanics/Molecular-Mechanics (QM/MM) method, within the ONIOM scheme.[38−40]

The *cluster* approach consists of extracting finite-size models from a given crystal structure: specific clusters are constructed for each crystallographically inequivalent [29]Si and [17]O sites, with the atom of interest at the center of the structural model made by its coordination shells. Obviously, the computed molecular properties depend to some extent on the size of the cluster model; therefore, the convergence of NMR shielding constants against the number of coordination shells was tested: Si- and O- centered clusters were built up to three complete tetrahedral shells (five complete atomic shells) embedding the site under investigation. In the following these cluster models are referred to as shell-n, where n denotes the number of complete atomic shells. The geometrical parameters were taken from experimental data: namely, single crystal X-ray or neutron scattering experiments.[22,34−37] Previous studies[14,22] demonstrated that this choice of reference geometries ensures better agreement between computed and experimental chemical shifts than the use of the structural parameters determined by powder X-ray diffraction experiments. Consistently with recent literature,[14,19,21,22,49] the clusters are terminated with H atoms whose positions are kept fixed along the direction of Si−O and O−Si bonds of the parent crystal structure, using O−H and Si−H distance of 0.96 and 1.48 A, respectively. These parameters correspond to the optimum O−H bond length, for non-hydrogen-bonded Si−O−H linkages in various silicate clusters, and to the optimum Si−H bond length, for Si_2OH_6 dimer, as computed at the B3LYP/6−311+G(2df,p) level of theory. In some cases, at the boundary of the cluster, when the atoms of the outer coordination shells form four-membered rings, an additional Si atom has been included together with two terminating H atoms.

Complementary to the cluster approach, the importance of including long-range electrostatic effects when computing the [29]Si and [17]O shielding tensor has been investigated by making use of the well-known ONIOM method.[38−40] In this framework, the system under study is subdivided into two (or more) regions of concentric layers, each of which is treated at a different level of theory. Following the original

^{29}Si and ^{17}O NMR Parameters in SiO$_2$ Polymorphs

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2133**

nomenclature proposed by Morokuma,[38−40] the whole system is called the *real system*. The most important part of the system forms the innermost layer, the *model system*, which is described at the highest degree of accuracy: the *model system* usually includes the most chemically relevant portion of the entire system, i.e. the region where the local phenomena of interest take place. Subsequent layers are described with progressively less accurate methods, i.e. lower-level QM models or molecular mechanics. The boundary between the *model system* and the exterior layer could also be along covalent bonds; in this case, the consistency of the model system is attained by adding a set of fictitious atoms, usually H atoms, along the directions of pre-existing chemical bonds. The two-layer scheme has been employed, performing QM/MM computations with the so-called 'electronic embedding' scheme, namely including the point charges of the MM layer in the QM Hamiltonian of the *model system* according to the scheme

$$\widehat{H}_{el}(\text{model:MM}) = \widehat{H}_{el}(\text{model}) - \sum_i^Y \sum_J^X \frac{q_J}{r_{iJ}} \qquad (3)$$

where $H_{el}(\text{model:MM})$ and $H_{el}(\text{model})$ are the electronic Hamiltonians for the QM region with and without the external field, Y is the number of the electrons in the model system, and X is the number of the point charges in the MM region. This scheme allows the QM wave function to be properly polarized by the electrostatic properties of the surroundings.

In our calculations, the *real system* consists of a sphere made of ca. 1200−1300 atoms, which have been extracted from the experimental structures of α-quartz and α-cristobalite with the Si and O sites of interest at the center of each sphere. The outer Si and O atoms have been saturated with OH and H atoms with the Si−O−H angles and O−H bond lengths set to 115.0° and 0.96 Å, respectively. The Unified Force Field (UFF)[50] parameters have been exploited, and the partial atomic charges have been obtained according to the Qeq scheme developed by Rappe and Goddard.[51] Regarding the *model system*, the HF and PBE levels of theory on structures with equivalent shell-n size as for the aforementioned cluster approach have been adopted. In order to avoid spurious effects by the electronic embedding, the MM point charges of the atoms at the boundary between *model* and *real systems* have been properly scaled by using the default values set in the Gaussian code.

The dependence of shielding constants both on the Si-centered cluster and *model system* sizes has been tested by using shell-n clusters with $n = 1$ (1 Tetrahedron, 1T), 3 (2T), 5 (3T), and 7 (4T) by using the 6−311+G(2d) basis set for the central silicon and oxygen atoms of the first shell, the 6−31+G(d) basis set for the atoms of the second tetrahedron, and the 6−31G* basis set on all the further atoms. Shell-n clusters with $n = 1$, 2, 4, and 6 have been used in the case of O-centered clusters.

Besides the ^{29}Si and ^{17}O chemical shifts, the ^{17}O electric field gradients and quadrupolar parameters have been also calculated. The primary quantum-mechanical quantity is the traceless electric field gradient (efg) tensor, whose elements
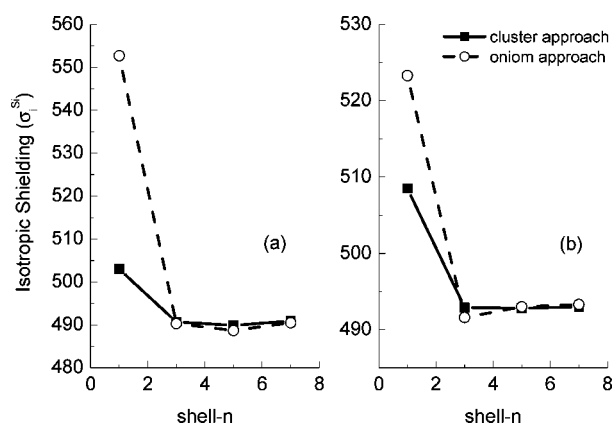
Figure 3. Comparison between ^{29}Si isotropic shielding calculated by using the cluster approach (solid line) and the oniom approach (dashed line) for (a) α-quartz and (b) α-cristobalite as a function of cluster and model system size. High level calculations performed at the Hartree−Fock level of theory by using the 6−311+G(2d) basis set for the central silicon and oxygens of the first shell ($n = 1,2$). The 6−31+G(d) basis set has been used for the atoms of the second shell ($n = 3,4$) and 6−31G* basis set for the atoms of the third shell and beyond ($n > 4$). In the ONIOM approach the low level calculations were performed by using the UFF embedded charges (Qeq scheme).

are related to the quadrupolar coupling constant, C_Q, and quadrupolar coupling asymmetry parameter, η_Q, according to

$$C_q = e^2 Q \langle q_{zz} \rangle / h, \text{ and } \eta_Q = \frac{\langle q_{xx} \rangle - \langle q_{yy} \rangle}{\langle q_{zz} \rangle} \qquad (4)$$

where eQ is the nuclear quadrupole moment of the nucleus of interest, and $e\langle q_{xx} \rangle$, $e\langle q_{yy} \rangle$, and $e\langle q_{zz} \rangle$ are the components of the efg tensor at the nucleus in the principal axis system, with $|\langle q_{zz} \rangle| > |\langle q_{yy} \rangle| > |\langle q_{xx} \rangle|$. Since the nuclear quadrupole moment eQ cannot be measured experimentally, it has been usually derived from the experimental C_Q value and the calculated efg of the free atom, ground-state atomic O ($^3P_2$) in the case of ^{17}O. The resultant eQ is thus dependent on the employed level of theory. In order to achieve internal consistency and a better agreement with experimental data, we used an eQ value calculated at the same level as for the studied clusters. Under such circumstances, the eQ values were derived from the accurate experimental ^{17}O C_Q of the H$_2$O molecule (10.175 ± 0.067 MHz)[52] and the $e\langle q_{zz} \rangle$ values computed for the same water molecule at the same level of theory as that exploited for the silica clusters.

Results and Discussion

^{29}Si NMR. CLUSTER vs ONIOM Performances and System-Size Convergence.
Figure 3 reports a comparison between ^{29}Si isotropic shielding calculated by using the cluster approach (solid line) and the ONIOM approach (dashed line) for representative Si sites (namely, α-quartz and α-cristobalite) as a function of cluster and model system size. The high level calculations (cluster approach and model system in the ONIOM approach) were performed at the HF level of theory. In the ONIOM approach the low level

Table 1. Calculated [29]Si Isotropic Chemical Shift for Shell-3 Cluster Models of Various SiO$_2$ Polymorphs Using Different Methods[a]

	HF	B3LYP	PBE0	M05	CAM-B3LYP	exp.[b]	Profeta et al.[28]
			Cristobalite				
Si	−109.1 (0.6)	−108.9 (0.4)	−108.9 (0.4)	−109.3 (0.8)	−109.3 (0.8)	−108.5	−109.1 (0.6)
			Coesite				
Si1	−113.6 (0.3)	−115.0 (1.1)	−114.9 (1.0)	−114.0 (0.1)	−114.5 (0.6)	−113.9	−114.7 (0.8)
Si2	−108.7 (0.6)	−109.8 (1.7)	−109.8 (1.7)	−109.0 (0.9)	−109.4 (1.3)	−108.1	−108.4 (0.3)
			Sigma2				
Si1	−115.2 (0.6)	−117.8 (2.0)	−115.5 (0.3)	−115.9 (0.1)	−117.3 (1.5)	−115.8	
Si2	−113.2 (0.4)	−114.7 (1.1)	−112.3 (1.3)	−114.1 (0.5)	−114.7 (1.1)	−113.6	
Si3	−119.0 (0.7)	−121.2 (1.5)	−118.8 (1.1)	−119.3 (0.4)	−121.0 (1.3)	−119.7	
Si4	−108.9 (0.4)	−110.4 (1.9)	−108.3 (0.2)	−109.6 (1.1)	−110.3 (1.8)	−108.5	
			Ferrierite				
Si1	−116.3 (0.2)	−119.9 (3.4)	−119.3 (2.8)	−117.7 (1.2)	−119.3 (2.8)	−116.5	−117.7 (1.2)
Si2	−112.5 (0.2)	−116.9 (4.6)	−116.1 (3.8)	−114.2 (1.9)	−115.5 (3.2)	−112.3	−113.7 (1.4)
Si3	−112.1 (0.2)	−114.0 (2.1)	−113.7 (1.8)	−112.7 (0.8)	−113.7 (1.8)	−111.9	−112.2 (0.3)
Si4	−117.4 (0.2)	−121.2 (4.0)	−120.7 (3.5)	−118.8 (1.6)	−120.4 (1.6)	−117.2	−119.5 (2.3)
Si5	−116.1 (0.1)	−119.2 (3.0)	−118.7 (2.5)	−117.1 (0.9)	−118.5 (0.9)	−116.2	−116.3 (0.1)
$\langle\Delta\delta\rangle$	0.4	2.2	1.7	0.9	1.8		0.9

[a] The 6-311+G(2df,p) basis set has been applied to all atoms. The Si sites in α-quartz,[34] α-cristobalite,[35] coesite,[36] Sigma-2,[22] and ferrierite[37] all-silica zeolites have been cut out from the experimental framework structures. The errors between experimental and calculated data are reported in brackets. [b] Experimental ^{29}Si NMR data taken from ref 3 for cristobalite and coesite, ref 22 for Sigma-2, and ref 37 for ferrierite.

calculations were performed by using the UFF embedded charges (Qeq scheme).[51] The results clearly show that the [29]Si isotropic shielding (σ_i^{Si}) converges when three complete atomic shells around Si are included (shell-3 cluster), with a scattering among the values for the shell-3 and shell-7 clusters of only 0.2 ppm for both α-quartz and α-cristobalite. In agreement with these results, Xue et al.[19] compared the calculated σ_i^{Si} for Si-centered shell model clusters of increasing size for cristobalite and coesite (Si1 site) and found that convergence occurred at the shell-3 cluster model, by using the 6−311+G(2df,p) basis set. The same was found by Brouwer et al.[22] by using the 6−311G(2df) basis set on Si-centered clusters of α-quartz.

Figure 3 also shows that computed [29]Si isotropic shielding is not really affected by long-range Coulomb interactions included by the ONIOM(QM:MM) approach. Indeed, the scattering values of the shell-7 model are within 0.3−0.4 ppm.

Our results demonstrate that large errors arise from small-sized cluster/model systems, due to the lack of Pauli repulsion and magnetic susceptibility of the nearby atoms in the QM/MM calculations. These QM effects could not be safely described by electrostatic potentials, and quite large systems should be accounted for in the QM calculations of shielding tensors. Therefore, provided that the cluster sizes are sufficiently large, the use of cluster models for QM calculations of the [29]Si shielding tensor is relatively accurate, and a QM/MM scheme is not particularly helpful in reducing the computational effort.

^{29}Si NMR. Chemical Shifts. Table 1 lists the ^{29}Si NMR chemical shifts predicted by HF together with those computed by DFT, with B3LYP, PBE0, M05−2X, and CAM-B3LYP density functionals. Calculations of ^{29}Si NMR chemical shifts of different Si sites in α-quartz,[34] α-cristobalite,[35] coesite,[36] Sigma-2,[22] and ferrierite[37] all-silica zeolites have been performed using the shell-3 clusters (see Figures S1, S2, and S3 in the Supporting Information for

pictures of the clusters) using α-quartz as secondary chemical shift standard. The experimentally observed isotropic chemical shift for α-quartz referred to tetramethylsilane (TMS) was δ_{iso}^{TMS}(α-quartz) = −107.4 ppm.[53] The calculated shielding values for σ_{iso}(α-quartz) were 491.68, 446.83, 452.94, 444.20, and 446.18 ppm for HF, B3LYP, PBE0, M05−2X, and CAM-B3LYP methods, respectively. The chemical shifts reported in Table 1 have been compared with experimental data and chemical shifts calculated by Profeta et al.[28] by using the PBE density functional and the periodic approach implemented in the PARATEC code.[54] In this approach the gauge including projector augmented wave (GIPAW) method[55] ensures the reproduction of all electron results using pseudopotentials and plane-waves basis sets.

The computed [29]Si chemical shifts in closest agreement with experiments are those obtained by HF: the average errors ($\langle\Delta\delta\rangle$) on the ^{29}Si NMR chemical shifts are 0.4, 2.2, 1.7, 0.9, and 1.8 ppm for HF, B3LYP, PBE0, M052X and CAM-B3LYP, respectively. The largest errors are 0.7, 4.6, 3.8, 1.9, and 3.2 ppm for HF, B3LYP, PBE0, M05−2X, and CAM-B3LYP, respectively. The HF-cluster approach is of even superior accuracy than the GIPAW-PBE periodic approach;[28] the latter having $\langle\Delta\delta\rangle$ = 0.9 ppm and the largest error equal to 2.3 ppm, in line with the performances of the density-functionals here exploited.

Figures 4 and 5 compare the results of the shell-3 HF/6−311+G(2df,p) calculations of the [29]Si chemical shifts with the experimental spectra for Sigma-2 and ferriereite all-silica zeolites, respectively. The assignments according to the experimental structures of Brouwer et al.[22] for Sigma-2 and Lewis et al.[37] for ferrierite are also given above each peak, and the heights of the peaks reflect the multiplicities of each of the sites. It is worth noting that while Sigma-2 has four crystallographic unique Si sites well resolved in the experimental spectrum, ferrierite has five crystallographic unique sites where Si1 and Si5 are separated in the experimental NMR spectrum by only 0.2 ppm. Such a feature has been
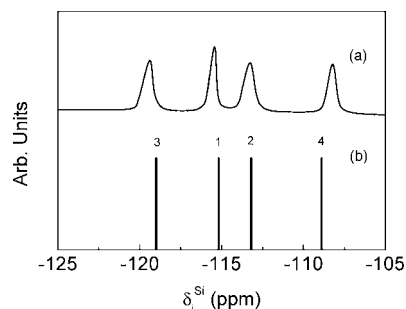
Figure 4. The experimental ^{29}Si NMR isotropic chemical shifts (a) of Sigma-2 taken by Brouwer[23] compared with the (b) calculated shifts (HF/6−311+G(2df,p) with shell-3 Si-centered clusters) of the experimental crystal structure. The numbering scheme corresponds to the Si sites in the crystal structure of Sigma-2.[65]
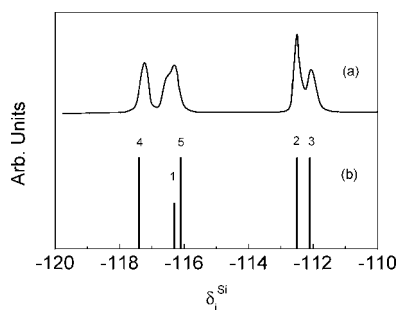
Figure 5. The experimental ^{29}Si NMR isotropic chemical shifts of (a) Sil-ferrierite taken by Lewis[37] compared with the (b) calculated shifts (HF/6−311+G(2df,p) with shell-3 Si-centered clusters) of the experimental crystal structure. The numbering scheme corresponds to the Si sites in the crystal structure of Sil-ferrierite.[37]

remarkably well described by our calculations, and it was not predicted by recent calculations reported by Bull et al.,[14] where an overlap of sites 1 and 5 was found, while a lower accurate shift of 1.3 ppm was obtained by the GIPAW-PBE periodic approach.[28]

A meaningful test for the accuracy of *ab initio* calculations consists of comparing the computed principal components, the span (Ω) and skew (κ) values of the ^{29}Si chemical shift tensor. These quantities can be extracted from recoupling and slow MAS experiments.[22] The principal components of the ^{29}Si chemical shift tensor are the eigenvalues, with $\delta_{11} > \delta_{22} > \delta_{33}$. The span ($\Omega$) describes the breadth of the tensor and the skew (κ) describes the asymmetry of tensor components and are defined as follows

$$\Omega = \delta_{11} - \delta_{33}, \text{ and } \kappa = 3(\delta_{22} - \delta_{iso})/\Omega \qquad (5)$$

where $\delta_{iso} = (\delta_{11} + \delta_{22} + \delta_{33})/3$.

Table 2 lists the principal components, the span (Ω) and skew (κ) values of the ^{29}Si chemical shifts tensor of Sigma-2 calculated at the HF, B3LYP, PBE0, M05−2X, and CAM-B3LYP levels of theory, together with those determined by recoupling and slow MAS experiments.[22] Slow MAS data are more accurate (smaller experimental uncertainties) and are taken as reference in Figure 6 where correlation plots between calculated and observed principal components of the ^{29}Si chemical shifts tensor of Sigma-2 are reported. The

average errors of computed principal components are 0.6, 1.9, 1.6, 1.0, and 1.7 ppm for HF, B3LYP, PBE0, M05−2X, and CAM-B3LYP, respectively. A more detailed analysis shows that while HF well reproduces all the trends of the principal components, B3LYP, PBE0, and CAM-B3LYP wrongly predict that $\delta_{11}(Si4) > \delta_{11}(Si2)$ and $\delta_{33}(Si2) > \delta_{33}(Si3)$, and the M05−2X method wrongly predicts that $\delta_{33}(Si2) > \delta_{33}(Si3)$. An even poorer agreement with experiment has been found in the case of the span (Ω) and skew (κ) values (see Figure 7 and Table 2). However, it is worth noting the presence of high experimental uncertainties: different techniques yield quite dissimilar values.

The excellent agreement with the experiments indicates that the HF level of theory on cluster models extracted from the highest quality crystal structures is adequate for calculating ^{29}Si shielding tensors in zeolites and silica polymorphs.

^{17}O NMR. CLUSTER vs ONIOM Performance and System-Size Convergence. Figure 8 reports the ^{17}O isotropic shielding constants calculated by using the cluster approach (solid line) and the ONIOM approach (dashed line) for the single O site of α-cristobalite as a function of cluster and model system size. The high level calculations (cluster approach and model system in the ONIOM approach) were performed at the PBE level of theory by using the 6−311+G(2d) basis set for the central oxygen and silicon of the first two atomic shells ($n = 1,2$) and the 6−31+G(d) basis set for the remaining atoms ($n > 2$). In the ONIOM approach the low level calculations were performed by using the UFF embedded charges (Qeq scheme),[51] in close resemblance with the calculations carried out in the case of ^{29}Si.

The results clearly show that within the cluster approach the ^{17}O isotropic shielding (σ_i^O) converges when four complete atomic shells around O are included (shell-4 cluster), with a scattering among the values for the shell-4 and shell-6 clusters of only 0.3 ppm. Figure 8 also shows that σ_i^O is slightly affected by long-range Coulomb interactions, as included in the ONIOM approach, even though not in a systematic way. However, since the shell-6 model system contains more than 130 atoms a further expansion of the cluster size would require too high computational effort. Moreover, in such circumstances the simplest cluster approach would be preferred since the ONIOM approach is usually applied with the aim of decreasing the size of the system to be included in QM calculations.

Table 3 lists the principal components of the electric field gradients calculated for the O-site in α-cristobalite. Our results clearly show that the electric field gradients are not affected by long-range Coulomb interactions and they converge at the shell-4 model size.

^{17}O NMR. Chemical Shifts. The calculations of ^{17}O NMR chemical shifts were performed on shell-4 cluster models of α-quartz,[34] α-cristobalite,[35] and coesite[36] by using the 6−311+G(2df,p) basis set. In the present study, α-cristobalite, δ_{iso}^{H2O}(α-cristobalite) = 40 ± 2 ppm,[56] has been used as internal secondary chemical shift reference: the computed values for σ_{iso}(α-cristobalite) were 272.0, 241.8, 244.2, 238.0, and 233.1 ppm for HF, B3LYP, PBE0, M05−2X, and PBE methods, respectively. These absolute values can be com-

Table 2. Experimental and Calculated Principal Components of the ^{29}Si Chemical Shift Tensors, Span Ω, and Skew κ Values of Sigma-2

site	δ_{11} (ppm)	δ_{22} (ppm)	δ_{33} (ppm)	Ω (ppm)	κ
		Experiment (CSA Recoupling)			
Si1	-108.9 ± 1.1	-113.0 ± 1.0	-125.3 ± 1.1	16.4 ± 1.2	0.50 ± 0.12
Si2	-104.5 ± 1.5	-113.3 ± 1.2	-123.0 ± 1.5	18.5 ± 1.6	0.05 ± 0.18
Si3	-115.3 ± 1.4	-119.6 ± 0.9	-124.1 ± 1.4	8.8 ± 1.7	0.03 ± 0.25
Si4	-103.5 ± 0.9	-105.8 ± 0.9	-116.2 ± 0.9	12.6 ± 0.8	0.65 ± 0.14
		Experiment (Slow MAS)			
Si1	-108.1 ± 0.5	-113.5 ± 0.5	-125.8 ± 0.5	17.7 ± 0.5	0.39 ± 0.06
Si2	-104.3 ± 0.6	-112.8 ± 0.5	-123.6 ± 0.6	19.3 ± 0.6	0.12 ± 0.05
Si3	-114.4 ± 0.3	-119.6 ± 0.3	-125.0 ± 0.3	10.7 ± 0.2	0.02 ± 0.05
Si4	-102.8 ± 0.7	-106.0 ± 0.8	-116.7 ± 0.7	14.0 ± 0.5	0.53 ± 0.13
		Hartree−Fock			
Si1	-107.9	-113.5	-124.2	16.3	0.31
Si2	-103.1	-112.9	-123.7	20.6	0.05
Si3	-114.7	-118.1	-124.0	9.3	0.27
Si4	-102.7	-106.4	-117.5	14.8	0.50
		B3LYP			
Si1	-110.0	-115.9	-127.6	17.6	0.33
Si2	-102.9	-113.8	-127.4	24.5	0.11
Si3	-116.1	-120.3	-127.3	11.2	0.25
Si4	-103.8	-107.6	-119.9	16.2	0.53
		PBE0			
Si1	-109.8	-115.6	-127.1	17.3	0.33
Si2	-102.8	-113.5	-126.8	24.0	0.11
Si3	-115.9	-119.9	-126.6	10.7	0.25
Si4	-103.7	-107.5	-119.7	16.0	0.53
		M052X			
Si1	-108.4	-114.0	-125.3	16.9	0.34
Si2	-102.8	-113.4	-126.1	23.3	0.09
Si3	-114.7	-118.1	-125.0	10.3	0.34
Si4	-102.5	-107.3	-118.9	16.4	0.41
		CAM-B3LYP			
Si1	-109.5	-115.3	-127.2	17.6	0.34
Si2	-103.0	-114.2	-127.0	24.0	0.07
Si3	-116.0	-120.1	-126.9	10.9	0.25
Si4	-103.7	-107.4	-119.7	16.0	0.54

pared with the estimated value of the experimental absolute shielding of 248 ± 2 ppm, obtained from the aforementioned isotropic chemical shift, $\delta_{\text{iso}}^{\text{H2O}}$($\alpha$-cristobalite), and the
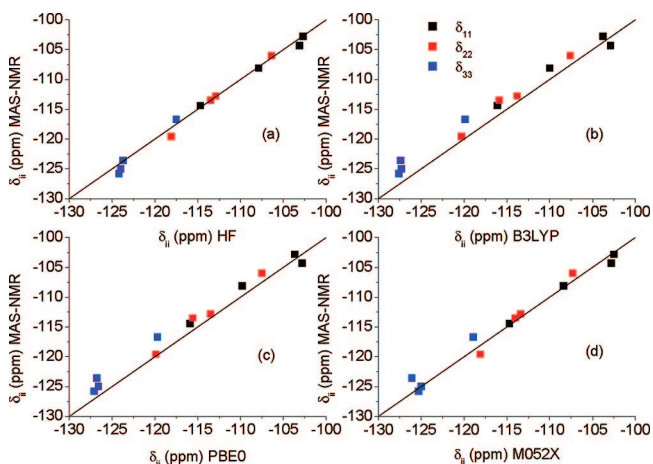
Figure 6. Correlation plots between the experimental (slow MAS experiments) principal components of ^{29}Si chemical shift tensors and the calculated ones at the (a) Hartree−Fock, (b) B3LYP, (c) PBE0, and (d) M052X level of theory for Sigma-2. If the experimental data were perfectly reproduced, the data points would lie on the solid lines (ideal lines of equation $y = x$).

reported absolute shielding value of 324 ppm for water molecule,[57] plus the shift difference of -36 ppm required to reference molecular to liquid water.[58] The experimental ^{17}O NMR isotropic chemical shift of gaseous water compares fairly well with its theoretical counterpart, as computed at different levels of theory on the experimental structure of gas-phase water (R(O−H)=0.9572 Å and ∠HOH = 104.52°).

