J. Chem. Theory Comput. 2008, 4, 1029–1031

JCTC Journal of Chemical Theory and Computation

Revised Basis Sets for the LANL Effective Core Potentials

Lindsay E. Roy, P. Jeffrey Hay, and Richard L. Martin*

Theoretical Division, Los Alamos National Laboratory, MS B268, Los Alamos, New Mexico 87545

Received February 5, 2008

Abstract: We suggest a new contraction of the basis sets associated with the Hay-Wadt relativistic effective core potentials (RECPs) for the main group and transition metal atoms. These bases are more suitable for density functional theory investigations than the previous 'double-ζ' contractions based upon Hartree–Fock atomic results. The original Hay-Wadt primitives are now contracted [5*s*5*p*3*d*], [4*s*4*p*3*d*], and [4*s*4*p*3*d*] for the first, second, and third transition series, respectively, and denoted as LANL2TZ basis sets. For the main group atoms, we advocate using a completely uncontracted basis denoted LANL08. While modestly extending the size of the basis, the resulting sets should be suitable for both DFT and wave function based approaches. The valence bases for the transition metal atoms can be supplemented with the polarization functions determined by Frenking.

It has become a common practice in the quantum chemistry community to utilize relativistic core potentials (RECP) determined from atomic Hartree-Fock (HF) calculations in the context of a Density Functional Theory (DFT) investigation. In the physics community, however, it is common to have separate RECPs or pseudopotentials generated for each functional of interest. Some years ago, concerned with the possibility that HF derived potentials might lead to significant errors when coupled with DFT, we compared HF versus functional specific RECPs and found little difference.[1] This transferability was especially true with the small-core RECPs. With some large-core RECPs, however, transferability of HF-based potentials appears to be an issue. An example occurs in the 5*f* actinides series.[2,3]

Table 1. Diffuse *d* Exponent for the First-Row Transition Metals

Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
0.0250	0.0349	0.0464	0.0587	0.0670	0.0706	0.0822	0.0898	0.101	0.170

Table 2. Theoretical and Experimental First Bond Dissociation Energy (kcal·mol^{-1}) for $M(CO)_6$ (M = Cr, Mo, W)

	$Cr(CO)_6$	$Mo(CO)_6$	$W(CO)_6$
	B3LYP		
LANL2DZ	35.8	34.3	42.2
LANL2TZ	36.9	35.3	43.0[12]
LANL08	36.9	35.3	43.0
LANL2TZ+	38.6	–	–
LANL08+	38.6	–	–
LANL2TZ(f)	38.6	37.1	43.4[13]
LANL08(f)	38.5	37.1	43.4
SDD	38.6	38.8	43.1
6–311+G*	38.2	–	–
	BP86		
LANL2DZ	41.8	38.2	46.0
LANL2TZ	43.2	39.3	46.9[12]
LANL08	43.1	39.3	46.9
LANL2TZ+	43.3	–	–
LANL08+	43.2	–	–
LANL2TZ(f)	43.2	39.6	47.3[13]
LANL08(f)	43.1	39.6	47.3
SDD	43.2	41.2	45.4
6–311+G*	42.9	–	–
Exp[14]	36.8 ± 2	40.5 ± 2	46.0 ± 2

We did note at the time that a more significant problem lay with the associated basis sets. The issue is that the coefficients for the contractions that describe the near valence or semicore regions can be quite different between HF and DFT calculations. This is usually a small effect, but we have seen examples where this leads to a 3–4 kcal·mol^{-1} difference in a bond energy or barrier height. For that reason, we have routinely uncontracted LANL2DZ main group and transition metal basis sets in our research. However, this is a more draconian solution than is necessary. While the computational cost of the additional *s* and *p* functions is not dramatic, there are larger consequences in the transition metal series associated with a completely uncontracted *d* space. We have found that a slight enlargement of the valence *d* space to triple-ζ quality is generally sufficient to ensure that the basis is adequate for both wave function and DFT based approaches.

In this Letter, we present these new contracted bases for the transition metal series. They result from the original Hay-Wadt[4] set by simply recontracting them to [5*s*5*p*3*d*], [4*s*4*p*3*d*], and [4*s*4*p*3*d*] for the first three transition series, respectively. This is straightforward, except that uncontraction of the *p*-space in the second and third rows leads to two nearly linearly dependent primitives. In all cases, we have retained the more diffuse primitive and deleted the tighter one. The new sets are available

---

* Corresponding author e-mail: rlmartin@lanl.gov.

10.1021/ct8000409 CCC: $40.75

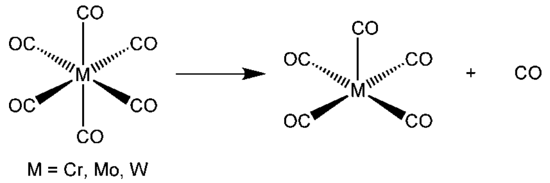
Table 3. Comparison of the Relative Energies (kcal·mol^{-1}) vs Basis Set for the Stationary Points of the [3 + 2] Cycloaddition of $CpMO_3$ (M = Mn, Tc, Re) to Ethylene Using the B3LYP Functional

	basis	**1** + C_2H_4 → **TS**	**TS** → **2**	net exothermicity
$CpMnO_3$	LANL2DZ	5.2	−63.8	−58.6
	LANL2TZ	7.2	−61.8	−54.6
	LANL08	7.2	−61.8	−54.6
	LANL2TZ+	7.5	−62.3	−54.8
	LANL08+	7.5	−62.3	−54.8
	LANL2TZ(f)	7.5	−61.2	−53.7
	LANL08(f)	7.5	−61.2	−53.7
	SDD	8.0	−61.9	−53.9
	6−311+G*	8.6	−60.8	−52.2
$CpTcO_3$	LANL2DZ	10.4	−49.7	−39.3
	LANL2TZ	12.1	−47.9	−35.7
	LANL08	12.6	−46.5	−33.9
	LANL2TZ(f)	14.7	−44.1	−29.4
	LANL08(f)	14.7	−44.1	−29.4
	SDD	13.1	−46.7	−33.6
$CpReO_3$	LANL2DZ	21.3	−35.1	−13.8
	LANL2TZ/LANL08[12]	24.2	−33.6	−9.4
	LANL2TZ(f)/LANL08(f)[13]	27.8	−29.8	−2.0
	SDD	21.9	−35.1	−13.2

Table 4. Theoretical First Bond Dissociation Energy (kcal·mol^{-1}) vs Basis Set for MC_2H_4 (M = Ni, Pd, Pt) Using the B3LYP Functional

	NiC_2H_4	PdC_2H_4	PtC_2H_4
	d^{10}	d^{10}	d^{10}
LANL2DZ	83.4	30.9	63.7
LANL2TZ	85.9	32.5	66.1[12]
LANL08	86.0	32.5	66.1
LANL2TZ+	66.3	–	–
LANL08+	66.2	–	–
LANL2TZ(f)	86.1	33.0	67.8[13]
LANL08(f)	86.2	33.0	67.8
SDD	67.0	30.5	59.1
6−311+G*	63.6	–	–

Scheme 1. M−CO Bond Dissociation



online on the EMSL Basis Set Exchange and will be referred to as LANL2TZ in the next release of the Gaussian suite of electronic structure codes.[5] We advocate their use over the LANL2DZ sets. In addition, as in all-electron basis sets, greater flexibility is sometimes required in the first row metals when treating '*d*-rich' configurations such as d^{10}. We have therefore added an additional diffuse *d* function for the first row transition metals (see Table 1) and denote as LANL2TZ+. The diffuse primitives were chosen by an even-tempered criterion.

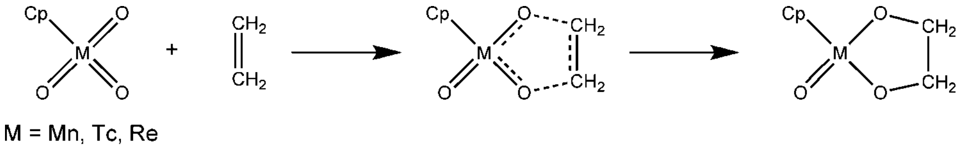
We also provide completely uncontracted sets, denoted as LANL08, for the transition metal and main group series. The LANL08 basis sets for main group atoms were derived from the original Hay-Wadt set[6] and correspond to triple-ζ valence orbital quality. For typical DFT applications we recommend LANL08 basis sets for main group atoms with RECPs and LANL2TZ for transition metals. The uncontracted LANL08 basis sets may also be useful at times for wave function based correlation treatments of transition metal complexes. Diffuse extensions available for the first transition series are denoted as LANL08+. Note that while the transition metals have outer cores, the main group atoms do not.

When polarization *f* functions are needed, we suggest the transition metal sets be augmented with the single primitive determined by Frenking.[7] Because these were optimized in wave function based CISD calculations where they serve to polarize the density *and* to provide angular correlation, they are not the optimal exponents for DFT, where they need only to polarize the density. However, these are used successfully and routinely in the literature, and the limited tests we have run suggest little is to be gained with an alternative set optimized in the DFT context. We follow the nomenclature of placing polarization functions in parentheses, e.g. LANL2TZ(f) and LANL08(f). Gilbert and co-workers determined a set of polarization functions for the main group atoms, and these will be denoted LANL08(d).[8]

A few examples for the transition metal series that are indicative of the changes to be expected are presented in Tables 2–4. In all cases the 6−311G* basis has been used for all ligands. In addition, we compared our results to the Stuttgart RSC 1997 (SDD) RECP which contains an *f* function for the first-row transition elements.[9,10] Finally, these RECP approaches are compared with Wachters' all-electron basis augmented with Hay's diffuse *d* function and Raghavachari and Trucks' *f* functions for the first-row complexes, known in Gaussian as 6−311+G*.[11] The Gaussian03 suite of electronic structure codes has been used to compute the energies.[5]

Table 2 reports the first bond dissociation energy in the series $M(CO)_6$, M = Cr, Mo, W. The metal atoms here are formally neutral, d^6 species. The differences between LANL2DZ and LANL2TZ are of the order of 1 kcal·mol^{-1}. This is the case for both the hybrid functional B3LYP and a pure generalized gradient approximation like BP86. Note also that the triple-ζ *d* orbital description is adequate, as the differences between LANL2TZ and the completely uncontracted LANL08 sets are quite small. Addition of a polarization function on the metal increases the bond dissociation energy by only 0.2 − 0.4

Scheme 2. [3 + 2] Cycloaddition



1 **TS** **2**

kcal·mol^{-1}. Polarization appears to be more important for the heavier transition metals.

As a test of metal atoms in a high valence state, we examined the olefin addition reaction shown in Scheme 2. We have restricted our investigation to the concerted pathway in which ethylene undergoes [3 + 2] addition to $CpMO_3$, with M = Mn, Tc, and Re. These systems are all formally M(VII), d^0 complexes. Table 3, column 3, reports the barrier height for the reaction, and the net exothermicity is given in column 5. The differences associated with the more contracted LANL2DZ basis are larger here, on the order of 2−4 kcal·mol^{-1} for the heavier transition metals. Once again the triple-ζ basis LANL2TZ is essentially saturated with respect to the fully uncontracted LANL08 case in the first row. For the second and third row, the completely uncontracted LANL08 basis recovers an additional 1−2 kcal·mol^{-1} in the energy differences compared with LANL2TZ. Finally, we note that the Stuttgart SDD RECP and associated basis gives results quite similar to LANL08.

Finally, we examine a case involving metals toward the right side of each row. In Table 4 we present the bond dissociation energies for the ethylene adducts MC_2H_4, M = Ni, Pd, Pt. The metals are all formally neutral and may have d^8s^2, d^9s, or d^{10} configuration in the ground state depending on the metal. For our purposes, we compared results using the d^{10} reference state. Once again, uncontraction has an influence of a few kcal·mol^{-1}. In this case, the importance of the diffuse *d* function in describing the d^{10} asymptote properly is evident in the huge reduction in bond dissociation energy with LANL2TZ+. Polarization functions once again have an influence of 1−2 kcal·mol^{-1}. The RECP results are quite similar between the comparable SDD and LANL08+ for NiC_2H_4.

These basis sets will be available in the next version of Gaussian and are provided on the EMSL Basis Set Exchange Web site (https://bse.pnl.gov/bse/portal).

Acknowledgment. This work was sponsored by the Laboratory Directed Research and Development program and the Heavy Element Chemistry and Materials Science programs of the DOE Office of Science at Los Alamos National Laboratory under the auspices of Los Alamos National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396.

References

(1) Russo, T. V.; Martin, R. L.; Hay, P. J. *J. Phys. Chem.* **1995**, *99*, 17085–7.

(2) De Jong, W. A.; Harrison, R. J.; Nichols, J. A.; Dixon, D. A. *Theor. Chem. Acc.* **2001**, *107*, 22–26.

(3) Batista, E. R.; Martin, R. L.; Hay, P. J.; Peralta, J. E.; Scuseria, G. E. *J. Chem. Phys.* **2004**, *121*, 2144–2150.

(4) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299–310.

(5) 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 E.01*; Wallingford, CT, 2003.

(6) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284–98.

(7) Ehlers, A. W.; Bohme, M.; Dapprich, S.; Gobbi, A.; Hollwarth, A.; Jonas, V.; Kohler, K. F.; Stegmann, R.; Veldkamp, A.; Frenking, G. *Chem. Phys. Lett.* **1993**, *208*, 111.

(8) Check, C. E.; Faust, T. O.; Bailey, J. M.; Wright, B. J.; Gilbert, T. M.; Sunderlin, L. S. *J. Phys. Chem. A* **2001**, *105*, 8111–8116.

(9) Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1987**, *86*, 866–72.

(10) Andrae, D.; Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123–41.

(11) Raghavachari, K.; Trucks, G. W. *J. Chem. Phys.* **1989**, *91*, 1062.

(12) LANL2TZ is equivalent to LANL08 for the third row transition elements.

(13) LANL2TZ(f) is equivalent to LANL08(f) for the third row transition elements.

(14) Rosa, A.; Baerends, E. J.; van Gisbergen, S. J. A.; van Lenthe, E.; Groeneveld, J. A.; Snijders, J. G. *J. Am. Chem. Soc.* **1999**, *121*, 10356–10365.

CT8000409

1032 *J. Chem. Theory Comput.* **2008,** *4,* 1032–1039

Journal of Chemical Theory and Computation

Combined Electrostatically Embedded Multiconfiguration Molecular Mechanics and Molecular Mechanical Method: Application to Molecular Dynamics Simulation of a Chemical Reaction in Aqueous Solution with Hybrid Density Functional Theory

Masahiro Higashi and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, 207
Pleasant Street SE, University of Minnesota,
Minneapolis, Minnesota 55455-0431

Received March 07, 2008

Abstract: We here combine the electrostatically embedded multiconfiguration molecular mechanics (EE−MCMM) method for generating global potential energy surfaces in the presence of an electrostatic potential with molecular mechanics (MM). The resulting EE−MCMM/MM method is illustrated by applying it to carry out a molecular dynamics simulation for the symmetric bimolecular reaction $Cl^- + CH_3Cl' \rightarrow ClCH_3 + Cl'^-$ in aqueous solution with hybrid density functional theory as the quantum mechanical level. The potential of mean force is calculated, and the free energy barrier is found to be 25.3 kcal/mol, which is in good agreement with previous work. The advantage of the combined EE−MCMM and MM method is that the number of quantum mechanical calculations required for the active subsystem is very small compared to straight direct dynamics.

1. Introduction

Combined quantum mechanical and molecular mechanical (QM/MM) methods have provided powerful means for studying chemical reactions in solution, enzymes, and solids.[1−35] In this approach, an active zone, which can be a solute molecule or the reaction center involved in the formation and breaking of chemical bonds, is described quantum mechanically, while the surroundings (e.g., the solvent or protein environment) are treated by using a molecular mechanics (MM) force field. When the system contains a large number of atoms, a statistical sampling method such as molecular dynamics (MD) or Monte Carlo simulation is required. However, the high computational cost of high-level *ab initio* or density functional QM calculations prevents carrying out QM/MM MD simulations for most catalytic and other condensed-phase reactions with reliable accuracy and adequate sampling.

Recently, we proposed a new method called electrostatically embedded multiconfiguration molecular mechanics (EE−MCMM) for generating global or semiglobal potential energy surfaces (PESs) in the presence of an electrostatic potential.[36] The new method is based on the QM/MM methodology, and it extends the domain of applicability of the multiconfiguration molecular mechanics (MCMM) method,[37] which has been successful[37−44] in describing semiglobal potential energy surfaces of gas-phase reactions and calculating their reaction rates with multidimensional tunneling contributions. Because the method is efficient, we can use DF/MM, that is, QM/MM with the QM level being density functional theory. We applied the new method to the symmetric bimolecular reaction $Cl^- + CH_3Cl' \rightarrow ClCH_3 + Cl'^-$ in aqueous solution; this reaction is a standard test case that has been investigated with various theoretical methods.[2,45−68] We compared the EE−MCMM potential energy with the directly evaluated electrostatically embedded QM energy, for which we used geometries and electrostatic potentials calculated by the reference interaction site model−self-consistent field method,[69−71] and we found that the potential energy in aqueous solution calculated by the new method is very close to that calculated directly without any fitting, although the EE−MCMM calculations required only a limited amount of electronic structure information for the gas-phase reaction in a field and no liquid-phase electronic structure calculations.

In the present paper, we develop the EE−MCMM method further. We apply the EE−MCMM method to full QM/MM MD simulations by replacing the electrostatically embedded QM energy with the EE−MCMM energy; we label the resulting potential energy surface as EE−MCMM/MM. EE−MCMM can reproduce global PESs calculated by high-level QM calculations in the presence of an electrostatic potential with a computational cost that is much lower than the cost of direct dynamics. Therefore, EE−MCMM/MM MD simulations can be carried out with high-level QM calculations, and adequate sampling is possible even when one uses high-level QM methods for a process requiring rare-event sampling (such as umbrella sampling).

The organization of the article is as follows. In the next section, we describe the theoretical methods employed here and derive the equations needed for the EE−MCMM/MM MD simulations. We then apply the EE−MCMM/MM method to the $Cl^- + CH_3Cl' \rightarrow ClCH_3 + Cl'^-$ reaction in aqueous solution.

---

* Corresponding author e-mail: truhlar@umn.edu.

10.1021/ct8000816 CCC: $40.75

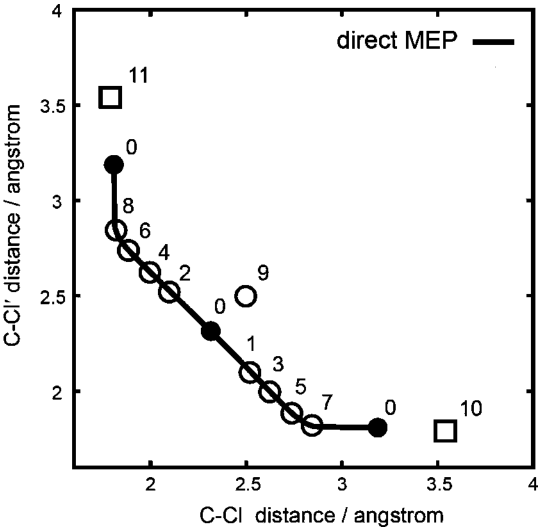


Figure 1. Gas-phase calculations: two-dimensional representation of the direct MEP and the location of Shepard points for the EE−MCMM calculation. Filled circles are stationary points; open circles are nonstationary Shepard points used in the previous and present studies; open squares are nonstationary Shepard points used in the present study.

In section 3, we present the computational details, and in section 4 we present and discuss the results of the calculations; the potential of mean force (PMF) is calculated, and the free energy of activation is compared with previous work. We also calculate the interaction energy and the radial distribution function between the QM and MM regions at the reactant and transition state. The conclusions are summarized in section 5.