These results are listed in Table 4 together with the efg, the ^{17}O efg asymmetry parameter (η) for water and the nuclear quadrupole moment eQ values, derived from the
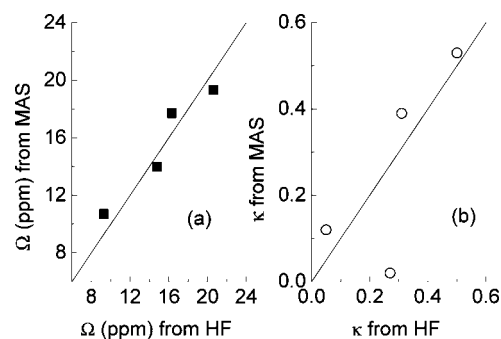
Figure 7. Comparison of (a) the span (Ω) and (b) the skew κ values determined from slow MAS experiments[22] and the ones calculated by the HF/6−311+G(2df,p) method on shell-3 cluster models of Si sites in Sigma-2.

^{29}Si and ^{17}O NMR Parameters in SiO₂ Polymorphs

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2137**
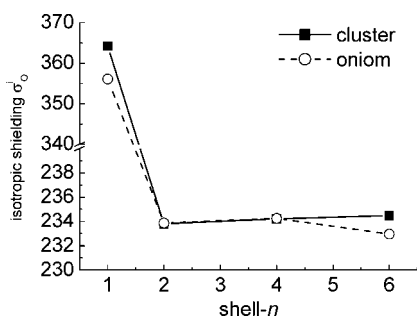
Figure 8. Comparison between ^{17}O isotropic shielding calculated by using the cluster approach (solid line) and the oniom approach (dashed line) for α-cristobalite as a function of cluster and model system size. High level calculations performed at the Hartree−Fock level of theory by using the 6−311+G(2d) basis set for the central oxygen and silicon atoms of the first two atomic shells ($n = 1,2$). The 6−31+G(d) basis set has been used for the atoms of the third shell and beyond ($n > 2$). In the ONIOM approach the low level calculations were performed by using the UFF embedded charges (Qeq scheme).

Table 3. Electric Field Gradients of the O-Site of α-Cristobalite As a Function of System Size within the Cluster and ONIOM Approaches

	$e<q_{xx}>$ (au)		$e<q_{yy}>$ (au)		$e<q_{zz}>$ (au)	
	Cluster	ONIOM	Cluster	ONIOM	Cluster	ONIOM
shell-2	0.418308	0.418271	0.572242	0.572290	−0.99055	−0.99056
shell-4	0.416158	0.416228	0.566511	0.566386	−0.98267	−0.98261
shell-6	0.425327	0.420848	0.511205	0.563811	−0.98493	−0.98466

accurate experimental ^{17}O C_Q of the H$_2$O molecule (10.175 ± 0.067 MHz)[52] and the $e\langle q_{zz}\rangle$ values computed for the same water molecule at the same levels of theory as those exploited for the silica clusters. The computed η values for water are in the range between 0.80 and 0.76, which compares well with the experimental value of 0.75 ± 0.01 determined by Verhoeven et al.[52] Good agreement is also found for the nuclear quadrupole moment eQ, which ranges between 2.38E-30 and 2.53E-30 m^2 with respect to the experimental value of 2.55E-30 m^2 determined by Pyykko.[59]

Table 5 summarizes the ^{17}O NMR parameters computed by HF, B3LYP, PBE0, M05−2X, and PBE methods for various shell-4 O-centered clusters (see Figures S4 and S5 of the Supporting Information for the pictures of the cluster models used). Moreover, Figure 9 displays a comparison of the calculated C_Q^O with experimental data. The best correlation is found with B3LYP, which reproduces the quadrupolar coupling constant by an average error of 0.04 MHz, with respect to the average error of 0.14, 0.11, 0.36, and 0.10 MHz for HF, PBE0, M05−2X, and PBE, respectively. However, a more in-depth analysis shows that while the PBE method well reproduces all the trends of the C_Q^O of coesite, both B3LYP and PBE0 wrongly predict that C_Q^O(O2) = C_Q^O(O3), and both HF and M05−2X as well as the GIPAW-PBE periodic approach[28] wrongly predict that C_Q^O(O2) > C_Q^O(O3). The calculated η_Q^O values show a reasonable agreement of all the methods with experiments (mostly within 0.05), see Figure 10 and Table 4. The reported

deviations fall well within the uncertainties of experimental estimations.

The agreement between calculations and experiments observed in the relative ^{17}O chemical shifts among different oxygen sites is not as good as that for δ_i^{Si}. Indeed, the average errors ($<\Delta\delta>$) of computed δ_i^O are 5.7, 4.5, 4.4, 4.7, and 3.9 ppm for HF, B3LYP, PBE0, M05−2X, and PBE methods, respectively; the largest errors being 9.8, 8.4, 7.8, 7.1, and 7.4 for HF, B3LYP, PBE0, M05−2X, and PBE, respectively.

A closer examination of the results listed in Table 5 suggests that our calculations reproduce well the relative δ_i^O for O1 and O2 sites of coesite, with differences between computed and measured values well below the experimental uncertainties of 1 ppm. Both these sites link neighboring four-membered rings. The worse results are obtained for the O3, O4, and O5 sites, all of which are part of four-membered rings. Figure S5 of the Supporting Information shows the exploited coesite shell-4 O centered cluster structures. These have been cut out from the bulk structure (reported in Figure 1) by following the bond connectivity. However, in this way the central oxygen does not feel the effects of the atoms towering above, which are about 3 Å faraway. Thereafter, the low accuracy of predicted δ_i^O values for O3, O4, and O5 sites seems to result from the convergence of cluster size, which is not fulfilled in these cases. To further support such a statement, further calculations on the O4 site (which shows errors between 6.6 and 8.2 ppm depending on the method used) have been performed by using two bigger clusters, namely, the shell-6 cluster (132 atoms) made by the bond connectivity method and a new cluster containing 147 atoms made by including atoms inside a spherical region with a radius of about 7.0 Å (see Figure S6 of the Supporting Information). The results listed in Table 6 show that a shell-4 and a shell-6 O4 clusters of coesite do not reach convergence; the cluster system containing 147 atoms provides the best results compared to experiment, with an error on isotropic chemical shift of only 1 ppm at the PBE/6−311+G(2df,p) level of theory.

In conclusion, these results show that oxygens in some silicate topologies present some difficulties to reach full convergence with respect to the cluster size. Preliminary tests on crystal silicates containing alkali and alkaline-earth cations performed in our laboratories showed that the problem of the convergence of ^{29}Si and ^{17}O NMR parameters with respect to the cluster size is even worse than in SiO₂ polymorphs. This could be probably the reason why previous calculations of ^{29}Si-, ^{27}Al-, and ^{17}O NMR chemical shifts of silicate and alumino-silicate glasses yielded quite poor results.[17,60−62]

Concluding Remarks

In the present paper a comparative study on the performance of quantum chemical techniques in computing the magnetic parameters of SiO₂ polymorphs has been reported. Several all silica systems (α-quartz, α-cristobalite, coesite, Sigma-2, and ferrierite zeolites) and several *ab initio* methods (HF, B3LYP, CAM-B3LYP, PBE, PBE0, and M05−2X) have been tested.

Table 4. ^{17}O Isotropic Shielding ($\sigma_i{}^O$), Diagonal Element of the Electric Field Gradient Tensor ($<q_{ii}>$), Electric Quadrupole Moment of the Oxygen Nucleus Calculated by Using the Experimentally Quadrupolar Coupling Constants of Oxygen in Water ($C_Q{}^O = 10.175 \pm 0.067$ MHz), and the Asymmetry Parameter for Water Calculated at Different Levels of Theory

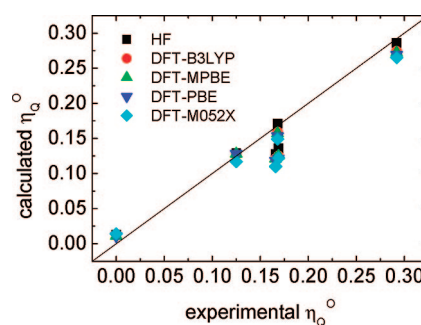|	$\sigma_i{}^O$ (ppm)	$e<q_{xx}>$ (au)	$e<q_{yy}>$ (au)	$e<q_{zz}>$ (au)	eQ (m^2)	$\eta_Q{}^O$
HF	328.2	−0.179811	−1.642521	1.822333	2.38E-30	0.802658
B3LYP	326.9	−0.200025	−1.566409	1.766434	2.45E-30	0.773527
PBE0	329.3	−0.192725	−1.543946	1.736671	2.49E-30	0.778052
M05−2X	338.8	−0.172313	−1.575157	1.74747	2.48E-30	0.802786
PBE	328.5	−0.201431	−1.512331	1.713762	2.53E-30	0.764925

Table 5. Calculated ^{17}O Isotropic Chemical Shift and EFG Parameters for OH Ended 2T Shell Clusters Models (4 Atomic Shells around Each O Center) of Various SiO$_2$ Polymorphs Using Different Methods[a]

O site	HF	B3LYP	PBE0	M052X	PBE	exp.	Profeta et al.
				$\delta_i{}^O$ (ppm)			
coesite							
O1	29.7	28.5	28.1	26.1	28.3	29 ± 1	25.8
O2	39.7	40.3	40.1	39.2	41.0	41 ± 1	39.2
O3	47.2	48.6	49.2	49.9	49.5	57 ± 1	56.0
O4	44.8	45.7	46.2	45.9	46.4	53 ± 1	52.4
O5	49.4	51.6	52.4	53.3	53.1	58 ± 1	57.8
cristobalite	40	40	40	40	40	40 ± 2	39.3
quartz	42.5	43.3	43.5	44.2	43.8		44.3
$\Delta\delta$	5.7	4.5	4.4	4.7	3.9		1.4
				$C_Q{}^O$ (MHz)			
coesite							
O1	5.89	6.04	6.17	5.67	6.15	6.05 ± 0.05	6.24
O2	5.30	5.43	5.53	5.08	5.51	5.43 ± 0.05	5.56
O3	5.28	5.43	5.53	5.05	5.52	5.45 ± 0.05	5.45
O4	5.41	5.56	5.66	5.19	5.66	5.52 ± 0.05	5.73
O5	5.11	5.25	5.34	4.88	5.33	5.16 ± 0.05	5.23
cristobalite	5.11	5.25	5.35	4.87	5.34	5.3 ±0.1	5.30
quartz	5.16	5.28	5.38	4.91	5.36		5.31
ΔC_Q	0.14	0.04	0.11	0.36	0.10		0.10
				$\eta_Q{}^O$			
coesite							
O1	0.013	0.012	0.011	0.014	0.011	0.000 ± 0.005	0.040
O2	0.128	0.121	0.121	0.110	0.118	0.166 ± 0.005	0.190
O3	0.171	0.159	0.157	0.149	0.153	0.168 ± 0.005	0.190
O4	0.136	0.125	0.125	0.122	0.121	0.169 ± 0.005	0.166
O5	0.286	0.274	0.272	0.265	0.269	0.292 ± 0.005	0.296
cristobalite	0.129	0.127	0.128	0.117	0.128	0.125 ± 0.005	0.145
quartz	0.169	0.169	0.171	0.159	0.171		0.202
$\Delta\eta_Q$	0.016	0.022	0.022	0.029	0.025		0.019

[a] The 6-311+G(2df,p) basis set has been applied to all atoms.
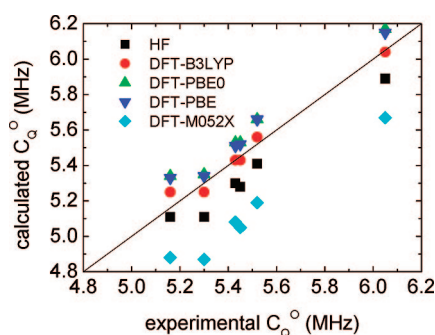
Figure 9. Calculated ^{17}O quadrupolar coupling constants, $C_Q{}^O$, for shell-4 O-centered clusters of α-cristobalite and coesite vs experimental values.

Figure 10. Calculated ^{17}O electric field gradient asymmetry parameter, $\eta_Q{}^O$, for shell-4 O-centered clusters of α-cristobalite and coesite vs experimental values.

First, we consider the convergence of computed parameters with respect to cluster size, with and without the inclusion of partial atomic charges, by the ONIOM(QM:MM) scheme, to account for long-range Coulomb interactions. In the case of ^{29}Si, the computed shielding constants converged at system sizes still amenable to full QM descriptions. Unfortunately,

the situation is not the same for ^{17}O parameters: the results show that full convergence is system dependent; in particular, it is hard to achieve for oxygen atoms that are part of small n-membered rings ($n < 6$). However, from our work it is at least evident that the inclusion of long-range electrostatic effects is not mandatory for well-behaving systems and not

^{29}Si and ^{17}O NMR Parameters in SiO$_2$ Polymorphs

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2139**

Table 6. DFT-PBE//6-311+G(2df) Calculated [17]O Isotropic Chemical Shift and efg Parameters for the O4 Site in Coesite as a Function of the Cluster Size Dimension[a]

	shell-4	shell-6	cluster-147 atoms	exp.
σ_i^O	226.65	223.65	219.06	–
δ_i^O (ppm)	46.43	49.43	54.02	53 ± 1
C_Q^O (MHz)	5.66	5.74	5.82	5.52 ± 0.005
η^O	0.1214	0.1861	0.160	0.169 ± 0.005

[a] See Figure S6 of the Supporting Information for pictures of the clusters used.

useful for those systems whose convergence is questionable. Use of QM:QM embedding techniques[63] or periodic boundary conditions[64] seem to be viable solutions for the systems whose complexity prevent the safe cutting out of cluster models.

Regarding the quality of our results, an excellent agreement with experiments has been found for the [29]Si chemical shifts of Sigma-2 and ferrierite zeolites. In this second case, the characteristic features of the experimental spectrum have been predicted with an accuracy that has never been achieved before.

In conclusion, the present work shows that the HF model could be safely exploited for computing [29]Si magnetic parameters, while methods rooted into DFT perform slightly better in the more complex case of [17]O nucleus. It is worth noting that the calculations of magnetic shielding tensors result in being challenging for the currently available density functionals: the slightly better or comparable performance of HF than DFT highlights the importance of developing and testing new exchange-correlation approximations for an accurate prediction of molecular parameters.

Therefore, beyond the intrinsic relevance of the investigated silicate systems, the many theoretical results reported here represent, in our opinion, valuable references for computational chemists approaching the theoretical study of NMR parameters for SiO$_2$ polymorphs.

Supporting Information Available: Pictures of all the clusters considered. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Stebbins, J. F. Nuclear magnetic resonance spectroscopy of silicates and oxides in geochemistry and geophysicsIn *Mineral physics and crystallography, a handbook of physical constants*; American Geophysical Union: Washington, DC, 1995; p 303.

(2) Engelhardt, G.; Michel, D. *High-Resolution Solid-State NMR of Silicates and Zeolites*; Wiley: New York, 1987; p 106.

(3) Smith, J. V.; Blackwell, C. S. *Nature* **1983**, *303*, 223.

(4) Engelhardt, G.; Radeglia, R. *Chem. Phys. Lett.* **1984**, *108*, 271.

(5) Ramdas, S.; Klinowski, J. *Nature* **1984**, *308*, 525.

(6) Janes, N.; Oldfiels, E. *J. Am. Chem. Soc.* **1985**, *107*, 6769.

(7) Oestrike, R.; Yang, W.-H.; Kirkpatric, R. J.; Hervig, R. L.; Navrotsky, A.; Montez, B. *Geochim. Cosmochim. Acta* **1987**, *51*, 2199.

(8) Sherriff, B. L.; Grundy, H. D. *Nature* **1988**, *332*, 819.

(9) Stebbins, J. F.; Kanzaki, M. *Science* **1991**, *251*, 294.

(10) Mauri, F.; Pasquarello, A.; Pfrommer, B. G.; Yoon, Y. G.; Louie, S. G. *Phys. Rev. B* **2000**, *62*, R4786.

(11) Grandinetti, P. J.; Baltisberger, J. H.; Farnan, I.; Stebbins, J. F.; Werner, U.; Pines, A. *J. Phys. Chem.* **1995**, *99*, 12341.

(12) Grandinetti, P. J.; Clark, T. M. *Mod. Magn. Reson.* **2006**, *3*, 1543.

(13) Clark, T. M.; Grandinetti, P. J.; Florian, P.; Stebbins, J. F. *Phys. Rev. B* **2004**, *70*, 064202/1.

(14) Bull, L. M.; Bussemer, B.; Anupold, T.; Reinold, A.; Samoson, A.; Sauer, J.; Cheetham, A. K.; Dupree, R. *J. Am. Chem. Soc.* **2000**, *122*, 4948.

(15) Bull, L. M.; Cheetham, A. K.; Anupold, T.; Reinold, A.; Samoson, A.; Sauer, J.; Bussemer, B.; Lee, Y.; Gann, S.; Shore, J.; Pines, A.; Dupree, R. *J. Am. Chem. Soc.* **1998**, *120*, 3510.

(16) Xue, X.; Kanzaki, M. *Phys. Chem. Miner.* **1998**, *26*, 14.

(17) Xue, X.; Kanzaki, M. *J. Phys. Chem. B* **1999**, *103*, 10816.

(18) Clark, T. M.; Grandinetti, P. J. *J. Phys.: Condens. Matter* **2003**, *15*, S2387.

(19) Xue, X.; Kanzaki, M. *Solid State Nucl. Magn. Reson.* **2000**, *16*, 245.

(20) Azizi, S. N.; Rostami, A. A.; Godarzia, A. A. *J. Phys. Soc. Jpn.* **2005**, *74*, 1609.

(21) Bussemer, B.; Schroder, K.-P.; Sauer, J. *Solid State Nucl. Magn. Reson.* **1997**, *9*, 155.

(22) Brouwer, D. H.; Enright, G. D. *J. Am. Chem. Soc.* **2008**, *130*, 3095.

(23) Brouwer, D. H. *J. Am. Chem. Soc.* **2008**, *130*, 6306.

(24) Civalleri, B.; Garrone, E.; Ugliengo, P. *Chem. Phys. Lett.* **1999**, *299*, 443.

(25) Casanovas, J.; Illas, F.; Pacchioni, G. *Chem. Phys. Lett.* **2000**, *326*, 523.

(26) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(27) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(28) Profeta, M.; Mauri, F.; Pickard, C. J. *J. Am. Chem. Soc.* **2003**, *125*, 541.

(29) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(30) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396.

(31) Adamo, C.; Barone, V. *J. Chem. Phys.* **2002**, *116*, 5933.

(32) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51.

(33) Zhao, Y.; Schultz, E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.

(34) Will, G.; Bellotto, M.; Parrish, W.; Hart, M. *J. Appl. Crystallogr.* **1988**, *21*, 182.

(35) Schmahl, W. W.; Swainson, I. P.; Dove, M. T.; Graeme-Barber, A. *Z. Kristallogr.* **1992**, *201*, 125.

(36) Smyth, J. R.; Smith, J. V.; Artioli, G.; Kvick, A. *J. Phys. Chem.* **1987**, *91*, 988.

(37) Lewis, J. E.; Freyhardt, C. C.; Davis, M. E. *J. Phys. Chem.* **1996**, *100*, 5039.

(38) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.

(39) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.

(40) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959.

(41) Meier, W. M.; Olson, D. H. *Atlas of Zeolite Structure Types*; Butterworth-Heinemann: Boston, 1992.

(42) Vaughan, P. A. *Acta Crystallogr.* **1966**, *21*, 983.

(43) Wise, W. S.; Tschernich, R. W. *Am. Mineral.* **1976**, *61*, 60.

(44) Attfield, M. P.; Weigel, S. J.; Cheetham, A. K. *J. Catal.* **1997**, *172*, 274.

(45) Haggin, J. C&EN 1993, October 25, p 30.

(46) 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. *Gaussian03, Revison C.02*; Gaussian, Inc.: Wallingford CT, 2004.

(47) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, *104*, 5497.

(48) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789.

(49) Barone, V. *Surf. Sci.* **1987**, *189−190*, 106.

(50) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

(51) Rappe, A. K.; Goddard, W. A. I. *J. Phys. Chem.* **1991**, *95*, 3358.

(52) Verhoeven, J.; Dynamus, A.; Bluyssen, H. *J. Chem. Phys.* **1969**, *15*, 3330.

(53) Magi, M.; Lippmaa, E.; Samoson, A.; Engelhardt, G.; Grimmer, A.-R. *J. Phys. Chem.* **1984**, *88*, 1518.

(54) Pfrommer, B.; Raczkowski, D.; Canning, A.; Louie, S. G. *PARATEC (PARAllel Total Energy Code)*; Lawrence Berkeley National Laboratory: with contributions from Mauri, F. C. M., Yoon, Y., Pickard, C., Haynes, P., Eds.; for more information see www.nersc.gov/projects/paratec. Page last modified: Thu, 03 Jun 2004 21:03:20 GMT.

(55) Pickard, C. J.; Mauri, F. *Phys. Rev. B* **2001**, *63*, 245101.

(56) Spearing, D. R.; Farnan, I.; Stebbins, J. F. *Phys. Chem. Miner.* **1992**, *19*, 307.

(57) Vaara, J.; Lounila, J.; Ruud, K.; Helgaker, T. *J. Chem. Phys.* **1998**, *109*, 8388.

(58) Wasylishen, R. E.; Mooibroek, S.; Macdonald, J. B. *J. Chem. Phys.* **1984**, *81*, 1057.

(59) Pyykko, P. *Mol. Phys.* **2001**, *99*, 1617.

(60) Cody, G. D.; Mysen, B.; Sàghi-Szabò, G.; Tossell, J. A. *Geochim. Cosmochim. Acta* **2001**, *65*, 2395.

(61) Tossell, J. A.; Horbach, J. *J. Phys. Chem. B* **2005**, *109*, 1794.

(62) Tossell, J. A. *Phys. Chem. Miner.* **2004**, *31*, 41.

(63) Huang, P.; Carter, E. A. *J. Chem. Phys.* **2006**, *125*, 084102.

(64) Ferrero, M.; Rerat, M.; Orlando, R.; Dovesi, R. *J. Comput. Chem.* **2008**, *29*, 1450.

(65) McCusker, L. B. *J. Appl. Crystallogr.* **1988**, *21*, 305.

JCTC Journal of Chemical Theory and Computation

Characterization of Chitin and Chitosan Molecular Structure in Aqueous Solution

Eduardo F. Franca,[†,‡] Roberto D. Lins,*,[†] Luiz C. G. Freitas,[‡] and T. P. Straatsma[†]

Pacific Northwest National Laboratory, Richland, Washington 99352, and Departamento de Química, Universidade Federal de São Carlos, São Carlos, SP, Brazil

Received July 24, 2008

Abstract: Molecular dynamics simulations have been used to characterize the structure of single chitin and chitosan chains in aqueous solutions. Chitin chains, whether isolated or in the form of a β-chitin nanoparticle, adopt the 2-fold helix with ϕ and φ values similar to its crystalline state. In solution, the intramolecular hydrogen bond $HO3_{(n)}\cdots O5_{(n+1)}$ responsible for the 2-fold helical motif in these polysaccharides is stabilized by hydrogen bonds with water molecules in a well-defined orientation. On the other hand, chitosan can adopt five distinct helical motifs, and its conformational equilibrium is highly dependent on pH. The hydrogen bond pattern and solvation around the O3 atom of insoluble chitosan (basic pH) are nearly identical to these quantities in chitin. Our findings suggest that the solubility and conformation of these polysaccharides are related to the stability of the intrachain $HO3_{(n)}\cdots O5_{(n+1)}$ hydrogen bond, which is affected by the water exchange around the O3-HO3 hydroxyl group.

Introduction

Chitin is the second most abundant organic material in nature after cellulose. It is a naturally occurring glucose-derivative polymer forming the primary component of fungal cell walls, arthropods, and certain algaes. Chitosan, its main derivative, is the partly N-deacetylated chitin. Deacetylation level almost never reaches 100%. It is generally accepted that this biopolymer is considered as chitosan if chitin is N-deacetylated to such a degree that it becomes soluble in dilute acidic medium, which is typically on the order of 40−60%.[1] The amine groups of the chitosan glucosamine units ($pK_a \sim$ 6.3−6.5) are important chelation sites for metal ions in neutral to alkaline conditions and form stable complexes with anions by electrostatic interactions in acidic conditions.[2−5] Unlike chitin, chitosan is positively charged in acidic to neutral conditions with a net charge density, and therefore its adsorption capacity is dependent on pH and N-deacetylation level.[6−8] The overall metal- and anion-binding capacities of these polymers result from the interplay between the degree of deacetylation, molecular weight, surface area, and crystallinity. Because of chitin and chitosan widespread availability, and of their strong affinity for ions, molecules, macromolecules and even microorganisms, a great variety of chitosan particles is engineered for an impressively wide range of applications including the following: removal of toxic radionuclide and heavy metals, recovery of precious metals, nitrogen fertilization in agriculture, pesticide removal, fine clay particle stabilization/coagulation (e.g. on some U.S. Department of Transport highways), alkaline fuel cell, tissue-engineering (e.g., fibers, textiles), biomaterial stabilization, catalysis support, food emulsification, paper food, fungal and bacterial disease prevention, drug delivery, pharmaceutical excipients, and as blood plasma cholesterol and intestinal triglycerides regulation in the human body.[9−15]

Natural chitin has three anhydrous crystalline polymorphs, α-, β-, γ-chitins, in the native state, which is found in the skeletal structure of crustaceans, insects, mushrooms, and the cell wall of fungi.[16] The structure of α and β forms differ only in that the piles of chains are arranged alternately antiparalel in a α-chitin, whereas they are all parallel in β-chitin.[17] The γ-chitin form has characteristics of both α and β forms, where two chains run in one direction and another chain in the opposite direction; however, it is considered only a variant of the α family, because it has the same properties as the α-chitin.[18] α-Chitin is the most abundant and also the

* Corresponding author phone: (509)375-2755; fax: (509)372-4720; e-mail: roberto.lins@pnl.gov.
† Pacific Northwest National Laboratory.
‡ Universidade Federal de São Carlos.

most stable thermodynamically,[19,20] and the β- and γ-chitin forms can be irreversibly converted into the α-form.[16] Chitosan has anhydrous and hydrated forms, with piles chains arranged in an antiparallel fashion.[21–23]

Experimental results obtained using various methods of polymer structure investigations (IR spectroscopy, NMR, X-ray scattering, microscopy, sorption techniques) showed that chitin and chitosan, just as cellulose, are characterized by an ordered fibrillar structure, a developed system of intra- and intermolecular hydrogen bonds, a high degree of crystallinity, and polymorphism.[13,24] According to X-ray crystallography experiments, chitin is a linear polysaccharide where the chains composed of 2-acetamido-2-deoxy-D-glucopyranose and linked by β-(1→4) glycoside bonds on a 2_1 screw axis have a repeating unit of 1.030 to 1.043 nm.[25–27] Similarly to chitin, both hydrated and the anhydrous forms of chitosan have a conventional extended 2-fold helical conformation with a repeating pattern every ca. 1.0 nm,[21,28] although acid salts of chitosan can favor different structural arrangements.[23] Three types of structural complexes have been observed for chitosan and can be classified according to their chain repeating unit. The first type, so-called type I form, is the anhydrous form, where in these crystals the backbone chitosan chains retain the extended 2-fold helix of the unreacted chitosan molecule.[21] The type II is a hydrated crystal with a chain repeat about 4.08 nm and asymmetric unit (repeating units) consisting of four glucosamine residues. Since two tetrasaccharides make a chain repeat, this is also a 2-fold helix even though the conformation is completely different to that of type I form, where the asymmetric unit is one glucosamine residue. Type II form is also known as relaxed 2-fold helix because it is almost four times longer than unreacted chitosan.[29,30] A type II salt variant, called type IIa, has a similar chain repeat (4.05 nm), but the molecular conformation is a 4/1 helix, with an asymmetric unit consisting of a glucosamine dimer. This right-handed helix comprises four asymmetric units.[30] The most recently discovered form, type III, has a chain pattern with a chain repeat of 2.55 nm, a 5-fold helix, and an asymmetric unit of one glucosamine residue.[31]

It is well-known that hydrogen bonds play an important part in the stabilization of secondary structures of proteins. Likewise, it has been proposed that the structure of chitin and chitosan chains are also stabilized by two hydrogen bonds, one intrachain and one interchain.[32] An intramolecular hydrogen bond is formed between atoms HO3 and O5 of consecutive linked glucosamine units. This intramolecular hydrogen bond ($HO3_{(n)} \cdots O5_{(n+1)}$) is responsible for the length of the chain repeating distance and to keep the chitin chain in a 2-fold helical pattern.[25,33] There are two possibilities of intermolecular hydrogen bonds in the solid state, between two N-acetyl groups (e.g., $N-H_{(i)} \cdots O=C_{(j)}$) and between one N-acetyl and a hydroxymethyl group. In α-chitin, each amounts for roughly 50% of the interchain interactions.[13] The latter pattern has not been reported in the crystal structures of β-chitin and is thought to account for the increased susceptibility to crystalline swelling of

β-chitin.[13] The hydrogen bond network is responsible for linking the chitin chains in arranged bonded piles or sheets.[19,26]

Other studies using X-ray crystallography,[27] ^{13}C solid-state NMR,[20] and molecular dynamics simulations[34] showed further information about the hydrogen bonds and conformations of the hydroxymethyl and the N-acetyl groups in solid state. However, there are no studies that described the dynamical properties of these hydrogen bonds and the possible orientations of these groups in chitin upon solvation. The only molecular study involving the presence of water as solvent is the X-ray structure of the monohydrate form of the chitin $[C_8H_{13}O_5N(H_2O)]_n$.[35] This study showed that water molecules taking part in the hydrogen bond system of chitin are stable and not unusual. Hydroxyl groups were found to be the main partners in the hydrogen bonds involving water molecules, which have a typical distance of 2.8 Å. (A hydrogen bond is assumed to be present when the hydrogen-acceptor distance is less than 3.5 Å and the donor-hydrogen-acceptor angle is larger than 135°.) Like chitin, the structure of chitosan also displays hydrogen bonds between the $O3_{(n)}$ and $O5_{(n+1)}$ oxygen atoms across the glycosidic linkages and by interchain hydrogen bonds involving the oxygen atom of the hydroxymethyl group (O6) of one chain and the amine nitrogen atom (N2) of the D-glucosamine unit of another chain. The patterns in polymer packing are therefore highly dependent on deacetylation levels. In addition, water molecules also play an important role in the packing, conformation, and mechanical properties of chitin and chitosan-based materials.[36,37] In hydrated forms of chitosan, the packing structure can be stabilized by several hydrogen bonds mediated by water molecules without direct interactions between sheets.[23] However, the solution conformation and the influence of the solvent on these polymers remain unclear and under debate.[36,38,39]

Toward a better understanding of these interactions at a molecular level, we have performed molecular dynamics simulations of chitin and chitosan chains under different conditions to address the role of solvation in the hydrogen bond pattern, structure, and solubility of these polymers. Within this context, the influence of ionic strength on the conformation of chitin and the influence of pH in the conformation of chitosan are also discussed.

Methodology

Molecular Systems and Simulation. A modeled 10-mer polysaccharide filament with $\phi \sim -60°$ and $\varphi \sim +110°$ (where ϕ and φ are defined by atoms O5−C1−O1−C4 and C1−O1−C4−C3, respectively) was used as starting structural framework in all simulations. The systems are summarized in Table 1. The polymer chains were placed in a rectangular simulation box with dimensions of $x = 3.0$, $y = 3.0$, and $z = 5.0$ nm, where a covalent glycosidic bond, with the necessary bonded terms, was defined across the periodic box, hence treating the filaments as infinite chains. The nanoparticle was formed by 9 chitin filaments arranged in a 3×3 matrix in a β-chitin configuration. The systems were then solvated by filling the box with SPC water model molecules.[40] Sodium and chloride ions were used to achieve

Characterization of Chitin/Chitosan Molecular Structure

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2143**

Table 1. Description of the Simulated Systems

system[a]	no. of solute atoms	no. of ions	no. of solvent atoms	degree of acetylation (%)	ionic strength (mol.Kg^{-1})	pH[b]
CHT	170	0	1365	100	0.0	n/a
CHT$_{0.4}$	170	10	1355	100	0.4	n/a
CHT$_{1.1}$	170	30	1335	100	1.1	n/a
CHS$_{LOW}$	160	10	1473	0	0.4	low
CHS$_{NEUTRAL}$	155	10	1394	0	0.4	neutral
CHS$_{HIGH}$	150	10	1477	0	0.4	high
nanochitin[c]	1530	0	6310	100	0.0	n/a

[a] CHT: chitin; CHS: chitosan. [b] pH values corresponds to low <4; neutral = 6.5; high >10. [c] Corresponds to a 9-chain β-chitin nanoparticle arranged into a 3 × 3 matrix comprising a total of 90 N-acetyl-glucosamine residues.

the desired ionic strengths. Each system was energy minimized using 10,000 steps of the steepest descent method. After minimization the solvent was equilibrated by performing 10 ps molecular dynamics simulation at 50, 150, and 300K, with non-hydrogen atoms positionally restrained (force constant 1.0×10^3 kJ.mol^{-1}.nm^{-2}). Following the 30 ps solvent equilibration, a total of 30 ns for the chains and 20 ns for the nanoparticles molecular dynamics simulations were performed in an isothermal−isobaric (NPT) ensemble using the leapfrog algorithm[41] with a 1-fs time step, therefore allowing structural flexibility of the filament along its axis. The configurations were recorded every 1 ps for analysis. During the MD simulation, at every time step, the translational and rotational motion of the center of mass was removed. The temperature was kept at 300 K by coupling the solutes and the solvent separately to Berendsen thermostats[42] with a relaxation time of 0.1 ps. The pressure was maintained at 1 bar by coupling to a Berendsen barostat[42] via semi-isotropic coordinate scaling with a relaxation time of 10 ps and a compressibility of 4.5×10^{-6} (kJ.mol^{-1}.nm^{-3})$^{-1}$. Water stretching and bending motions were constrained using the LINCS algorithm.[43] A 1.4 nm cutoff was used for the short-range electrostatics and van der Waals interactions. Long-range electrostatic contributions were treated via the generalized reaction field[44] with ε=66. All simulations were carried out using extensions of the GROMOS carbohydrate force field[45] within the GROMACS 3.3.2 program.[46]

Electrostatic Potential and Solvation Free Energy Calculations. Electrostatic potential and solvation free energy calculations were obtained by solving numerically the nonlinear Poisson−Boltzmann equation and applying a finite-difference procedure,[47−49] which was performed using the APBS (Adaptive Poisson−Boltzmann Solver) program[50] in conjunction with the GROMOS point charge parameter set for carbohydrates.[45] A dielectric constant for solvent of 78.54 C^2/N.m^2 with solvent radius of 1.4 nm, surface tension of 0.105 N/m, and ionic strength of 0.4 mM was used to describe the structures in aqueous solution. The internal dielectric constant of the solute was set to 1 C^2/N.m^{2}, and the apolar contribution to the solvation free energy was calculated using gamma equal to 0.105 kJ/mol. The three-

dimensional potentials were obtained using 129 grid points in the x, y, and z directions.

Results and Discussion

Ionic strength is well-known to play an important role on the ability of chitin and chitosan to aggregate forming different gels, beads, and nanoparticles in solution.[51] For example, a decrease of the persistence length of chitosan in gels has been proposed to be correlated with the increase of the ionic strength in the media.[51] Nevertheless, its effect on the chains of these polysaccharides remains elusive. Due to the nature of our setup, we are able to isolate the influence of the ionic strength on single chains of chitin in aqueous solution. It is worth noting that values of ϕ −60° and φ +110° were used as the starting point for all the initial structures. A time evolution analysis of chitin's ϕ and φ torsional angles shows the influence of the ionic strength on the structural convergence of the chitin chain, from the initial conformation (Figure 1a,b). It is clear from Figure 1 that an increase in ionic strength results in a slower convergence to its equilibrium structure. Chitin in the absence of salt reached its equilibrium structure (i.e., a 2-fold helix) after ca. 5 ns. To achieve this same configuration, a total of 25 ns of sampling were needed for the system with an ionic strength of 1.1 M NaCl. In all simulations the final values for ϕ and φ dihedral angles of fully solvated chains converge to the similar values of crystalline chitin. This is illustrated by the ϕ/φ maps for the last sampled 5 ns (Figure 1c-e). Available experimental values obtained by X-ray diffraction for two independent measurements for β-chitin and one for α-chitin crystal packing are also plotted for reference (Figure 1c-e). This result shows that fully solvated chitin chains tend to assume a typical nanoparticle crystal-like configuration and indicates that ionic strengths up to 1.1 M NaCl do not alter the equilibrium conformation of chitin in water at room temperature.

In order to further characterize the structure of chitin chains in aqueous solution, we have calculated its helicity as a function of the simulation time. It is well-known that the crystalline structure of chitin displays less conformational variability than chitosan.[13,23] The former maintains its configuration in the 2-fold helix and ϕ and φ values around −90° and +90°, respectively. On the other hand, chitosan can assume four main helical conformations: extended 2-fold helix, relaxed 2-fold helix, 4/1 helix, and 5/3 helix. The ϕ and φ values for chitosan in a 2-fold helix shifts slightly to $\phi \cong$ −98 and $\varphi \cong$ 92;[21] the 4/1 helix can exist in two forms, $\phi_1 =$ −66.4, $\omega_1 =$ 121.6 and $\phi_2 =$ −75.0, $\omega_2 =$ 126.9[28] and has a repetition unit composed of two residues, thus generating the full 4-fold helix using the space group P4$_1$ (2 residues with a translation after a 90° of rotation). No specific values of ϕ and φ are associated with 5-fold and relaxed 2-fold helices since they can adopt a wide range of values.[30,31] Due to the effects of solvation and higher temperature (300 K) on the polymer, compared to a crystal structure, we have chosen not to take into account the all crystalline parameters that define chitin and chitosan helical motifs.[23] Instead, the helical periodicity (helicity and chain repeat) was used as criterion, with a ±15°-variation allow-
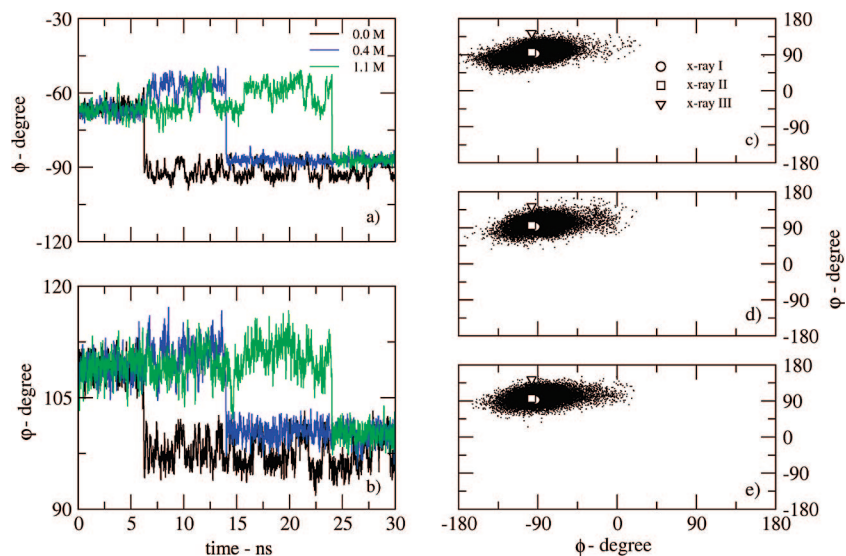
Figure 1. ϕ and φ torsional angles for chitin at different ionic strengths: a) and b) as a function of simulation time and c-e) distribution map for the last 5 ns for the systems at c) 0.0 M, d) 0.4 M, and e) 1.1 M ionic strength. Three independent experimentally determined values of ϕ and φ from α- (X-ray II[27]) and β-chitin (X-ray I;[34] X-ray III[26]) crystals are plotted for reference in c-e). a) and b) values are shown as a 20-ps interval averages.
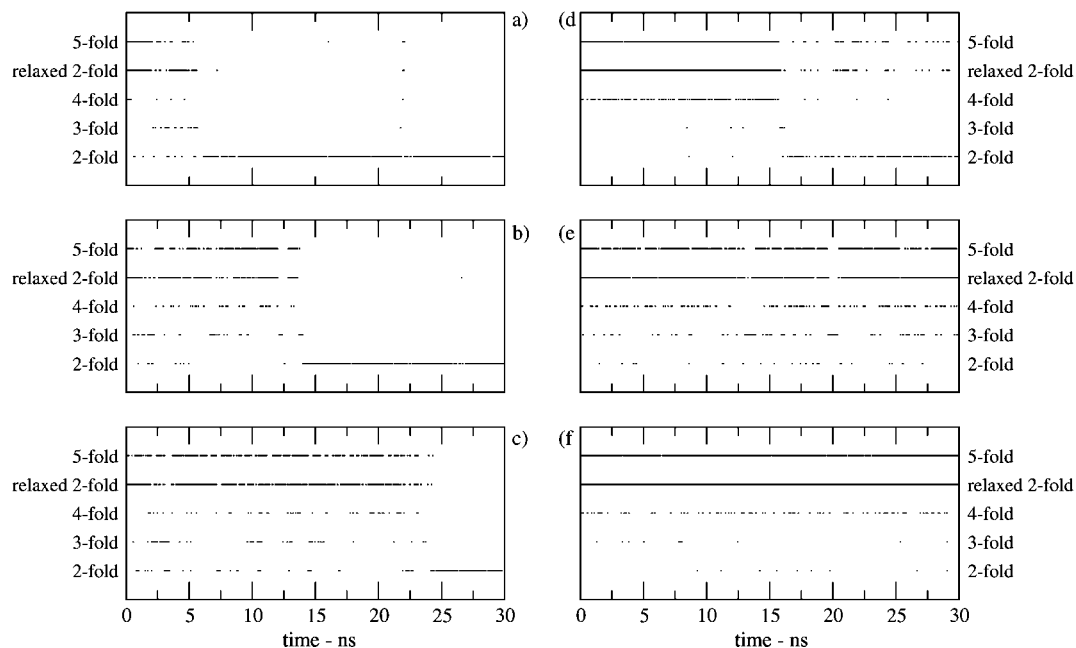
Figure 2. Structural pattern for a-c) chitin and d-f) chitosan chains in aqueous solution as a function of time: a) chitin, 0.0 M; b) chitin, 0.4 M; c) chitin 1.1 M; d) chitosan at high pH; e) chitosan at neutral pH; and f) chitosan at low pH.

ance. The algorithm employed here was based on the virtual bond method developed by Zugenmaier and Sarko to analyze the structure of fibrous polymers such as polysaccharides.[52] Thanks to the nature of our setup (i.e., an infinite chain along the periodic box) chitosan conformation was described according to Ogawa's classification of the crystal structure of chitosans.[23,53−55] According to this approach, five motifs have been probed: 2-fold helix (2 residues per turn forming a *zigzag* motif), 3-fold helix (right-handed helix with 3 residues per turn), 4-fold helix (right-handed helix with 8 residues per turn where four sets of two residues are rotated by 90° around the principal axis to complete one turn), relaxed 2-fold (right-handed helix with 8 residues per turn where two sets of four residues are rotated by 180° around

the principal axis to complete one turn), and 5-fold helix (left-handed helix with 5 residues per turn). If the configuration of the entire chain, at any given time, fails to fulfill the requirements of each motif, it is then classified as having a random configuration, and it is not displayed in the presented graphs. (A graphical representation of the X-ray structures for the chitin and chitosan helical motifs can be found in a comprehensive review by Ogawa and co-workers).[23]

Time dependent evolution of helical configuration preferences are shown in Figure 2 for the chains at 0.0, 0.4, and 1.1 M ionic strength. Once again the structural behavior of the chains shows that chitin dynamics is slowed down with the increase of ionic strength, which seems to stabilize the

Characterization of Chitin/Chitosan Molecular Structure

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2145**

initial configuration (Figure 2). As expected, all the chains converge to the 2-fold helix, and no significant difference is observed in helical motif distributions for the last 5 ns of simulation. It is worth noting that all chains have visited all possible configurations before reaching equilibrium indicating that simulation setup is insensitive to the initial conformation (with the exception of the 2-fold helix motif). These findings indicate that chitin dynamics and conformation may be transiently altered by interactions with ionic species. This is supported by a recent experimental work where a solution of NaOH/urea (8 wt%/4 wt%, respectively) was shown to solubilize chitin at temperatures around −10 °C.[56] In the proposed mechanism, Na$^+$ ions would facilitate water molecules to enter and disrupt the crystalline structure of chitin by breaking interchain hydrogen bonds. The presence of urea was verified to be necessary to stabilize solubility as gelation would take place in the presence of the salt only. In the currently presented simulations, the highest ionic strength setup is equivalent to a solution of roughly 4 wt% NaCl concentration.

It is well-known that interchain hydrogen bonds in chitin are responsible for its highly stable crystalline state.[19,25,57] It is also known that N-acetylglucosamine is soluble in water and slightly soluble in ethanol. However, our results show that a chitin chain in aqueous solution shows low conformational exchange and similar degree of rigidity to the crystalline state (based on ϕ/φ distributions and helical conformation). This cannot be explained solely by the high torsional barrier around the glycosidic bond resulting from the anomeric effect, otherwise chitosan should also behave similarly, which is clearly not the case. Figure 2d-f shows that chitosan presents a higher conformational diversity than chitin, and it becomes less structured in acidic medium. At high pH, the 2-fold helix is clearly the dominant configuration followed by the 5-fold and relaxed 2-fold conformations. The 2-fold helix is rarely visited at neutral pH, and the less structured 5-fold and relaxed 2-fold are the two preferred conformations at low pH.

Analyses of the intrachain hydrogen bond pattern in chitin reveal only one persistent hydrogen bond, namely HO3$_{(n)}$···O5$_{(n+1)}$, which seems to be the driving force behind the stabilization of chitin in a 2-fold helix motif (details in Figure 3). This interaction favors φ to adopt values locking the chain into the 2-fold configuration. A similar behavior was observed for the simulation of a β-chitin nanoparticle (Table 2). The β-nanochitin setup comprised 9 chains assembled in a 3 × 3 arrangement solvated by water molecules. Therefore the central chain in the system is completely shielded from water molecules. Analysis of the interactions of the central chain with the other filaments showed three types of persistent hydrogen bond. The intramolecular hydrogen bond HO3$_{(n)}$···O5$_{(n+1)}$ and intermolecular hydrogen bonds between the N-acetyl groups of different chains have been both described to occur in crystals and are responsible for the bonded piles or sheet arrangement in chitin.[19,26] In addition to these interactions, our simulation shows that ca. 20% of the intermolecular hydrogen bonds are between the hydroxymethyl group of one chain with the N-acetyl group of another chain of the nanoparticle (Table
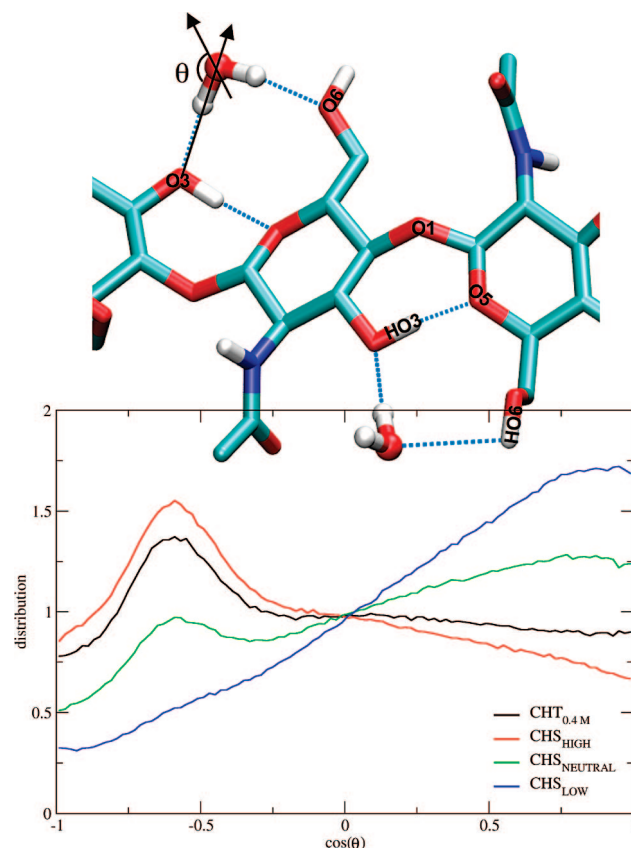
Figure 3. Solvent (water) orientation around the O3 oxygen atom of chitin and chitosan chains. Property is shown as the average orientation (cos θ) over the entire simulation for a 5-Å radius around the O3 atom. θ is defined by the angle formed between the vector formed by O3 and the water oxygen atoms and the water dipole vector (as illustrated by the black arrows). Hydrogen bonds are represented by dashed blue lines. Water molecules are displayed in a ball-and-stick model and the sugar units of chitin in sticks (atoms are color coded as cyan: carbon; blue: nitrogen, red: oxygen, and white: hydrogen).

2). To date, these interactions have not been observed in crystals of β-chitin. This discrepancy may be due to i) the relatively low occurrence (20%) and weak character of these hydrogen bonds (lifetime is half of that of hydrogen bonds between two N-acetyl groups); ii) experimental measurements were obtained from anhydrous crystals, which typically display very low or no conformational variability; or iii) a combination of both. For the internal chain of the nanoparticle, the frequency of O3-HO3$_{(n)}$···O5$_{(n+1)}$ intramolecular hydrogen bond reaches nearly 90%, and its lifetime increases almost 60% compared to the solvated chain. If the O3-HO3$_{(n)}$···O5$_{(n+1)}$ intramolecular hydrogen bond is responsible for maintaining a chitin chain in a 2-fold helix motif and absence of solvation in these regions strengthens such interactions, it is reasonable to propose that solubility of chitin is inversely related to the stability of this interaction.

In order to verify this hypothesis, three 30 ns simulations of fully solvated chitosan chains (100% N-deacetylated chitin) in water were carried out at low, ca. neutral, and high pH. It is worth noting that chitosan is only fully soluble at pH below 6.[13] The helical propensities and hydrogen-bond patterns were calculated for the three simulations and are

***Table 2.*** Number and Lifetime of Hydrogen Bonds in Chitin and Chitosan Averaged for the Entire Trajectory

system	$HO3_{(n)} \cdots O5_{(n+1)}$		$O3\text{-}HO3 \cdots O_{WATER}$		$H_{WATER} \cdots O3$	
	time[a]	number[b]	time[a]	number[b]	time[a]	number[b]
			Chitin			
CHT	6.13	0.69 ± 0.14	2.73	0.19 ± 0.11	3.10	1.08 ± 0.16
$CHT_{0.4}$	5.88	0.63 ± 0.15	3.32	0.24 ± 0.12	3.08	1.08 ± 0.16
$CHT_{1.1}$	5.34	0.58 ± 0.15	3.25	0.25 ± 0.13	3.02	1.08 ± 0.17
			Chitosan			
CHS_{LOW}	2.71	0.53 ± 0.14	1.69	0.27 ± 0.13	1.10	0.26 ± 0.14
$CHS_{NEUTRAL}$	4.83	0.48 ± 0.12	2.23	0.29 ± 0.13	2.26	0.72 ± 0.16
CHS_{HIGH}	5.60	0.59 ± 0.15	1.84	0.19 ± 0.12	2.66	1.25 ± 0.18

nanochitin (central chain)	$HO3_{(n)} \cdots O5_{(n+1)}$		$NAc \cdots NAc$		$NAc \cdots OH_{MET}/OH_{MET} \cdots NAc$	
	time[a]	number[b]	time[a]	number[b]	time[a]	number[b]
			Chitin Nanoparticle			
	9.77	0.86 ± 0.11	7.61	0.51 ± 0.09	4.23	0.22 ± 0.14

[a] Time given in ps. [b] Average number per sugar unit; NAc: N-acetyl group; OH_{MET}: carbohydrate hydroxymethyl group.

illustrated in Figure 2d-f and Table 2. The average number of the $HO3_{(n)} \cdots O5_{(n+1)}$ intrachain hydrogen bond in chitosan is not significantly different from the observed values in chitin or as a function of the pH. However, dramatic changes are observed in i) the number of water molecules hydrogen bonding O3 and ii) the lifetime of the $HO3_{(n)} \cdots O5_{(n+1)}$ interaction. The higher the pH the closer these numbers are to the values observed in chitin. For example, for chitosan in low pH the $HO3_{(n)} \cdots O5_{(n+1)}$ hydrogen bond lifetime is only half of the value of chitin (3 versus 6 ps, respectively), while for chitosan at high pH the values for the average number and residence time of a water molecule interacting with O3 is nearly identical to the values observed in chitin. These differences in residence times suggest that chain conformation and solubility are affected by microsolvation entropy.