2. Theoretical Method

Our goal is to calculate the Born−Oppenheimer potential energy surface V of a large system divided into a QM zone with N^{QM} atoms and a MM zone with N^{MM} atoms. The first component of V is the electrostatically embedded QM energy V^{EEQM}. We adopt a site−site representation of the QM/MM electrostatic interaction.[67,69,72–75] Then, we have

$$V^{\mathrm{EEQM}}(\mathbf{R}, \mathbf{\Phi}) = \langle \Psi | \hat{H}_0 + \hat{\mathbf{Q}}^{\mathrm{T}} \mathbf{\Phi} | \Psi \rangle \tag{1}$$

where $\mathbf{R}$ stands for the collection of the coordinates $\mathbf{R}_a$ ($a = 1, 2, \ldots, N^{\mathrm{QM}}$) of atoms in the QM region, Ψ is the electronic wave function, $\hat{H}_0$ is the electronic Hamiltonian (including nuclear repulsion) of the QM region, $\hat{\mathbf{Q}}$ is the population operator vector of order N^{QM} whose components $\hat{Q}_a$ are the population operators that generate the partial charges Q_a on QM atomic sites a:

$$Q_a = \langle \Psi | \hat{Q}_a | \Psi \rangle \tag{2}$$

and $\mathbf{\Phi}$ is the electrostatic potential distribution, which is a vector of order N^{QM}, each of whose components Φ_a is the electrostatic potential at atom a from the MM region,

$$\Phi_a(\mathbf{R}, \mathbf{R}^{\mathrm{MM}}) = \sum_{A=1}^{N^{\mathrm{MM}}} \frac{Q_A^{\mathrm{MM}}}{|\mathbf{R}_a - \mathbf{R}_A^{\mathrm{MM}}|} \tag{3}$$

where $\mathbf{R}^{\mathrm{MM}}$ is the collection of the coordinates $\mathbf{R}_A^{\mathrm{MM}}$ of atom A in the MM region, and Q_A^{MM} is the effective charge of MM atom A. Note that the first derivative of V^{EEQM} with respect to Φ_a is given by[73]

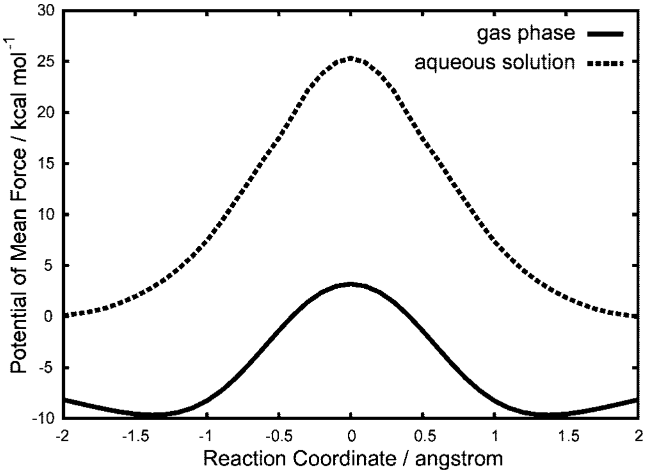


Figure 2. Potential energy profile of the $\mathrm{Cl}^- + \mathrm{CH_3Cl'} \rightarrow \mathrm{ClCH_3} + \mathrm{Cl'}^-$ reaction in the gas phase (solid line) and potential of mean force of the reaction in the aqueous solution (dashed line). The gas-phase curve is relative to its value at $z = -\infty$, and the aqueous curve is relative to the value at reactants, at $z = -2.0$ Å.

$$\frac{\partial V^{\mathrm{EEQM}}}{\partial \Phi_a} = \langle \Psi | \hat{Q}_a | \Psi \rangle = Q_a \tag{4}$$

Since details of the EE−MCMM method are presented elsewhere,[36] we describe the method only briefly in this Letter. The potential energy in EE−MCMM is the lowest eigenvalue of a 2 × 2 diabatic Hamiltonian matrix,

$$\mathbf{U}^{\mathrm{EE-MCMM}}(\mathbf{q}, \mathbf{\Phi}) = \begin{pmatrix} U_{11}(\mathbf{q}, \mathbf{\Phi}) & U_{12}(\mathbf{q}, \mathbf{\Phi}) \\ U_{12}(\mathbf{q}, \mathbf{\Phi}) & U_{22}(\mathbf{q}, \mathbf{\Phi}) \end{pmatrix} \tag{5}$$

where $U_{11}(\mathbf{q},\mathbf{\Phi})$ and $U_{22}(\mathbf{q},\mathbf{\Phi})$ are analytic functions that describe V^{EEQM} in the regions of reactants and products, $U_{12}(\mathbf{q},\mathbf{\Phi})$ is explained in the next paragraph, and we use nonredundant or redundant internal coordinates[76] $\mathbf{q}$ to represent the nuclear coordinates of the QM subsystem. The lowest eigenvalue of eq 5 is given by

$$V^{\mathrm{EE-MCMM}}(\mathbf{q}, \mathbf{\Phi}) = \frac{1}{2}\{[U_{11}(\mathbf{q}, \mathbf{\Phi}) + U_{22}(\mathbf{q}, \mathbf{\Phi})] - [(U_{11}(\mathbf{q}, \mathbf{\Phi}) - U_{22}(\mathbf{q}, \mathbf{\Phi}))^2 - 4U_{12}(\mathbf{q}, \mathbf{\Phi})^2]^{\frac{1}{2}}\} \tag{6}$$

We evaluate $V^{\mathrm{EE-MCMM}}$ and its derivatives in terms of the internal coordinates $\mathbf{q}$;[37,77] then, we transform the derivatives to the Cartesian coordinate system $\mathbf{R}$: $\mathbf{V}^{\mathrm{EE-MCMM}}(\mathbf{q},\mathbf{\Phi}) \rightarrow \mathbf{V}^{\mathrm{EE-MCMM}}(\mathbf{R},\mathbf{\Phi})$.

The evaluation of $U_{12}(\mathbf{q},\mathbf{\Phi})$ is the key feature of the EE−MCMM algorithm. It is based on a set of interpolation nodes called Shepard points $(\mathbf{R}^{(k)},\mathbf{\Phi}^{(k)})$, where $k = 1, 2, \ldots, N$. We evaluate $[U_{12}(\mathbf{q},\mathbf{\Phi};k)]^2$ by a second-order Taylor expansion around each Shepard point $(\mathbf{R}^{(k)},\mathbf{\Phi}^{(k)})$, where the Taylor series coefficients are determined such that $V^{\mathrm{EE-MCMM}}$ reproduces V^{EEQM} and its first and second derivatives with respect to $\mathbf{q}$ and $\mathbf{\Phi}$ at Shepard point $(\mathbf{R}^{(k)},\mathbf{\Phi}^{(k)})$. Then, we construct $U_{12}(\mathbf{q},\mathbf{\Phi})$ at any arbitrary geometry by Shepard interpolation[78,79] of these expressions.

The total potential energy V is obtained in the EE−MCMM/MM method by replacing the electrostatically embedded QM energy V^{EEQM} in the conventional QM/MM method with the EE−MCMM potential energy $V^{\mathrm{EE-MCMM}}$; this yields

$$V(\mathbf{R},\mathbf{R}^{\mathrm{MM}}) = V^{\mathrm{EE-MCMM}}(\mathbf{R},\mathbf{\Phi}(\mathbf{R},\mathbf{R}^{\mathrm{MM}})) + V_{\mathrm{vdW}}^{\mathrm{QM/MM}}(\mathbf{R},\mathbf{R}^{\mathrm{MM}}) + V_{\mathrm{val}}^{\mathrm{QM/MM}}(\mathbf{R},\mathbf{R}^{\mathrm{MM}}) + V^{\mathrm{MM}}(\mathbf{R}^{\mathrm{MM}}) \quad (7)$$

where $V_{\mathrm{vdW}}^{\mathrm{QM/MM}}$ and $V_{\mathrm{val}}^{\mathrm{QM/MM}}$ are respectively the van der Waals and valence interaction energies between the QM and MM regions, and V^{MM} is the MM potential energy. MD simulations require the first derivatives of V with respect to $\mathbf{R}$ and $\mathbf{R}^{\mathrm{MM}}$. These derivatives are given by

$$\frac{\partial V}{\partial \mathbf{R}_a} = \frac{\partial V^{\mathrm{EE-MCMM}}}{\partial \mathbf{R}_a} + \frac{\partial V^{\mathrm{EE-MCMM}}}{\partial \Phi_a}\frac{\partial \Phi_a}{\partial \mathbf{R}_a} + \frac{\partial V_{\mathrm{vdW}}^{\mathrm{QM/MM}}}{\partial \mathbf{R}_a} + \frac{\partial V_{\mathrm{val}}^{\mathrm{QM/MM}}}{\partial \mathbf{R}_a} \quad (8)$$

and

$$\frac{\partial V}{\partial \mathbf{R}_A^{\mathrm{MM}}} = \sum_{a=1}^{N^{\mathrm{QM}}} \frac{\partial V^{\mathrm{EE-MCMM}}}{\partial \Phi_a}\frac{\partial \Phi_a}{\partial \mathbf{R}_A^{\mathrm{MM}}} + \frac{\partial V_{\mathrm{vdW}}^{\mathrm{QM/MM}}}{\partial \mathbf{R}_A^{\mathrm{MM}}} + \frac{\partial V_{\mathrm{val}}^{\mathrm{QM/MM}}}{\partial \mathbf{R}_A^{\mathrm{MM}}} + \frac{\partial V^{\mathrm{MM}}}{\partial \mathbf{R}_A^{\mathrm{MM}}} \quad (9)$$

The terms involving $V_{\mathrm{vdW}}^{\mathrm{MM}}$, $V_{\mathrm{val}}^{\mathrm{QM/MM}}$, and V^{MM} are the same as in any other QM/MM method and need not be discussed. The derivatives of Φ_a are obtained analytically from eq 3. The first term of eq 8 is obtained analytically.[37,77] $\partial V^{\mathrm{EE-MCMM}}/\partial \Phi_a$ is obtained by differentiating $V^{\mathrm{EE-MCMM}}$ of eq 6 with respect to Φ_a as in eq 4 and equals the partial charge $Q_a^{\mathrm{EE-MCMM}}$. Therefore, we can regard the electrostatic interaction between the QM and MM regions as

$$V_{\mathrm{ele}}^{\mathrm{QM/MM}} = \sum_{a}^{N^{\mathrm{QM}}} Q_a^{\mathrm{EE-MCMM}} \Phi_a \quad (10)$$

where

$$Q_a^{\mathrm{EE-MCMM}} \equiv \frac{\partial V^{\mathrm{EE-MCMM}}}{\partial \Phi_a} \quad (11)$$

3. Computational Details

We applied the EE−MCMM/MM method to the reaction $\mathrm{Cl^- + CH_3Cl' \rightarrow ClCH_3 + Cl'^-}$ in aqueous solution. We took the difference between the two C−Cl distances as the reaction coordinate:

$$z = R_{\mathrm{CCl'}} - R_{\mathrm{CCl}} \quad (12)$$

Since the computational details of the EE−MCMM calculation in the present study are almost the same as those in the previous study,[36] we here describe them only briefly. We used the MPW1K density functional[80] for the electronic structure calculations on the QM subsystem. The basis set is 6-31G(d,p) for C and H atoms and 6-31+G(d,p) for Cl. We choose the population operator $\hat{Q}_a$ according to Charge Model 4 (CM4).[81] All of the electronic structure calculations were performed by a modified GAMESSPLUS[82] computer code based on the GAMESS quantum chemistry package.[83] In the EE−MCMM calculations, we used a modified[36] MM3 force field[84–86] for the diagonal elements. All of the EE−MCMM calculations were carried out by a new version of the MC-TINKER program.[77]

Table 1. Contributions to the QM/MM Electrostatic Interaction Energy From Individual Solute Atoms (in kcal/mol). Values in Parentheses are Standard Deviations.

	reactant	transition state
$\langle Q_{\mathrm{C}}^{\mathrm{EE-MCMM}}\Phi_{\mathrm{C}}\rangle$	−14.3 (3.8)	0.8 (2.1)
$\langle Q_{\mathrm{C}}^{\mathrm{EE-MCMM}}\Phi_{\mathrm{H}}\rangle$	6.3 (1.3)	10.0 (1.1)
$\langle Q_{\mathrm{C}}^{\mathrm{EE-MCMM}}\Phi_{\mathrm{Cl'}}\rangle$	−3.4 (2.7)	−67.7 (9.4)
$\langle Q_{\mathrm{C}}^{\mathrm{EE-MCMM}}\Phi_{\mathrm{Cl}}\rangle$	−149.6 (10.0)	−67.7 (9.4)
total	−148.4 (10.4)	−104.7 (8.6)

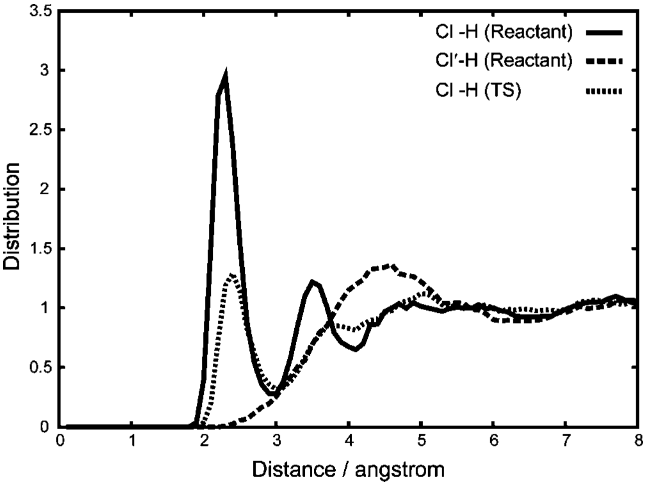


Figure 3. Radial distribution functions between the chloride in the solute and the hydrogen in solvent water: reactant Cl−H (solid line), reactant Cl′−H (dashed line), and TS Cl−H (dotted line).

The only difference between the present and previous studies is the number of Shepard points used in the EE−MCMM calculation. In the previous study, we used the electronic structure information at 12 Shepard points: three stationary points (the precursor ion−dipole complex, the saddle point, and the successor ion−dipole complex), eight nonstationary points along the minimum energy path (MEP), and one nonstationary point on the concave side of the reaction path. The ion−dipole complexes in the gas phase are located at $z = \pm 1.378$Å in the MPW1K calculation, and these 12 Shepard points are placed at $|z| \leq 1.378$ Å. To reproduce the PES at $|z| \geq 1.378$ Å, two additional Shepard points were added at $z = \pm 1.75$ Å; the remaining coordinates for these two additional points were optimized by direct gas-phase calculations. (Without these two additional Shepard points, total energy in the EE−MCMM/MM MD simulation was not conserved well at large $|z|$.) The locations of all 14 Shepard points and the direct MEP are shown in Figure 1. Note that six of the points are related to six others by symmetry, so one needs to calculate only eight Hessians as input to EE−MCMM. Note that, although the application is to a condensed-phase reaction, only electronic structure information for the gas-phase reaction is used.

For the QM/MM van der Waals interaction energy $V_{\mathrm{vdW}}^{\mathrm{QM/MM}}$, we used Lennard-Jones potentials. The Lennard-Jones param-

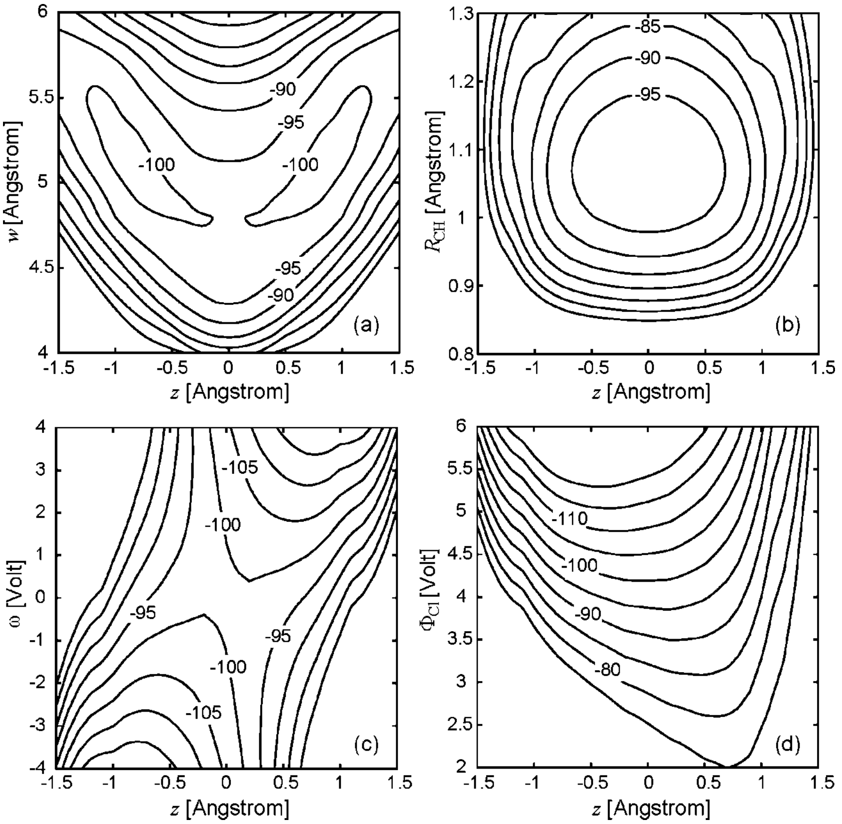


Figure 4. Equipotential contours of the EE−MCMM potential energy. The abscissa is the reaction coordinate *z*. The ordinate is (a) the sum of the length of the two C−Cl bonds, $w = R_{CCl} + R_{CCl'}$, (b) the length of the C−H bond, R_{CH}, (c) the difference between the electrostatic potential on the Cl atom and that on the Cl′ atom, $\omega = \Phi_{Cl'} - \Phi_{Cl}$, and (d) the electrostatic potential on the Cl atom, Φ_{Cl}. Contour labels are in kilocalories per mole. Contours are spaced (a) from −100 to −70, (b) from −95 to −65, (c) from −115 to −80, and (d) from −120 to −75 by 5 kcal/mol. The zero of energy is at infinitely separated reagents in the gas phase.

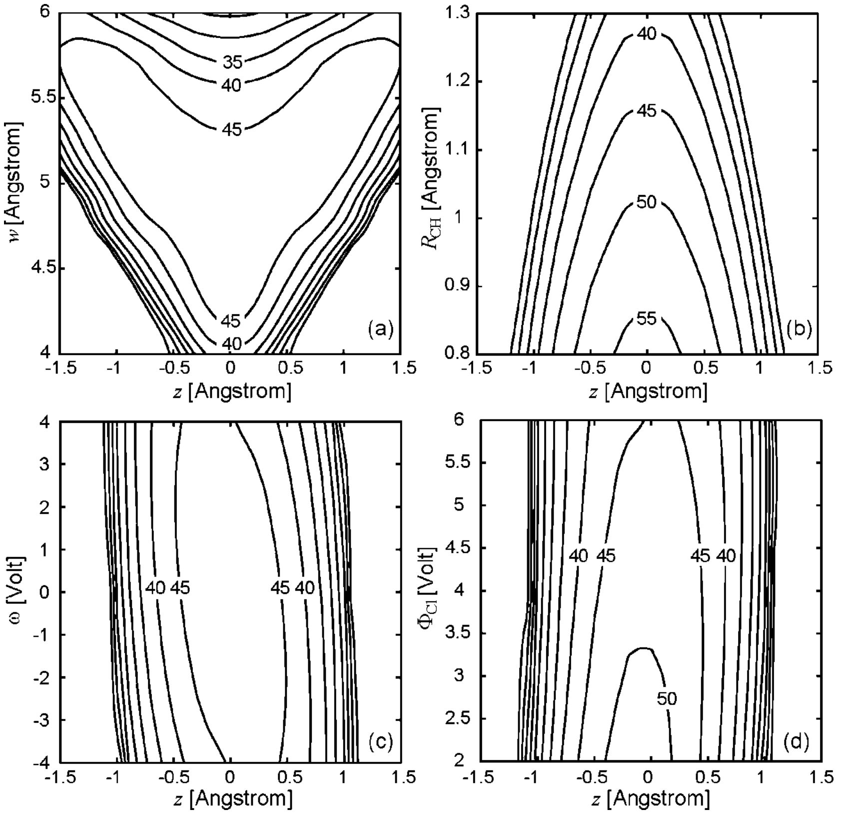


Figure 5. Equipotential contours of $U_{12}(\mathbf{q},\mathbf{\Phi})$. The axes are the same as in Figure 4. Contour labels are in kilocalories per mole. Contours are spaced (a) from 15 to 45, (b) from 25 to 55, (c) from 10 to 45, and (d) from 5 to 50 by 5 kcal/mol.

eters for the solute atoms were taken from the AMBER force field,[87] and the TIP3P model[88] was used for solvent−water molecules.

The EE−MCMM/MM MD simulation calculations were performed with periodic boundary conditions by the AMBER 9 package[89] combined with the MC-TINKER program. A cubic unit cell was used with a box length of 31.2 Å; this contains one solute ($Cl^- + CH_3Cl'$) and 1021 water molecules with a density of 1.0 g/cm^3 for water molecules. For the long-range electrostatic interaction, we used the tapering method implemented in the TINKER program,[90] where the electrostatic interaction smoothly becomes zero at 15 Å. A cutoff of 15 Å was also employed for the Lennard-Jones interactions. The equations of motion were integrated by the velocity Verlet method[91] with a time step of 0.5 fs at a temperature of 300 K. The SHAKE algorithm[92] was used to fix the intramolecular distances of the water solvent molecules.

To calculate the PMF along the reaction coordinate, we used the umbrella sampling method. We employed 21 umbrella sampling windows along the reaction coordinate z with harmonic restraining force constants 30−90 kcal mol^{-1} Å^{-2}. For each umbrella sampling window, we began with a 10 ps MD trajectory calculation for equilibration, followed by a 40 ps calculation for statistical sampling. The probability distributions for each window were pieced together with the weighted histogram analysis method[93–95] to compute the PMF.

4. Results and Discussion

The calculated aqueous PMF is presented in Figure 2. We also show the gas-phase potential energy profile along the reaction coordinate of the gas-phase reaction in Figure 2. The two curves are quite different because of the solute−solvent interaction. The free energy barrier between the reactant ($z = -2.0$ Å) and the transition state (TS; $z = 0.0$ Å) is found to be 25.3 kcal/mol, which agrees well with the experimental[96] activation free energy, 26.6 kcal/mol, and with previous[46] theoretical work.

In order to understand the differences between the QM/MM electrostatic interaction energy at the reactant and at the TS, we compared the contributions from the individual solute atoms according to eq 10. Note that the partial charges as well as the electrostatic potentials fluctuate in the EE−MCMM/MM MD simulation; therefore we averaged $Q_a^{\mathrm{EE-MCMM}}\Phi_a$ over the trajectories corresponding to a finite interval around a given value of z. (In particular, the trajectories with -0.05 Å $\leq z \leq +0.05$ Å were averaged to obtain the value for the TS, $z = 0.0$ Å, and the trajectories with $z = -2.0$ Å $\pm$ 0.05 Å were averaged for the reactant, $z = 2.0$ Å.) The results are shown in Table 1. The table shows that the chloride anion in the reactant interacts strongly with the solvent. In fact, the radial distribution functions between the chloride in the solute and the hydrogen atoms of water show much different character for the reactant and the TS (Figure 3). The sharp peak between the chloride anion and hydrogen at the reactant is a signature of the strong hydrogen bonding. There are no peaks in the first solvation shell between the chlorine atom of CH_3Cl and hydrogen. The standard deviations of the QM/MM electrostatic interaction energies are also shown in Table 1. The ratio of the standard deviation to the value of $\langle Q_{\mathrm{Cl}}^{\mathrm{EE-MCMM}}\Phi_{\mathrm{Cl}}\rangle$ at the TS is larger than that at the reactant. These fluctuations of the interaction energies are related to the fluctuations of the charges; the charge fluctuation at the reactant is smaller because the charge of the chloride anion is almost equal to −1 in the MD simulation, with a standard deviation of only 0.005, whereas the average value of the charge on either chlorine at the TS is −0.703, with a standard deviation of 0.034. This may be compared to the gas-phase charge on the chlorine at the gas-phase TS, which is −0.645.

Table 2. Average Values and Standard Deviations of Matrix Elements and EE−MCMM Energy at the TS in the EE−MCMM/MM MD Simulation.

	average	standard deviation
U_{11}, U_{22}	−47.7	11.7
U_{12}	49.0	2.1
$V^{\mathrm{EE-MCMM}}$	−97.2	8.6

We present equipotential contour plots of the EE−MCMM potential energy and $U_{12}(\mathbf{q},\mathbf{\Phi})$ in Figures 4 and 5. The abscissa is taken as the reaction coordinate z, and the ordinate is (a) the sum of the length of the two C−Cl bonds, $w = R_{\mathrm{CCl}} + R_{\mathrm{CCl'}}$, (b) the length of the C−H bond, R_{CH}, (c) the difference between the electrostatic potential on the Cl atom and that on the Cl′ atom, $\omega = \Phi_{\mathrm{Cl'}} - \Phi_{\mathrm{Cl}}$, and (d) the electrostatic potential on the Cl atom, Φ_{Cl}. The remaining coordinates are taken from the gas-phase TS geometry ($R_{\mathrm{CH}} = 1.065$ Å, $w = 4.594$ Å, and $\angle \mathrm{HCCl} = \angle \mathrm{HCCl'} = 90°$), and the remaining electrostatic potentials in the electrostatic potential distribution are taken from the average electrostatic potential during the EE−MCMM/MM MD simulation at the TS ($\Phi_{\mathrm{C}} = 3.444$ V, $\Phi_{\mathrm{H}} = 3.314$ V, $\Phi_{\mathrm{Cl}} = \Phi_{\mathrm{Cl'}} = 4.162$ V, and, in case c, $\Phi_{\mathrm{Cl}} + \Phi_{\mathrm{Cl'}} = 8.324$ V). The change of U_{12} is smaller than that of the EE−MCMM potential energy because the diagonal terms U_{11} and U_{22} can describe the main change of the EE−MCMM potential energy at the reactant and product. Although the effects of the electrostatic potential $\mathbf{\Phi}$ on U_{12} are smaller, U_{12} surely depends on $\mathbf{\Phi}$. Therefore, it is important to consider the dependence of U_{12} on the external electrostatic potential $\mathbf{\Phi}$. Table 2 shows average values and standard deviations of the matrix elements of $\mathbf{U}^{\mathrm{EE-MCMM}}$ and the EE−MCMM potential energy $V^{\mathrm{EE-MCMM}}$ at the TS in the EE−MCMM/MM MD simulation. The standard deviation of U_{12} is smaller than that of $V^{\mathrm{EE-MCMM}}$, which indicates that the change of U_{12} is smaller than that of $V^{\mathrm{EE-MCMM}}$ as in the case of Figures 4 and 5.

In previous work,[42] two different strategies, called QM/MM−MCMM and MCMM/MM, were proposed for combining MCMM with MM, and QM/MM−MCMM was applied to the reaction of complex molecules in the gas phase. The number, N^{MM}, of MM atoms in these examples was 4 and 24, respectively. The present EE−MCMM/MM approach is more similar to MCMM/MM; however, it eliminates the chief drawback of MCMM/MM, namely, that it corresponds to mechanical embedding, whereas EE−MCMM/MM corresponds to electrostatic embedding. This was an unfortunate disadvantage of MCMM/MM because it had the attractive feature of requiring one to handle only $3N^{\mathrm{QM}} \times 3N^{\mathrm{QM}}$ Hessians, whereas the Hessian size in QM/MM–MCMM is $3(N^{\mathrm{QM}} + N^{\mathrm{MM}}) \times 3(N^{\mathrm{QM}} + N^{\mathrm{MM}})$, which would be cumbersome (unless simplifying approximations were made) for simulating enzyme-catalyzed reactions for which N^{MM} is larger than N^{QM} by a factor of 10^2 or 10^3. The need for electrostatic embedding rather than mechanical embedding has been emphasized elsewhere,[5,8,10,15,17,33,34,97] and the arguments need not be repeated here. EE−MCMM/MM includes electrostatic embedding, but it keeps the QM Hessian size manageable, at $4N^{\mathrm{QM}} \times 4N^{\mathrm{QM}}$. Note that the simulations reported here involve 1.05 ns of simulation time with a time step of 0.5 fs. Since the velocity Verlet algorithm requires one gradient per step, the total number of gradient calculations is 2.1×10^6. In a direct dynamics calculation, this would require 2.1×10^6 electronic structure gradients. However, the EE−MCMM input is only eight 24×24 Hessians. If the Hessians with respect to coordinates were evaluated by central differences of gradients and the Hessians with respect to electrostatic potentials were evaluated by central differences of charges, this would require only 2.7×10^3 gradient calculations and 2.1×10^3 charge calculations. The effort for QM calculations is thereby reduced by more than a factor 500. (Note that, because the first derivatives with respect to the electrostatic potentials are the partial charges, as shown in eq 4, the computational cost to calculate them is much lower than that to calculate the first derivatives with respect to the coordinates. Note also that if the symmetry of the solute molecule is considered, the number of gradient calculations is much lower.) If analytic Hessians[98] or partial Hessians[39] are employed, the savings are even greater, and if simulation times longer than 1.05 ns are considered, the electronic structure cost for direct dynamics is proportional to simulation time, whereas the electronic structure cost of EE−MCMM is fixed.

5. Conclusion

In the present study, we presented a combined EE−MCMM and MM method by replacing the electrostatically embedded QM energy in QM/MM with the EE−MCMM potential energy. We illustrated this method by applying it to carry out a molecular dynamics simulation of the reaction $\mathrm{Cl^-} + \mathrm{CH_3Cl'} \rightarrow \mathrm{ClCH_3} + \mathrm{Cl'^-}$ in aqueous solution using hybrid DFT for the QM region. We used these EE−MCMM/MM MD simulations to compute the PMF. The free energy barrier is calculated to be 25.3 kcal/mol, which is in good agreement with the experimental estimates. We also compared the contribution from the individual solute atoms of the QM/MM electrostatic interaction energy.

On the basis of the present results, we conclude that the EE−MCMM/MM method is a very powerful tool for studying reactions in the condensed phase. It is noteworthy that it is straightforward (using, for example, link atom methods that have been extensively developed[2,3,6,8,10,17,20,29,97,99–101] in the context of previous QM/MM methods) to apply this method to a system that involves covalent bonds between the QM and MM regions, such as many reactions catalyzed by enzymes or heterogeneous catalysts.

Acknowledgment. This work is supported by the National Science Foundation by grant no. CHE07-04974.

References

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

(2) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.

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

(4) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298.

(5) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.

(6) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum Chem.* **1996**, *60*, 1189.

(7) Truong, T. N.; Truong, T.-T.; Stefanovich, E. V. *J. Chem. Phys.* **1997**, *107*, 1881.

(8) Antés, I.; Thiel, W. In *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M. A., Eds.; American Chemical Society: Washington, DC, 1998; ACS Symposium Series 712, p 50.

(9) Tongraar, A.; Liedl, K. R.; Rode, B. M. *J. Phys. Chem. A* **1998**, *102*, 10340.

(10) Burton, N. A.; Harrison, M. J.; Hart, J.; Hillier, I. H.; Sheppard, D. W. *Faraday Discuss.* **1998**, *110*, 463.

(11) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.

(12) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468.

(13) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452.

(14) Woo, T. K.; Blöchl, P. E.; Ziegler, T. *J. Phys. Chem. A* **2000**, *104*, 121.

(15) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.

(16) Gogonea, V.; Westerhoff, L. M.; Merz, K. M., Jr *J. Chem. Phys.* **2000**, *113*, 5604.

(17) Sherwood, P. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, Germany, 2000; Vol. NIC Series 3, p 285.

(18) Chalmet, S.; Rinaldi, D.; Ruiz-Lopez, M. F. *Int. J. Quantum Chem.* **2001**, *84*, 559.

(19) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Theor. Chem. Acc.* **2001**, *105*, 207.

(20) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.

(21) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941.

(22) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.

(23) Vreven, T.; Morokuma, K. *Theor. Chem. Acc.* **2003**, *109*, 125.

(24) Kerdcharoen, T.; Birkenheuer, U.; Krüger, S.; Woiterski, A.; Rösch, N. *Theor. Chem. Acc.* **2003**, *109*, 285.

(25) Nemukhin, A. V.; Grigorenko, B. L.; Topol, I. A.; Burt, S. K. *J. Comput. Chem.* **2003**, *24*, 1410.

(26) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *J. Mol. Biol.* **2003**, *327*, 549.

(27) Toniolo, A.; Ciminelli, C.; Granucci, G.; Laino, T.; Persico, M. *Theor. Chem. Acc.* **2004**, *111*, 270.

(28) Bathelt, C. M.; Zurek, J.; Mulholland, A. J.; Harvey, J. N. *J. Am. Chem. Soc.* **2005**, *127*, 12900.

(29) Pu, J.; Gao, J.; Truhlar, D. G. *ChemPhysChem* **2005**, *6*, 1853.

(30) Sundararajan, M.; Hillier, I. H.; Burton, N. A. *J. Phys. Chem. A* **2006**, *110*, 785.

(31) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458.

(32) To, J.; Sherwood, P.; Sokol, A. A.; Bush, I. J.; Catlow, C. R. A.; van Dam, H. J. J.; French, S. A.; Guest, M. F. *J. Mater. Chem.* **2006**, *16*, 1919.

(33) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.

(34) Senn, H. M.; Thiel, W. *Curr. Opinion Chem. Biol.* **2007**, *11*, 182.

(35) Woodcock, H. L., III; Hodošček, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer, H. F., III; Brooks, B. R. *J. Comput. Chem.* **2007**, *28*, 1485.

(36) Higashi, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 790.

(37) Kim, Y.; Corchado, J. C.; Villa, J.; Xing, J.; Truhlar, D. G. *J. Chem. Phys.* **2000**, *112*, 2718.

(38) Albu, T. V.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 8465.

(39) Lin, H.; Pu, J.; Albu, T. V.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 4112.

(40) Kim, K. H.; Kim, Y. *J. Chem. Phys.* **2004**, *120*, 623.

(41) Kim, Y.; Kim, Y. *J. Phys. Chem. A* **2006**, *110*, 600.

(42) Lin, H.; Zhao, Y.; Tishchenko, O.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 1237.

(43) Tishchenko, O.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13530.

(44) Tishchenko, O.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 938.

(45) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 3049.

(46) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154.

(47) Bash, P. A.; Field, M. J.; Karplus, M. *J. Am. Chem. Soc.* **1987**, *109*, 8092.

(48) Kozaki, T.; Morihashi, K.; Kikuchi, O. *J. Am. Chem. Soc.* **1989**, *111*, 1547.

(49) Huston, S. E.; Rossky, P. J.; Zichi, D. A. *J. Am. Chem. Soc.* **1989**, *111*, 5680.

(50) Tucker, S. C.; Truhlar, D. G. *J. Am. Chem. Soc.* **1990**, *112*, 3347.

(51) Zhao, X. G.; Tucker, S. C.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 826.

(52) Basilevsky, M. V.; Chudinov, G. E.; Napolov, D. V. *J. Phys. Chem.* **1993**, *97*, 3270.

(53) Mathis, J. R.; Bianco, R.; Hynes, J. T. *J. Mol. Liq.* **1994**, *61*, 81.

(54) Truong, T. N.; Stefanovich, E. V. *J. Phys. Chem.* **1995**, *99*, 14700.

(55) Pomelli, C. S.; Tomasi, J. *J. Phys. Chem. A* **1997**, *101*, 3561.

(56) Cossi, M.; Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1998**, *297*, 1.

(57) Mo, Y.; Gao, J. *J. Comput. Chem.* **2000**, *21*, 1458.

(58) Safi, B.; Choho, K.; Geerlings, P. *J. Phys. Chem. A* **2001**, *105*, 591.

(59) Ohmiya, K.; Kato, S. *Chem. Phys. Lett.* **2001**, *348*, 75.

(60) Gao, J.; Garcia-Viloca, M.; Poulsen, T. D.; Mo, Y. *Adv. Phys. Org. Chem.* **2003**, *38*, 161.

(61) Mo, S. J.; Vreven, T.; Mennucci, B.; Morokuma, K.; Tomasi, J. *Theor. Chem. Acc.* **2004**, *111*, 154.

(62) Vayner, G.; Houk, K. N.; Jorgensen, W. L.; Brauman, J. I. *J. Am. Chem. Soc.* **2004**, *126*, 9054.

(63) Sato, H.; Sakaki, S. *J. Phys. Chem. A* **2004**, *108*, 1629.

(64) Freedman, H.; Truong, T. N. *J. Phys. Chem. B* **2005**, *109*, 4726.

(65) Song, L.; Wu, W.; Hiberty, P. C.; Shaik, S. *Chem.—Eur. J.* **2006**, *12*, 7458.

(66) Casanova, D.; Gusarov, S.; Kovalenko, A.; Ziegler, T. *J. Chem. Theory Comput.* **2007**, *3*, 458.

(67) Su, P.; Wu, W.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A*, In press.

(68) Hu, H.; Lu, Z.; Parks, J. M.; Burger, S. K.; Yang, W. *J. Chem. Phys.* **2008**, *128*, 034105.

(69) Ten-no, S.; Hirata, F.; Kato, S. *Chem. Phys. Lett.* **1993**, *214*, 391.

(70) Ten-no, S.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1994**, *100*, 7443.

(71) Sato, H.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1996**, *105*, 1546.

(72) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.

(73) Morita, A.; Kato, S. *J. Am. Chem. Soc.* **1997**, *119*, 4021.

(74) Morita, A.; Kato, S. *J. Chem. Phys.* **1998**, *108*, 6809.

(75) Hayashi, S.; Ohmine, I. *J. Phys. Chem. B* **2000**, *104*, 10678.

(76) Wilson, E. B., Jr.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; Dover: New York, 1955.

(77) Tishchenko, O.; Higashi, M.; Albu, T. V.; Corchado, J. C.; Kim, Y.; Villà, J.; Xing, J.; Lin, H.; Truhlar, D. G. *MC-TINKER*, version 2008; University of Minnesota: Minneapolis, MN, 2008. The correct analytical gradient of the MCMM energy with respect to the coordinates is given in the manual of this program [Tishchenko and Truhlar, Apendix: *MCMM Equations*, July 11, 2007, http://comp.chem.umn.edu/mc-tinker/MCMM-APPENDIX.pdf (accessed April 19, 2008)].

(78) Ischtwan, J.; Collins, M. A. *J. Chem. Phys.* **1994**, *100*, 8080.

(79) Nguyen, K. A.; Rossi, I.; Truhlar, D. G. *J. Chem. Phys.* **1995**, *103*, 5522.

(80) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811.

(81) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.

(82) Higashi, M.; Chamberlin, A. C.; Pu, J.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Lynch, B. J.; Liotard, D. A.; Rinaldi, D.; Gao, J.; Cramer, C. J.; Truhlar, D. G. *GAMESSPLUS*, version 2008; University of Minnesota: Minneapolis, MN, 2008.

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

(84) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(85) Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566.

(86) Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576.

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

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

(89) Case, D. A.; Darden, T. A.; Cheatham, I. T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, CA, 2006.

(90) Ponder, J. W. *TINKER*, version 3.5; Washington University: St. Louis, MO, 1997.

(91) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637.

(92) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952.

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

(94) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339.

(95) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275.

(96) McLennan, D. J. *Aust. J. Chem.* **1978**, *31*, 1897.

(97) Zhang, Y.; Lin, H.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1378.

(98) Burant, J. C.; Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Chem. Phys. Lett.* **1996**, *258*, 45.