To probe such a hypothesis, the orientation of all the water molecules within a 5-Å radius from all O3 atoms was calculated and averaged over the entire simulation. It is displayed in Figure 3 as the cosine of the angle (θ) formed by the dipole vector of the water molecule with the vector defined by O3 with the water oxygen. The averaged water molecule orientation profile for chitin at 0.4 M is shown along with the profiles for chitosan chains at the same ionic strength but at different pH values. No significant difference was observed for the solvent orientation profile of chitin as a function of ionic strength (data not shown for conciseness). The peak observed for chitin around −0.7 is due to a hydrogen atom of a water molecule that interacts with the oxygen atom O3 of the sugar ring, leaving the HO3 available to hydrogen bond the O5 oxygen atom of the next sugar unit. The spatial distribution of the N-acetyl and the hydroxylmethyl groups define a pocket, which is occupied by a water molecule. The position of this water molecule is stabilized by a hydrogen bond with either the hydrogen or oxygen atom of the hydroxymethyl group of the neighboring sugar (Figure 3). The NAc groups were not observed to take part in this arrangement.

Accordingly for chitosan, the solvent orientation profile around the region becomes increasingly similar to chitin's profile as pH increases, which leads to a decrease in
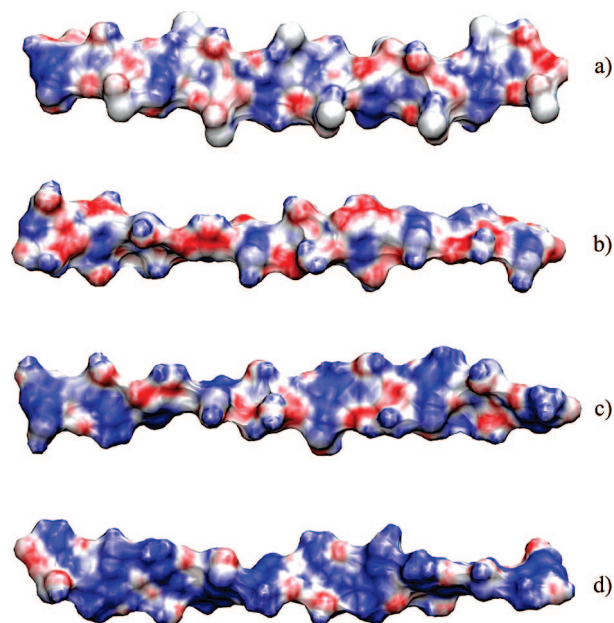
Figure 4. Electrostatic potential plotted onto the molecular surface of a) chitin and b-d) chitosan; b) chitosan at high pH; c) chitosan at neutral pH; and d) chitosan at low pH. Positive (blue) and negative (red) potentials correspond to the range of −50 $k_B T/e$ to + 50 $k_B T/e$.

solubility. A nearly identical profile for the orientation and distribution of the water molecules around this region is observed for chitin and chitosan at high pH (the two most insoluble species). The analysis of the electrostatic potential reveals an increased availability of negatively charged sites at the surface of chitosan with the increase of pH (Figure 4). It does not come as a surprise the fact that the molecular surfaces of chitin and chitosan at high pH exhibit a similar electrostatic profile (Figure 4). The visualization of this property allows the understanding of how water molecules are solvating the polymer. The waters in the first solvation shell of chitosan at low pH interact with the sugars mainly via their oxygen atoms. In contrast, a considerably large amount of water molecules would hydrogen bond chitin via their hydrogen atoms (as is shown in Figure 3). The conformational variation of the chitosan chains is also in

Characterization of Chitin/Chitosan Molecular Structure

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2147**

Table 3. Residue Average Contribution (kcal/mol) to the Free Energy of Solvation of Chitin and Chitosan Chains

system	$\Delta G_{ELECTROSTATIC}$	ΔG_{APOLAR}	ΔG
$CHT_{0.4}$	−15.0	6.1	−8.9
CHS_{LOW}	−50.6	5.2	−45.4
$CHS_{NEUTRAL}$	−27.4	5.2	−22.2
CHS_{HIGH}	−15.2	5.3	−9.9

accordance with the hypothesis that the stability of the intramolecular $HO3_{(n)}\cdots O5_{(n+1)}$ hydrogen bond favors the 2-fold helical configuration. Time dependent conformational analysis of the chitosan simulations show that the 2-fold helix motif is rarely visited by the chain at low pH (soluble). This trend is altered by pH increase, and the 2-fold helix motif becomes the most representative configuration for the setup at high pH (Figure 2d-f).

A number of contributing components to chitosan solubility, such as intramolecular hydrogen bonding, van der Waals forces, and hydrophobic interactions, have been reported.[58–60] However, a recent experimental study suggests that the solubility of chitosan in aqueous solution is primarily attributed to electrostatics forces between the polymer and the medium.[61] To shed some light on the problem, the electrostatics and apolar contributions to the free energy of solvation of chitin and chitosan chains were calculated via Poisson–Boltzmann electrostatics using the APBS program[50] (see Methodology). Representative structures having their corresponding equilibrium conformation were selected from the last 5 ns of each simulation with the same ionic strength, so that pH was the only variable among the different chitosan chains. The contributions to the free energy of solvation were calculated and are presented in Table 3 as the average contribution per residue.

The estimated solvation free energy values cannot be compared to experimental values due to the absence of such measurements in the literature. However, in agreement with experimental observations on chitin and chitosan solubility, a clear trend shows that chitosan at low pH is the most soluble form, whereas chitin and chitosan at basic pH are the least soluble chains. This suggests that these calculations can be potentially used to predict relative solubility for the different forms of chitosan. Once again, remarkable similarities can be seen for the calculated solvation free energy contributions for chitin and chitosan chains at basic pH. It is also clear that solubility in chitin and chitosan chains is mainly driven by electrostatic contributions. Moreover, it is interesting to note that the electrostatic contributions to the solvation free energy are inversely proportional to the residence times of the intramolecular $HO3_{(n)}\cdots O5_{(n+1)}$ hydrogen bond and highly correlated with the average orientation of water molecules around the O3 oxygen. We have also investigated the influence of different ionic strengths on the electrostatic contribution. The latter is restricted to a few percentile units for salt concentrations from 0.15 to 1.10 mM (data not shown for conciseness).

Conclusion

We have shown from a series of molecular dynamics simulations of chitin and chitosan chains that the flexibility of the polysaccharide chain is inversely related to the stability of the intramolecular hydrogen bond between the atom HO3 of one sugar unit and the O5 atom of the next monosaccharide. As experimentally observed for crystals, solvated chitin chains also assume mostly the 2-fold helix motif, and its equilibrium conformation was not affected by ionic strengths up to 1.1 M, in agreement with rheological studies of chitin in solution that portrays it as a semirigid polysaccharide.[13] The stability of the intrachain hydrogen bond $HO3_{(n)}\cdots O5_{(n+1)}$ seems to be heavily affected by water exchange and residency times in its neighborhood. It presents a higher persistence in the internal chains of nanoparticles where the presence of the solvent is excluded. Although weak and of low occurrence, our simulation suggests that intermolecular hydrogen bonds between the hydroxymethyl and N-acetyl groups are also present in solvated β-chitin nanoparticles. The charged amino group (NH_3^+) at the neighboring C2 atom of chitosan at low pH values seems to increase water exchange in the region of the O3 atom destabilizing the $HO3_{(n)}\cdots O5_{(n+1)}$ hydrogen bond. This interaction in chitin and chitosan is responsible for locking the polymer conformation into a 2-fold helical motif, reducing conformational variability. The calculated solvation free energies indicate that solubility in chitosan is controlled mainly by electrostatic interactions. The high correlation of the electrostatic contributions to the solvation free energies with hydrogen bond patterns suggest that the solubility of these polymers is inversely related to their ability to form intramolecular $HO3_{(n)}\cdots O5_{(n+1)}$ hydrogen bonds. The presented data draw for the first time a parallel between the intrachain hydrogen bond pattern in chitin and chitosan chains and their solubility. These findings are expected to provide insights into the rationale of chitin and chitosan-based nanomaterial design.

References

(1) Sannan, T.; Kurita, K.; Iwakura, Y. Studies on Chitin. 2. Effect of Deacetylation on Solubility. *Makomol. Chem.* **1976**, *177* (12), 3589–3600.

(2) Gibbs, G.; Tobin, J. M.; Guibal, E. Sorption of Acid Green 25 on Chitosan: Influence of Experimental Parameters on Uptake Kinetics and Sorption Isotherms. *J. Appl. Polym. Sci.* **2003**, *90* (4), 1073–1080.

(3) Knorr, D. Dye Binding-Properties of Chitin and Chitosan. *J. Food Sci.* **1983**, *48* (1), 36–37.

(4) Maghami, G. G.; Roberts, G. A. F. Studies on the Adsorption of Anionic Dyes on Chitosan. *Makromol. Chem.* **1988**, *189* (10), 2239–2243.

(5) Uragami, T.; Yoshida, F.; Sugihara, M. Studies of Synthesis and Permeabilities of Special Polymer Membranes 0.51. Active-Transport of Halogen Ions through Chitosan Membranes. *J. Appl. Polym. Sci.* **1983**, *28* (4), 1361–1370.

(6) Liu, C. X.; Bai, R. B. Adsorptive Removal of Copper Ions with Highly Porous Chitosan/Cellulose Acetate Blend Hollow Fiber Membranes. *J. Membr. Sci.* **2006**, *284* (1−2), 313–322.

(7) Vieira, R. S.; Beppu, M. M. Dynamic and Static Adsorption and Desorption of Hg(II) Ions on Chitosan Membranes and Spheres. *Water Res.* **2006**, *40* (8), 1726–1734.

(8) Zhou, D.; Zhang, L.; Guo, S. L. Mechanisms of Lead Biosorption on Cellulose/Chitin Beads. *Water Res.* **2005**, *39* (16), 3755–3762.

(9) Dutta, P. K.; Dutta, J.; Tripathi, V. S. Chitin and Chitosan: Chemistry, Properties and Applications. *J. Sci. Ind. Res.* **2004**, *63* (1), 20–31.

(10) Kim, I. Y.; Seo, S. J.; Moon, H. S.; Yoo, M. K.; Park, I. Y.; Kim, B. C.; Cho, C. S. Chitosan and Its Derivatives for Tissue Engineering Applications. *Biotechnol. Adv.* **2008**, *26* (1), 1–21.

(11) Ravi Kumar, M. N. V. A Review of Chitin and Chitosan Applications. *React. Funct. Polym.* **2000**, *46* (1), 1–27.

(12) Revoredo, O. B.; Nieto, O. M.; Suarez, Y.; Garcia, V.; Fernandez, M.; Iraizoz, A.; Henriques, R. D. Applications of Chitin and Chitosan in Pharmacy and Cosmetology. *Eur. J. Pharm. Sci.* **2006**, *28*, S7–S8.

(13) Rinaudo, M. Chitin and Chitosan: Properties and Applications. *Prog. Polym. Sci.* **2006**, *31* (7), 603–632.

(14) Ruel-Gariepy, E.; Leroux, J. C. *Chitosan: A Natural Polycation with Multiple Applications*; American Chemical Society: Montreal, 2006; p 384.

(15) Smitha, B.; Sridhar, S.; Khan, A. A. Chitosan-Poly(Vinyl Pyrrolidone) Blends as Membranes for Direct Methanol Fuel Cell Applications. *J. Power Sources* **2006**, *159* (2), 846–854.

(16) Rudall, K. M.; Kenching, W. Chitin System. *Biol. Rev. Camb. Philos. Soc.* **1973**, *48* (4), 597.

(17) Sugiyama, J.; Boisset, C.; Hashimoto, M.; Watanabe, T. Molecular Directionality of Beta-Chitin Biosynthesis. *J. Mol. Biol.* **1999**, *286* (1), 247–255.

(18) Atkins, E. Conformations in Polysaccharides and Complex Carbohydrates. *J. Biosci. (Bangalore)* **1985**, *8* (1−2), 375–387.

(19) Jang, M. K.; Kong, B. G.; Jeong, Y. I.; Lee, C. H.; Nah, J. W. Physicochemical Characterization of Alpha-Chitin, Beta-Chitin, and Gamma-Chitin Separated from Natural Resources. *J. Polym. Sci., Part A: Polym. Chem.* **2004**, *42* (14), 3423–3432.

(20) Kameda, T.; Miyazawa, M.; Ono, H.; Yoshida, M. Hydrogen Bonding Structure and Stability of Alpha-Chitin Studied by C-13 Solid-State Nmr. *Macromol. Biosci.* **2005**, *5* (2), 103–106.

(21) Yui, T.; Imada, K.; Okuyama, K.; Obata, Y.; Suzuki, K.; Ogawa, K. Molecular and Crystal-Structure of the Anhydrous Form of Chitosan. *Macromolecules* **1994**, *27* (26), 7601–7605.

(22) Okuyama, K.; Noguchi, K.; Miyazawa, T.; Yui, T.; Ogawa, K. Molecular and Crystal Structure of Hydrated Chitosan. *Macromolecules* **1997**, *30* (19), 5849–5855.

(23) Ogawa, K.; Yui, T.; Okuyama, K. Three D Structures of Chitosan. *Int. J. Biol. Macromol.* **2004**, *34* (1−2), 1–8.

(24) Mogilevskaya, E. L.; Akopova, T. A.; Zelenetskii, A. N.; Ozerin, A. N. The Crystal Structure of Chitin and Chitosan. *Polym. Sci., Ser. A* **2006**, *48* (2), 116–123.

(25) Blackwel, J. Structure of Beta-Chitin or Parallel Chain Systems of Poly-Beta-(1•4)-N-Acetyl-D-Glucosamine. *Biopolymers* **1969**, *7* (3), 281–298.

(26) Gardner, K. H.; Blackwell, J. Refinement of Structure of Beta-Chitin. *Biopolymers* **1975**, *14* (8), 1581–1595.

(27) Minke, R.; Blackwell, J. Structure of Alpha-Chitin. *J. Mol. Biol.* **1978**, *120* (2), 167–181.

(28) Lertworasirikul, A.; Yokoyama, S.; Noguchi, K.; Ogawa, K.; Okuyama, K. Molecular and Crystal Structures of Chitosan/ Hi Type I Salt Determined by X-Ray Fiber Diffraction. *Carbohydr. Res.* **2004**, *339* (4), 825–833.

(29) Okuyama, K.; Noguchi, K.; Kanenari, M.; Egawa, T.; Osawa, K.; Ogawa, K. Structural Diversity of Chitosan and Its Complexes. *Carbohydr. Polym.* **2000**, *41* (3), 237–247.

(30) Lertworasirikul, A.; Tsue, S.; Noguchi, K.; Okuyama, K.; Ogawa, K. Two Different Molecular Conformations Found in Chitosan Type II Salts. *Carbohydr. Res.* **2003**, *338* (11), 1229–1233.

(31) Kawahara, M.; Yui, T.; Oka, K.; Zugenmaier, P.; Suzuki, S.; Kitamura, S.; Okuyama, K.; Ogawa, K. Fourth 3d Structure of the Chitosan Molecule: Conformation of Chitosan in Its Salts with Medical Organic Acids Having a Phenyl Group. *Biosci. Biotechnol. Biochem.* **2003**, *67* (7), 1545–1550.

(32) Carlstrom, D. The Crystal Structure of Alpha-Chitin (Poly-N-Acetyl-D-Glucosamine). *J. Biophys. Biochem. Cytol.* **1957**, *3* (5), 669–683.

(33) Ramarkrishnan, C.; Prasad, N. Study of Hydrogen Bonds in Amino Acids and Peptides. *Int. J. Protein Res.* **1971**, *3* (4), 209–231.

(34) Yui, T.; Taki, N.; Sugiyama, J.; Hayashi, S. Exhaustive Crystal Structure Search and Crystal Modeling of Beta-Chitin. *Int. J. Biol. Macromol.* **2007**, *40* (4), 336–344.

(35) Dweltz, N. E. Structure of Beta-Chitin. *Biochim. Biophys. Acta* **1961**, *51* (2), 283–294.

(36) Li, Q. X.; Song, B. Z.; Yang, Z. Q.; Fan, H. L. Electrolytic Conductivity Behaviors and Solution Conformations of Chitosan in Different Acid Solutions. *Carbohydr. Polym.* **2006**, *63* (2), 272–282.

(37) Mazeau, K.; Rinaudo, M. The Prediction of the Characteristics of Some Polysaccharides from Molecular Modeling. Comparison with Effective Behavior. *Food Hydrocolloids* **2004**, *18* (6), 885–898.

(38) Pedroni, V. I.; Gschaider, M. E.; Schulz, P. C. Uv Spectrophotometry: Improvements in the Study of the Degree of Acetylation of Chitosan. *Macromol. Biosci.* **2003**, *3* (10), 531–534.

(39) Pedroni, V. I.; Schulz, P. C.; Gschaider, M. E.; Andreucetti, N. Chitosan Structure in Aqueous Solution. *Colloid Polym. Sci.* **2003**, *282* (1), 100–102.

(40) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.

Characterization of Chitin/Chitosan Molecular Structure

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2149**

(41) Hockney, R. W. The Potential Calculation and Some Applications. In *Methods in Computational Physics*; Alder, B., Fernbach, S., Rotenberg, M., Eds.; Academic Press: New York/London, 1970; Vol. 9.

(42) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.

(43) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. Lincs: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.

(44) Tironi, I. G.; Sperb, R.; Smith, P. E.; Van Gunsteren, W. F. A Generalized Reaction Field Method for Molecular-Dynamics Simulations. *J. Chem. Phys.* **1995**, *102* (13), 5451–5459.

(45) Lins, R. D.; Hunenberger, P. H. A New Gromos Force Field for Hexopyranose-Based Carbohydrates. *J. Comput. Chem.* **2005**, *26* (13), 1400–1412.

(46) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. Gromacs: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.

(47) Davis, M. E.; McCammon, J. A. Calculating Electrostatic Forces from Grid-Calculated Potentials. *J. Comput. Chem.* **1990**, *11* (3), 401–409.

(48) Antosiewicz, J.; Gilson, M. K.; McCammon, J. A. Acetylcholinesterase - Effects of Ionic-Strength and Dimerization on the Rate Constants. *Isr. J. Chem.* **1994**, *34* (2), 151–158.

(49) Nicholls, A.; Honig, B. A Rapid Finite-Difference Algorithm, Utilizing Successive over-Relaxation to Solve the Poisson-Boltzmann Equation. *J. Comput. Chem.* **1991**, *12* (4), 435–445.

(50) Holst, M.; Baker, N.; Wang, F. Adaptive Multilevel Finite Element Solution of the Poisson-Boltzmann Equation I. Algorithms and Examples (Vol 21, Pg 1319, 2000). *J. Comput. Chem.* **2001**, *22* (4), 475–475.

(51) Vachoud, L.; Zydowicz, N.; Domard, A. Physicochemical Behaviour of Chitin Gels. *Carbohydr. Res.* **2000**, *326* (4), 295–304.

(52) Zugenmaier, P.; Sarko, A. In *The Variable Virtual Bond, Fiber Diffraction Methods, 1980*; French, A. D., Gardner, K. H., Eds.; American Chemical Society: Washington, DC, ACS Symposium Series, 1980; pp 225–237.

(53) Ogawa, K. Chain Conformation of Chitin and Chitosan. *J. Agric. Chem. Soc. Japan* **1988**, *62* (8), 1225–1228.

(54) Ogawa, K.; Yui, T. Crystallinity of Partially N-Acetylated Chitosans. *Biosci. Biotechnol. Biochem.* **1993**, *57* (9), 1466–1469.

(55) Ogawa, K.; Yui, T. Effect of Explosion on the Crystalline Polymorphism of Chitin and Chitosan. *Biosci. Biotechnol. Biochem.* **1994**, *58* (5), 968–969.

(56) Hu, Y.; Ding, Y.; Ding, D.; Sun, M. J.; Zhang, L. Y.; Jiang, X. Q.; Yang, C. Z. Hollow Chitosan/Poly(Acrylic Acid) Nanospheres as Drug Carriers. *Biomacromolecules* **2007**, *8* (4), 1069–1076.

(57) Mazeau, K.; Winter, W. T.; Chanzy, H. Molecular and Crystal Structure of a High-Temperature Polymorph from Chitosan and Electron Diffraction Data. *Macromolecules* **1994**, *27*, 7606–7612.

(58) Amiji, M. M. Pyrene Fluorescence Study of Chitosan Self-Association in Aqueous Solution. *Carbohydr. Polym.* **1995**, *26*, 211–213.

(59) Nystrom, B.; Kjoniksen, A. L.; Iversen, C. Characterization of Association Phenomena in Aqueous Systems of Chitosan of Different Hydrophobicity. *Adv. Colloid Interface Sci.* **1999**, *79*, 81–103.

(60) Philippova, O. E.; Volkov, E. V.; Sitnikova, N. L.; Khokhlov, A. R.; Desbrieres, J.; Rinaudo, M. Two Types of Hydrophobic Aggregates in Aqueous Solutions of Chitosan and Its Hydrophobic Derivative. *Biomacromolecules* **2001**, *8* (2), 483–490.

(61) Lavertu, M.; Filion, D.; Buschmann, D. Heat-Transfer of Protons from Chitosan to Glycerol Phosphate Chitosan Precipitation and Gelation. *Biomacromolecules* **2008**, *9*, 640–650.

JCTC Journal of Chemical Theory and Computation

Milestones in the Activation of a G Protein-Coupled Receptor. Insights from Molecular-Dynamics Simulations into the Human Cholecystokinin Receptor-1

Christophe Chipot*

Equipe de dynamique des assemblages membranaires, UMR No 7565, Nancy Université BP 239, 54506 Vandœuvre-lès-Nancy cedex, France

Abstract: Activation of G protein-coupled receptors (GPCRs) obeys an allosteric mechanism triggered by ligand binding. To understand how the signal is transduced in the cell, identification of the milestones paving the pathway between the active and the inactive states of the receptor is necessary. A model of the human cholecystokinin receptor-1 (CCK1R) has been proposed recently. The complex formed by CCK1R and an agonist ligand will serve as a paradigm of an active conformation to capture milestones in GPCR activation. To reach this goal, assuming microreversibility, the initial step toward the inactivation of CCK1R was modeled using free energy calculations, whereby the ligand is removed from the binding pocket. However accurate the reproduction of the experimental affinity constant, this simulation only represents an embryonic stage of the inactivation process. Starting from the *apo* receptor, an unprecedented 0.1-μs molecular dynamics trajectory was generated, bereft of experimental biases, bringing into the light key events in the inactivation of CCK1R, chief among which the hydration of its internal cavity, concomitant with the spatial rearrangement of the transmembrane helical segments. Hydration is intimately related to the isomerization of the highly conserved residue W326 of helix VI, acting as a two-state toggle switch, and of residue M121 of helix III. In the active state, the former residue obstructs the crevice, thereby preventing water leakage, which would otherwise trigger the disruption of an ionic lock between helices II and III involving the signature E/DRY motif ubiquitous to GPCRs.

Introduction

G protein-coupled receptors (GPCRs) constitute the largest family of membrane proteins responsible for signal transduction across the biological membrane, mediating the cellular response to a host of environmental stimuli.[1,2] A variety of pathologies are rooted in the malfunction of these proteins, which, not too surprisingly, have rapidly become privileged targets for drug discovery.[3] Activation of GPCRs follows an allosteric mechanism actuated by ligand binding and resulting in a conformational modification of its seven-helix transmembrane (TM) domain.[4] A complete understanding of how the signal is transduced across the cell membrane

requires at its core the identification of the milestones that pave the pathway connecting the active state of the receptor to its inactive counterpart. A convenient framework for apprehending the activation of GPCRs is provided by the extended ternary complex model, which assumes that active and inactive conformations of the receptor coexist in an equilibrium.[5] The complexity of the activation process is further magnified by the possibility of the receptor to adopt a variety of active conformations,[6,7] to the extent that different agonist ligands can bind distinct active forms of the same GPCR, triggering in turn different signaling pathways.[8]

The dearth of structural information readily available for GPCRs has imparted a new momentum to the computational investigation of these biological systems.[9] Access to the

* Corresponding author phone: +33.3.83.68.40.97; e-mail: Christophe.Chipot@edam.uhp-nancy.fr.

three-dimensional structure of bovine rhodopsin,[10,11] the paradigm of family-A GPCRs, has opened new vistas for the modeling of structurally related membrane proteins. Within the past ten years, considerable effort has been invested to predict the structure and the function of GPCRs, based on an arsenal of modeling tools ranging from first principles[12] to knowledge-based approaches.[13,14] Models built following the latter route can be used profitably for the design of novel site-directed mutagenesis experiments, which, in turn, can serve to refine the model of the receptor and decipher the mechanisms whereby it accomplishes its cellular function. Combining synergistically theory and experiment, a model of the human cholecystokinin receptor-1 (CCK1R)—a GPCR pertaining to family A, was put forth recently[15] and probed by means of large-scale molecular dynamics (MD) simulations and free energy calculations targeted at reproducing the binding affinities from whence the model had been constructed.[16] Cholecystokinin[17] (CCK) is a hormone ubiquitous to the gastrointestinal system, where it mediates digestion, while in the central nervous systems, it acts as a neurotransmitter, which, among others, stimulates satiety. The proposed model[15] consists of a complex, henceforth referred to as CCK1R*:CCK9, and formed by the receptor, presumably in an active state, and the agonist ligand CCK9, a nonapeptide analogue of the endogenous ligand. It is the fruit of several years of a relentless exchange between theory and experiment, confronting hypotheses and inferences to obtain a self-consistent picture of the membrane receptor (see refs 18−21).