(99) König, P. H.; Hoffmann, M.; Frauenheim, Th.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082.

(100) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632.

(101) Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991.

CT8000816

JCTC Journal of Chemical Theory and Computation

Ab Initio Molecular Dynamics Study of a Highly Concentrated LiCl Aqueous Solution

L. Petit,[†] R. Vuilleumier,*,[‡] P. Maldivi,[†] and C. Adamo[§]

Laboratoire de Reconnaissance Ionique et de Chimie de Coordination, CEA - INAC/LCIB (UMRE 3 CEA-UJF), 17 rue des Martyrs, F-38054 Grenoble Cedex 9, France, Laboratoire de Physique Théorique de la Matiere Condensée, UMR7600, Université Pierre et Marie Curie, Paris, Tour 24 Boite 121, 4 place Jussieu, F-75252 Paris CEDEX 05, France, and Laboratoire d'Electrochimie et de Chimie Analytique, CNRS UMR-7575, Ecole Nationale Supérieure de Chimie de Paris, 11 rue P. et M. Curie, F-75231 Paris Cedex 05, France

Received January 5, 2008

Abstract: The properties of a highly concentrated aqueous lithium chloride solution (|LiCl| = 14 mol L^{-1}) are investigated using Car−Parrinello molecular dynamics. The coordination spheres of lithium ions, chloride ions, and water molecules are described successively. On the whole, our simulation provides results—distances and coordination numbers—in very good agreement with experimental data. The lithium solvation shell is found to exhibit a tetrahedral configuration on average, with three stable clusters observed during the simulation: Li^+-4H_2O, $Li^+(H_2O)_3Cl^-$, and $Li^+(H_2O)_2(Cl^-)_2$. The chloride coordination sphere is logically formed by strong Cl−H hydrogen bonds with neighboring water molecules, for a mean coordination number of 4.4. The structuring of water molecules is strongly affected by the high concentration in LiCl. The hydrogen bond network is globally broken down, but little variation is calculated on water dipoles (μ = 3.07 D) because of the strong polarization from Li^+ and Cl^- ions. We also point out some of the characteristic features of such a highly concentrated solution: water bridging between Li^+ and Cl^- hydration spheres, Li^+-Cl^- ion-pairing, and intermediate behavior between dilute solutions and molten salts. Finally, the reliability of our simulation to describe ion-pairing is discussed.

1. Introduction

Lithium chloride aqueous solutions have been extensively studied because of their implication in many fields: biology, medicine, or electrochemistry. LiCl also dissolves outstandingly well in water because of the strong difference in the ionic radius between lithium and chloride ions,[1,2] making them suitable prototypes to investigate the properties of ionic complexes in highly concentrated aqueous solutions. To date, there has been considerable experimental work on the properties of LiCl solutions, with various methods: viscosimetry measurements,[3,4] NMR and Raman spectroscopies,[5,6] X-ray diffraction,[7–10] and especially neutron diffraction.[9,11–21] Yet, the isotopic substitution technique used to analyze neutron diffraction patterns[22,23] has been several times called into question because of large statistical errors.[17,23] More generally, all of these methods do not provide access to the detailed microscopic structure of LiCl, and the structuring of water surrounding the ions is usually clouded by the bulk water response. Theoretical methods, based on QM/MM,[24,25] Monte Carlo,[26,27] classical molecular dynamics approaches,[2,28–31] have been proposed to overcome such limitations, successfully or not. For instance, classical molecular dynamics seem to be dramatically dependent on the type of model used to describe the atomic interactions and, in particular, whether polarizability effects are considered. Egorov et al.[28] have compared the results from three different interaction potentials on LiCl solutions over a wide range of concentrations.

---

* Corresponding author e-mail: vuilleum@lptl.jussieu.fr.

[†] Laboratoire de Reconnaissance Ionique et de Chimie de Coordination.

[‡] Université Pierre et Marie Curie.

[§] Ecole Nationale Supérieure de Chimie de Paris.

10.1021/ct800007v CCC: $40.75

They found that the structure of the Li^+-H_2O interaction is indeed strongly dependent on the potential, with coordination numbers ranging from 4 to 6.

For a few years now, first principles molecular dynamics have been carried out to study ion solvation in water at low concentrations, and results are globally in very good agreement with experimental data.[32–44] They provide a fine description of the microstructure of both the solute and solvent and should thus enable the treatment of LiCl solutions in a more unbiased way than classical molecular dynamics. These simulations pave the way for the study of electrolyte solutions at higher concentrations as well as the study of solvation effects of metal ions in electrolyte solutions.

For instance, highly concentrated solutions of lithium chloride have been used to model the complexation of chloride ions to trivalent f elements,[45] complexes that are assumed to be formed in geological salt formations when nuclear wastes are being stored. Yet, the coordination sphere of f elements in LiCl aqueous solutions is still uncertain (see ref 46 and references therein for more details), and theoretical studies could thus help to clarify this aspect. In the following, we study the local structure of a highly concentrated lithium chloride solution (|LiCl| = 14 M) with Car–Parrinello molecular dynamics[47] based on density functional theory (DFT). Previous works have demonstrated the ability of DFT to describe the solvation of Li^+ and Cl^- ions at infinite dilution.[34–36,41,43,48] Yet, to our knowledge, such a technique has never been applied to LiCl aqueous solutions. Our ultimate goal is to use ab initio molecular dynamics to investigate the coordination sphere of trivalent f elements in such a solution. The present work must thus be considered as a preliminary step to check the reliability of our calculations as well as a way to solve some of the uncertainties that still remain about the local structure of LiCl solutions at high concentrations.

2. Computational Details

The simulation was performed with the CPMD code.[49] The BP86 GGA functional[50] was applied throughout as this functional has already proved to provide a reliable description of f elements complexes,[51,52] while it leads to a reasonable description of pure water.[53] We have thus kept this functional in order to ensure a consistent study with the next work including f elements (see the Introduction). Troullier–Martins norm-conserving pseudopotentials[54] were used along with the Kleinman–Bylander decomposition for all Li, Cl, O, and H atoms.[55] The reference electronic configurations for O, H, and Cl were the neutral atoms. Pseudization radii of s = p = 1.52 au for chlorine, 2s = 2p = 3d = 1.05 au for oxygen, and 1s = 2p = 0.5 au for hydrogen atoms were used. The pseudopotentials for BP are similar to those used in ref 36 for the simulation of Cl^- in liquid water, where they were shown to correctly describe the chloride–water interaction. For lithium, we have employed a semicore pseudopotential starting from Li^+ as a reference configuration. The pseudization radius of the 1s electrons was set to 0.51 au. In the CPMD simulation, the electronic wave functions have been expanded in a plane-waves basis set up to the energy cutoff of 70 Ry. Pseudopotentials and the cutoff were checked on Cl^--H_2O and Li^+-H_2O systems by comparing the CPMD optimized structures and interaction energies with that obtained from Gaussian 03 (MP2 and BP86, 6-311++G(d,p); ref 56). Discrepancies are very low between CPMD and G03/BP86 calculations, and they do not exceed 0.09 Å on distances and 1.5 kcal/mol on energies with respect to MP2 results (see Table 1).

Table 1. Comparison of Geometries and Interaction Energies (E_{int}) between Gaussian 03 (DFT or MP2) and CPMD (BP86, Troullier–Martins Pseudopotentials) Calculations[a]

	G03/MP2/ 6-311++G(d,p)	G03/BP86/ 6-311++G(d,p)	CPMD/ BP86
	Li^+-H_2O		
d(Li–O), Å	1.86	1.86	1.86
d(O–H), Å	0.96;0.96	0.97;0.97	0.98;0.98
$(Li\hat{O}H)$, deg	127.6;127.6	127.4;127.4	126.7;128.3
E_{int}, kcal/mol	−33.4	−32.5	−32.4
	Cl^--H_2O		
d(Cl–H), Å	2.15	2.10	2.06
d(O–H), Å	0.98;0.96	1.01;0.97	1.02;0.97
$(H\hat{O}Cl)$, deg	165.4	168.8	174.1
E_{int}, kcal/mol	−13.7	−15.2	−15.2

[a] Gaussian interaction energies are BSSE-corrected.

A preliminary classical molecular dynamics step was performed with the Moldy package[57] to get a reliable initial configuration for the CPMD simulation as well as to determine an adequate size for the periodic box. The Lennard-Jones 12-6 potential was taken from Dang[58] following the previous simulations by Egorov et al.[28] Simulations were carried out for 400 ps on three different sizes of box with a density of 1.196 corresponding to a concentration of 14 M (box 1, 243 H_2O, 50 LiCl; box 2, 60 H_2O, 12 LiCl; box 3, 30 H_2O, 6 LiCl). The system containing 30 H_2O, 6 Cl^-, and 6 Li^+ was found to give reasonable results compared to larger boxes and was thus kept for the CPMD simulation.

The Car–Parrinello simulation was performed in a cubic box of size 10.33 Å with periodic boundary conditions. To ensure an accurate description of structural properties, we have taken the fictitious mass $\mu = 200$ au, while the time step was set to 3.0 au (0.072 fs). This choice should allow for negligible dependency of our results with respect to the unphysical parameter μ.[59] The temperature was first fixed to 300 K for 5 ps to make the system equilibrate, and then the trajectories were sampled for 25 ps in the NVE ensemble. To compute molecular dipole moments, the maximally localized Wannier functions[60] were collected every 10 steps for the first 15 ps of the simuation.

In the following, the program VMD[61] was used for visualizations. Radial distribution functions $g(r)$ were also analyzed with VMD. Corresponding $g(r)$ results are presented from 0 to 5 Å to avoid double-counting due to periodic boundary conditions.

3. Results and Discussion

3.1. Lithium Ions Coordination Sphere. The radial distribution functions of lithium ions are depicted in Figure

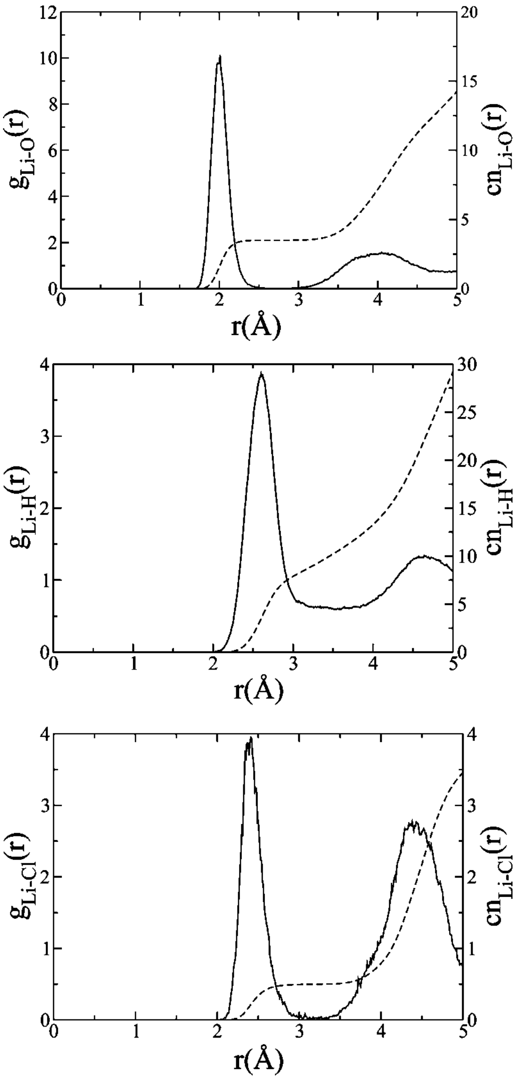


Figure 1. Radial distribution function *g* (full line) and coordination number cn (dashed line) for the Li^+-O, Li^+-H, and Li^+-Cl^- interactions.

1, while the first peaks are detailed in Table 2. We first focus on the interaction between Li^+ and water molecules. The first Li−O peak at 2.01 Å compares nicely with the experimental values determined by neutron diffraction. This first peak is quite sharp. It is followed by a well-defined minimum, which indicates that the lithium first hydration sphere is fairly rigid and clearly separated from the second one. This may be the result of the small size and high electronegativity of the Li^+ ion, which makes the Li−O interaction quite strong and thus prevents the exchange of water molecules with outer hydration spheres. Actually, comparisons with studies at lower concentrations in LiCl show that the Li−O distance is rather constant, whereas the coordination number is much more dependent on concentration.[17] From the integration of the first peak of $g(r)$, we found 3.5 water molecules on average around each lithium ion (Table 2). Such a value does not mean that the lithium coordination sphere is mobile but comes from a balance between three stable clusters: Li^+-4H_2O, $Li^+(H_2O)_3Cl^-$, and $Li^+(H_2O)_2(Cl^-)_2$. There are indeed several occurences of Li^+-Cl^- ion pairing, as will be discussed in section 3.5. For all three clusters, the total coordination number of Li^+ ions is 4. Actually, we observe four Li^+-4H_2O clusters, one $Li^+(H_2O)_3Cl^-$ cluster, and one $Li^+(H_2O)_2(Cl^-)_2$ cluster, leading to the calculated Li−O coordination number of 3.5. Such a value is in agreement with experimental data, but longer and probably larger simulations are needed to provide a definitive picture (see also the discussion in section 3.5). Snapshots illustrating these three clusters are displayed in Figure 2. The oxygen atoms of the water molecules and the chloride ions occupy the vertices of a slightly distorted tetrahedron. The angular distribution ($O-L\hat{i}-O$) is centered at 108.6° and shows that a tetrahedral-like configuration (109°) is indeed found.

The impact of the lithium ion on the second coordination sphere is much less pronounced. As can be seen from Figure 1, the second peak of g_{LiO} is located around 4 Å but is rather flat and not well-resolved. It also partially penetrates into the following peak, suggesting an exchange between the water molecules of the second hydration shell and the bulk. The agreement with the experimental value (4.4−4.7 Å; ref 7) becomes poorer, but this is the only experimental result we have found describing the second solvation shell. Note that the concentration in this work is lower than ours (10 M).

Even so, the second peak shows evidence of the structuring of the lithium second coordination sphere, as sketched in Figure 3. The Li−O distance is 2.0 Å, while, as will be shown in section 3.3, the O−O distance is around 2.8 Å (Table 2). If the solution was totally unstructured, we should observe the second peak for the Li−O interaction around 4.8 Å. In our simulation, this distance is reduced to 4.0 Å. From Figure 3, it is then clear that the ($Li-\hat{O}-O$) angle is near 109°, that is, close to a tetrahedral configuration around oxygen atoms of the first solvation shell.

3.2. Chloride Ions Coordination Sphere. As for the lithium coordination sphere, we give in Figure 4 the radial distribution functions for the Cl^- ion, while more details on the main peaks can be found in Table 3. The positively charged hydrogen atoms of the water molecules are oriented toward the negatively charged chloride ions, forming hydrogen bonds. The first peak is thus due to the Cl−H interaction and is calculated at 2.13 Å. It is well-defined, but the nonzero minimum of the radial distribution function indicates that the water molecules in the first coordination sphere are more mobile than with the lithium ion. Integration yields to a coordination number of 4.4 hydrogen atoms around each chloride ion, but this is not straightforward to interpret. Indeed, because of ion pairing, the radial distribution function of Cl−O interaction takes into account water molecules that belong to the lithium first coordination sphere. This could explain how we can find more oxygen atoms (5.7) than hydrogen atoms (4.4) around each chloride ion.

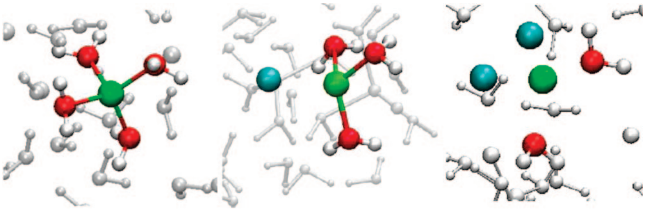
The second Cl−H peak is not sharply defined, lying around 3.6 Å. In Figure 4, we observe that it is actually divided into two lower peaks, showing that the different hydration shells can interpenetrate each other.

Even though the chloride first hydration sphere is found to be rather flexible, hydrogen bonds should be nevertheless quite strong, as it is well-known that hydrogen bonds involving charged ions feature high binding energies. This

Table 2. Structural Parameters for the Li^+ Hydration[a]

	CPMD calc.		experiment				
	r, Å	cn	*r*, Å	cn	|LiCl|	technique	ref
Li−O, 1st peak	2.01	3.5	1.95 ± 0.02	3.3 ± 0.5	9.95 m	ND	15
			1.96 ± 0.02	3.2 ± 0.2	14 m	ND	17
			2.01 ± 0.05	4 ± 1	9.5 m	ND	19
			2.22 ± 0.02	4	13.75 M	ND	9
			1.95 ± 0.02	3.3 ± 0.5	9.95 m	ND	20
Li−O, 2nd peak	≃4.0		4.4−4.7		10 M	XRD	7
			4.5		9.95 m	ND	15
Li−H, 1st peak	2.60	≃8.7	2.50 ± 0.2		9.95 m	ND	15
			2.52 ± 0.02		14 m	ND	17
			2.61 ± 0.05		9.5 m	ND	19
			2.68 ± 0.01		13.75 M	ND	9
			2.50 ± 0.02		9.95 m	ND	20
Li−Cl, 1st peak	2.41	0.5	2.76	1.5	12.8 m	ND	74
Li−Cl, 2nd peak	4.38		3.28	7.4	12.8 m	ND	74

[a] *r* is the interatomic distance, and cn is the coordination number. ND, neutron diffraction; XRD, X-ray diffraction. M = mol L^{-1}; m = mol kg^{-1}.



$[Li(H_2O)_4]^+$ $[Li(H_2O)_3(Cl)]$ $[Li(H_2O)_3(Cl)_2]^-$

Figure 2. Representation of the three stable clusters for the lithium coordination sphere. Lithium is in green, chloride in blue, oxygen in red, and hydrogen in white.

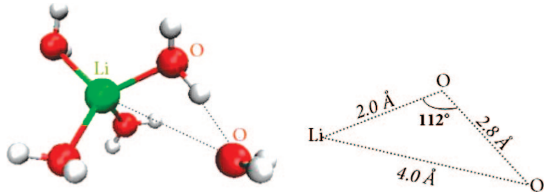


Figure 3. Sketch of the interaction between one lithium ion and two water molecules.

can be assessed through intermolecular distances, as shorter interatomic distances correspond to stronger hydrogen bonds.[62] In our simulation, the Cl−H distance (2.13 Å) is clearly shorter than the sum of the Cl^- and H radii (r_{Cl^-} = 1.81 Å, r_H = 0.53 Å), and Cl−H hydrogen bonds must thus be rather strong and stable. As given in Table 3, the first oxygen atoms are found at 3.13 Å, which exactly corresponds to the sum of the Cl−H distance (2.13 Å) and the O−H distances (1.01 Å). This indicates that the Cl−H_2O interaction is quite directional with $(Cl-\hat{H}-O) \simeq 180°$. Such an intermediate position of the hydrogen atom along the connecting line between Cl and O atoms is also very favorable for the hydrogen bond strength.[63]

To get further insights into chloride ions' behavior, the g_{Cl-Cl} radial distribution function is also plotted in Figure 4. In agreement with neutron diffraction patterns,[11,64] there are two dominant peaks. They are badly resolved because the total number of chloride ions in the simulation box is too low, and the statistics are thus rather inaccurate. The first peak is calculated at 3.81 Å, versus 3.75 Å and 4.2 Å for the experimental counterparts. In our simulation as well as in experimental measurements, this peak is assigned to the interaction between the chloride ions involved in ion pairing. It is actually quite characteristic of LiCl molten salts, for which it is indeed attributed to Li^+−Cl^- direct contact.[11]

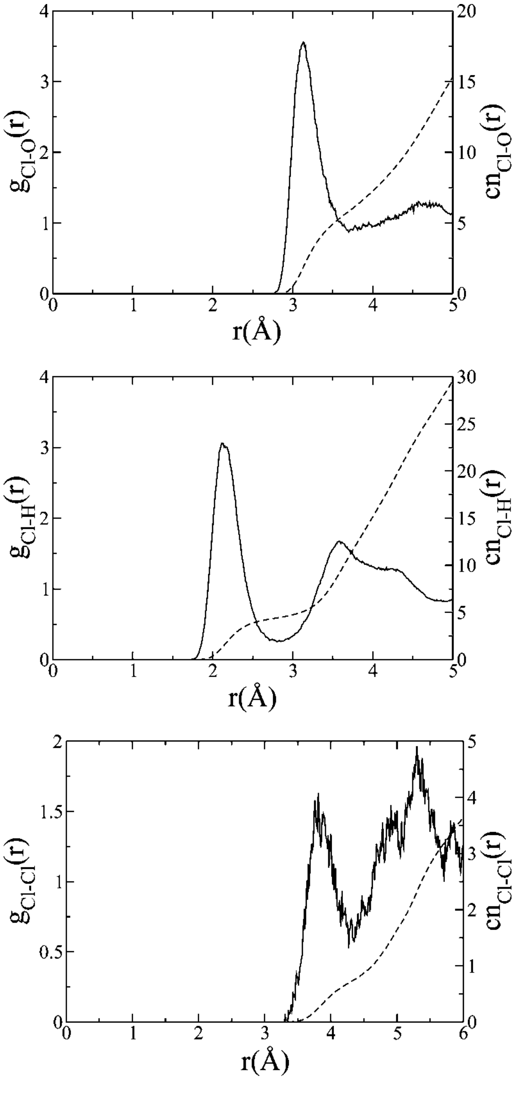


Figure 4. Radial distribution functions *g* (full line) and coordination number cn (dashed line) for the Cl^-−O, Cl^-−H, and Cl^-−Cl^- interactions.