The number of theoretical studies aimed at characterizing the activation of GPCRs has remained hitherto limited, primarily due to the paucity of experimentally validated three-dimensional structures, and, to a somewhat lesser extent, the formidable computational effort involved in their simulation. A common feature shared by these investigations is the use of experimental biases, viz. usually in the form of harmonic restraints, to convert the inactive conformation into an active one and *vice versa*.[22−25] In the present work, a different strategy is followed to dissect the milestones of GPCR activation, whereby, assuming microreversibility, the inactivation of the receptor will be modeled, bereft of external biases, starting from the structure of the CCK1R*:CCK9 complex. Inactivation of the receptor is an intricate process, anticipated to span time scales that are not routinely amenable to classical, all-atom MD simulations. This process subsumes the dissociation of the agonist ligand from the native binding pocket, assumed here to be a rapid event, and the ensuing internal relaxation of the receptor, which constitutes the rate-limiting step of the overall transformation. The first step, consisting of removing the agonist ligand CCK9 from the binding pocket of the receptor, will be examined employing free energy perturbation (FEP) calculations. As will be seen in the following sections, provided that appropriate sampling is performed, accurate reproduction of experimental binding affinities can be achieved, even for very large, supramolecular assemblies. The second step, during which the constitutively active *apo* receptor, CCK1R*, relaxes toward the inactive conformation, CCK1R° will be tackled in a 0.1-μs MD simulation.

Computational investigations of GPCR activation based on three-dimensional models remain admittedly scarce[24] and are burdened by the pervasive dogmatic assumption that, in general, no relevant qualitative, let alone quantitative information can be inferred from calculations that rest on an unknown parameter: The structure. To a large extent, this assertion explains why efforts to disentangle the intricate activation mechanism of GPCRs[22,25−27] have remained focused on structures solved experimentally, namely the bovine rhodopsin[10,11] and the human β_2-adrenergic[28] receptors. Does it mean that attempts to gain new insights from numerical simulations of GPCR models are necessarily vain as long as the structure of the membrane protein is not solved? Whereas simulations over appreciably long time scales evidently cannot validate a three-dimensional receptor model, they may, nonetheless, be used fruitfully to probe hypotheses drawn from experiments.[16] In the particular instance of CCK1R, accepting the view that refined homology- or template-based constructs are inevitably wrong would be tantamount to ignoring the wealth of data accrued in the past ten years,[15,18−21,29−34] and during which the structural model of the CCK1R*:CCK9 complex has been built, progressively improved, and thoroughly tested. On the other hand, it is fair to recognize that the aforementioned tenet can be fueled and, thus, reinforced by controversies relying upon conflicting models—e.g. the hypothesis of a reverted agonist ligand interacting at its C-terminus end with residue W39[35] of the receptor, challenged by NMR studies of CCK bound to a fragment of CCK1R.[36]

The primary thrust of this work is not an unrealistic claim that an inactive conformation of CCK1R has been isolated, but rather the demonstration that unrestrained, all-atom MD can be used beneficially to delineate how the structural features of the active receptor evolve as its conformation slowly interconverts into an inactive one. Shedding new light on the molecular mechanisms that underlie activation, large-scale MD simulations complement experimental advances by rationalizing the lower affinity of the ligand for the receptor upon conformational transition. As will be seen in the present contribution, central to this structural modification is the role played by the hydration of the internal cavity triggering a variety of events, which can evidently only be captured in a biologically realistic environment.

Methods

Three-Dimensional Model and Molecular Assay. The initial three-dimensional molecular assembly consisting of the complex formed by CCK1R* and agonist ligand CCK9,[15] inserted in a fully hydrated POPC bilayer, which altogether represents a total of 72,255 atoms, was equilibrated in a 31 ns MD simulation.[16] Over that period, it was observed that the distance rmsd computed over the TM α-helices never exceeded *ca.* 2.5 Å, revealing no alteration of their secondary structure.[16] The essential of the key protein-ligand interactions brought to light experimentally were preserved throughout the trajectory (see refs 18−21).

Free Energy Calculation. The free energy change associated with the binding of CCK9 to CCK1R was estimated

2152 *J. Chem. Theory Comput., Vol. 4, No. 12, 2008*

Chipot
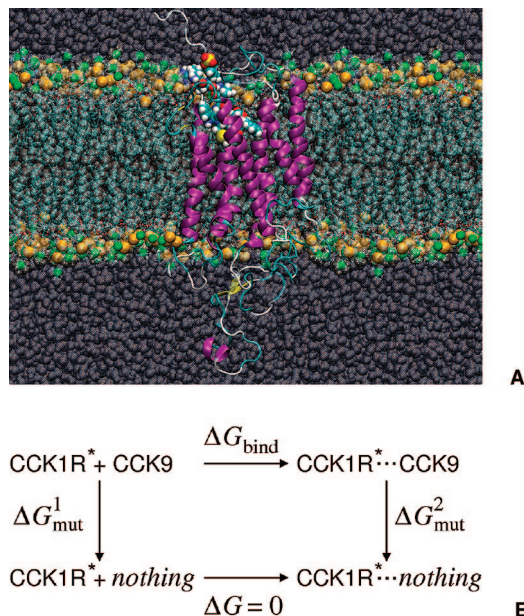
Figure 1. **A**. CCK1R*:CCK9 assembly in a fully hydrated POPC bilayer after 31 ns.[16] TM α-helices are shown as purple ribbons. The agonist ligand CCK9 is depicted as a space-filling molecular model. A semitransparent representation is used for water and lipid units, the phosphate and choline groups of which are featured as orange and green van der Waals spheres, respectively. The image rendering was done with Vmd.[65] **B**. Thermodynamic cycle used for the double annihilation of agonist ligand CCK9 in the free, hydrated state (left leg) and in the receptor (right leg). The binding free energy, ΔG_{bind}, which is determined experimentally, corresponds to the difference $\Delta G_{mut}^2 - \Delta G_{mut}^1$.

by means of a double annihilation[37] of the ligand, as described in Figure 1. Transformations in bulk water and in the receptor were performed employing the free energy perturbation (FEP) method,[38] wherein the Gibbs free energy difference between two thermodynamic states connected by $M-2$ intermediate, nonphysical states is expressed as

$$\Delta G = -\frac{1}{\beta} \sum_{i=1}^{M-1} \ln\langle exp\{-\beta[U(x;\lambda_{i+1}) - U(x;\lambda_i)]\}\rangle_{\lambda_i} \quad (1)$$

where $\beta = 1/k_BT$, k_B is the Boltzmann constant, T is the temperature, and $U(x;\lambda_i)$ is the potential energy function that depends upon the Cartesian coordinates of the system $\{x\}$ and the coupling parameter, λ_i, which connects the initial and the final states of the annihilation. $\langle \cdots \rangle_{\lambda_i}$ denotes an ensemble average over configurations representative of intermediate state i. Because in the course of the alchemical transformations charged amino acids vanish, the reaction path was stratified into 110 stages of uneven widths.[37] Narrow intermediate states were defined toward the end point of the simulation to avoid singularities. Each transformation, whether in the aqueous medium or in the receptor, involved 10 ps of equilibration followed by 100 ps of data collection, corresponding to a total of 12.1 ns. The associated error was estimated from a first-order expansion of the free energy (see the Supporting Information).
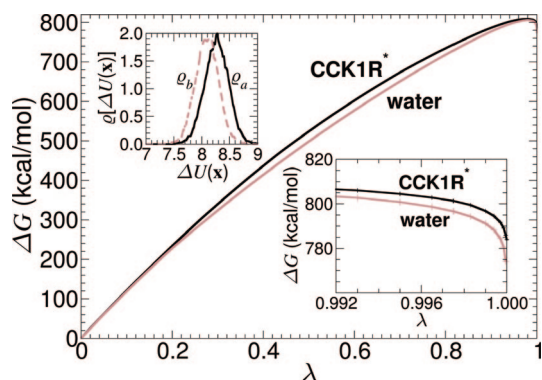
MD Simulation. All simulations were performed using the Namd simulation package[39] in the isobaric-isothermal ensemble. The pressure and the temperature were fixed at 1 bar and 310 K, respectively, employing the Langevin piston algorithm[40] and softly damped Langevin dynamics. The molecular assays were replicated in the three directions of Cartesian space by means of periodic boundary conditions. The particle-mesh Ewald method[41] was employed to compute electrostatic interactions. The *r*-RESPA multiple time-step integrator[42] was used with a time step of 2 and 4 fs for for short- and long-range forces, respectively. Covalent bonds involving a hydrogen atom were constrained to their equilibrium length. The CCK1R*:CCK9 complex and its environment were described by the all-atom Charmm27 force field.[43,44] The main MD simulation, wherein the initial conformation of the *apo* receptor corresponds to the end point of the FEP calculation during which the agonist ligand was annihilated in the CCK1R*:CCK9 complex, was conducted over a period of 0.1 μs. In addition, a shorter, 50-ns control simulation was performed, starting with the equilibrated structure of the complex,[16] from whence the nonapeptide, CCK9, was removed abruptly. Running on an array of twelve 2.40-GHz Intel Xeon processors communicating via a gigabit network, the wall clock time was equal to 37.6 h per nanosecond.

Results and Discussion

Binding of the Agonist Ligand to the Receptor. Direct measurement of the free energy difference associated with the inactivation of the receptor goes far beyond the current scope of classical, all-atom MD simulations. Under the assumption that the conformation of the membrane protein is preserved upon dissociation of the ligand—i.e. the structural features of the constitutively active *apo* receptor, CCK1R*, and the ligand-bound receptor, CCK1R*:CCK9, remain unperturbed—this early event in the inactivation process can be tackled using free energy methods. Microreversibility further suggests that the binding free energy of the agonist ligand to the receptor can be estimated by simulating the dissociative process, whereby CCK9 is annihilated in the CCK1R*:CCK9 complex depicted in Figure 1 and in a bulk aqueous environment, following the thermodynamic cycle[37] in the same figure.

This so-called alchemical transformation is considerably more challenging than the simple *in silico* point mutations of selected amino-acid residues reported in ref 16, due to the larger perturbation involved.[37] Convergence of the ensemble average featured in eq 1 imposes that all accessible configurations of the solvent—whether the membrane protein or the aqueous medium—around the vanishing ligand be sampled thoroughly. As can be seen in Figure 2, appropriate overlap of the thermodynamic ensembles embodied in their density of states indicates that this requirement is likely to be fulfilled[45] (see the Supporting Information). Noteworthily, the binding free energy, ΔG_{bind}, consists of a difference of two large terms, the magnitude of which is dictated primarily by the ionic solvation of the participating charged residues. Based on two 12.1-ns FEP calculations, viz. in CCK1R* and in water, and employing the acceptance ratio method of Bennett that combines forward and reverse transformations,[46,45] ΔG_{bind} is found to be equal to -11.5 ± 1.2 kcal/mol. This

Activation of a G Protein-Coupled Receptor

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2153**

Figure 2. Evolution of the Gibbs free energy as a function of the general extent parameter, λ, utilized to annihilate CCK9. Upper inset: Overlapping configurational ensembles embodied in the density of states, $\varrho[\Delta U(x)]$, of contiguous intermediate states (here, at $\lambda = 0.5$)—where $\Delta U(x)$ denotes the variation of the potential energy between state a and state b, which is a function of Cartesian coordinates $\{x\}$ (see the Supporting Information). Lower inset: Closeup on the ultimate intermediate states of the transformation. ΔG_{bind} corresponds to the difference between the dark curve (bound state) and the light curve (free state).

estimate compares reasonably well with the experimental value of -12.6 ± 1.4 kcal/mol, which relies upon repeated radio-activity measurements[16,19] of transfected cells expressing the receptor after exposure to the radio-iodinated agonist ligand.[47]

The binding free energy of -11.5 ± 1.2 kcal/mol can be interpreted as a sum of contributions of different natures. Association of CCK9 to CCK1R proceeds through strong electrostatic interactions involving (i) down in the crevice, aspartate D8 of the agonist ligand and residue R336 of the receptor,[19] and (ii) above, in the upper region of the binding pocket, sulfated tyrosine sY3 of CCK9 and amino acids M195 and R197 of CCK1R. As was shown earlier, point mutations of residues D8 and sY3 into alanine yield a loss in the binding free energy equal to 3.2 ± 0.3 and 2.7 ± 0.1 kcal/mol, respectively[16,19]—i.e. roughly speaking, one-half of the total binding free energy. The affinity of CCK9 toward CCK1R can be further explained in terms of robust van der Waals contacts[21] formed between residues M7 and F9 of the agonist ligand and a network of hydrophobic residues of the receptor, which includes L53, V125, I329, and I352. Additional interactions of lesser strength involve the N-terminus of nonapeptide CCK9 and residues W39 and Q40 of CCK1R[48] and appear to be complemented by ancillary, water-mediated interactions that embrace amino acids T4, G5, W6, and F9 of the agonist ligand and residues F107, T117, T118, and A341 of the receptor.[16]

The present free energy calculation raises, however, several key issues, chief among which is the level of confidence that can be assigned to the estimate of the binding affinity. From a dogmatic point of view, such a computation represents a numerically challenging endeavor, based on a seemingly fragile construct, and, hence likely to yield a series of numbers that have unaccountable reliability and validity. Even for smaller ligands binding a structure determined experimentally,[49,50] accurate predictions of free energy

differences by means of FEP calculations, which are notoriously plagued by convergence issues,[37] has proven to remain a difficult undertaking. Yet, although the present simulations rely upon a model of the receptor that cannot be indisputably validated until the three-dimensional structure of the latter is solved, it ought to be reminded that a host of experimental data[15,18−21,29−34] lends support to the putative placement of the ligand in the binding pocket. Moreover, precise reproduction of relative binding affinities for the CCK1R*:CCK9 complex[16] not only strengthens this contention but also suggests that the proposed methodology and sampling strategy are sufficiently robust to measure the free energy characterizing the association of the agonist ligand to the receptor with appreciable confidence.

Structural Features of the *apo* Receptor over the 0.1-μs Time Scale. It is interesting to observe that as the agonist ligand vanishes in the binding pocket, water molecules from the bulk environment rapidly seep in the receptor. Figure 3 reveals that after 0.1 μs of MD sampling, the *apo* structure is essentially flooded, which, as will be seen shortly, necessarily bears some implications on the conformational equilibrium of the membrane protein. The density profiles characterizing the location of the different components of the molecular assembly are shown in Figure 3. The nonzero probability to find water molecules across the protein is suggestive of a possible communication between the periplasmic and the cytoplasmic sides of the model membrane.

The density profiles also shed light on the organization of the receptor in its membrane environment. The width of the density characterizing the transmembrane (TM) domain of CCK1R roughly coincides with that of the acyl chains of the lipid bilayer. The marked mobility of the extracellular loops is mirrored in the broad distribution of their position along the normal to water-membrane interface.[16] This is especially true for the cytoplasmic loops, which, in the absence of the G protein subunits, drift without much hindrance. The distance root mean-square deviation (rmsd) of the receptor highlighted in Figure 3 increases by approximately 1 Å within the first 50 ns of the simulation and stays roughly constant at *ca.* 2 Å over the remaining 50 ns. The same figure indicates that the secondary structure of the receptor remains unaltered on the 0.1-μs time scale, thereby suggesting that the moderate deviation monitored with respect to the initial conformation, viz. CCK1R*, should be ascribed preferentially to internal motions within the TM scaffold.

Spatial Rearrangement of the Receptor toward Its Inactive Conformation. A distance rmsd represents a global, average measure of the deformations undergone by the membrane protein. It, therefore, does not indicate where the structural changes are localized. As underlined previously, the time scale spanned by classical, all-atom MD is *a priori* too short to capture the complete inactivation process. On the other hand, the only known structure of an activated GPCR, the bovine rhodopsin, reveals that the modifications with respect to its dark, inactivated form[10] are much smaller than hitherto expected.[11] Intricate local movements within the TM α-helix bundle can be disentangled in terms of
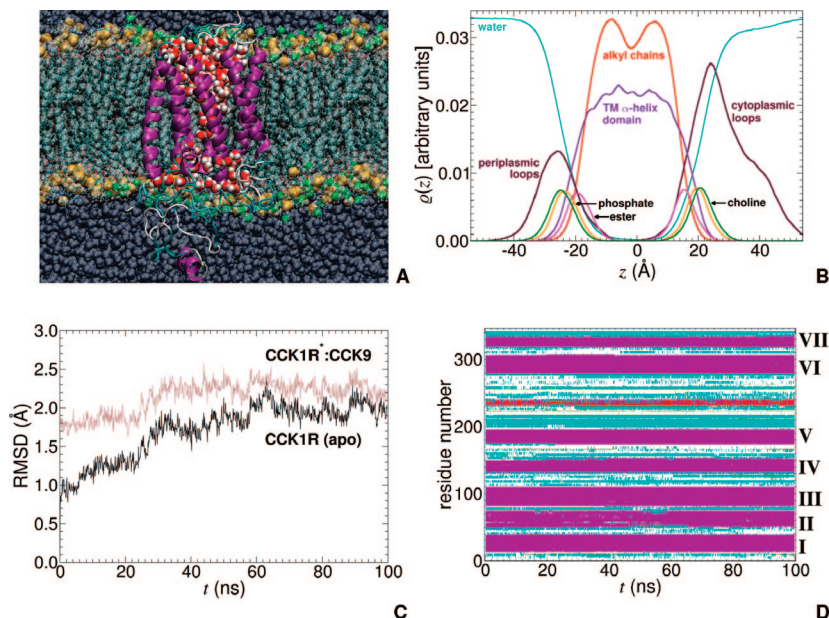
Figure 3. **A**. *apo* CCK1R in a fully hydrated POPC bilayer after 100 ns. Color coding is identical to that of Figure 1. Hydration water molecules flooding into the receptor are highlighted as van der Waals spheres (oxygen atoms in red and hydrogen atoms in white). **B**. Number density profiles of the *apo* CCK1R assembly averaged over the entire trajectory. **C**. Time-evolution of the distance rmsd over backbone atoms of *apo* CCK1R with respect to the initial structure at $t = 0$ (dark curve) and the CCK1R*:CCK9 complex (light curve). **D**. Time-evolution of the secondary structure of *apo* CCK1R. α-, 3_{10}-, and π-helices are shown in purple, pink, and red, respectively. Coils and turns are depicted as white and cyan tubes.

translational and rotational motions of the individual α-helical segments. A quantification of these movements is provided in Figure 4, distinguishing between interhelical separations, Δr, and rotation, Ω, of individual TM α-helices about their longitudinal axis as a function of time.

Using the CCK1R*:CCK9 complex as a reference, the geometrical analysis of the 0.1-μs trajectory illuminates a sequence of events in the reorganization of the receptor deprived of its agonist ligand. The most striking feature of the latter is the lateral translation of TM III, TM IV, TM V, and TM VI, while TM I, TM II, and TM VII remain virtually fixed in Cartesian space. In a seemingly concerted fashion, within the first 30 ns of the simulation, the interhelical separation of TM IV and TM V and of TM IV and TM VI increases, while TM V and TM VI move closer to each other. During the same period, TM II and TM III are initially pulled apart, prior to returning to their original separation, and eventually diminishing the latter. The distance between the centers of mass of TM I and TM VII is virtually unchanged over the entire simulation, in line with the site-directed spin labeling experiments on bovine rhodopsin.[51] Past the first 50 ns of the MD trajectory, only marginal fluctuations of Δr are measured. It is noteworthy that on the 0.1-μs time scale, variations of the different interhelical separations remain moderate, but in the most glaring examples of the TM II and TM V and the TM IV and TM VI pairs of α-helices depicted in Figure 4, they can be as large as *ca.* 2 Å. The continuous translation of TM III toward TM VI by *ca.* 1.5 Å is of somewhat lesser amplitude yet reflects previous measurements on the dark state of rhodopsin.[52] There is evidently a structural rationale for these broader displacements, as will be seen in the following sections.

Rotation of the α-helical segments is in general confined within a $\pm 25°$ range. In reference with the CCK1R*:CCK9 complex, it is apparent that preliminary rotation of the individual TM α-helices occurred in the course of the alchemical transformation, whereby the agonist ligand was annihilated in the binding pocket. In the second half of the simulation, TM II rotates clockwise, albeit intermittently, congruent with the approach of TM III and TM V. No perceptible rotation is monitored for TM III, in line with previous observations for another cholecystokinin receptor, CCK2R, in light of biased MD simulations.[24] In contrast, TM IV rotates rapidly, yet in a transient fashion, within the first 40 ns, before returning to its original state. Rotation of TM IV upon activation has been previously described, for instance, in the case of the β_2 adrenergic receptor.[53,54] Reminiscent of bovine rhodopsin,[55] variations of Ω is significant for TM VI, undergoing an anticlockwise writhing motion as it moves closer to TM V and away from TM IV. A symmetrical behavior has been witnessed in the activation process of CCK2R.[24]

Hydration of the Binding Pocket Is an Early Event in the Inactivation of the Receptor. Figure 3 sheds light on the internal hydration of the receptor, conducive to a possible communication between the aqueous phases on the cytoplasmic and the periplasmic sides of the lipid bilayer. How can this communication be established? Initial simulations of the CCK1R*:CCK9 complex confirmed that the conduit formed by the binding pocket is too narrow to allow water molecules to crawl through when the agonist ligand is locked in.[16] They also revealed higher in the binding pocket the existence of a cluster of water molecules with residence times spanning the nanosecond time scale and communicating with the extracellular environment. It be-

Activation of a G Protein-Coupled Receptor

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2155**
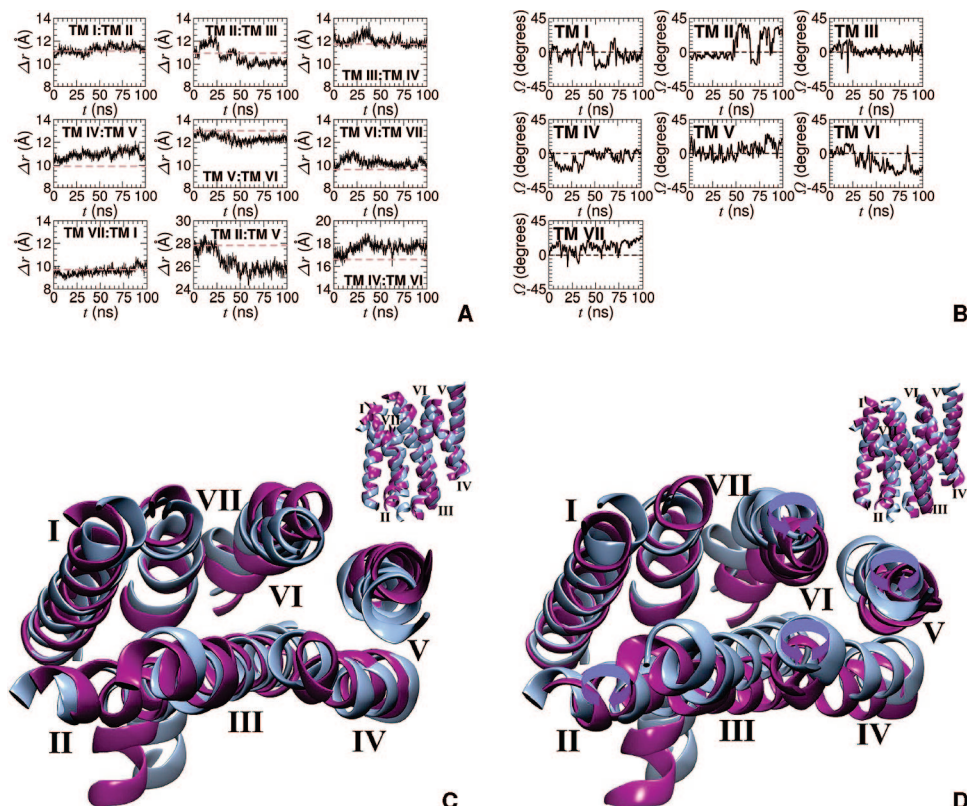
Figure 4. **A**. Time-evolution of the distance separating the center of mass of the α-helices forming the TM domain of *apo* CCKR1. Comparison with the corresponding separations in the CCK1R*:CCK9 complex (dashed line). **B**. Rotation of the TM helices of *apo* CCKR1 as a function of time. **C**. Top and side views of the TM region of *apo* CCK1R at *t* = 0 (purple). Comparison with the CCK1R*:CCK9 complex (pale blue). **D**. Top and side views of the TM region of *apo* CCK1R at *t* = 100 ns (purple).

comes obvious that in the absence of CCK9 blocking the access to the crevice, these water molecules can diffuse freely and accumulate at the bottom of the pocket.

As depicted in Figure 5, a well-conserved tryptophan residue, W326, is located underneath the binding pocket of CCK1R. Its position, within the simulation time, obeys a two-state regime driven by torsional angle χ_2, which is either equal to *ca.* 0° in the constitutively active receptor or to *ca.* 90° otherwise. Remarkably enough, an abrupt, first-order transition of this dihedral angle is observed around 27 ns, coinciding with the maximum separation of both the TM III and TM IV and the TM IV and TM VI α-helix pairs. Reorientation of the indole ring is accompanied by a concerted rearrangement of TM V and TM VI, leading to double π-π stacking, W326 being intercalated between F187 and F330. On the basis of a model of phenylalanine, a π-π interaction consisting of two optimally stacked aromatic rings has been shown to amount to *ca.* −3.4 kcal/mol.[56] This π-π interaction is expected to be reinforced in the case of tryptophan, as a result of an increased dispersion contribution. In addition, cooperative, Axilrod-Teller-like effects are envisioned to tighten even further the interaction of the two α-helical segments, thereby contributing to the overall stability of the TM scaffold.

This toggle-switch[57] isomerization in the side chain of W326 corresponds to virtually the maximum hydration of the receptor, whereby an excess of 300 water molecules occupy the interior of the α-helix bundle. As a basis of

comparison, using the same criteria, the number of water molecules in the CCK1R*:CCK9 complex was found to be *ca.* three times less (see Figure 5). That the internal hydration of GPCRs is modified substantially in the course of their activation appears to be a ubiquitarian property, as was recently underlined by Grossfield et al. in the light of extended simulations of bovine rhodopsin[58]—albeit in the latter, the marked flow of water molecules follows the isomerization of the Schiff base, which is distinct in essence from the actual removal of a noncovalently bound ligand from its designated binding pocket. The present result suggests that flooding of the receptor occurred in the course of the free energy calculation, concomitantly as the agonist ligand vanishes. It is also reflected in the initial translational and rotational motions of the TM α-helices. It is probable that reorientation of the conserved tryptophan side chain of TM VI is a signature mechanism ubiquitous to the activation or inactivation of GPCRs, as has been observed previously in the case of bovine rhodopsin, both at the experimental[59,60] and the theoretical level.[23] Yet, it is worth noting that in metarhodopsin II, isomerization of W265[61,27] occurs through its χ_1 torsional angle, and not χ_2, as described here for CCK1R.