Table 3. Structural Parameters for the Cl$^-$ Hydration[a]

	CPMD calc.		experiment				
	r, Å	cn	*r*, Å	cn	[LiCl]	method	ref
Cl−O, 1st peak	3.13	5.7	3.10		11 M	XRD	10
			3.29 ± 0.04	5.3 ± 0.2	9.95 m	ND	14
Cl−H, 1st peak	2.13	4.4	2.22 ± 0.02		9.95 m	ND	14
			2.24	4.4 ± 0.3	14.9 m	ND	11
			2.225		9.2 M	ND	18
			2.238 ± 0.006		13.75 M	ND	9
Cl−H, 2nd peak	3.59		3.5−3.7		9.95 m	ND	14
			3.56 ± 0.04		14.9 m	ND	11
			3.625		9.2 M	ND	18
Cl−Cl, 1st peak	3.81	≃0.7	3.75 ± 0.03	2.3 ± 0.3	14.9 m	ND	11
			4.2 ± 0.1	1.2 ± 0.5	8.6 m	ND	64
Cl−Cl, 2nd peak	5.3		6.38 ± 0.03	10.0 ± 0.5	14.9 m	ND	11
			6.4 ± 0.2	7.0 ± 1.0	8.6 m	ND	64

[a] *r* is the interatomic distance, and cn is the coordination number. ND, neutron diffraction; XRD, X-ray diffraction. M = mol L^{-1}; m = mol kg^{-1}.

Table 4. Structural Parameters for the Water Molecules[a]

	CPMD, calcd		experiment, LiCl					experiment, pure water		
	r, Å	cn	*r*, Å	cn	[LiCl]	method	ref	*r*, Å	cn	ref
O−H, 1st peak, intra	1.01	2.0	0.97 ± 0.02		10 m	ND	12	1.02 ± 0.02		76
			0.97		9.2 M	ND	18	0.966 ± 0.006		77
			0.970 ± 0.002		13.75 M	ND	16			
			0.940 ± 0.005		13.75 M	ND	9			
O−H, 2nd peak, H-bond	1.71	3.0 (1.0)	1.8 ± 0.1		10 m	ND	12	1.87 ± 0.04		76
O−H, 3rd peak, inter	3.20		3.5 ± 0.4		10 m	ND	12	3.3 ± 0.2		76
H−H, 1st peak, intra	1.60	1.0	1.55 ± 0.02		10 m	ND	12	1.50 ± 0.02	1.2 ± 0.1	76
			1.57		9.2 M	ND	18	1.51 ± 0.03		77
			1.59 ± 0.01		13.75 M	ND	16			
			1.487 ± 0.006		13.75 M	ND	9			
H−H, 2nd peak, H-bond	2.26	3.8 (2.8)	2.3 ± 0.06		10 m	ND	12	2.44 ± 0.04	6.0 ± 0.4	76
H−H, 3rd peak, inter	≃3.65		3.8 ± 0.2		10 m	ND	12	3.8 ± 0.2	30 ± 4	76
O−O, 1st peak, H-bond	2.71	3.0	2.85		13.75 M	ND	9			
O−O, 2nd peak, inter	3.26	5.3 (2.3)						3.6	5.2 ± 0.04	78
O−O, 3rd peak, inter	≃4.25							4.5		78

[a] *r* is the interatomic distance, and cn is the coordination number. Coordination numbers in parentheses correspond to the real integration; i.e., they take into account previous peaks. ND, neutron diffraction. M = mol L^{-1}; m = mol kg^{-1}.

Degreve and Mazzé[2] have noticed that a Cl−Cl distance around 3.8 Å is close to the sum of the ionic radii of two chlorides (r_{Cl^-} =1.81 Å) and that such a short distance between two anions is only possible if a cation is close enough to provide electrostatic stabilization. This is particularly the case for the observed Li$^+$(H$_2$O)$_2$(Cl$^-$)$_2$ cluster, although a coordination number around 0.7 indicates that this is not the only occurrence of Cl−Cl close contact. In contrast, the second peak found in the interval from 4.5 to 5.7 Å (as explained in the Computational Details, the radial distribution function is strictly valid only between 0 and 5 Å, so the reliability of the second peak at 5.3 Å is questionable), with a maximum at 5.3 Å, is consistent with the usual distance between two solvated chloride ions at a lower dilution. This has already been suggested by Neilson and Enderby[65] for a dilute solution of NiCl$_2$. It is interesting to compare these two peaks, as they underline the intermediate character of our concentrated LiCl solution (14 M) between a dilute solution (2nd peak) and a molten salt (1st peak). The first peak, experimentally observed, is a signature of Li−Cl ion pairing in highly concentrated LiCl solutions.

3.3. Water Structure. Distances and coordination numbers for water molecules in concentrated lithium chloride (14 M) are listed in Table 4. Radial distribution functions for the O−H and O−O interactions are also displayed in Figure 5. They are compared to available experimental results for LiCl solutions at close concentration. In the last columns, data for pure liquid water are also presented for comparison. In order to make the analysis easier, the main peaks are classified according to their nature: intramolecular ("intra"), intermolecular ("inter"), or hydrogen-bonding ("H-bond") interactions.

In Table 4, the peak centered at 1.01 Å is logically associated with the intramolecular O−H bond, in agreement with experimental values. As observed by Jal et al.[18] with neutron diffraction, no discrepancy is found in the intramolecular water structure with respect to pure water. There is in contrast a small increase in the H−H intramolecular distance with respect to pure water, but there is no straightforward link with the interaction with Li$^+$ and Cl$^-$ ions. In particular, the angular distribution (H−$\hat{O}$−H) is centered at 104.4° and remains close to the value in pure water (104.5°, ref 66).

As could be expected, the hydrogen-bonding network is more affected. We globally observe a strong reorganization of the bulk solvent with respect to dilute solutions. In pure water, molecules are usually arranged in a tetrahedral network: each oxygen atom accepts two hydrogen bonds

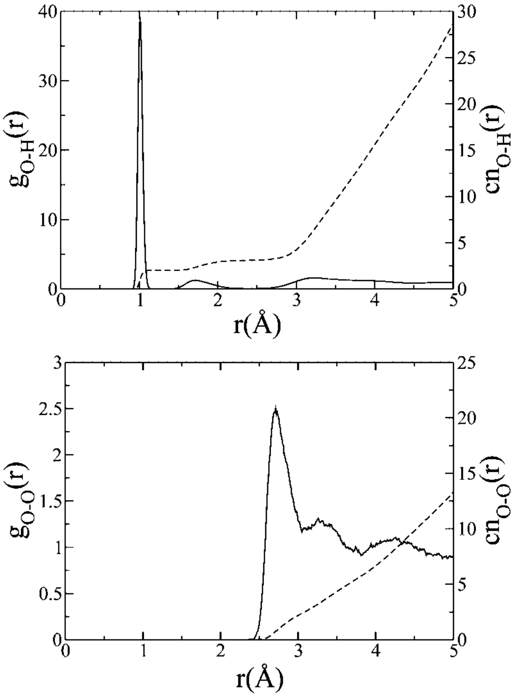


Figure 5. Radial distribution functions *g* (full line) and coordination number cn (dashed line) for the O−H and O−O interactions.

from neighboring water molecules, and nearly every hydrogen atom interacts with another oxygen atom. In our simulation, oxygen atoms accept only one hydrogen bond on average, as the H-bond peak of g_{OH} integrates for 1.0. Disruption of the water structure is further suggested by the mixing of g_{HH} peaks at a high LiCl concentration, which are normally well-defined at 2.44 and 3.8 Å in pure water. The length of O•••H hydrogen bonds is more difficult to understand. We can indeed expect from the general breakdown of the solvent structuring that hydrogen bonds should be weakened in our solution with respect to pure water, and so their length should increase. In Table 4, we observe the opposite trend: the O−H hydrogen bonds goes from 1.87 Å in pure water to 1.71 Å in our calculations. This shortening may be overestimated in our simulation, but the global trend is actually correct, as can be seen from the available experimental data on LiCl solutions (1.8 Å, see Table 4). We think this could be the result of a positive cooperation effect between ions and water molecules. Oxygen atoms interacting with lithium ions are strongly polarized; their hydrogen atoms become more acidic, and thus hydrogen bonds can be locally strengthened.

Finally, because of the high concentration in LiCl, there is a deficiency in water molecules in our solution. Lithium and chloride ions must then share their coordination sphere, and indeed we observe several occurrences of water-bridging configurations in our simulation. Figure 6 shows how they arrange one another, the negatively charged oxygen atoms interacting with the cation Li^+, while the positive hydrogen atoms are oriented toward chlorides. No bifurcated hydrogen bond, that is, one chloride anion forms two hydrogen bonds with the two hydrogen atoms of a water molecule, is nevertheless observed during the simulation.

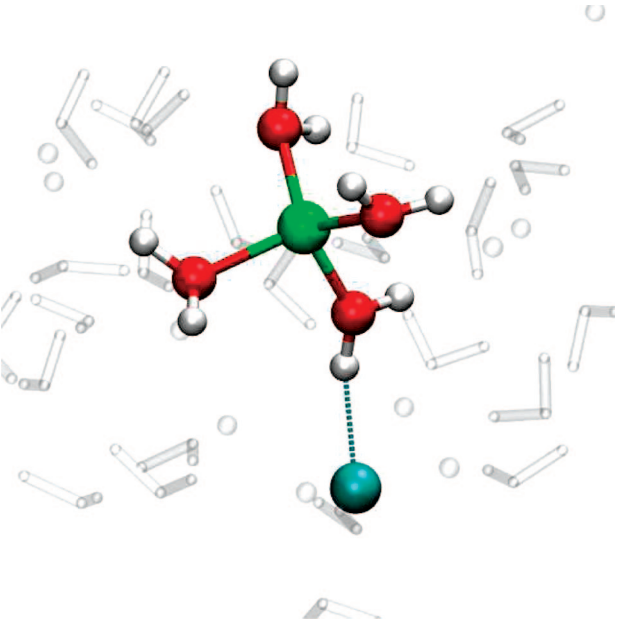


Figure 6. Water bridging between Li^+ and Cl^- ions: lithium in green, chloride in blue, oxygen in red, and hydrogen in white.

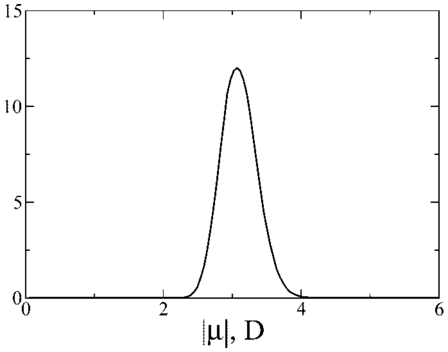


Figure 7. Distribution of the dipole moment of water molecules in |LiCl| = 14 M.

3.4. Dipole Moments. In order to assess these effects on the electronic structure of water molecules, we have estimated their dipole moments. The molecular charge distribution was studied using maximally localized Wannier functions. These provide an unambiguous way to assign dipole moments to molecules. The dipole moment of water molecules in pure water has been previously estimated with Car−Parrinello molecular dynamics as being on the order of 3.0 D by Silvestrelli and Parrinello.[2,68] This value is very close to the value proposed by Badyal et al.,[69] 2.9 ± 0.6 D, for pure water, from an analysis of the experimental X-ray diffraction form-factor. Other theoretical values were also proposed; see, for instance, refs 68− 72. The dipole moment distribution found in our simulation features a main peak at 3.07 D (Figure 7). We thus do not find a strong influence of the electrolytes on the dipole moment of water molecules. The effect of broken hydrogen bonds is compensated by the strengthening of the remaining bonds and polarization by the Li^+ and Cl^- ions.

3.5. Discussion. In order to rationalize the structuring of water in electrolyte solutions, Collins[1] has proposed a general principle in which ions are classified as being either structure-maker or structure-breaker ions. Structure-maker ions bind tightly to water molecules and thus form well-defined ion−water complexes. In contrast, structure-breaker ions give weak interactions with water molecules, and their hydration sphere is thus quite mobile. The lithium ion is small, with a high charge density, and is thus classified as a structure-maker ion. Its interaction with surrounding water molecules is quite strong, and its hydration sphere is oriented so as to give favorable electrostatic interactions.[73] Chlorides are usually considered to be structure-breaker ions because of their fairly large ionic radii, and their hydration sphere is thus much more flexible than for lithium. Hribar et al.[73] have shown with Monte Carlo simulations that the organization of water molecules results from a balance between two competitive effects: the ion−water interaction, which is governed by electrostatic effects (charge densities), and water−water interactions that are stabilized by hydrogen bonding. They confirmed that, for small cations like Li^+, the high charge density favors strong electrostatic interactions with nearby water molecules, but such a strong ordering of water molecules also decreases their ability to form a well-structured hydrogen-bond network. This is indeed what we observe in the present simulation.

According to Collins' classification also,[1] ion pairing mainly occurs between similarly sized ions that belong to the same category, either structure-maker or structure-breaker. Thus, lithium and chloride ions should have a low tendency to get in direct contact, and they best remain apart in solution, explaining their high solubility in water. Even so, there has been a long debate about the occurrence of ion pairing between Li^+ and Cl^- ions for highly concentrated solutions.

A very recent simulation with the reverse Monte Carlo method has shown that there was no chloride ion within the lithium first coordination sphere, even at saturation.[27] Yet, most neutron diffraction works,[11,17,64,74] many of them quite recent, have evidenced ion pairing for concentrated solutions, either directly through the Li^+-Cl^- radial distribution function or indirectly with the Cl^--Cl^- interaction around 3.8 Å, as discussed in section 3.1.[11,64] Rudolph et al.[6] have studied, with Raman spectroscopy, aqueous LiCl solutions from saturation to low concentration. They also showed that ion pairing occurs at high concentrations through the Li−Cl vibration around 376 cm^{-1}, which was not observed for a more dilute solution. X-ray Compton scattering experiments[75] as well as molecular dynamics simulations[2,30] also support the existence of ion pairing.

In our simulation, there is clear evidence that chloride can enter the lithium first coordination sphere, as we find the Li−Cl mean distance to be 2.41 Å, versus 2.01 Å for the first oxygen atoms and 2.60 Å for hydrogen atoms (see Table 2). This peak is badly resolved but is quite thin and integrates for 0.5 chloride on average (see Figure 1). Only one piece of experimental data is available for this distance, from Ichikawa et al.,[74] and exhibits a Li^+-Cl^- direct contact around 2.76 Å. Degreve and Mazzé[2] have obtained from molecular dynamics a distance of 2.20 Å. With increasing LiCl concentration, the number of water molecules available to interact with the lithium ion becomes smaller, and at 14 M, there are not enough water molecules left to complete the lithium cation coordination sphere. Chlorides can then penetrate the lithium hydration shell to form ion pairs, as represented in Figure 2.

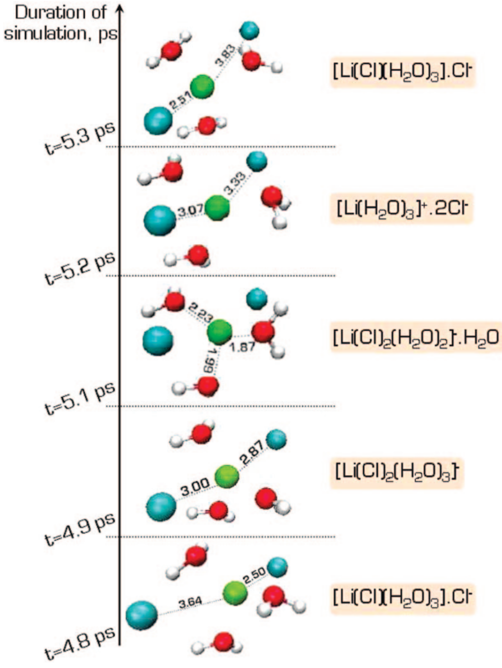


Figure 8. Exchange of chloride ions in lithium first coordination sphere. Distances are given in angtröms. Lithium is in green, chlorides are in blue, oxygen atoms in red, and hydrogen atoms in white.

In the Computational Details, we have specified that the starting configuration for the ab initio MD simulation is based on a classical molecular dynamics calculation performed with the Lennard-Jones potential, as proposed by Dang.[58] This preliminary step already predicts direct contact between Li^+ and Cl^- ions, and we logically wondered if this initial configuration could have biased our results. Indeed, the concentration in our simulation is so high that atoms could hardly be displaced from their position at the end of the classical MD step. We have thus compared the Li^+-Cl^- radial distribution function patterns in our simulation with that obtained by Egorov et al.[28] from the Dang potential with classical MD for the Li−Cl interaction. Results are rather close since both plots show a quite pronounced first peak around 2.4 Å, followed by a smoother peak between 4 and 5 Å. The residence time of water molecules around lithium ions is usually assessed in the range 30−100 ps,[30,31] so no full reorganization of water in the lithium coordination sphere is expected from an ab initio simulation. In contrast, the Li^+-Cl^- interaction has been calculated to be much shorter, on the order of 4.5 ps, and our simulation should thus be long enough (30 ps total) to allow the breakdown of ion pairing. And indeed, we do observe chloride exchange in the lithium first coordination sphere during our simulation. In Figure 8, we have represented how one of these exchanges occurs: starting at the end of the thermalization (4.9 ps), one chloride ion enters a lithium coordination sphere. There are then five neighbors around the lithium ion ($2Cl^- + 3H_2O$), whereas the common coordination number is 4. Because of the steric hindrance, two transient configurations are then observed: in the first one, at 5.1 ps, one water molecule tends to move away from the cation (d(Li−O) = 2.23 Å), while in the next configuration (t = 5.2 ps), both chloride ions are pulled away. Only one chloride finally moves to the second coordination sphere, and the initial coordination shell is restored. It is also of importance to underline that the mean square displacement of chloride ions and oxygens atoms is about 1.5 Å^2. This indicates that limited rearrangement occurs but that Cl and O atoms do move away from their initial coordination cage. This is further confirmed by the strong modification in distances with respect to the classical MD results. Of course, a longer simulation could provide a more clear-cut vision of ion pairing and of the related number of water molecules in the solvation shell of Li$^+$. Starting from another initial configuration could also be a relevant way to check our results. Even so, as most experimental data, our simulation agrees with the occurrence of ion pairing.