The Role of the E/DRY and NPxxY Motifs in the Activation of CCK1R. In light of site-directed mutagenesis experiments, it has been suggested that one of the important events in the activation of class A GPCRs is the protonation
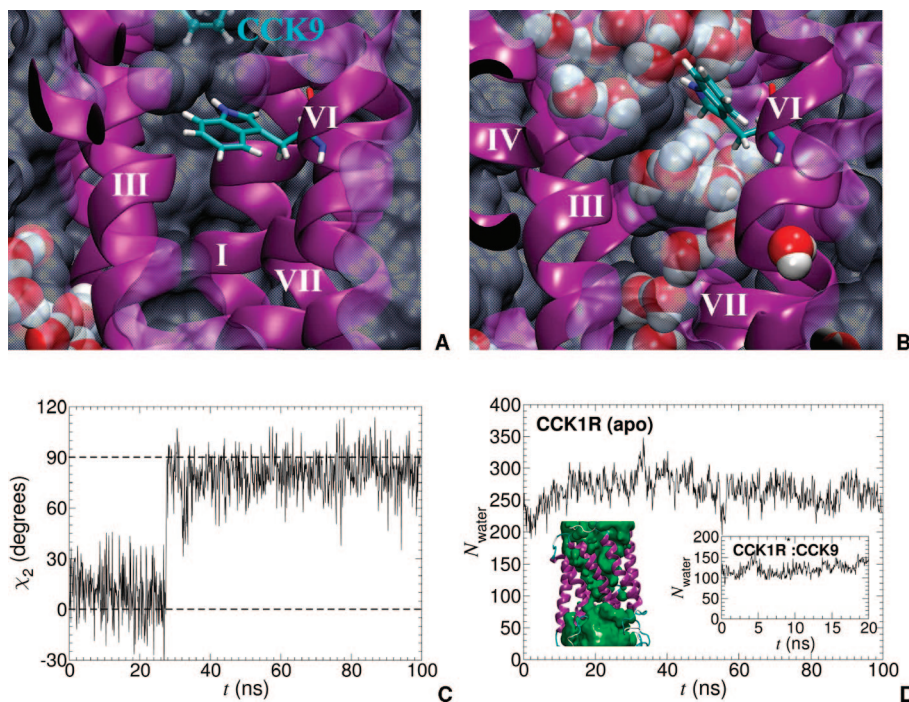
Figure 5. **A**. Cavity accommodating agonist ligand CCK9 in CCK1R*:CCK9 complex. The toggle switch W326 residue of TM helix VI is down. **B**. After 100 ns, the cavity in *apo* CCK1R is filled with water. W326 is up. **C**. Time-evolution of the χ_2 torsional angle of W326 in *apo* CCK1R. **D**. Instantaneous number of water molecules in the cavity of *apo* CCK1R. Inset: mass-weighted density of water in *apo* CCK1R measured over a period of 20 ns and highlighting a possible conduction between the cytoplasmic and the periplasmic face of the membrane (left). Number of water molecules in the cavity of CCK1R*:CCK9 (right). The density was computed with the VolMap module of Vmd.[65]

of the aspartic acid pertaining to the highly conserved E/DRY motif located at the bottom of TM III, on the cytoplasmic side of the membrane.[62] Computational studies have lent support to this conjecture, proposing, for instance, that during activation of the gonadotropin-releasing hormone receptor, an aspartate residue of TM III is protonated and subsequently replaced by another one in TM II to form a salt bridge with an arginine residue of TM III.[63] Experiment has also demonstrated that D83 of rhodopsin is hydrogen bonded upon activation, suggestive of a potential interaction with a neighboring residue in the active state of the protein.[64]

In the CCK1R*:CCK9 complex, E138 and R139 of TM III point in the same direction, albeit interact only mildly on account of their promiscuity. On the other hand, R139 forms a steady salt bridge with D87 of TM II, following a canonical C_{2v} symmetry pattern,[15] as highlighted in Figure 6. In the *apo* receptor, this ionic lock is preserved for approximately 15 ns, seemingly unaffected by the early separation of TM II and TM III. Its strength, however, subsides as the internal cavity of *apo* CCK1R reaches its maximum hydration, within roughly speaking 40 ns. In the meantime, the salt bridge breaks and reforms intermittently, until the two residues are sufficiently screened by the aqueous environment to no longer interact.

Another region of paramount importance for the activation of GPCRs is the NPxxY motif of TM VII. In the particular instance of CCK1R, mutation of N366 has been shown to abolish activation.[15] Over the 0.1-μs time scale, the present simulation reveals only marginal fluctuations about the original position of N366. Moreover, the hydrogen bond formed between N366 and R310 of the third intracellular

loop is fully preserved. The same, however, cannot be said for residue M121 of TM III, which has been recognized to be involved indirectly in the activation of CCK1R through its hydrophobic environment.[29] Pointing toward the periplasm in the CCK1R*:CCK9 complex, M121 undergoes a conformational transition around 27 ns from *ca.* -50 to $-180°$. Interestingly enough, this transition occurs concomitantly with the flip of W326 and the approach of TM II and TM III or TM V highlighted in Figures 5 and 4, respectively.

Relevance of the Molecular Dynamics Trajectories. At this stage of the study, is it legitimate to claim that after 0.1 μs, CCK1R has fully relaxed toward its inactive state? Although the receptor has undergone structural modifications, conversion from CCK1R* to CCK1R° is anticipated to span a commensurably longer time scale, not amenable to classical, all-atom MD simulations. While ascertaining without ambiguity that conformations of the inactive receptor have been sampled goes evidently beyond the scope of the present work, it is, nonetheless, possible to determine how the occupation of the conformational space has evolved over 0.1 μs. In addition to inferring displacements along anharmonic modes, essential dynamics (ED) will be employed to measure the overlap of the essential subspaces formed by the TM domain of CCK1R, explored at the beginning and at the end of the simulation, viz. in its first and last 20-ns leg (see the Supporting Information). Noteworthily, this principal component analysis reveals that modes corresponding to collective motions of large amplitude are suggestive of a functional movement, rather than a random diffusion. Of particular interest, projection onto the first essential mode of the 0.1-

Activation of a G Protein-Coupled Receptor

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2157**
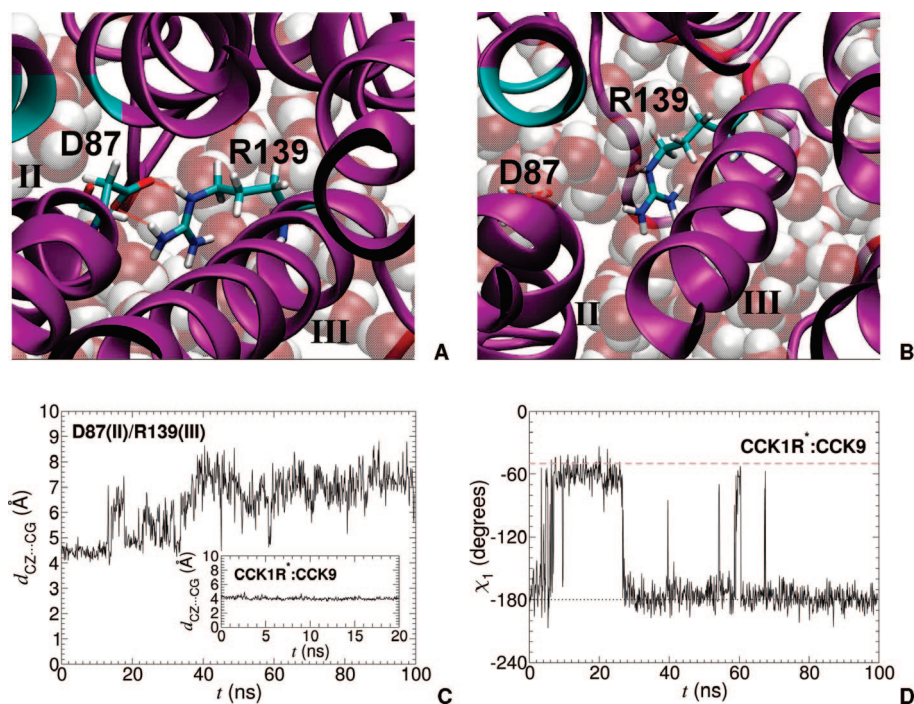
Figure 6. **A**. Salt bridge formed between residues D87 of TM helix II and R139 of TM helix III in the CCK1R*:CCK9 complex. **B**. Structural rearrangement in *apo* CCK1R and hydration of its internal cavity cause the rupture of the D87···R139 salt bridge. **C**. Time-evolution of the distance separating residues D87 and R139 in *apo* CCK1R. **D**. Time-evolution of the χ_1 torsional angle of M121 in *apo* CCK1R. The dashed light line depicts the value of χ_1 in the native CCK1R*:CCK9 complex.

μs trajectory suggests a possible tilting motion at the top of TM VI—i.e. seen from the extracellular side, consistent with previous observations in CCK2R[24] and seemingly concomitant with the displacement of TM III. This rationalizes the marked differences monitored in the interhelical separations when the latter are measured as distances between centers of mass or between selected residues.[25] More importantly, the overlap of the chosen essential subspaces—equal to 0.47, indicates that although these subspaces share analogous features, CCK1R can occupy markedly distinct regions of the conformational space over a 0.1-μs period.

Of equal concern is the choice of an initial conformation of CCK1R* to investigate by means of numerical simulations how the membrane protein evolves toward an inactive form. In other words, is the relaxation of the receptor in the course of an unphysical alchemical transformation relevant for the present study? To address this question, a second molecular assay was considered, wherein the agonist ligand, CCK9, bound to the hypothesized active receptor was removed abruptly from the latter. The additional, control, 50-ns MD trajectory highlights noteworthy differences in how the structural properties of CCK1R change as a function of time, compared to the reference, 0.1-μs simulation. Extraction of nonapeptide CCK9 from the binding pocket of CCK1R* results in the rapid reorganization of receptor, as suggested by the variations of the different interhelical separations, Δr (see the Supporting Information). In a number of instances, displacement of the TM segments appear to be at variance with the conclusions drawn from the 0.1-μs trajectory—e.g. whereas in the reference simulation, TM IV shifts away from TM VI, these two segments move toward each other in the shorter, control run. Similarly, rotation of the α-helices are somewhat different in the reference and in the control

trajectories, barring TM IV, which in both cases, rapidly undergoes a temporary writhing motion. Abrupt removal of the agonist ligand perturbs significantly the structure of the crevice and is, thus, anticipated to drive the entire molecular assembly out of equilibrium. In sharp contrast with the long 0.1-μs simulation, hydration of the membrane protein is only partial (see the Supporting Information), viz. about 30% higher than in the CCK1R*:CCK9 complex. Yet, as has been seen previously, flooding of the internal cavity of CCK1R* constitutes an early event in its inactivation. Among the milestones paving the road toward CCK1R°, concerted reorganization of the TM scaffold triggers the signature toggle-switch isomerization of the side chain of W326 (see Figure 5), an event that is not observed in the 50-ns control simulation. From the present set of results, it may be legitimately contended that smooth removal of the agonist ligand ought to be preferred for investigating the inactivation of the receptor, allowing the latter to relax and adapt as annihilation of CCK9 proceeds in the binding pocket. That the key events recorded within *ca.* 27 ns of the reference trajectory remain undetected in the shorter, 50-ns control simulation does not necessarily imply that they will never be captured, but rather that this simulation is too short to allow a substantial perturbation—i.e. the extraction of the ligand, to be absorbed entirely by the molecular assembly.

Conclusions

On the long and winding road toward GPCR inactivation, FEP calculations have been carried out as a preliminary step, during which agonist ligand CCK9 was removed progressively from the binding pocket of CCK1R*, the human cholecystokinin receptor-1 in its putative activated conforma-

2158 *J. Chem. Theory Comput., Vol. 4, No. 12, 2008*

Chipot

tion. Whereas the noteworthy accuracy of these unprecedented calculations does not validate *per se* the model of the receptor, the present results are, nonetheless, suggestive of an appropriate positioning of the ligand in the TM α-helix bundle, which further reinforces previous computations of relative binding affinities.[16] By shedding light onto the amino-acid residues participating in the membrane protein-ligand association, they also congrue with the wealth of experimental data accrued hitherto for CCK1R[15,18−21,29−34] thus challenging the pervading dogma that essentially no information of practical interest can be drawn from numerical simulations relying upon three-dimensional models of GPCRs. The additional MD simulation performed over the 0.1-μs time scale unveils early events of the inactivation process, sampling conformations of the receptor distinct from those adopted by the initial, native CCK1R*:CCK9 complex. The interplay of hydration and structural rearrangement of the TM scaffold is underscored by signature events, like the toggle-switch action of the well-conserved tryptophan residue of TM VI, the isomerization of the methionine residue of TM III, or the rupture of the salt bridge connecting TM II and TM III and involving the E/DRY motif ubiquitous to GPCRs. The latter are not random events but, on the contrary, constitute milestones toward the inactive conformation of the receptor, which are envisioned to be triggered by the flow of water molecules upon removal of the agonist ligand from the binding pocket. The subtle topological changes embodied in the fluctuations of the interhelical separations or the rotation of the TM segments appear to be sufficient to perturb durably the internal organization of the receptor. Intermediate conformations populating the path that leads to the inactive receptor, CCK1R°, constitute alternative targets of potential interest for rational drug design, which should help explain differences in binding affinities. As sizable as the assembly formed by *apo* CCK1R* and its membrane environment may be from the perspective of numerical simulations, it is still an incomplete model, from which the heterotrimeric G protein is clearly absent. Whether or not the latter component can be legitimately ignored in the simulation of GPCR activation remains an intriguing question likely to guide future modeling studies. The present contribution, nevertheless, provides valuable information on the important events that lead to activation, emphasizing the key role played by the environment and the necessity to explore these phenomena over sufficiently long time scales.

Supporting Information Available: Details of the free energy calculation and convergence criteria thereof, the essential dynamics post-treatment, and the measure of key observables characterizing the spatial rearrangement of the receptor toward its inactive conformation together with the analyses performed from the control simulation. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Bockaert, J.; Pin, J. P. *EMBO J.* **1999**, *18*, 1723–1728.

(2) Lefkowitz, R. J. *Nat. Cell Biol.* **2000**, *2*, E133−E136.

(3) Sautel, M.; Milligan, G. *Curr. Med. Chem.* **2000**, *7*, 889–896.

(4) Gether, U. *Endocr. Rev.* **2000**, *21*, 90–113.

(5) Lefkowitz, R. J.; Cotecchia, S.; Samama, P.; Costa, T. *Trends Pharmacol. Sci.* **1993**, *14*, 303–307.

(6) Kenakin, T. *Trends Pharmacol. Sci.* **1997**, *18*, 416–417.

(7) Kenakin, T. *Trends Pharmacol. Sci.* **2004**, *25*, 186–192.

(8) Swaminath, G.; Deupi, X.; Lee, T. W.; Zhu, W.; Thian, F. S.; Kobilka, T. S.; Kobilka, E. *J. Biol. Chem.* **2005**, *280*, 22165–22171.

(9) Fanelli, F.; Benedetti, P. G. D. *Chem. Rev.* **2005**, *105*, 3297–3351.

(10) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739–745.

(11) Salom, D.; Lodowski, D. T.; Stenkamp, R. E.; Trong, I. L.; Golczak, M.; Jastrzebska, B.; Harris, T.; Ballesteros, J. A.; Palczewski, K. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16123–16128.

(12) Vaidehi, N.; Floriano, W. B.; Trabanino, R.; Hall, S. E.; Freddolino, P.; Choi, E. J.; Zamanakos, G.; Goddard, W. A., III *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12622–12627.

(13) Ballesteros, J.; Palczewski, K. *Curr. Opin. Drug Discovery Dev.* **2001**, *4*, 561–574.

(14) Becker, O. M.; Shacham, S.; Marantz, Y.; Noiman, S. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 353–361.

(15) Archer-Lahlou, E.; Tikhonova, I.; Escrieut, C.; Dufresne, M.; Seva, C.; Clerc, P.; Pradayrol, L.; Moroder, L.; Maigret, B.; Fourmy, D. *J. Med. Chem.* **2005**, *48*, 180–191.

(16) Hénin, J.; Maigret, B.; Tarek, M.; Escrieut, C.; Fourmy, D.; Chipot, C. *Biophys. J.* **2006**, *90*, 1232–1240.

(17) Villanueva, M. L.; Collins, S. M.; Jensen, R. T.; Gardner, J. D. *Am. J. Physiol.* **1982**, *242*, G416–G422.

(18) Gigoux, V.; Escrieut, C.; Silvente-Poirot, S.; Maigret, B.; Gouilleux, L.; Fehrentz, J. A.; Gully, D.; Moroder, L.; Vaysse, N.; Fourmy, D. *J. Biol. Chem.* **1998**, *273*, 14380–14386.

(19) Gigoux, V.; Escrieut, C.; Fehrentz, J. A.; Poirot, S.; Maigret, B.; Moroder, L.; Gully, D.; Martinez, J.; Vaysse, N.; Fourmy, D. *J. Biol. Chem.* **1999**, *274*, 20457–20464.

(20) Gigoux, V.; Maigret, B.; Escrieut, C.; Silvente-Poirot, S.; Bouisson, M.; Fehrentz, J. A.; Moroder, L.; Gully, D.; Martinez, J.; Vaysse, N.; Fourmy, D. *Protein Sci.* **1999**, *8*, 2347–2354.

(21) Escrieut, C.; Gigoux, V.; Archer, E.; Verrier, S.; Maigret, B.; Behrendt, R.; Moroder, L.; Bignon, E.; Silvente-Poirot, S.; Pradayrol, L.; Fourmy, D. *J. Biol. Chem.* **2002**, *277*, 7546–7555.

(22) Gouldson, P. R.; Kidley, N. J.; Bywater, R. P.; Psaroudakis, G.; Brooks, H. D.; Diaz, C.; Shire, D.; Reynolds, C. A. *Proteins* **2004**, *56*, 67–84.

(23) Niv, M. Y.; Skrabanek, L.; Filizola, M.; Weinstein, H. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 437–448.

Activation of a G Protein-Coupled Receptor

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2159**

(24) Marco, E.; Foucaud, M.; Langer, I.; Escrieut, C.; Tikhonova, I. G.; Fourmy, D. *J. Biol. Chem.* **2007**,

(25) Tikhonova, I.; Best, R.; Engel, S.; Gershengorn, M.; Hummer, G.; Costanzi, S. *J. Am. Chem. Soc.* **2008**, *130*, 10141–10149.

(26) Martínez-Mayorga, K.; Pitman, M. C.; Grossfield, A.; Feller, S. E.; Brown, M. F. *J. Am. Chem. Soc.* **2006**, *128*, 16502–16503.

(27) Crozier, P. S.; Stevens, M. J.; Woolf, T. B. *Proteins* **2007**, *66*, 559–574.

(28) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. *Nov* **2007**, *318* (5854), 1258–1265.

(29) Fourmy, D.; Escrieut, C.; Archer, E.; Galès, C.; Gigoux, V.; Maigret, B.; Moroder, L.; Silvente-Poirot, S.; Martinez, J.; Fehrentz, J. A.; Pradayrol, L. *Mol. Pharmacol.* **2002**, *91*, 313–320.

(30) Archer, E.; Maigret, B.; Escrieut, C.; Pradayrol, L.; Fourmy, D. *Trends Pharmacol. Sci.* **2003**, *24*, 36–40.

(31) Archer-Lahlou, E.; Escrieut, C.; Clerc, P.; Martinez, J.; Moroder, L.; Logsdon, C.; Kopin, A.; Seva, C.; Dufresne, M.; Pradayrol, L.; Maigret, B.; Fourmy, D. *J. Biol. Chem.* **2005**, *280*, 10664–10674.

(32) Mart Jourdan, M.; Escrieut, C.; Archer, E.; González-Muniz, R.; García-López, M. T.; Maigret, B.; Herranz, R.; Fourmy, D. *J. Med. Chem.* **2005**, *48*, 4842–4850.

(33) Tikhonova, I. G.; Marco, E.; Lahlou-Archer, E.; Langer, I.; Foucaud, M.; Maigret, B.; Fourmy, D. *Curr. Top. Med. Chem.* **2007**, *7* (12), 1243–1247.

(34) Foucaud, M.; Archer-Lahlou, E.; Marco, E.; Tikhonova, I. G.; Maigret, B.; Escrieut, C.; Langer, I.; Fourmy, D. *Regul. Pept.* **2008**, *145*, 17–23.

(35) Dawson, E. S.; Henne, R. M.; Miller, L. J.; Lybrand, T. P. *Pharmacol. Toxicol.* **2002**, *91*, 290–296.

(36) Giragossian, C.; Mierke, D. F. *Biochemistry* **2001**, *40*, 3804–3809.

(37) Chipot, C.; Pohorille, A., Eds., *Free energy calculations. Theory and applications in chemistry and biology*; Springer Verlag: 2007.

(38) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(39) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(40) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(41) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(42) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. *J. Phys. Chem. B* **1992**, *97*, 1990–2001.

(43) MacKerell, 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, B.; Reiher, W. E.; 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*3586–3616.

(44) Feller, S. E.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2000**, *104*, 7510–7515.

(45) Lu, N.; Kofke, D. A.; Woolf, T. B. *J. Comput. Chem.* **2004**, *25*, 28–39.

(46) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.

(47) Fourmy, D.; Lopez, P.; Poirot, S.; Jimenez, J.; Dufresne, M.; Moroder, L.; Powers, S. P.; Vaysse, N. *Eur. J. Biochem.* **1989**, *185*, 397–403.

(48) Kennedy, K.; Gigoux, V.; Escrieut, C.; Maigret, B.; Martinez, J.; Moroder, L.; Fréhel, D.; Gully, D.; Vaysse, N.; Fourmy, D. *J. Biol. Chem.* **1997**, *272*, 2920–2926.

(49) Miyamoto, S.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8402–8406.

(50) Dixit, S. B.; Chipot, C. *J. Phys. Chem. A* **2001**, *105*, 9795–9799.

(51) Altenbach, C.; Cai, K.; Klein-Seetharaman, J.; Khorana, H. G.; Hubbell, W. L. *Biochemistry* **2001**, *40*, 15483–15492.

(52) Farrens, D. L.; Altenbach, C.; Yang, K.; Hubbell, W. L.; Khorana, H. G. *Science* **1996**, *274*, 768–770.

(53) Ghanouni, P.; Steenhuis, J. J.; Farrens, D. L.; Kobilka, B. K. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5997–6002.

(54) Jensen, A. D.; Guarnieri, F.; Rasmussen, S. G.; Asmar, F.; Ballesteros, J. A.; Gether, U. *J. Biol. Chem.* **2001**, *276*, 9279–9290.

(55) Gether, U.; Kobilka, B. K. *J. Biol. Chem.* **1998**, *273*, 17979–17982.

(56) Chipot, C.; Jaffe, R.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 11217–11224.

(57) Visiers, I.; Ballesteros, J. A.; Weinstein, H. *Methods Enzymol.* **2002**, *343*, 329–371.

(58) Grossfield, A.; Pitman, M. C.; Feller, S. E.; Soubias, O.; Gawrisch, K. *J. Mol. Biol.* **2008**, *381*, 478–486.

(59) Lin, S. W.; Sakmar, T. P. *Biochemistry* **1996**, *35*, 11149–11159.

(60) Patel, A. B.; Crocker, E.; Reeves, P. J.; Getmanova, E. V.; Eilers, M.; Khorana, H. G.; Smith, S. O. *J. Mol. Biol.* **2005**, *347*, 803–812.

(61) Crocker, E.; Eilers, M.; Ahuja, S.; Hornak, V.; Hirshfeld, A.; Sheves, M.; Smith, S. O. *J. Mol. Biol.* **2006**, *357*, 163–172.

(62) Arnis, S.; Fahmy, K.; Hofmann, K. P.; Sakmar, T. P. *J. Biol. Chem.* **1994**, *269*, 23879–23881.

(63) Ballesteros, J.; Kitanovic, S.; Guarnieri, F.; Davies, P.; Fromme, B. J.; Konvicka, K.; Chi, L.; Millar, R. P.; Davidson, J. S.; Weinstein, H.; Sealfon, S. C. *J. Biol. Chem.* **1998**, *273*, 10445–10453.

(64) Rath, P.; DeCaluwé, L. L.; Bovee-Geurts, P. H.; DeGrip, W. J.; Rothschild, K. J. *Biochemistry* **1993**, *32*, 10277–10282.

(65) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

CT800313K

JCTC Journal of Chemical Theory and Computation

Flat-Bottom Strategy for Improved Accuracy in Protein Side-Chain Placements

Victor Wai Tak Kam and William A. Goddard III*

Materials and Process Simulation Center (MC-139-74), California Institute of Technology, Pasadena, California 91125

Abstract: We present a new strategy for protein side-chain placement that uses flat-bottom potentials for rotamer scoring. The extent of the flat bottom depends on the coarseness of the rotamer library and is optimized for libraries ranging from diversities of 0.2 Å to 5.0 Å. The parameters reported here were optimized for forcefields using Lennard-Jones 12−6 van der Waals potential with DREIDING parameters but are expected to be similar for AMBER, CHARMM, and other forcefields. This *Side-Chain Rotamer Excitation Analysis Method* is implemented in the SCREAM software package. Similar scoring function strategies should be useful for ligand docking, virtual ligand screening, and protein folding applications.

1. Introduction

In developing general predictive approaches for structures of membrane proteins[1−3] (Membstruk), we found that current available side-chain placement methods, e.g. SCWRL, did not provide sufficiently accurate results to determine the helix-helix relative orientations within the membrane. Consequently, we developed the SCREAM approach reported here, which we have found to lead to dramatically improved protein structures. In this paper, we validate SCREAM against standard libraries of crystal structures. In a subsequent paper, we will report the accuracy of SCREAM in predicting stable membrane structures (where unfortunately there are very few accurate X-ray structures).

Side-chain placement methods play a major role in recent applications in the field of computational molecular biology: from protein design,[4−6] flexible ligand docking,[7] and loop-building[8] to prediction of protein structures.[9] Much attention has been paid to this important problem, which is difficult because it is in a category of problems known as NP-hard,[10] for which no efficient algorithm is known to exist. Since the groundbreaking work by Ponder and Richards,[11] many approaches have been developed, including mean-field approximation,[12,13] Monte Carlo algorithms,[14,15] and Dead-End Elimination (DEE).[16−19] In practice, however, studies have also concluded that the combinatorial issue may not be as severe as originally thought.[20,21] Compared to the placement methods and rotamer libraries, scoring functions have not been studied as extensively.[22−24] The focus of this paper is on the scoring function.