4. Conclusion

We have studied, with ab initio molecular dynamics, the local structure and behavior of a highly concentrated aqueous solution of lithium chloride. The structure of the ion hydration spheres is very well reproduced, with a long-lived tetrahedral configuration for the lithium ion, and six water molecules around chloride anions. The water structure is deeply affected for such a high concentration: hydrogen bonds are massively disrupted, and there is also evidence of water-bridging configurations between Li$^+$ and Cl$^-$ hydration spheres because of the low content in water molecules. Few occurrences of ion pairing have also been observed, showing the intermediate character between solution and ionic liquid. Interatomic distances are in very good agreement with both experimental and theoretical results, as are the observed coordination numbers. The latter, especially around Li$^+$, may be affected by the simulation length because of the long time scale of water exchange around Li$^+$. However, this is very encouraging for a future study on the coordination sphere of f elements in concentrated LiCl.

Acknowledgment. We would like to thank the Centre de Calcul Recherche et Technologie (CCRT) for a generous allocation of computer time.

References

(1) Collins, K. D. *Biophys. J.* **1997**, *72*, 65.

(2) Degreve, L.; Mazzé, F. M. *Mol. Phys.* **2003**, *101*, 1443.

(3) Martin, J. M.; Berntsson, T. S. *J. Chem. Eng. Data* **1994**, *39*, 68.

(4) Moynihan, C. T.; Balitactac, N.; Boone, L.; Litovitz, T. A. *J. Chem. Phys.* **1971**, *55*, 3013.

(5) Bryant, R. G. *J. Phys. Chem.* **1969**, *73*, 1153.

(6) Rudolph, W.; Brooker, M. H.; Pye, C. C. *J. Phys. Chem.* **1995**, *99*, 3793.

(7) Lawrence, R. M.; Kruh, R. F. *J. Chem. Phys.* **1967**, *47*, 4758.

(8) Brady, G. W. *J. Chem. Phys.* **1958**, *28*, 464.

(9) Narten, A. H.; Vaslow, F.; Levy, H. A. *J. Chem. Phys.* **1973**, *58*, 5017.

(10) Yamanaka, K.; Yamagami, M.; Takamuku, T.; Yamaguchi, T.; Wakita, H. *J. Phys. Chem.* **1993**, *97*, 10835.

(11) Copestake, A. P.; Neilson, G. W.; Enderby, J. E. *J. Phys. C: Solid State Phys.* **1985**, *18*, 4211.

(12) Tromp, R. H.; Neilson, G. W.; Soper, A. K. *J. Chem. Phys.* **1992**, *96*, 8460.

(13) Ohtomo, N.; Arakawa, K. *Bull. Chem. Soc. Jpn.* **1979**, *52*, 2755.

(14) Cummings, S.; Enderby, J. E.; Neilson, G. W.; Newsome, J. R.; Howe, R. A.; Howells, W. S.; Soper, A. K. *Nature* **1980**, *287*, 714.

(15) Newsome, J. R.; Neilson, G. W.; Enderby, J. E. *J. Phys. C: Solid State Phys.* **1980**, *13*, L923.

(16) Ichikawa, K.; Kameda, Y. *J. Phys.: Condens. Matter* **1989**, *1*, 257.

(17) Howell, I.; Neilson, G. W. *J. Phys.: Condens. Matter* **1996**, *8*, 4455.

(18) Jal, J. F.; Soper, A. K.; Carmona, P.; Dupuy, J. *J. Phys.: Condens. Matter* **1991**, *3*, 551.

(19) Yamagami, M.; Yamaguchi, T.; Wakita, H. *J. Chem. Phys.* **1994**, *100*, 3122.

(20) Enderby, J. E.; Neilson, G. W. *Rep. Prog. Phys.* **1981**, *44*, 38.

(21) Kameda, Y.; Suzuki, S.; Ebata, H.; Usuki, T.; Uemura, O. *Bull. Chem. Soc. Jpn.* **1997**, *70*, 47.

(22) Ohtaki, H.; Radnai, T. *Chem. Rev.* **1993**, *93*, 1157.

(23) Ansell, S.; Barnes, A. C.; Mason, P. E.; Neilson, G. W.; Ramos, S. *Biophys. Chem.* **2006**, *124*, 171.

(24) Tongraar, A.; Liedl, K. R.; Rode, B. M. *Chem. Phys. Lett.* **1998**, *286*, 56.

(25) Loeffler, H. H.; Mohammed, A. M.; Inada, Y.; Funahashi, S. *Chem. Phys. Lett.* **2003**, *379*, 452.

(26) Asada, T.; Nishimoto, K. *Chem. Phys. Lett.* **1995**, *232*, 518.

(27) Harsànyi, I.; Pusztai, L. *J. Chem. Phys.* **2005**, *122*, 124512.

(28) Egorov, A. V.; Komolkin, A. V.; Chizhik, V. I.; Lyubartsev, A. P.; Laaksonen, A. *J. Phys. Chem. B* **2003**, *107*, 3234.

(29) Hermansson, K.; Wojcik, M. *J. Phys. Chem. B* **1998**, *102*, 6089.

(30) Du, H.; Rasaiah, J. C.; Miller, J. D. *J. Phys. Chem. B* **2007**, *111*, 209.

(31) Impey, R. W.; Madden, P. A.; McDonald, I. R. *J. Phys. Chem.* **1983**, *87*, 5071.

(32) White, J. A.; Schwegler, E.; Galli, G.; Gygi, F. *J. Chem. Phys.* **2000**, *113*, 4688.

(33) Vuilleumier, R.; Sprik, M. *J. Chem. Phys.* **2001**, *115*, 3454.

(34) Lyubartsev, A. P.; Laasonen, K.; Laaksonen, A. *J. Chem. Phys.* **2001**, *114*, 3120.

(35) (a) Laasonen, K.; Klein, M. L. *J. Phys. Chem. A* **1997**, *101*, 98. (b) Laasonen, K.; Klein, M. L. *J. Am. Chem. Soc.* **1994**, *116*, 11620.

(36) Heuft, J. M.; Meijer, E. J. *J. Chem. Phys.* **2003**, *119*, 11788.

(37) Sillanp, A. J.; Laasonen, K. *Phys. Chem. Chem. Phys.* **2004**, *6*, 555.

(38) Cavallari, M.; Cavazzoni, C.; Ferrario, M. *Mol. Phys.* **2004**, *102*, 959.

(39) Ramaniah, L. M.; Bernasconi, M.; Parrinello, M. *J. Chem. Phys.* **1999**, *111*, 1587.

(40) Krekeler, C.; Hess, B.; Delle Site, L. *J. Chem. Phys.* **2006**, *125*, 054305.

(41) Izvekov, S.; Philpott, M. R. *J. Chem. Phys.* **2000**, *113*, 10676.

(42) (a) Raugei, S.; Klein, M. L. *J. Am. Chem. Soc.* **2001**, *123*, 9484. (b) Raugei, S.; Klein, M. L. *J. Chem. Phys.* **2002**, *116*, 196.

(43) Pagliai, M.; Cardini, G.; Schettino, V. *J. Phys. Chem. B* **2005**, *109*, 7475.

(44) Kirchner, B.; Seitsonen, A. P.; Hutter, J. *J. Phys. Chem. B* **2006**, *110*, 11475.

(45) Allen, P. G.; Bucher, J. J.; Shuh, D. K.; Edelstein, N. M.; Craig, I. *Inorg. Chem.* **2000**, *39*, 595.

(46) Petit, L.; Borel, A.; Daul, C.; Maldivi, P.; Adamo, C. *Inorg. Chem.* **2006**, *45*, 7382.

(47) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.

(48) Loeffler, H. H.; Rode, B. M. *J. Chem. Phys.* **2002**, *117*, 110.

(49) *CPMD*, v.3.9; IBM Corp.: Armonk, NY, 1990−2004; MPI für Festkörperforschung: Stuttgart, Germany, 1997−2001.

(50) (a) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098. (b) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822.

(51) (a) Vetere, V.; Maldivi, P.; Adamo, C. *J. Comput. Chem.* **2003**, *24*, 850. (b) Vetere, V.; Maldivi, P.; Adamo, C. *Int. J. Quantum Chem.* **2003**, *91*, 321.

(52) Guillaumont, D. *J. Phys. Chem. A* **2004**, *108*, 6893.

(53) Sprik, M.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1996**, *105*, 1142.

(54) Troullier, N.; Martins, J. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 1993.

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

(56) *Gaussian 03*, revision D.01; 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, Inc.: Wallingford, CT, 2004.

(57) Refson, K. *Comput. Phys. Commun.* **2000**, *126*, 310.

(58) Dang, L. X. *J. Chem. Phys.* **1992**, *96*, 6970.

(59) Grossman, J. C.; Schwegler, E.; Draeger, E. W.; Gygi, F.; Galli, G. *J. Chem. Phys.* **2004**, *120*, 300.

(60) (a) Silvestrelli, P. L.; Marzari, N.; Vanderbilt, D.; Parrinello, M. *Solid State Commun.* **1998**, *107*, 7.

(61) (b) Marzari, N.; Vanderbilt, D. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *56*, 12847. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.

(62) Gilli, P.; Bertolasi, V.; Ferretti, V.; Gilli, G. *J. Am. Chem. Soc.* **1994**, *116*, 909.

(63) Scheiner, S. *Hydrogen Bonding, A Theoretical Perspective*; Oxford University Press: New York, 1997.

(64) Ansell, S.; Neilson, G. W. *J. Chem. Phys.* **2000**, *112*, 3942.

(65) Neilson, G. W.; Enderby, J. E. *Proc. R. Soc. London, Ser. A* **1983**, *390*, 353.

(66) Clough, S. A.; Beers, Y.; Klein, G. P.; Rothman, L.S. *J. Chem. Phys.* **1973**, *59*, 2254.

(67) Silvestrelli, P. L.; Parrinello, M. *J. Chem. Phys.* **1999**, *111*, 3572.

(68) (a) Silvestrelli, P. L.; Parrinello, M. *Phys. Rev. Lett.* **1999**, *82*, 3308. (b) Silvestrelli, P. L.; Parrinello, M. *Phys. Rev. Lett.* **1999**, *82*, 5415.

(69) Badyal, Y. S.; Saboungi, M.-L.; Price, D. L.; Shastri, S. D.; Haeffner, D. R.; Soper, A. K. *J. Chem. Phys.* **2000**, *112*, 9206.

(70) Batista, E. R.; Xantheas, S. S.; Jonsson, H. *J. Chem. Phys.* **1998**, *109*, 4546.

(71) Laasonen, K.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **1993**, *99*, 9080.

(72) Coudert, F.-X.; Vuilleumier, R.; Boutin, A. *Chem. Phys. Chem.* **2006**, *7*, 2464.

(73) Hribar, B.; Southall, N. T.; Vlachy, V.; Dill, K. A. *J. Am. Chem. Soc.* **2002**, *124*, 12302.

(74) Ichikawa, K.; Kameda, Y.; Matsumoto, T.; Misawa, M. *J. Phys. C: Solid State Phys.* **1984**, *17*, L725.

(75) Nygard, K.; Hakala, M.; Manninen, S.; Hämäläinen, K.; Itou, M.; Andrejczuk, A.; Sakurai, Y. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *73*, 024208.

(76) Soper, A. K.; Phillips, M. G. *Chem. Phys.* **1986**, *107*, 47.

(77) Thiessen, W. E.; Narten, A. H. *J. Chem. Phys.* **1982**, *77*, 2656.

(78) Soper, A. K. *J. Chem. Phys.* **1994**, *101*, 6888.

CT800007V

J. Chem. Theory Comput. **2008,** *4,* 1049–1056 **1049**

JCTC Journal of Chemical Theory and Computation

First Principles Study of Alkali−Tyrosine Complexes: Alkali Solvation and Redox Properties

Francesca Costanzo,*,† Marialore Sulpizi,‡ Raffaele Guido Della Valle,† and
Michiel Sprik‡

Dipartimento di Chimica Fisica e Inorganica and INSTM-UdR Bologna, Universita di Bologna, Viale Risorgimento 4, I-40136 Bologna, Italy, and, Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, U.K.

Received February 6, 2008

Abstract: The oxidation of alkali-tyrosinate to radical alkali−tyrosine in aqueous solution is studied using ab initio Car−Parrinello molecular dynamics (CPMD). The aim is to investigate the cation-π interactions between alkali cations M and the aromatic ring of tyrosine, in gas phase and in aqueous solution, using the influence of the cation M on the reaction $M^+(Tyr^-) \rightarrow M^+(Tyr^*) + e^-$ as a probe. To this end, we calculate the redox potential and the reorganization free energy using a CPMD-based method derived from the Marcus theory of electron transfer. We discuss the redox properties of Tyr, Na^+(Tyr), and K^+(Tyr), in reduced and oxidized states, by analyzing selected interatomic distances, coordination numbers, and charge populations. Our results confirm the known inversion in the relative stabilities of Na^+(Tyr) and K^+(Tyr) in going from gas phase to solution and point to a stronger cation-π affinity of K^+ in solution.

1. Introduction

Cation-π interactions[1] involving aromatic amino acids are now recognized as a type of noncovalent-binding forces important for structure and function of proteins. Cation-π interactions have been shown to stabilize protein geometries[1] and to facilitate molecular recognition[1,2] at protein−protein interfaces[3] and in biological receptors.[4–6] The π system is provided by the aromatic rings of Phe, Tyr, and Trp, while the cations come from the positively charged side chains of Lys, Arg, and His. Cation-π interactions between aromatic amino acids and alkali cations are of particular interest, because of their role in ion selectivity.[3] K^+ channels, for example, show a highly conserved Gly-Tyr-Gly sequence, which is thought to be essential for the K^+ selectivity.[7,8]

Resolving the physical chemical nature of cation-π interactions and characterizing their properties has become a very active field of research,[4,9] both experimental and theoretical. Theoretical work[10–16] has focused on small model systems, by analyzing in detail the electrostatic, charge transfer, polarization, and covalent contributions to the binding.[4,17–19] Experimental and theoretical (computational) studies agree that in vacuum alkali cations are strongly bound, with the binding affinity for benzene-like systems following the electrostatic rules ($Li^+ > Na^+ > K^+$). In water solution, a dramatic reordering occurs,[20] and the cation-π affinity follows the order $K^+ > Na^+ > Li^+$. This is the same sequence seen in voltage-gated K^+ channels.[21]

In this work we present Car−Parrinello molecular dynamics (CPMD) simulations of metal−tyrosine complexes in water. The aim is to clarify the effect of the solvent on the stability of the cation-π complex, by comparing the binding of Na^+ and K^+. The phenol-alkali ion motif is a most relevant and popular model system for cation-π interactions in proteins and has been the subject of several computational studies.[1,10,14,15] The focus of these and related computational work has been on the neutral protonated closed shell state of the aromatic group. However, two of the three aromatic amino acid π electron donors, namely Tyr and Trp, are also well-known redox active amino acid residues. In the oxidized deprotonated state, these residues can stabilize unpaired electrons in radical enzymes (see for example refs 22−24). In particular, the neutral tyrosine radical is essential to the catalytic activity of class I ribonucleotide reductase and

---

* Corresponding author phone: +39-051-2093710; e-mail: costanzo@ms.fci.unibo.it.

† Universita di Bologna.

‡ University of Cambridge.

10.1021/ct8000415 CCC: $40.75

similarly plays a central role in the generation of O_2 by plants in photosystem II.[25] Tryptophan cation radicals are involved, for example, in cytochrome oxidase. The redox chemistry of Tyr and Trp has been studied using relatively simple model complexes in solution[26–28] as well as de novo designed proteins.[29–31]

The link to radical enzyme catalysis prompted us to view the tyrosine-alkali ion pair from an electrochemical perspective and use the aqueous redox reaction as a probe of the interaction with Na^+ and K^+. The tyrosine residue in our model system is therefore either the deprotonated (oxidized) neutral radical (tyrosil), which will be indicated by Tyr*, or its anionic reduced form indicated by Tyr^-. Anticipating our results, we can confirm that a stronger affinity for the K^+ (Tyr) complex in aqueous solution correlates with an increase of the redox potential, due to a stronger stabilization of reduced K^+ (Tyr^-). While our study is limited to a single amino acid residue in solution, in an attempt to include some rudimentary aspects of the protein structure, we also discuss the role of the Tyr backbone in terms of both electrostatic and steric effects.

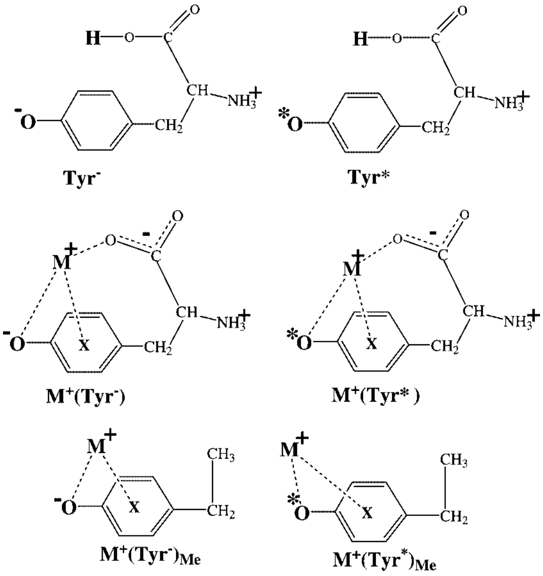


Figure 1. Structures of the reduced (left) and oxidized (right) states of tyrosine (top) and alkali cation-tyrosine complexes (middle). In the $M^+(Tyr^-)_{Me}$ structures (bottom), the zwitterionic backbone -CH(NH_3^+)-COO^- is replaced by methyl -CH_3.

2. Methods

2.1. Model Complexes. As explained in the Introduction, we will use the redox properties of aqueous alkali cation complexes to probe the cation affinity of tyrosine in aqueous solution. The following three redox half-reactions will be compared

$$Tyr^- \rightarrow Tyr^* + e^- \quad (1)$$

$$Na^+(Tyr^-) \rightarrow Na^+(Tyr^*) + e^- \quad (2)$$

$$K^+(Tyr^-) \rightarrow K^+(Tyr^*) + e^- \quad (3)$$

We have first investigated the structural properties of the various molecular species, in vacuum and in aqueous solution, and then computed the redox potentials of each reaction. The species involved, shown in Figure 1, are the reduced (anionic) or oxidized (radical) states of deprotonated tyrosine, i.e. the acid phenolic proton is missing in all our model systems. While formally a tyrosyl, these species will also be referred to as tyrosine in the discussion below and indicated by the symbol Tyr. The metal-tyrosine (tyrosyl) complexes, either in reduced $M^+(Tyr^-)$ state or in oxidized $M^+(Tyr^*)$ state, are in zwitterionic form, since water solvation stabilizes the charge separation in the terminal groups -NH_3^+ and -COO^-. The metal M is either Na or K. For plain tyrosine (tyrosine without metal) we have chosen the form in which the carboxyl group is protonated. The charge of reduced Tyr^- or oxidized Tyr* is thus the same (0 or +1, respectively) as for the corresponding metal−tyrosine complexes (this is also a technical requirement in our redox free energy calculation scheme, see section 2.4). In the complexes labeled $M^+(Tyr^-)_{Me}$ and $M^+(Tyr^*)_{Me}$ (bottom of Figure 1) we have replaced the zwitterionic backbone -CH(NH_3^+)-COO^- with a methyl group -CH_3. By comparison with the unsubstituted complexes, we obtain information on the importance of the backbone.