The scoring function is based on the all-atom forcefield DREIDING[25] which includes an explicit hydrogen bond term. The use of a rotamer library is widely used in side-chain prediction methods, and many authors have introduced quality rotamer libraries[21,26,27] since the Ponder library. To account for the discreteness of rotamer libraries, several approaches have been introduced, such as reducing van der Waals radii,[28,29] capping of repulsion energy,[30] rotamer minimization,[14,31] and the use of subrotamer ensembles for each dominant rotamer.[32] We introduce a flat-bottom region for the van der Waals (VDW) 12−6 potential and the DREIDING hydrogen bond term (12−10 with a cosine angle term). The width of the flat bottom depends on the specific atom of each side chain as well as the coarseness of the underlying rotamer library used.

We show in this study that accuracy can be improved substantially by introducing the flat-bottom potential and in a systematic way. In addition to showing that placement accuracy is dependent upon the number of rotamers used in a library, we find that it is possible for suitably chosen energy functions to compensate the use of coarser rotamer libraries. We demonstrate a high overall accuracy in side-chain placement and make a comparison to the popular side-chain placement program SCWRL.[33]

---

* Corresponding author e-mail: wag@wag.caltech.edu.

Flat-Bottom Strategy

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2161**

Table 1

diversity	starting	0.2 Å	0.6 Å	1.0 Å	1.4 Å	1.8 Å	2.2 Å	3.0 Å	5.0 Å	All-Torsion
rotamer count	35828	14755	3195	1014	378	214	136	84	44	382

2. Materials and Methods

2.1. Preparation of Rotamer Libraries. Rotamer libraries of various diversities are derived from the complete coordinate rotamer library of Xiang.[21] We added hydrogens to the rotamers and considered both δ and ε versions in the case for histidines. CHARMM charges are used throughout.[34] Since the Xiang library was based on crystal structure data, we minimized each of the conformations so that the internal energies will be consistent with subsequent energy evaluations of the proteins. To do this we placed each side chain on a template backbone (Ala-X-Ala in the extended conformation) and did 10 steps conjugate gradient minimization using the DREIDING forcefield.

We generated rotamer libraries of varying coarseness by a clustering procedure, using the heavy atom rmsd between minimized rotamers as the metric. Starting with the closest rotamers, we eliminated those within the specific threshold rmsd value choosing always the rotamer with the lowest minimized DREIDING energy. This threshold rmsd value is defined as the *diversity* of the resulting library. To ensure that rotamers can make proper hydrogen bonds, each side-chain conformation for serine, threonine, and tyrosine was repeated with each possible polar hydrogen position. Thus, for serine and threonine, the three sp^3 position hydrogens were added to the hydroxyl oxygen, while for tyrosine, we add the out-of-place OH bonds 90 degrees from the phenyl ring in addition to two sp^2 positions in the plane. The final number of rotamers for libraries of different diversities is shown in Table 1.

In addition, we constructed the "*all-torsion*" rotamer library in which one rotamer for each major torsional angle (120 degrees for sp^3 anchor atoms, 180 degrees for sp^2 anchor atoms) was included. The angles were obtained from the backbone independent rotamer library from Dunbrack[35] and built using the some procedure as described above.

All our rotamer libraries are backbone independent.

2.2. Preparation of Structures for Validation of SCREAM. We considered three sets of protein for validating and training SCREAM.

• Xiang: Xiang[21] considered 33 proteins for testing their method for developing libraries of side-chain conformations: 1aac, 1aho, 1b9o, 1c5e, 1c9o, 1cbn, 1cc7, 1cex, 1cku, 1ctj, 1cz9, 1czp, 1d4t, 1eca, 1igd, 1ixh, 1mfm, 1plc, 1qj4, 1ql0, 1qlw, 1qnj, 1qq4, 1qtn, 1qtw, 1qu9, 1rcf, 1vfy, 2pth, 3lzt, 5p21, 5pti, and 7rsa. We have tested SCREAM for exactly these cases.

• Liang: Liang[22,36] considered 15 proteins for testing their method for scoring functions for choosing side-chain conformations. Of these, the 10 were not in the Xiang set are denoted as the Liang set: 1bpi, 1isu, 1ptx, 1xnb, 256b, 2erl, 2hbg, 2ihl, 5rxn, and 9rnt. The proteins that overlap with the Xiang set are not included.

• Other: In addition we included 10 proteins with resolution not worse than 1.8 Å from the SCWRL data set: 1a8d, 1bfd,

1bgf, 1c3d, 1ctf, 1ctj, 1moq, 1rzl, 1svy, and 1yge. Here we ignored structures with ligands or missing residues or which had a sequence identity of more than 50% with the Xiang or Liang sets. As will be described in later sections, this set is used only for deriving the σ-values and side-chain placement parameters.

For each of these 53 proteins, the raw atom coordinates were downloaded from the PDB database. Hydrogens were added using WHATIF[37] and ligands were typed using PRODRUG.[38] Manual typing of ligands were carried out in cases where they cannot be typed by PRODRUG ($\sim$10 cases). Waters, solvents, and metals were kept when present.

These structures were then minimized (100 conjugate gradient steps) using the DREIDING forcefield. In all cases, the minimized structures differed by less than 0.3 Å total rmsd compared to the original crystal structures. All metals, prolines, cysteines in disulfide bonds, and side chains in coordination with metals were kept fixed throughout side-chain placement calculations.

2.3. Surface Area Calculations. Which residues were considered as buried or exposed was determined from the Solvent Accessible Surface Area (SASA), using a probe of radius 1.4 Å. The reference for fully exposed surface area for each side-chain type is a fully extended tripeptide in the form of Ala-X-Ala. A side chain with >20% SASA compared with the reference SASA was considered exposed. This percentage is smaller than the typical 50% level in the literature—around 25% for the Xiang set and 39% for the Liang set because we include solvent molecules as part of the structure.

2.4. Positioning of Side Chains. Placement of the rotamers on the backbone is decided by the coordinates of the C, C_α, N backbone atoms plus the C_β atom. To specify the position of the C_β atom we use the coordinates with respect to C, C_α, N based on the statistics gathered from the HBPLUS protein set (see above). This involves three parameters:

1. The angle of the $C\alpha$-C_β bond from the bisector of the C-$C\alpha$-N angle: 1.81° (from the HBPLUS protein set)

2. The angle of the $C\alpha$-C_β bond with the C-C_α-N plane: 51.1° (from the HBPLUS protein set); and

3. The $C\alpha$-C_β bond length: 1.55 Å (average value from the other protein set).

Thus the C_β atom will generally have a different position from the crystal C_β position. As in common practice in the literature, we did not include this C_β deviation in the rmsd calculations.

2.5. Combinatorial Placement Algorithm. The SCREAM combinatorial placement algorithm consists of three stages: self-energy calculation for rotamers, clash elimination, and further optimization of side chains.

2.5.1. Stage 1: Rotamer Self-Energy Calculation. The all atom forcefield DREIDING[25] was used to calculate the interactions between atoms, with a modification to be

described in the scoring function section. The internal energy contributions $E_{internal}$ (bond, angle, and torsion terms and nonbonds that involve only the side-chain atoms) were precalculated and stored in the rotamer library. For each residue to be replaced, the interaction energy ($E_{sc-fixed}$) was calculated for each rotamer interacting with just the protein backbone and fixed residues (all fixed atoms). The sum of these two terms is the empty lattice energy (E_{EL}) of a rotamer in the absence of all other side chains to be replaced

$$E_{EL} = E_{internal} + E_{sc-fixed}$$

We use the term ground-state to refer to the rotamer with the lowest E_{EL} energy. All other rotamer states are termed excited states. Excited states with an energy 50 kcal/mol above the ground-state were discarded from the rotamer list for the remaining calculations.

2.5.2. Stage 2: Clash Elimination. Eisenmenger et al.[20] showed that the side-chain-backbone interaction accounts for the geometries of 74% of all core side chains and 53% of all side chains. Thus, the ground-state of each side chain was taken as the starting structure. Of course, this structure might have severe VDW clashes between side chains since no interaction between side chains has been included. Elimination of these clashes was done as follows. A list of clashes of all ground-state pairs, above a default threshold of 25 kcal/mol, was sorted by their clashing energies. The pair (A, B) with the worst clash was then subjected to rotamer optimization by considering all pairs of rotamers and selecting the lowest energy to form a super-rotamer with a new energy

$$E_{tot}(A, B) = E_{self}(A) + E_{self}(B) + E_{Int.}(A, B) \equiv E_{self}(AB)$$

where E_{Int} indicates the interaction energy between rotamer A and rotamer B, which was the only energy calculation done at this step since the E_{EL} terms were calculated in Stage 1. The ground-state for this super rotamer now replaced the rotamer pair in the original structure. Since large side chains such as ARG and LYS may have as many as Y rotamers for the 1.0 Å library, we limited the number of pairs to be calculated explicitly to 1000, which we selected based upon the sum of the empty lattice energies. Of these interaction pairs we kept the ones with interaction energies below Z.

After resolving a clash, we considered the lowest X rotamer pairs from the above calculation as a super residue. Thus, subsequent clash resolution, say between residue C and residue A, will consider interactions of all side chains of C with the X (A,B) rotamer pairs. Now the spectrum of interaction energies treats (A,B) as a super rotamer so that the (C, (A,B)) energy spectrum is treated the same as for a simple rotamer pair with the spectrum:

$$E_{tot}(A, B, C) = E_{self}(A) + E_{self}(B) + E_{self}(C) + E_{Int}(A, B) +$$
$$E_{Int}(A, C) + E_{Int}(B, C)$$
$$= E_{self}(AB) + E_{self}(C) + E_{Int}(A, C) + E_{Int}(B, C)$$
$$\equiv E_{self}(AB) + E_{self}(C) + E_{Int}(AB, C)$$

This process continued by generating a new list of clashing residue pairs including the new (A,B,C), resolving the next worst clash as above. The procedure was repeated until no further clashes were identified between two rotamers or superrotamers.

2.5.3. Stage 3: Final Doublet Optimization. It is possible for some clashes to remain after Stage 2, since the number of rotamers pair evaluations is capped (at 1000) and also the numbers of rotamers in a super-rotamer (20). To solve this problem, the structure from the end of stage 2 was further optimized. Side-chain pairs (termed *doublets*) were now ordered in decreasing energies in the presence of all other side chains, and one iteration round of local optimization on those residue pairs was performed in that order. Any residue that had already been examined in this stage as part of a doublet was eliminated from further doublet examination. Always, the doublet with the lowest overall energy was kept.

2.5.4. Stage 4: Final Singlet Optimizations. The structure would undergo one final round of optimization, where all residues were examined one at a time, again in order of decreasing energies for the rotamer currently placed in the structure. Again, the rotamer with the best overall energy was retained for the final structure. More iterations rounds on the final result improved the overall rmsd (unpublished results), but we did not pursue this path[39] for the purposes of this paper.

We illustrate the effects of the doublet and singlet optimization stages by giving a specific example—1aac, using the 1.0 Å rotamer library and optimal parameters (to be described in a later section). After the clash elimination stage, the rmsd between the predicted structure and the crystal structure was 0.733 Å. The pair clashes remaining in this case included the pairs F57 and L67, V37 and F82, and V43 and W45. Doublets optimization brought the rmsd down to 0.703 Å. The final singlet optimization stage brought the rmsd value further down to 0.622 Å.

For this case, doublet optimization took 3 s, while singlet optimization took 13 s. For comparison, clash elimination took 30 s to complete, while the rotamer self-energy calculation took 8 s.

2.6. The Flat-Bottom Scoring Function. Since our library is discrete, the best position for a side chain may lead to some contacts slightly too short. Since the VDW interactions become very repulsive very quickly for distances shorter than R_e, a distance too short by even 0.1 Å may cause a very repulsive VDW energy. This might lead to selecting an incorrect rotamer. In order to avoid this problem, we use a flat-bottom potential in which the attractive region is exactly the same down to R_e, but the repulsive region is displaced by some amount Δ so that contacts that are slightly too short by Δ will not cause a false repulsive energy. The form of this potential is shown in Figure 1.

We allow a different Δ for each atom of each residue of each diversity. The way this is done is by writing Δ as

$$\Delta = s \cdot \sigma$$

where s is a scaling factor, and the σ values are compiled as follows.

2.6.1. Compilation of σ Values. For each rotamer library we considered the 10 query protein structures in the HBPLUS set (see Materials and Methods). For each side

Flat-Bottom Strategy

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2163**

$$VDW_{flat}(r) = \begin{cases} VDW(r) & r > R_e \\ D_e & R_e - \Delta < r < R_e \\ VDW(r + \Delta) & r < R_e - \Delta \end{cases}$$
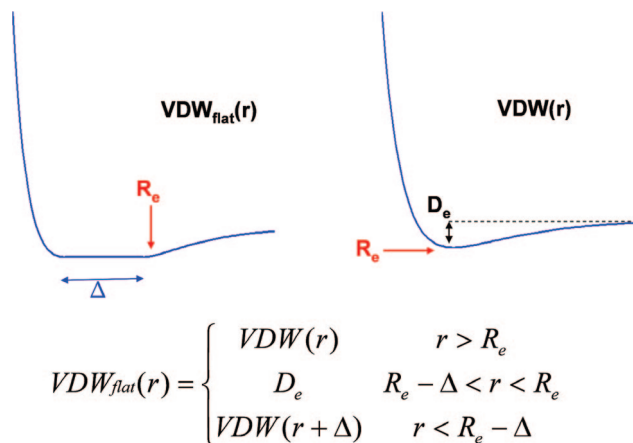
Figure 1. The flat-bottom potential. The inner wall is shifted by an amount Δ.

Table 2. δ and σ Values for Each Atom on the Arginine Side Chain, Listed in Order of Distance Away from the Main Chain[a]

dist. deviation (Å)	mean (δ)	corrected error (σ)
C_β	0.090	0.059
C_γ	0.245	0.153
C_δ	0.439	0.275
N_ϵ	0.502	0.315
C_ζ	0.588	0.369
$N_{\eta 1}, N_{\eta 2}$	0.858, 0.839	0.538, 0.526

[a] $N_{\eta}1$ and $N_{\eta}2$ are equivalent atoms; the average value is used in actual calculations. These numbers were obtained from the rotamer library of diversity 1.0 Å.

Table 3. δ and σ Values for Each Atom on the Lysine Side Chain, Listed in Order of Distance Away from the Main Chain[a]

dist. deviation (Å)	mean (δ)	corrected error (σ)
C_β	0.089	0.056
C_γ	0.259	0.162
C_δ	0.406	0.254
C_ϵ	0.596	0.373
N_ζ	0.803	0.503

[a] These numbers were obtained from the rotamer library of diversity 1.0 Å.

chain in each query structure, we picked the closest matching rotamer (in rmsd) from the library and record the distance deviation for each atom of the side chain of that residue. Thus, the atoms at the tip of the longer side chains such as arginine and lysine would have greater distance deviations than C_β atoms. The mean distance deviation (δ) for every atom of each amino-acid type over all 10 query proteins is then calculated. As an example, the δ values for arginine and lysine rotamers in the rotamer library of 1.0 Å diversity (rotamer libraries were described in section 2.1) are listed in Tables 2 and 3.

We assume that the error in positioning of any one atom of the side chain will have a Gaussian distribution of the form

$$f(r) \propto e^{-\frac{r^2}{2 \cdot \sigma^2}}$$

where r is the radial distance, and σ represents the standard deviation. Thus,

$$\rho(r) \propto 4\pi r^2 f(r)$$

is the probability of finding an atom at position r from the crystal position (which is weighted by a factor of $4\pi r^2$ from the x, y, and z distributions). The uncertainty δ in the Cartesian distance along the line between two atoms is related to σ by the form

$$\delta = 2 \cdot \sqrt{\frac{2 \cdot \sigma^2}{\pi}}$$

where δ is the value described above. This σ is listed for arginine and lysine in Tables 2 and 3.

2.6.2. Scaling Factor s. The Δ values for each side-chain atom type will depend on their σ values:

$$\Delta = s \cdot \sigma$$

The deviations for σ above provide a measure of relative uncertainties in the ability of a library to describe the correct position of the side-chain atoms. However, to obtain the absolute value of the flat-bottomness we allow an overall scaling factor for the flat-bottom portion of the potential for all atoms.

The value of *s* was optimized for the Xiang set of 33 proteins for libraries of diversities ranging from 0.2 Å to 5.0 Å as discussed in section 3.

2.6.3. Flat-Bottom Potential on Hydrogen Bond Terms. We use a flat bottom for the VDW interactions and not for the Coulomb interactions because the VDW inner wall potential becomes repulsive very quickly with distance (e.g., $1/r^{12}$). Such scaling is not important for Coulomb since it scales as 1/r. Most forcefields use a modified VDW interaction between hydrogen bonded atoms. Current version of AMBER and CHARMM do this between donor hydrogen and the acceptor heavy atom, treating the interaction as a standard 12−6 Lennard-Jones with modified parameters. The flat bottom for the other van der Waal interactions should apply equally well for these hydrogen bond terms. However, DREIDING uses an explicit 12−10 hydrogen bond term between the heavy atoms combined with a factor depending upon the linearity of the donor-hydrogen-acceptor triad

$$E_{hb} = D_{hb}[5(R_{hb}/R_{DA})^{12} - 6(R_{hb}/R_{DA})^{10}]\cos^4(\theta_{DHA})$$

where D_{hb} stands for the well-depth of the hydrogen bond potential, R_{hb} is the equilibrium distance, and θ_{DHA} is the angle between the hydrogen bond donor atom, hydrogen, and the acceptor atom. We use a flat-bottom potential for this DREIDING hydrogen bond term. However, we now allow both the inner and outer walls to shift by an amount Δ from the equilibrium point. The objective here is to also let the potential capture the polar contacts that would otherwise be missed, both when a donor−acceptor pair is too close or too far away from each other.

2.6.4. Charges. We use the CHARMM[34] charges for the protein and water, since these are standard and well-tested values. For ligands and other solvents, we use QEq[40] charges, which provide values similar to those from quantum mechanics.
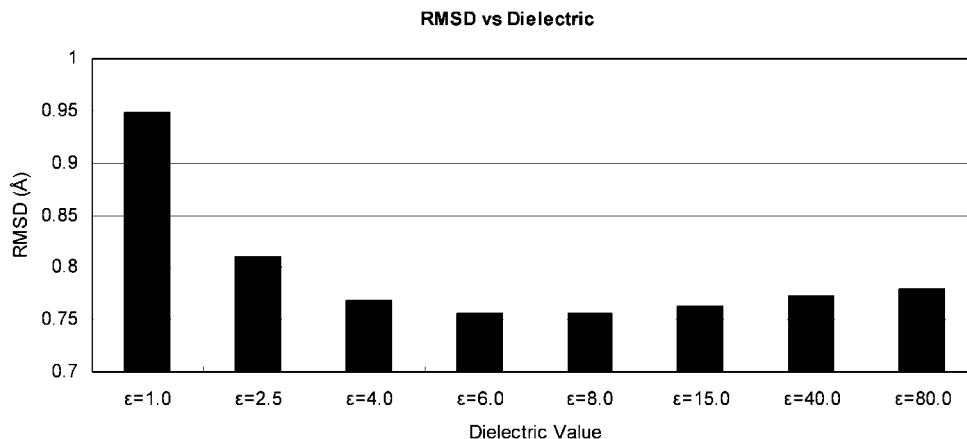
Figure 2. Effects on dielectric value on rmsd. The optimum value for the constant dielectric, $\varepsilon = 6.0$ shown here, was obtained by fitting results for the Xiang set with a diversity of 1.0 Å and a scaling factor s of 1.0.

The Coulomb interaction between atoms 1 and 2 is written as

$$E_{Coulomb} = \frac{c_0}{\varepsilon} \frac{q_1 q_2}{r_{12}}$$

where q_1 and q_2 are charges in electron units, r_{12} is in Å, ε is the dielectric constant, and $c_0 = 332.0637$ and converts to energies in kcal/mol. After optimization on a Xiang set of proteins using the 1.0 Å diversity rotamer library and a scaling factor $s = 1.0$, we chose the dielectric $\varepsilon = 6.0$ (see Figure 2). Our calculation of electrostatics used a cubic spline cutoff beginning at 8 Å and ending at 10 Å.

2.6.5. Total Rotamer Energies. The valence energies (bonds, angles, torsions, and inversion) plus the internal HB, Coulomb, and VDW energies of the rotamers were calculated beforehand and stored in the rotamer library.

The final form of the scoring function is thus

$$E_{Total} = \sum_i E_{EL} + \sum_{i<j} E_{Pair}$$

where E_{EL} is the sum over internal energies and the backbone interaction energies as described in section 2.1 and

$$E_{Pair} = E_{VDW} + E_{HB} + E_{Coulomb}$$

is the total nonbond energy between all pairs of atoms between a pair of residues.

For any particular atoms i and j, the total flat-bottom correction $\Delta_{i,j}$ for the VDW and HB terms is obtained from the individual Δ values of Δ_i and Δ_j using the relation

$$\Delta_{i,j} = \sqrt{\Delta_i^2 + \Delta_j^2}$$

This value corresponds to the standard deviation from the convolution of two normal distributions with standard deviations Δ_i and Δ_j.

3. Results and Discussion

3.1. Single Placement of Side Chains. To explore the effect on placement accuracy of using flat-bottom potentials, we increased the scaling factor s from 0.0 (no scaling) to 2.0 in 0.1 increments. To isolate the effects of the scaling, we placed side chains one at a time onto the protein, in the presence of all other side chains in their crystal positions. The values here represent the best possible results given a scoring function and a rotamer library.[24] The Xiang set of proteins described in Materials and Methods are used here.

Figure 3 shows that the best scaling factor is $s \sim 1$ for all rotamer libraries. Note that $s = 1$ for the 1.0 Å library leads to an accuracy of 0.665 Å which is much better than the accuracy of 0.71 Å obtained using $s = 0$ (no scaling) for the much bigger 0.6 Å library.

Taking the all-torsion rotamer library as an example, the rmsd improves from 0.94 Å for $s = 0$ (no flat bottom) to 0.80 Å for $s = 0.9$. This library with 378 rotamers leads to an accuracy of 0.80 Å, which compares with the accuracy of 0.75 Å obtained using the 1.4 Å library, which has 382 rotamers.

We optimized the scaling factors for rotamer libraries of diversities ranging from just 5.0 Å (44 rotamers) to 0.2 Å (13,000 rotamers). Tables 4 and 5 lists the optimum scaling factors and accuracies of these rotamer libraries, which lead to accuracies ranging from 0.47 Å (0.2 Å diversity) to 1.86 Å (5.0 Å diversity). We consider that the 1.0 Å library with an accuracy of 0.665 Å using 1014 total rotamers as a good compromise of efficiency and accuracy. These tables also list the results for the unscaled potential.

3.2. Effects of Buried vs Exposed Residues. The percentage of exposed residues considered in section 3.1 is only 25% because crystallographic waters and solvents were included in the calculation. We consider this as the best test of the scoring function. However, in practical applications, such water and solvent molecules will not be present. This creates additional uncertainties for the surface residues whose positions should be affected by the solvent and water. Without such solvent molecules, the energy functions will tend to distort the side chains to interact with other residues of the protein. Surface residues have more flexibility, and it would be better to have smaller scaling factors for these side chains. Thus, we optimized separate scaling factors for surface residues versus bulk. To do this, we calculated the SASA for the Xiang set and assigned all residues >20% exposed as surface. The resulting optimized scaling factors are in Table 6. In Figure 4, we see that the accuracy for the
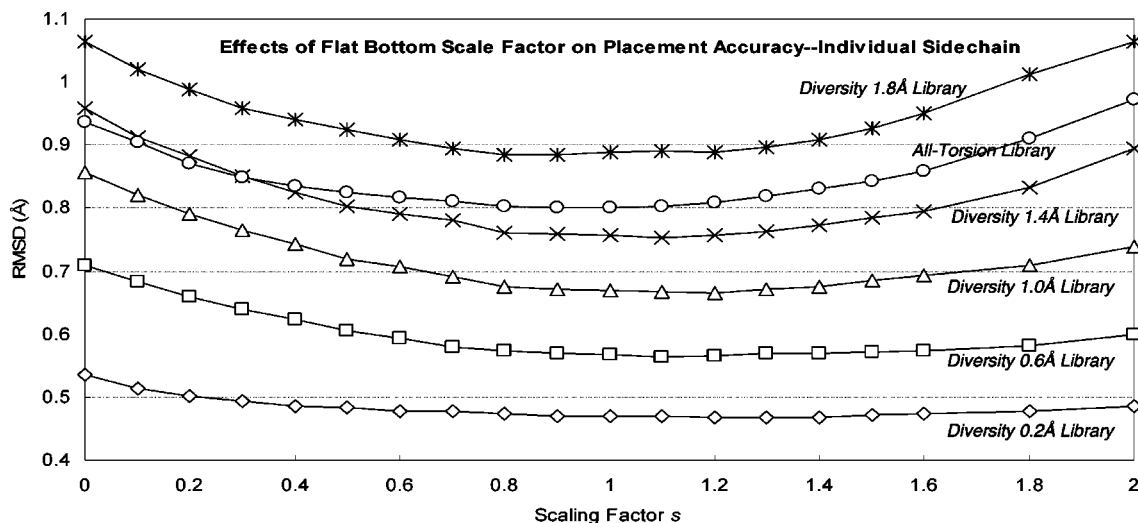
Figure 3. Single side-chain placement accuracy for various rotamer libraries at different *s* values. Shown are the libraries of 0.2 Å diversity (14755 rotamers), 0.6 Å diversity (3195 rotamers), 1.0 Å diversity (1014 rotamers), 1.4 Å diversity (378 rotamers), 1.8 Å diversity (218), and all-torsion (382 rotamers). The coarser the rotamer library is, the more pronounced the effect of *s* becomes.