2.2. Electronic Structure Calculation. Calculations were performed with the CPMD package.[32] This is a general ab initio Car−Parrinello molecular dynamics code,[33] based on density functional theory. The BLYP functional[34,35] was used in all calculations. Only valence electrons were treated explicitly, and interactions with the ionic cores were described by norm-conserving pseudopotentials generated following the prescription by Troullier and Martins.[36] For Na^+ and K^+ the formal partition in valence and core is ambiguous because ionization eliminates the 3*s* respectively 4*s* valence electron. For this reason, the states 2*s* and 2*p* (for Na^+) and 3*s* and 3*p* (for K^+) in the next inner shell were treated as semicore states, effectively including them in the valence (for more details see for example the discussion in ref 37). The Kleinman-Bylander[38] decomposition was used for all the atomic species, with *s* and *p* nonlocality (angular momentum dependence) for C, N, O, and Na atoms. H atoms are represented by a simple *s* type pseudopotential. For consistency with the calculations in aqueous solution, described below in section 2.3, we chose a cubic cell with sides of 12.426 Å. Isolated system conditions were used in the gas-phase calculations. The Kohn−Sham orbitals were expanded in plane waves up to an energy cutoff of 70 Ry, which is sufficient to yield converged structural properties for water.[39] Only the Γ point was considered in the Brillouin zone integration. With regard to the accuracy of binding energy calculation in the gas phase, we further note that the use of a plane wave basis set, while computationally expensive, has the advantage that basis set superposition errors can in practice be neglected.

2.3. Molecular Dynamics Simulation. The systems in aqueous solution were equilibrated by preliminary classical MD simulations using the ORAC program.[40] These simulations were run at a constant number of particles, volume and temperature (NVT ensemble), using a Nose-Hoover thermostat at 300 K. A cubic cell with periodic boundary conditions and sides of 12.426 Å was adopted. The space not occupied by the M^+(Tyr) solute was filled with SPC water[41,42] with the experimental density at 300 K. The chosen box size, which requires 50 water molecules, allows space for at least one and a half layers of water around the M^+(Tyr) complex (whose largest dimension is ≈8.15 Å). Following the AMBER protocol,[43] the atomic charges for the solute were fitted to the ab initio electrostatic potential (ESP) evaluated with Gaussian03[44] at the B3LYP/6−31G* level.[34,35]

CPMD simulations were started using the classical equilibrated systems as the initial configuration. The time step was 5 au (about 0.12 fs). A fictitious electronic mass of 800 au was used to maintain the system on the Born−Oppenheimer surface. To reduce a possible bias of the initial classical configurations on the CPMD simulation, about 2 ps of equilibration with a Nose thermostat[45] at 300 K were followed by ≈8 ps of production run without temperature control (NVE ensemble).

2.4. Redox Properties Calculations. Redox properties for the reactions 1−3 were calculated using a MD method based on Marcus theory,[46] originally proposed by Warshel.[47] In the CPMD implementation applied here,[48–52] reaction free energies and reorganization energies of half-reactions are obtained from the vertical ionization energy gaps in the linear response approximation. Here we just recall the equations employed in the calculations. From the simulations of the systems in the oxidized states Na^+(Tyr*), K^+(Tyr*), or Tyr* and in the reduced states Na^+(Tyr^-), K^+(Tyr^-), or Tyr $^-$, the free energy of reduction (redox potential) ΔA and the reorganization free energy λ for the oxidation reactions 1−3 are computed from ensemble averages of the vertical ionization energy ΔE, as $\Delta A = (\langle \Delta E \rangle_O + \langle \Delta E \rangle_R)/2$ and $\lambda = (\langle \Delta E \rangle_R - \langle \Delta E \rangle_O)/2$. $\langle \Delta E \rangle_O$ denotes the average of the vertical ionization energies, as obtained from MD trajectories on the potential energy surface of the oxidized state O. $\langle \Delta E \rangle_R$ is the corresponding quantity on the surface of the reduced state R. The variance of the energy describes the fluctuations of the energy gap and is also of interest. There are again two such quantities, one for the reduced state $\sigma_R^2 = \langle (\Delta E - \langle \Delta E \rangle_R)^2 \rangle_R$ and one for the oxidized state $\sigma_O^2 = \langle (\Delta E - \langle \Delta E \rangle_O)^2 \rangle_O$. In the linear approximation employed here, the variance is independent of the oxidation state, $\sigma_R^2 = \sigma_O^2 = 2K_BT\lambda$. In practice this relation is used to verify whether the linear approximation is valid for a specific system.

The ΔA and λ estimates are computed directly from the vertical energies for removal of electrons from (or addition to) the periodic MD systems and are therefore subject to substantial finite size errors. These errors largely cancel when half-reaction energies are combined to yield reaction free energies of full redox reactions, provided that the charges of reactant and product species are the same (isoCoulombic reaction), that all calculations are carried out in MD cell of the same fixed length *L,* and that the redox active species are of similar size.[53] These conditions are satisfied for the reactions 1−3, and comparison of the computed half-reaction (oxidation) free energies is therefore meaningful. Moreover, as argued in ref 53 on the basis of a periodic continuum model and demonstrated for a classical model system, the finite size correction for ΔA scales with the inverse volume ($1/L^3$) of the MD cell rather than inverse length ($1/L$). This is a result of screening by the solvent with high dielectric constant (water). Experience with the (limited) set of reactions investigated so far indicates that full reaction free energies agree with experiment within a margin of about 0.2 eV (for a recent compilation see ref 54). Unfortunately, the full $1/L$ long-range effect is recovered for the reorganization energy λ which, as a result, is seriously underestimated (in the order of one eV) in the small samples used here.[53]

Table 1. Binding Energies (kcal/mol) of the Metal−Tyrosine Complexes in the Gas Phase and Distances *r* (Å) from the Cation M to X_π, O_{Ph}, and O_{carb} [a]

system	binding energy	$r_{M\text{-}X_\pi}$	$r_{M\text{-}O_{Ph}}$	$r_{M\text{-}O_{carb}}$
$Na^+(Tyr^-)$	−152.22	3.07	2.29	2.18
Na^+(Tyr*)	−62.83	3.45	2.51	2.22
$K^+(Tyr^-)$	−133.09	3.30	2.68	2.58
K^+(Tyr*)	−49.32	3.90	2.80	2.58
$Na^+(Tyr^-)_{Me}$	−129.70	3.65	2.13	
$Na^+(Tyr^*)_{Me}$	−40.29	4.76	2.14	
$K^+(Tyr^-)_{Me}$	−112.00	4.01	2.39	
$K^+(Tyr^*)_{Me}$	−27.12	4.86	2.63	

[a] Center of phenolic ring and phenolic and carboxylic oxygens.

3. Results and Discussion

3.1. M^+(Tyr) Complexes in Gas Phase. Since gas-phase plain tyrosine is already well studied,[10,14] and, moreover, it is not stable in zwitterionic form, we have studied only the metal−tyrosine complexes (see Figure 1). The vertical binding energy of the various complexes and the distances from the cation M to X_π, O_{Ph}, and O_{carb} (center of phenolic ring, phenolic, and carboxylic oxygens, respectively) are shown in Table 1. The binding energies are calculated for the optimized geometry of the complex, as the optimized geometry of the complex, as $E_{M^+(Tyr)} - E_{M^+} - E_{Tyr}$. All E_{Tyr} energies are calculated with all atoms kept in the geometry of the complex and thus do not include the deformation energy required to bring the uncomplexed neutral molecule into its complexed geometry. The reduced complexes $Na^+(Tyr^-)$ and $K^+(Tyr^-)$ are significantly lower in energy than their oxidized counterparts, due to the increased Coulomb attraction between the negatively charged O_{Ph} and the positively charged metals. The $r_{M\text{-}O_{Ph}}$ and $r_{M\text{-}X_\pi}$ distances are also smaller in the reduced state. Comparing now the results for the two alkali cations, we see that the binding energy for $Na^+(Tyr^-)$ is about 20 kcal/mol stronger than for $K^+(Tyr^-)$. In the oxidized forms the binding energy difference between Na^+ and K^+ decreases to 14 kcal/mol, with the Na^+ complex remaining the most stable. This decreased stabilization might be related to a bending of about 20° observed for the phenyl ring in Na^+(Tyr*). For the M^+(Tyr) complexes containing neutral (protonated) tyrosine the published results[10,14] point in the same direction. There are no M-O_{Ph} interactions in these systems, and the difference between Na^+(Tyr) and K^+(Tyr) further decreases to 13 kcal/mol.

To quantify the contribution of the backbone to the binding with the metal, we have considered the $M^+(Tyr)_{Me}$ complex (Figure 1) in which the zwitterionic backbone -CH(NH_3^+)-COO^- is replaced by a methyl group -CH_3. The overall effect of removing the backbone is that the metal moves slightly away from the π-ring and comes closer to O_{Ph} (Table 1),

Table 2. Simulation Averages for $Na^+(Tyr^-)$, $Na^+(Tyr^*)$, $K^+(Tyr^-)$, $K^+(Tyr^*)$, Tyr^-, and Tyr^* in Aqueous Solution[a]

system	$r_{M\text{-}X_\pi}$	$r_{M\text{-}O_{Ph}}$	$r_{M\text{-}O_{carb}}$	$n_{M\text{-}Ow}$	$n_{O_{Ph}\text{-}Hw}$	$Q_{Mulliken}$	$Q_{Löwdin}$
$Na^+(Tyr^-)$	4.4 ± 0.4	5.6 ± 0.4	2.7 ± 0.6	3.95	3.14	−0.30 ± 0.02	−0.24 ± 0.02
$Na^+(Tyr^*)$	4.7 ± 1.0	6.1 ± 1.1	2.6 ± 0.5	3.90	2.60	−0.23 ± 0.01	−0.16 ± 0.01
$K^+(Tyr^-)$	3.6 ± 0.4	4.1 ± 0.4	3.2 ± 0.5	3.87	3.40	−0.31 ± 0.01	−0.25 ± 0.02
$K^+(Tyr^*)$	4.7 ± 0.6	5.5 ± 1.0	3.1 ± 0.6	4.19	2.24	−0.20 ± 0.02	−0.14 ± 0.02
Tyr^-	3.8 ± 0.2	5.2 ± 0.3		1.00	3.16	−0.30 ± 0.02	−0.25 ± 0.03
Tyr^*	3.9 ± 0.3	5.3 ± 0.4		1.00	2.46	−0.22 ± 0.01	−0.15 ± 0.01

[a] The table reports the distances r (Å) from the cation M to X_π, O_{Ph}, and O_{carb} (center of phenolic ring and phenolic and carboxylic oxygens), the coordination numbers $n_{M\text{-}Ow}$ and $n_{OPh\text{-}Hw}$ (where Ow or Hw indicate water oxygens or hydrogens), and the Mulliken and Löwdin charges Q for O_{Ph}. In the absence of alkali metal, the carboxylic hydrogen H_{carb} takes the role of the cation M. Standard deviations are given where appropriate.

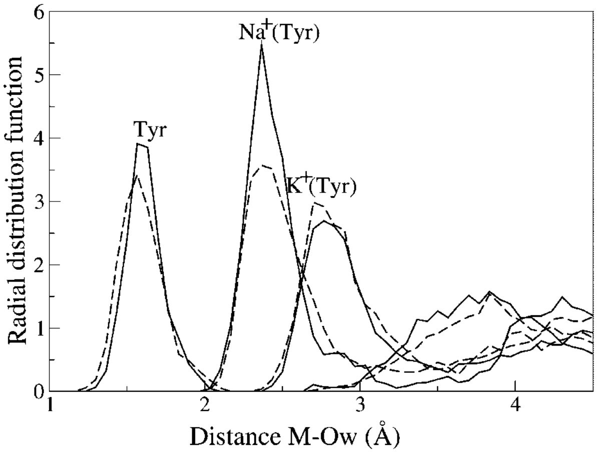


Figure 2. Radial distribution function $g_{M\text{-}Ow(r)}$ between the cation M and the water oxygens Ow for Na^+(Tyr), K^+(Tyr), and Tyr, as indicated. For Tyr, M corresponds to H_{carb}. Solid and dashed lines represent reduced and oxidized systems, respectively.

thus optimizing the electrostatics in the absence of M-O_{carb} interactions. The zwitterionic backbone is responsible for an energy stabilization of ≈22 kcal/mol for all structures, leaving unchanged the stability order $Na^+(Tyr^-) > K^+(Tyr^-) > Na^+(Tyr^*) > K^+(Tyr^*)$. In view of the importance of the -CH(NH_3^+)-COO^- backbone in the gas phase, we have included it in all the following simulations in aqueous solution. Presence of the backbone is a possibly significant difference with the only available CPMD simulation of a cation-π interaction in water, which is for a solvated ammonium−benzene complex.[16]

3.2. Cation Coordination in Aqueous Solution. All complexes shown in Figure 1 (namely $Na^+(Tyr^-)$, $Na^+(Tyr^*)$, $K^+(Tyr^-)$, $K^+(Tyr^*)$, Tyr^-, and Tyr^*) were investigated in aqueous solution by carrying out independent MD runs (for setting see section 2.3). The structural analysis will focus on the interactions between the cation M and its environment and on the effect of the redox induced change of charge of the phenolic oxygen O_{Ph}. In Table 2 we report a selection of relevant CPMD averages, which includes interatomic distances r and coordination numbers n around M or O_{Ph} and estimates of the charge Q on O_{Ph}. In Figure 2 we display the radial distribution function $g_{M\text{-}Ow(r)}$ between the cation M and the water oxygens Ow. For plain tyrosine, in the absence of metal, the carboxylic hydrogen H_{carb} takes the role of the cation M.

The $Na^+(Tyr^-)$ and $Na^+(Tyr^*)$ complexes have similar $g_{M\text{-}Ow(r)}$ peak shapes (Figure 2), coordination numbers $n_{Na\text{-}Ow}$, and distances $r_{Na\text{-}X_\pi}$ and $r_{Na\text{-}O_{Ph}}$ (Table 2). The oxidation state of tyrosine evidently has only very little influence on the solvation of Na^+. The effect is more pronounced for the K^+(Tyr) complex. Even though the $g_{M\text{-}Ow(r)}$ peak shapes (Figure 2) appear quite similar in the reduced and the oxidized states, the contraction of $r_{K\text{-}X_\pi}$ and $r_{K\text{-}O_{Ph}}$ upon reduction is larger for the K^+ complex (≈1 Å for K^+ compared to ≈0.5 Å for Na^+, see Table 2). Moreover, while the difference in coordination number $n_{K\text{-}Ow}$ for $K^+(Tyr^-)$ relative to $K^+(Tyr^*)$ is not very large (≈0.3), this decrease is significant in comparison to the Na^+ complex, for which coordination remains essentially unchanged in the redox reaction. We should mention here that the coordination numbers $n_{M\text{-}Ow}$ have been obtained by integrating $g_{M\text{-}Ow(r)}$ up to the first minimum, identified at $r_0 = 3.4$ and 3.5 Å for Na and K, respectively. The dependence of $n_{M\text{-}Ow}$ on the chosen cutoff r_0 is not a cause of concern, since the differences among the various $n_{M\text{-}Ow}$ are, within reasonable ranges, largely independent of r_0.

We interpret these results on coordination as evidence that reduction enhances the electrostatic interaction of K^+ with the π aromatic electrons and the O_{Ph} oxygen, whereas the effect for Na^+ is relatively unimportant. This suggests that the affinity of K^+ toward the π system is stronger compared to Na^+. This behavior is opposite to that for the affinity in the gas phase, where both distances and binding energies indicate that the complexes with Na^+ are more stable than the complexes with K^+. Our finding is in agreement with the reordering in solution predicted by Kumpf and Dougherty[21] and indicates that this reordering is still present in the deprotonated tyrosine and is not affected by its oxidation state.

Qualitative differences between aqueous Na^+ and K^+ are well-known. Aqueous solutions of these cations, even when of equal ionic strength, can have strikingly different effects on biopolymers, notably on folding and aggregation of proteins (Hofmeister effect, for a recent review see ref 55). These issues have attracted renewed attention in the recent literature. The explanation is generally sought in short-range effects such as the stability of the first hydration shell. A crucial feature distinguishing the two ions, according to this line of thought, is that Na^+ tends to hold onto its first solvation shell in inhomogeneous environments, while it is easier for K^+ to give up some of its coordinated water molecules in exchange for direct interactions with other

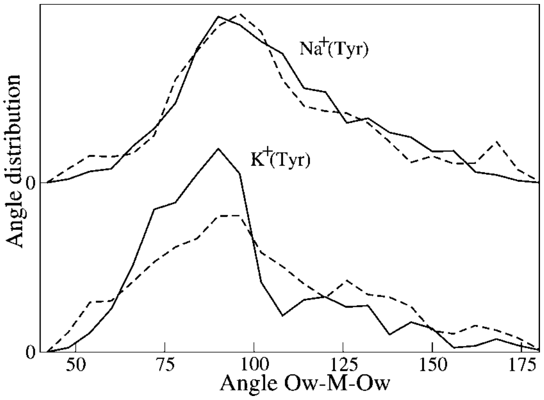


Figure 3. Ow-M-Ow bond angle distributions for Na^+(Tyr) and K^+(Tyr), as indicated. Solid and dashed lines represent reduced and oxidized systems, respectively.

species. Such a mechanism may also play a role in our system. This conjecture is supported by a comparison of the water coordination for our alkali ion-tyrosine complexes to the coordination in bulk solution. CPMD calculations on alkali cations[37,56–58] give for these ions fully solvated in water coordination numbers $n_{\text{Na-Ow}} = 5.2$ and $n_{\text{K-Ow}} = 7.1$. By comparison with our results for $M^+(Tyr^-)$ and M^+(Tyr*), we find that the cation-π interaction displaces about 1.3 water molecules in the case of Na^+ and 3.0 molecules for K^+. Further confirmation comes from the anisotropy of the hydration of the cations in the complex with tyrosine. We determined the angular distributions of the water molecules around Na^+ and K^+ by averaging over our MD trajectories. Figure 3 shows the distribution of the Ow-M-Ow angles formed by the metal and two water oxygens. For Na^+(Tyr) the angular distributions in the reduced and oxidized states have similar shape and angle range. For K^+(Tyr) the reduced form exhibits a sharper peak in the range 75−100°. This finding correlates with the shorter $r_{\text{M-X}_\pi}$. Since the K^+ is closer than Na^+ to the π ring, it has less space to coordinate water molecules.

Finally, we investigate plain tyrosine for a possible OH-π interaction involving the carboxyl hydrogen H_{carb}. This investigation is motivated by the examples, found in proteins, of configurations in which the OH side chain of an amino acid is arranged face-on to the aromatic face of a tyrosine.[59] For plain tyrosine, indeed, Table 2 indicates relatively short distances $r_{\text{M-X}_\pi}$ (where M correspond to H_{carb}). This may confirm a OH-π interaction between the H_{carb} proton and the ring. For both Tyr^- and Tyr*, we have noticed that H_{carb} is interacting with one water molecule only ($n_{\text{M-Ow}} = 1$) over the entire simulation.

3.3. Solvation of the Phenoxyl Oxygen. We now turn to the solvation of O_{Ph}, which is directly involved in the redox process. We find (Table 2) that the coordination numbers $n_{\text{O}_{\text{Ph}}\text{-Hw}}$ for the various complexes strongly correlate with the electrostatic charges on O_{Ph} (linear correlation coefficient $r \geq 0.98$, for both Mulliken and Löwdin). For both metals, the coordination number $n_{\text{O}_{\text{Ph}}\text{-Hw}}$ increases in going from the oxidized to the reduced complexes, due to the enhanced electrostatic interactions of the water hydrogens with the larger charge residing on the O_{Ph}. The K^+ complexes show the largest change in coordination numbers $n_{\text{O}_{\text{Ph}}\text{-Hw}}$ (1.16 versus 0.54 for the Na^+ complex). The changes in the charge of O_{Ph} are less pronounced (−0.11 for K^+ compared to −0.07 for Na^+, according to the data of Table 2). We note that for the K^+ complex the response of $n_{\text{O}_{\text{Ph}}\text{-Hw}}$ to reduction is opposite to that of $n_{\text{K-Ow}}$, discussed in section 3.2. While $n_{\text{K-Ow}}$ decreases, $n_{\text{O}_{\text{Ph}}\text{-Hw}}$ increases. There is no such a clear (anti) correlation for the Na^+ complex, consistently with our picture of a stronger reduction induced ion pairing for K^+ compared to Na^+.