Table 4. Optimized *s* Value for Rotamer Libraries of Size Ranging from 0.2 Å to 5.0 Å, Plus the All Torsion Rotamer Library[a]

library	number of rotamers	unmodified potential (rmsd, Å)	best *s* value	best rmsd (Å)
0.2 Å	14755	0.536	1.3	0.468
0.6 Å	3195	0.710	1.1	0.564
1.0 Å	1014	0.857	1.2	0.665
1.4 Å	378	0.958	1.1	0.753
1.8 Å	214	1.064	0.9	0.885
2.2 Å	136	1.343	0.8	1.175
3.0 Å	84	1.624	0.7	1.487
5.0 Å	44	1.890	0.7	1.860
all-torsion	382	0.937	0.9	0.800

[a] The *s values* for that gives the best RMSD value is listed.

Table 5. Effect of *s* Values on $\chi 1/\chi 1 + 2$ Accuracy[a]

library	number of rotamers	$\chi 1/\chi 1 + 2$ accuracy from unmodified scoring function	best scaling factor *s*	$\chi 1/\chi 1 + 2$ accuracy using best *s* value
0.2 Å	14755	95.0%/91.8%	1.3	96.3%/93.4%
0.6 Å	3195	92.6%/87.7%	1.1	95.6%/92.1%
1.0 Å	1014	90.0%/83.4%	1.2	95.3%/90.4%
1.4 Å	378	87.8%/80.0%	1.2	94.7%/88.9%
1.8 Å	214	84.3%/75.6%	1.2	91.5%/83.8%
2.2 Å	136	71.9%/61.0%	0.8	79.1%/68.0%
3.0 Å	84	63.4%/54.1%	0.7	68.4%/58.9%
5.0 Å	44	53.2%/44.9%	0.7	54.9%/45.8%
all-torsion	382	89.6%/81.3%	1.1	93.3%/86.8%

[a] Rotamer libraries of diversity ranging from 0.2 Å to 5.0 Å, plus the all-torsion rotamer library are used. The best $\chi 1 + 2$ accuracy is used to determine the most effective scaling factor *c*. A χ angle is considered correct if within 40° of the corresponding χ angle in the crystal side-chain conformation.

1.4 Å library increases from 0.809 (bulk) and 1.409 (surface) to 0.515 Å (bulk) and 1.107 Å (surface).

The current SCREAM software does not distinguish between surface and bulk residues. In order to predict the surface residues prior to assigning the side chains, we

Table 6. Accuracy Comparison in Single Side-Chain Placements for Buried and Exposed Residues for the Xiang Test Set

rotamer library	optimal scaling factor *s* for core residues	optimal scaling factor *s* for surface residues	core residue rmsd (Å) for optimal *s*	surface residue rmsd (Å) for optimal *s*
0.2 Å	1.4	0.6	0.309	0.939
0.6 Å	1.2	0.8	0.414	1.010
1.0 Å	1.2	0.9	0.515	1.107
1.4 Å	1.3	0.8	0.605	1.171
1.8 Å	1.2	0.7	0.742	1.227
2.2 Å	0.8	0.6	1.105	1.371
3.0 Å	0.7	0.6	1.439	1.625
5.0 Å	0.7	0.7	1.835	1.935
all-torsion	0.9	0.8	0.656	1.224

recommend using the alanized protein and rolling a ball of 2.9 Å instead of the standard 1.4 Å (Supporting Information).

3.3. Placement of All Side Chains on Proteins, Comparison with SCWRL. The effectiveness of the flat-bottom potential in the single-placement setting extends to multiple side-chain placements. Based on the same Xiang test set of 33 proteins, we report the placement accuracy shown in Figure 5. The optimal *s* values were similar to the values from single placement tests. For example, the 1.0 Å library had an optimum scaling factor *s*=1.0 leading to an accuracy of 0.747 Å (compared to 0.665 Å for single placement). Overall, the accuracy discrepancy in multiple placement and single placement setting comes to a 0.09 Å rmsd. Using the $\chi 1/\chi 2$ criterion leads to similar conclusions, as seen in Table 8.

The overall improvement in rmsd of the optimal *s* values over the exact Lennard-Jones potential, however, is more dramatic than in the single placement tests. For instance, by introducing the optimal *s value* for the float-bottom potential, in the single side-chain placement case, the accuracy improved from 0.834 Å to 0.663 Å, an improvement of 0.17
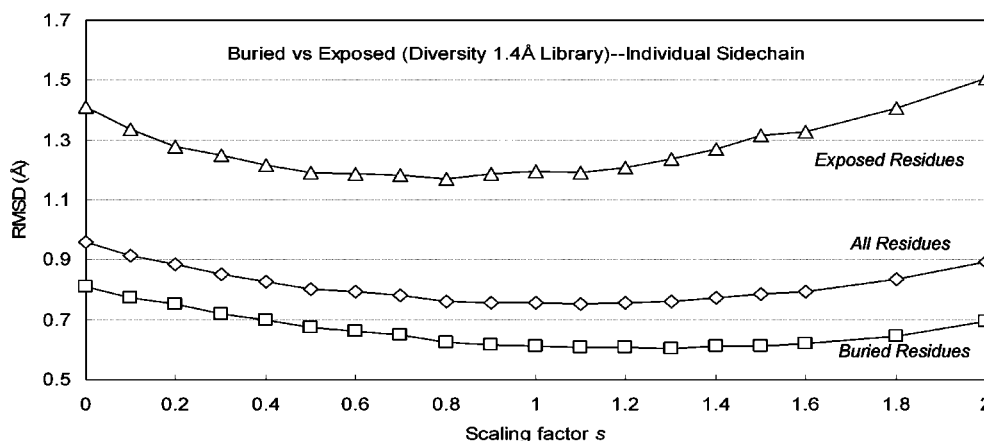
Figure 4. The effects of varying the scaling factor *s* on placement accuracies for the exposed and core residues. Shown are results from the 1.4 Å diversity rotamer library results. Exposed residues account for approximately 25% of all residues.
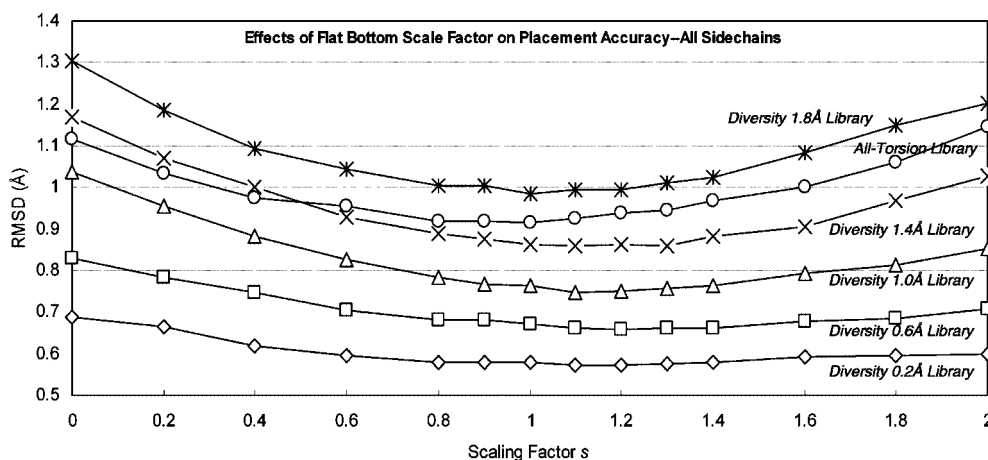
Figure 5. Accuracy for simultaneously replacing all side chains for various rotamer libraries at different *s* values. Shown are the libraries of 0.6 Å diversity (3195 rotamers), 1.0 Å diversity (1014 rotamers), 1.4 Å diversity (378 rotamers), 1.8 Å diversity (218), and all-torsion (382 rotamers).

Table 7. Optimized *s* Value for Rotamer Libraries of Size Ranging from 0.2 Å to 5.0 Å, Plus the All-Torsion Rotamer Library[a]

library	number of rotamers	unmodified potential (rmsd, Å)	best scale factor *s* value	best rmsd (Å)
0.2 Å	14755	0.689	1.2	0.571
0.6 Å	3195	0.830	1.2	0.657
1.0 Å	1014	1.036	1.1	0.747
1.4 Å	378	1.171	1.1	0.860
1.8 Å	214	1.303	1.0	0.985
2.2 Å	136	1.545	0.9	1.278
3.0 Å	84	1.756	0.8	1.565
5.0 Å	44	1.987	0.6	1.909
all-torsion	382	1.118	1.0	0.916
SCWRL		0.951 Å		

[a] The scaling factor *s* that gives the best RMSD value is included. For comparison, SCWRL gives a RMSD of 0.95 Å for the same residues and proteins tested in this set.

Å; in the all-side-chain placement case, the improvements went from 1.024 Å to 0.755 Å, an improvement of 0.27 Å.

To compare our results with SCWRL, we applied SCWRL3.0 on the Xiang set of proteins. We found an accuracy of 0.85 Å for SCWRL. A direct comparison between SCREAM and SCWRL is difficult since SCWRL

Table 8. Effect of *s* Values on $\chi 1/\chi 1+2$ Accuracy[a]

library	number of rotamers	$\chi 1/\chi 1+2$ accuracy from unmodified scoring function	optimal *s* value	$\chi 1/\chi 1+2$ accuracy using optimal *s*
0.2 Å	14755	91.4%/86.6%	1.3	94.1%/89.9%
0.6 Å	3195	89.7%/83.0%	1.1	93.8%/88.5%
1.0 Å	1014	84.5%/75.6%	1.1	92.9%/86.7%
1.4 Å	378	81.7%/71.4%	1.3	92.1%/84.3%
1.8 Å	214	77.4%/67.3%	1.2	88.6%/80.0%
2.2 Å	136	66.8%/55.0%	1.1	75.7%/64.6%
3.0 Å	84	60.6%/50.5%	0.8	66.2%/56.7%
5.0 Å	44	52.1%/43.9%	0.6	54.3%/45.7%
all-torsion	382	85.0%/73.4%	1.0	89.7%/81.5%
SCWRL		86.4%/79.7%		

[a] Rotamer libraries of diversity ranging from 0.2 Å to 5. 0 Å plus the all-torsion rotamer library are used. The best value for $\chi 1+2$ correctness is used to determine the most effective *s* value. A χ angle is considered correct if within 40° of the corresponding χ angle in the crystal side-chain conformation. The $\chi 1/\chi 1 + 2$ correctness for SCWRL is 86.4%/79.7%.

uses a backbone dependent rotamer library and a more sophisticated multiple side-chain placement algorithm. However, we note that the 1.8 Å SCREAM library, with just 214 rotamers, achieved an accuracy of 0.86 Å rmsd which is

Flat-Bottom Strategy

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2167**

Table 9. Average Energy Values for the 33 Proteins over Varying *s* Values[a]

s value	0.6 Å library		1.0 Å library		1.4 Å library		all-torsion library	
	Starting energy	minimized energy	starting energy	minimized energy	starting energy	minimized energy	starting energy	minimized energy
0	−1234.3	−3163.1	546.8	−2839.2	6957.0	−2544.8	1558154.0	−2317.1
0.2	−2237.0	−3225.5	530.7	−2969.3	2804.0	−2675.2	1260675.0	−2515.2
0.4	−2195.1	−3271.3	417.6	−3053.8	**2610.3**	−2790.4	34774.5	−2767.6
0.6	**−2364.8**	−3312.2	**−624.4**	−3102.8	3454.9	−2871.2	**34628.7**	−2826.2
0.8	−2227.6	−3328.1	−419.9	−3168.6	4970.1	−2929.7	41225.3	−2849.5
0.9	−2130.1	−3325.0	−166.4	−3165.1	10013.7	−2941.8	166369.5	−2836.7
1.0	−2041.5	−3331.6	143.2	−3166.3	132017.6	**−2952.7**	173157.0	**−2854.6**
1.1	−1952.9	**−3341.3**	1431.4	**−3177.5**	136424.5	−2945.5	53846.7	−2845.7
1.2	−1764.6	−3338.9	1885.2	−3171.0	146372.5	−2938.1	62057.7	−2794.9
1.3	−545.0	−3327.5	3278.3	−3161.9	161903.0	−2919.4	101904.8	−2783.0

[a] All energy values include valence and nonvalence terms, and the units are presented in kcal/mol. The energies do not include interaction terms between atoms that are not involved in the side-chain placement calculations. Numbers in bold are the minimum values for each category.

Table 10. Average RMSD Values (in Å) for the Xiang Set of 33 Proteins, before and after Minimization[a]

scaling factor	0.6 Å library		1.0 Å library		1.4 Å library		all-torsion library	
	starting rmsd	minimized rmsd	starting rmsd	minimized rmsd	starting rmsd	minimized rmsd	starting rmsd	minimized rmsd
0	0.830	0.737	1.036	0.930	1.171	1.061	1.112	1.003
0.2	0.784	0.694	0.954	0.848	1.071	0.962	1.035	0.916
0.4	0.746	0.658	0.884	0.773	**1.003**	0.887	0.975	0.848
0.6	**0.706**	0.615	**0.827**	0.718	0.930	0.814	**0.954**	0.823
0.8	0.681	0.591	0.784	0.668	0.888	0.767	0.920	0.787
0.9	0.682	0.591	0.766	0.651	0.877	0.752	0.917	0.786
1.0	0.672	0.581	0.764	0.647	0.863	**0.736**	0.916	**0.780**
1.1	0.662	**0.569**	0.747	**0.625**	0.860	0.729	0.923	0.786
1.2	0.657	0.562	0.752	0.629	0.861	0.727	0.937	0.799
1.3	0.662	0.568	0.758	0.632	0.860	0.724	0.946	0.803

[a] Entries in bold correspond to those with the lowest DREIDING energies before and after minimization; see Table 9 for details.

comparable to the 0.85 Å for SCWRL, which has a rotamer for each major torsion angle, coming to ∼370 rotamers. Of course, SCWRL uses a backbone dependent rotamer library, so the specific torsion angles of those rotamers depend on the backbone φ-Ψ angles.

3.4. Effects of Minimization on Structures from Different Scaling Factors. For efficiency in predicting the optimum combination of side-chain conformations, we use the discrete rotamers from the library with no minimization. Because of this, the closest rotamer in the library to the correct conformation may have short contacts. That is why we use the flat-bottom potential. Of course, after assigning the side chains we need to optimize the structures in preparation for docking and other applications. To assess how well this optimization improves the accuracy we have minimized the side chains for each structure for 100 steps (using DREIDING in vacuum) with the results in Table 9.

We see that the initial configurations often have very high energies, but after minimization these energies become fairly similar for different scaling factors with the same diversity. As expected, the best energies (in bold face) generally come from a scaling factor of 1.0 or 1.1. We note also that as the diversity of the library decreased, the energy of the final optimized configurations also decreased, indicating increased accuracy.

As expected, the rmsd also decreases as we minimize the structures. These results are shown in Table 10. For example, for the 1.0 Å library, accuracy improved from 0.747 Å to 0.625 Å.

3.5. Program Execution Performance. All tests have been run on Intel Xeon 2.33 GHz CPU single processors. The tradeoff in time vs rotamer library size is detailed in Table 11. Obviously, the size of rotamer libraries affects the time spent on side-chain placement. Compared to SCWRL, the time required by SCREAM is relatively slow. However, SCWRL does not explicitly include hydrogen atoms, and use of united atom should reduce the computational time by SCREAM by a factor of about three.[36]

It might appear that the increased accuracy of using SCREAM compared to SCWRL might not justify the increased expense. However, these test cases are all systems for which exact structures are available. We have found in applications involving predictions of new structures that the SCREAM procedure works better than SCWRL, in particular for predicting GPCRs, as will be presented elsewhere.[41]

3.6. Tests on the Liang Set Using The Optimized Scaling Factor. In the previous sections, we optimized the scaling factors for the Xiang set and discussed the accuracy for the Xiang set. As to better indicate how well SCREAM works for new systems we tested the predictions for the Liang set using the scaling factors optimized for the Xiang set.

Rotamer libraries of practical use, including those of diversities 0.6 Å, 1.0 Å, 1.4 Å, 1.8 Å, and the all-torsion rotamer library were used for this test. Results are shown in Table 12. For example, using the 1.4 Å library, we found an accuracy of 0.96 Å for all residues and 0.74 Å for the buried residues, which compares to 0.86 Å for all residues and 0.73 Å for the buried residues for the Xiang set. The

Table 11. Performance Measure of SCREAM, with Rotamer Libraries of Various Diversities[a]

library diversity	number of rotamers	time per protein	X₁ (%) buried	all	X₁₊₂ (%) buried	all	rmsd (Å) buried	all
0.2 Å	14755	554 s	96.7	93.8	93.7	89.7	0.43	0.58
0.6 Å	3195	291 s	96.1	93.5	91.6	88.0	0.53	0.67
1.0 Å	1014	146 s	95.5	92.4	89.8	85.9	0.62	0.76
1.4 Å	378	110 s	94.4	91.6	87.0	83.8	0.73	0.86
1.8 Å	214	91 s	90.9	87.8	83.4	80.0	0.85	0.99
all-torsion	382	147 s	92.4	89.7	85.2	81.5	0.78	0.92
SCWRL	n/a	3 s	90.3	86.4	84.4	79.7	0.79	0.95

[a] The timing statistics were taken from the runs that gave the best energy values.

Table 12. SCREAM Predictions on the Liang Test Set Using Optimized Scaling Factor for Rotamer Libraries of Various Diversities[a]

library diversity	number of rotamers	run time per protein	χ_1 (%) buried	all	χ_{1+2} (%) buried	all	rmsd (Å) buried	all
0.6 Å/s = 1.2	3195	78.9 s	96.4	90.8	92.6	84.3	0.52	0.80
1.0 Å/s = 1.1	1014	41.0 s	93.6	89.1	87.1	80.7	0.69	0.93
1.4 Å/s = 1.1	378	29.9 s	94.5	89.4	86.2	79.9	0.74	0.96
1.8 Å/s = 1.0	214	27.6 s	90.3	85.2	83.5	77.0	0.84	1.05
all-torsion /s = 1.0	382	32.5 s	93.4	87.6	87.3	79.4	0.77	0.99
SCWRL	n/a	2 s	90.5	83.7	84.3	75.5	0.82	1.10

[a] The percentage of buried residues in this test set is about 40%, greater than the 25% figure from the previous test set. We include crystal structure solvents in the predictions, and the increase in exposed residues is due to the fewer resolved solvents in those structures.

Table 13. Effect of Different Lennard-Jones Potentials and Their Optimal Scaling Factor s[a]

LJ type	unmodified potential (rmsd, Å)	best scale factor s value	best scale factor rmsd (Å)
7−6	0.831	0.4	0.767
8−6	0.845	0.6	0.752
9−6	0.855	0.7	0.752
10−6	0.911	0.8	0.749
11−6	0.963	1.0	0.741
12−6	1.036	1.1	0.747

[a] Tests were done on the Xiang protein set using the 1.0 Å rotamer library.

Table 14. Effects of VDW Scaling[a]

VDW radii scaling	rmsd (Å)
75%	0.959
80%	0.884
85%	0.866
90%	0.896
95%	0.956
100%	1.036

[a] Tests were done on the Xiang protein set using the 1.0 Å rotamer library.

Table 14. The improvement from using reduced VDW radii is not as pronounced as the improvement from using softer Lennard-Jones potential forms, described in the previous section.

reason for the decreased accuracy is that 40% of side chains in the Liang set are solvent exposed compared to 25% for the Xiang set. The prediction of core residues is approximately at the same level of accuracy as reported in previous sections.

3.7. Parameters for Other Lennard Jones Potentials. While the Lennard-Jones 12−6 potential is the most commonly used, it has been demonstrated that softer potentials improve placement accuracy.[42] Thus, we tested out Lennard-Jones potentials of the 7−6, 8−6, 9−6, 10−6, and 11−6 types on the 1.0 Å rotamer library for the Xiang protein set. As expected, the softer potentials performed better, but the results can be improved further by including a flat-bottom region in the potential. Results are shown in Table 13. The optimal value of the scaling factor s decreases with softer Lennard-Jones potentials, which was expected and was consistent with the flat-bottom potential approach. It is interesting to note that the 11−6 potential with optimized scaling factor s achieved the best overall rmsd value for this test, though the differences across the different Lennard-Jones potentials were small.

3.8. Comparison with VDW Radii Scaling. We also test out using reduced VDW radii values on the 1.0 Å rotamer library for the Xiang protein set. The results are shown in

3.9. Extension beyond the Natural Amino Acids. The σ values were calculated for the natural amino acids. To extend the flat-bottom potential approach for ligands and non-natural amino acids, a value for Δ or σ needs to be determined. These values clearly depend on how conformations were generated, but we recommend a simple scheme such as using $\Delta = 0.4$ Å for all atoms.

4. Conclusion

We show that side-chain placement using a flat-bottom potential leads to excellent side-chain placement results with a simple combinatorial side-chain placement algorithm. We present a straightforward method for deriving these parameters and applied this to rotamer libraries with a wide range of diversities (0.2 Å to 5.0 Å). The potential is a simple modification of a Lennard-Jones potential, making it easy to incorporable into existing software.

A particularly important application for side-chain placement is in protein folding applications where one wants to find rapidly the best side-chain positions for each backbone configuration. A first application of SCREAM for such

Flat-Bottom Strategy

J. Chem. Theory Comput., Vol. 4, No. 12, 2008 **2169**

problems is the recent development of the MembSCREAM methodology for predicting three-dimensional structures for G-Protein Coupled Receptors.[41]

Supporting Information Available: Contact information for the current SCREAM software and prediction of surface residues prior to side-chain assignment. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Trabanino, R. J.; Hall, S. E.; Vaidehi, N.; Floriano, W. B.; Kam, V. W. T.; Goddard, W. A. *Biophys. J.* **2004**, *86* (4), 1904–1921.

(2) Vaidehi, N.; Kalani, Y. S.; Hall, S. E.; Freddolino, P. L.; Trabanino, R. J.; Floriano, W. B.; Spijker, P.; Goddard, W. A. *Biophys. J.* **2005**, *88* (1), 357A–357A.

(3) Vaidehi, N.; Floriano, W. B.; Trabanino, R.; Hall, S. E.; Freddolino, P.; Choi, E. J.; Zamanakos, G.; Goddard, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (20), 12622–12627.

(4) Malakauskas, S. M.; Mayo, S. L. *Nat. Struct. Biol.* **1998**, *5* (6), 470–475.

(5) Kraemer-Pecore, C. M.; Lecomte, J. T. J.; Desjarlais, J. R. *Protein Sci.* **2003**, *12* (10), 2194–2205.

(6) Dwyer, M. A.; Looger, L. L.; Hellinga, H. W. *Science* **2004**, *304* (5679), 1967–1971.

(7) Brooijmans, N.; Kuntz, I. D. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.

(8) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 351–367.

(9) Al-Lazikani, B.; Jung, J.; Xiang, Z. X.; Honig, B. *Curr. Opin. Chem. Biol.* **2001**, *5* (1), 51–56.

(10) Pierce, N. A.; Winfree, E. *Protein Eng.* **2002**, *15* (10), 779–782.

(11) Ponder, J. W.; Richards, F. M. *J. Mol. Biol.* **1987**, *193* (4), 775–791.

(12) Koehl, P.; Delarue, M. *J. Mol. Biol.* **1994**, *239* (2), 249–275.

(13) Mendes, J.; Soares, C. M.; Carrondo, M. A. *Biopolymers* **1999**, *50* (2), 111–131.

(14) Vasquez, M. *Biopolymers* **1995**, *36* (1), 53–70.

(15) Kussell, E.; Shimada, J.; Shakhnovich, E. I. *J. Mol. Biol.* **2001**, *311* (1), 183–193.

(16) Desmet, J.; Demaeyer, M.; Hazes, B.; Lasters, I. *Nature* **1992**, *356* (6369), 539–542.

(17) Lasters, I.; Demaeyer, M.; Desmet, J. *Protein Eng.* **1995**, *8* (8), 815–822.

(18) Pierce, N. A.; Spriet, J. A.; Desmet, J.; Mayo, S. L. *J. Comput. Chem.* **2000**, *21* (11), 999–1009.

(19) Looger, L. L.; Hellinga, H. W. *J. Mol. Biol.* **2001**, *307* (1), 429–445.

(20) Eisenmenger, F.; Argos, P.; Abagyan, R. *J. Mol. Biol.* **1993**, *231* (3), 849–860.

(21) Xiang, Z. X.; Honig, B. *J. Mol. Biol.* **2001**, *311* (2), 421–430.

(22) Liang, S. D.; Grishin, N. V. *Protein Sci.* **2002**, *11* (2), 322–331.

(23) Peterson, R. W.; Dutton, P. L.; Wand, A. J. *Protein Sci.* **2004**, *13* (3), 735–751.

(24) Petrella, R. J.; Lazaridis, T.; Karplus, M. *Fold. Des.* **1998**, *3* (5), 353–377.

(25) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94* (26), 8897–8909.

(26) DeMaeyer, M.; Desmet, J.; Lasters, I. *Fold. Des.* **1997**, *2* (1), 53–66.

(27) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Genet.* **2000**, *40* (3), 389–408.

(28) Dahiyat, B. I.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (19), 10172–10177.

(29) Kuhlman, B.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (19), 10383–10388.

(30) Desjarlais, J. R.; Handel, T. M. *Protein Sci.* **1995**, *4* (10), 2006–2018.

(31) Wernisch, L.; Hery, S.; Wodak, S. J. *J. Mol. Biol.* **2000**, *301* (3), 713–736.

(32) Mendes, J.; Baptista, A. M.; Carrondo, M. A.; Soares, C. M. *Proteins: Struct., Funct., Genet.* **1999**, *37* (4), 530–543.

(33) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L. *Protein Sci.* **2003**, *12* (9), 2001–2014.

(34) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4* (2), 187–217.

(35) Dunbrack, R. L.; Karplus, M. *J. Mol. Biol.* **1993**, *230* (2), 543–574.

(36) Jain, T.; Cerutti, D. S.; McCammon, J. A. *Protein Sci.* **2006**, *15* (9), 2029–2039.

(37) Vriend, G. *J. Mol. Graph.* **1990**, *8* (1), 52-&.

(38) Schuttelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1355–1363.

(39) Holm, L.; Sander, C. *Proteins: Struct., Funct., Genet.* **1992**, *14* (2), 213–223.

(40) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95* (8), 3358–3363.

(41) Abrol, R.; Kam, V. W. T.; Jenelle, B.; Wienko, H.; Goddard, W. A., unpublished.

(42) Grigoryan, G.; Ochoa, A.; Keating, A. E. *Proteins: Struct., Funct., Bioinf.* **2007**, *68* (4), 863–878.