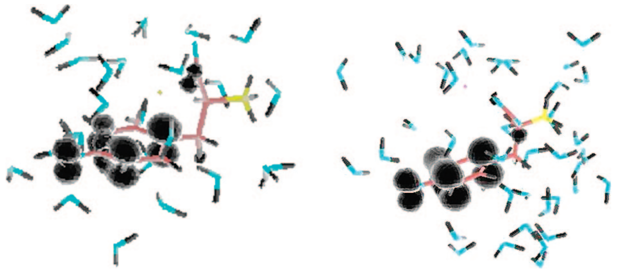


Figure 4. Unpaired spin density for Na^+(Tyr*) (left) and K^+(Tyr*) (right). The isosurface represents the spin density at 0.005 e au^{-3}.

Table 3. Averages $\langle \Delta E \rangle_R$ and $\langle \Delta E \rangle_O$ of the Vertical Transition Energies (eV) for the Reduced and Oxidized States, Respectively[a]

system	$\langle \Delta E \rangle_R$	$\langle \Delta E \rangle_O$	ΔA	λ
Na^+(Tyr)	1.25 ± 0.20	−0.11 ± 0.24	0.57 ± 0.16	0.68 ± 0.16
K^+(Tyr)	1.36 ± 0.19	−0.05 ± 0.19	0.66 ± 0.13	0.70 ± 0.13
Tyr	1.32 ± 0.20	0.00 ± 0.22	0.66 ± 0.15	0.66 ± 0.15

[a] Helmholtz free energy of reduction ΔA, reorganization energy λ, and all standard deviations σ are also reported.

The unpaired spin density for the oxidized systems K^+(Tyr*) and Na^+(Tyr*) is shown in Figure 4. The spin density is mostly localized on the O_{Ph}, C_{ortho}, and C_{para} positions around the phenoxyl ring, in agreement with previous calculations.[60] As expected for Na^+ and K^+ bound to the aromatic ring, even in aqueous solution, the unpaired spin density on the metal is negligible, supporting the known picture where electrostatics plays a predictive role in cation-π interactions.[19]

3.4. Redox Free Energies. The vertical ionization energies ΔE for the various reduced and oxidized systems in aqueous solution, calculated as described in section 2.4, are shown in Table 3. The table also reports the resulting free energies of reduction ΔA, the reorganization free energies λ, and all standard deviations σ. The agreement between the variances of the vertical ionization energies in the oxidized and the reduced state is sufficient to justify the use of the linear response approximation for the estimation of ΔA and λ (see section 2.4). The free energy change ΔA, in the definition used here, is the Helmholtz free energy of reduction. It is therefore tempting to interpret the positive sign of ΔA as evidence that the reduced forms of the cation−tyrosine complex are more stable than the oxidized forms. This would be consistent with the behavior of the distances, coordination numbers, and charges, discussed in the previous sections. Indeed, the stabilizing effect of the cation-π interaction on the reduced form (increasing the redox potential ΔA) has already been noticed by NMR experiments on tryptophan.[29–31]

Unfortunately, as repeatedly pointed out in earlier works,[48–51,53,54] using a similar half-reaction scheme, the utmost caution is required when interpreting half-reaction energies. Finite size effects in our small periodic systems can seriously bias energies if the net charge of the MD cell is changed. In ref 53 it was shown that the interactions with periodic images of the redox active species and the neutralizing homogeneous background charge artificially stabilize the species with the higher charge. This effect can even invert the sign the oxidation free energy ΔA. However, the reduced state, which is the state favored by our calculated half-reaction free energies for the cation−tyrosine complexes, has net charge zero. It is the oxidized state, the state with the higher free energy, which is charged. This observation, specific to our system, gives us some confidence that the sign of ΔA can be, after all, regarded as an indication that the reduced species with the anionic phenoxyl ring is the more stable, although the actual numerical value is not reliable for the reasons explained.

Of special interest in the present context are the relative values of ΔA in Table 3, which are far less sensitive to finite size effects. The differences in ΔA are in the 50−100 meV range. These values are comparable to the standard deviations of ΔA, indicated in Table 3, which are all around 150 meV. The statistical errors on the averages, which are computed over $N = 130$ configurations sampled from a time window of at least 5 ps, can therefore be assumed to be substantially smaller. A reasonable estimate of the uncertainty, obtained by taking into account[61] the statistical correlations between consecutive configurations, is 20 meV. The relative difference in ΔA of 90 meV between K^+ and Na^+, although small, therefore appears to be statistically significant. The difference is in the expected direction and confirms that K^+ interacts more strongly with the π aromatic electrons of tyrosine consistent with the analysis of the complex geometries of section 3.2. Finally, K^+(Tyr) and Tyr are found to have similar redox potentials ΔA (Table 3), suggesting that comparable strong cation-π interactions also occur for plain tyrosine, with the carboxylic hydrogen H_{carb} playing a role similar to K^+, as also indicated by the short $r_{M\text{-}X_\pi}$ distances (Table 3).

The reorganization energy λ, which is not experimentally accessible, describes the response of the solvent to the change of the charge distribution on the complex as a result of the oxidation process. The calculated reorganization energies λ are similar for all the species studied. A likely explanation is that the relaxation is dominated by an outer-sphere process, which is approximately the same for these closely related systems.

4. Conclusions

Using Car−Parrinello molecular dynamics (CPMD), we have investigated the cation-π interactions between Na^+ or K^+ alkali cations and the aromatic ring of tyrosine, in gas phase and in aqueous solution. The binding energy difference in gas phase and the free energy of reduction in solution (calculated within the Marcus theory of electron transfer) indicate that the reduced forms are, in all cases, more stable than the oxidized forms. This greater stability correlates with shorter distances $r_{M\text{-}X_\pi}$ and (in solution) with larger coordination numbers $n_{O_{Ph}\text{-}Hw}$ (where M, X_π, O_{Ph}, and Hw indicate alkali metal, center of phenolic ring, phenolic oxygen, and water hydrogens, respectively). Thus, we deduce that the cation-π interaction stabilizes the reduced forms of both metal−tyrosine complexes. A further observation is that in the absence of a metal ion, the carboxylic hydrogen H_{carb} can take over the role of the cation M.

The influence of the coordination by alkali cations (either Na^+ or K^+) on the reaction $M^+(Tyr^-) \rightarrow M^+(Tyr^*) + e^-$ is rather intriguing. In gas phase, in agreement with the literature[10,14] our calculations show the Na^+(Tyr) complexes to be more stable in both reduced and oxidized states, independently of any additional stabilization due to the zwitterionic backbone. A reordering of the cation affinities occurs in solution, again in agreement with literature.[11,21,62] Analysis of simulation observables shows the K^+ cation is partly desolvated in the reduced state, while Na^+ largely keeps its first solvation shell. The shorter distance $r_{K\text{-}X_\pi}$, the smaller coordination $n_{M\text{-}Ow}$, the enhanced hydrogen bonding of O_{Ph}, and the sharper Ow-M-Ow angular distribution, all show the K^+ to be more involved in the interaction with the π aromatic electrons.

The structural analysis has been backed up by a free energy calculation. Failing a full and, at the ab initio MD level, exceedingly expensive determination of the dissociation constant from a potential of mean force for the attraction between alkali atoms and the tyrosine ring, we have used the redox properties of the equilibrium complex as a probe of the energetics. We have found a difference of 90 meV between the redox potentials of K^+ and Na^+, with the reduced state of the K^+ forming the most stable complex (Table 3). This free energy difference is just above the lower margin of what can be considered statistically significant in our CPMD simulation scheme. Retrospectively, we could have expected such a difficulty, since energy increments important on the “bioscale” are typically in the order of one pK unit (60 meV). The experimental difference between the redox potentials of tyrosine and tryptophan, for example, is 110 meV.[63] This is at the limit of the current CPMD methods, with yield accuracies around 100 meV.[54]

The overall picture emerging from our calculations is that in solution K^+ has stronger cation-π interactions than Na^+. This is supported by our CPMD results for the distance and coordination observables and is consistent with our results for the redox properties. Our calculations therefore confirm experimental[62] and other theoretical[21] studies arriving at a similar conclusion. This result is important in the biological field, where the difference between the two alkali metals is fundamental in physiological processes related to the K channels, such as electrical signaling in the nervous system, regulation of cardiac excitability and regulation of insuline release.[8]

Acknowledgment. Work done with funds from MIUR (PRIN project and FIRB project through INSTM consortium). We warmly thank M. Boero for providing the pseudopotentials, P. Procacci for support with the ORAC program, Prof. Brillante for useful discussions, and the EPCC Supercomputer Center for an allocation of computer resources within the HPC-Europa program.

References

(1) Dougherty, D. A. Cation-π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr, and Trp. *Science* **1996**, *271*, 163–168.

(2) Dougherty, D. A.; Stauffer, D. A. Acetylcholine Binding by a Synthetic Receptor. Implications for Biological Recognition. *Science* **1990**, *250*, 1558–1560.

(3) Cabarcos, O. M.; Weinheimer, C. J.; Lisy, J. M. Size selectivity by cation-π interactions: Solvation of K^+ and Na^+ by benzene and water. *J. Chem. Phys.* **1999**, *110*, 8429–8435.

(4) Ma, J. C.; Dougherty, D. A. The Cation-π Interaction. *Chem. Rev.* **1997**, *97*, 1303–1324.

(5) Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van der Oost, J.; Smit, A. B.; Sixma, T. K. The crystal structure of AChBP, homolog of the N-terminal domain of the nicotinic acetylcholine receptor. *Nature* **2001**, *411*, 269–276.

(6) Gromiha, M. M.; Santhosh, C.; Ahmad, S. Structural analysis of cation-π interactions in DNA binding proteins. *Int. J. Biol. Macromol.* **2004**, *34*, 203–211.

(7) Heginbotham, L.; MacKinnon, R. The aromatic binding site for tetraethylammonium ion on potassium channels. *Neuron* **1992**, *8*, 483–491.

(8) Domene, C.; Sansom, M. S. P. Potassium channel, ions, and water: simulation studies based on the high resolution X-ray structure of KcsA. *Biophys. J.* **2003**, *85*, 2787–2800.

(9) Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angew. Chem. Int. Ed.* **2003**, *42*, 1210–1250.

(10) Dunbar, R. C. Complexation of Na^+ and K^+ to Aromatic Amino Acids: A Density Functional Computational Study of Cation-π Interactions. *J. Phys. Chem. A* **2000**, *104*, 8067–8074.

(11) Ryzhov, V.; Dunbar, R. C.; Cerda, B.; Wesdemiotis, C. Cation-π Effects in the Complexation of Na^+ and K^+ with Phe, Tyr, and Trp in the Gas Phase. *J. Am. Soc. Mass. Spectrom.* **2000**, *11*, 1037–1046.

(12) Siu, F. M.; Ma, N. L.; Tsang, C. W. Cation-π Interactions in Sodiated Phenylalanine Complexes: Is Phenylalanine in the Charge-Solvated or Zwitterionic Form? *J. Am. Chem. Soc.* **2001**, *123*, 3397–3398.

(13) Gapeev, A.; Dunbar, R. C. Na^+ affinities of gas-phase amino acids by ligand exchange equilibrium. *Int. J. Mass. Spectrom.* **2003**, *228*, 825–839.

(14) Ruan, C.; Rodgers, M. T. Cation-π Interactions: Structure and Energetics of Complexation of Na^+ and K^+ with the Aromatic Amino Acids, Phenylalanine, Tyrosine, and Tryptophan. *J. Am. Chem. Soc.* **2004**, *126*, 14600–14610.

(15) Reddy, A. S.; Sastry, G. N. Cation [$M = H^+, Li^+, Na^+, K^+, Ca^{2+}, Mg^{2+}, NH_4^+$, and NMe_4^+] Interactions with the Aromatic Motifs of Naturally Occurring Amino Acids: A Theoretical Study. *J. Phys. Chem. A* **2005**, *109*, 8893–8903.

(16) Sa, R.; Zhu, W.; Shen, J.; Gong, Z.; Cheng, J.; Chen, K.; Jiang, H. How Does Ammonium Dynamically Interact with Benzene in Aqueous Media? A First Principle Study Using the Car-Parrinello Molecular Dynamics Method. *J. Phys. Chem. B* **2006**, *110*, 5094–5098.

(17) Costanzo, F.; Della Valle, R. G.; Barone, V. MD simulation of the Na^+-phenylalanine complex in water: competition between cation-π interaction and aqueous solvation. *J. Phys. Chem. B* **2005**, *109*, 23016–23023.

(18) Cubero, E.; Luque, F. J.; Orozco, M. Is polarization important in cation-π interactions? *Proc. Natl. Acad. Sci.* **1998**, *95*, 5976–5980.

(19) Mecozzi, S.; West, A. P.; Dougherty, D. A. Cation-π Interactions in Simple Aromatics: Electrostatics Provide a Predictive Tool. *J. Am. Chem. Soc.* **1996**, *118*, 2307–2308.

(20) Hu, J.; Barbour, L. J.; Gokel, G. W. Probing alkali metal-π interactions with the side chain residue of tryptophan. *Proc. Natl. Acad. Sci*. **2002,** *99*, 5121−5126.

(21) Kumpf, R. A.; Dougherty, D. A. A Mechanism for Ion Selectivity in Potassium Channels: Computational Studies of Cation-π Interactions. *Science* **1993**, *261*, 1708–1710.

(22) Stubbe, J.; van der Donk, W. A. Protein Radicals in Enzyme Catalysis. *Chem. Rev.* **1998**, *98*, 705–762.

(23) Stubbe, J. Radicals with a controlled lifestyle. *Chem. Commun.* **2003**, *20*, 2511–2513.

(24) Himo, F.; Siegbahn, P. E. M. Quantum Chemical Studies of Radical-Containing Enzymes. *Chem. Rev.* **2003**, *103*, 2421–2456.

(25) Ishikita, H.; Knapp, E.-W. Function of Redox-Active Tyrosine in Photosystem II. *Biophys. J.* **2006**, *90*, 3886–3896.

(26) Sjödin, M.; Styring, S.; Åkermark, B.; Sun, L. Hammarström, L. Proton-coupled electron transfer from tyrosine in a tyrosine-ruthenium-tris-bipyridine complex: Comparison with $Tyrosine_z$ oxidation in photosystem II. *J. Am. Chem. Soc.* **2000**, *122*, 3932–3936.

(27) Sjödin, M.; Styring, S.; Wolpher, H.; Xu, Y.; Sun, L.; Hammarström, L. Switching the Redox Mechanism: Models for Proton-Coupled Electron Transfer from Tyrosine and Tryptophan. *J. Am. Chem. Soc.* **2005**, *127*, 3855–3863.

(28) Carra, C.; Iordanova, N.; Hammegs-Schiffer, S. Proton-Coupled Electron Transfer in a Model for Tyrosine Oxidation in Photosystem II. *J. Am. Chem. Soc.* **2003**, *125*, 10429–10436.

(29) Tommos, C.; Skalicky, J. J.; Pilloud, D. L.; Wand, A. J.; Dutton, P. L. De novo proteins as models of radical enzymes. *Biochemistry* **1999**, *38*, 9495–9507.

(30) Dai, Q.-H.; Tommos, C.; Fuentes, E. J.; Blomberg, M. R. A.; Dutton, P. L.; Wand, A. J. Structure of a de Novo Designed Protein Model of Radical Enzymes. *J. Am. Chem. Soc.* **2002**, *124*, 10952–10953.

(31) Westerlund, K.; Berry, B. W.; Privett, H. K.; Tommos, C. Exploring amino-acid radical chemistry: protein engineering and de novo design. *Biochim. Biophys. Acta* **2005**, *1707*, 103–116.

(32) CPMD version 3.11, Copyright IBM Corp 1990−2006, Copyright MPI für Festkrperforschung Stuttgart 1997−2001. http://www.cpmd.org/ (accessed March 7, 2008).

(33) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.

(34) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(35) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(36) Troullier, N.; Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **1991**, *43*, 1993–2006.

(37) Vuilleumier, R.; Sprik, M. Electronic properties of hard and soft ions in solution: Aqueous Na^+ and Ag^+ compared. *J. Chem. Phys.* **2001**, *115*, 3454–3468.

(38) Kleinman, L.; Bylander, D. M. Efficacious Form for Model Pseudopotentials. *Phys. Rev. Lett.* **1982**, *48*, 1425–1428.

(39) VandeVondele, J.; Mohamed, F.; Krack, M.; Hutter, J.; Sprik, M.; Parrinello, M. The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water. *J. Chem. Phys.* **2005**, *122*, 014515/1–6.

(40) Procacci, P.; Darden, T. A.; Paci, E.; Marchi, M. ORAC: A Molecular Dynamics Program to Simulate Complex Molecular Systems with Realistic Electrostatic Interactions. *J. Comput. Chem.* **1997**, *18*, 1848–1862. http://www.chim.unifi.it/orac/ (accessed March 7, 2008)

(41) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potential. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(42) Straatsma, T. P.; Berendsen, H. J. C. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J. Chem. Phys.* **1988**, *89*, 5876–5886.

(43) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(44) 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.; X. Li, Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P; Dannenberg, J. J.; Zakrzewski, V. G., Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O., Malick, D. K.; Rabuck, A. D., Raghavachari, K.; Foresman, J. B.; Ortiz, J. V., Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B., Chen, W., Wong, M. W.; Gonzalez., C.; Pople, J. A. *Gaussian 03, Revision B.05*; Gaussian, Inc.: Pittsburgh, PA, 2003.

(45) Nose, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

(46) Marcus, R. A.; Sutin, N. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta* **1985**, *811*, 265–322.

(47) Warshel, A. Dynamics of Reactions in Polar Solvents. Semiclassical Trajectory Studies of Electron Transfer and Proton Transfer Reactions. *J. Phys. Chem.* **1982**, *86*, 2218–2224.

(48) Tateyama, Y.; Blumberger, J.; Sprik, M.; Tavernelli, I. Density-functional molecular-dynamics study of the redox reactions of two anionic, aqueous transition-metal complexes. *J. Chem. Phys.* **2005**, *122*, 234505/1–17.

(49) Blumberger, J.; Sprik, M. Quantum versus classical electron transfer energy as reaction coordinate for the aqueous Ru^{+2} /Ru^{+3} redox reaction. *Theor. Chem. Acc.* **2006**, *115*, 113–126.

(50) Blumberger, J.; Tavernelli, I.; Klein, M.; Sprik, M. Diabatic free energy curves and coordination fluctuations for the aqueous Ag^+ /Ag^{+2} redox couple: A biased Born-Oppenheimer molecular dynamics investigation. *J. Chem. Phys.* **2006**, *124*, 064507/1–12.

(51) VandeVondele, J.; Sulpizi, M.; Sprik, M. From Solvent Fluctuations to Quantitative Redox Properties of Quinones in Methanol and Acetonitrile. *Angew. Chem., Int. Ed.* **2006**, *45*, 1936–1938.

(52) Sulpizi, M.; Raugei, S.; VandeVondele, J.; Carloni, P.; Sprik, M. Calculation of Redox Properties: Understanding Short- and Long-Range Effects in Rubredoxin. *J. Phys. Chem. B* **2007**, *111*, 3969–3976.

(53) Ayala, R.; Sprik, M. A Classical Point Charge Model Study of System Size Dependence of Oxidation and Reorganization Free Energies in Aqueous Solution. *J. Phys. Chem. B* **2008**, *112*, 257–269.

(54) VandeVondele, J.; Ayala, R.; Sulpizi, M.; Sprik, M. Redox free energies and one-electron energy levels in density functional theory based ab initio molecular dynamics. *J. Electroanal. Chem.* **2007**, *607*, 113–120.

(55) Collins, K. D.; Neilson, G. W.; Enderby, J. E. Ions in water: Characterizing the forces that control chemical processes and biological structure. *Biophys. Chem.* **2007**, *128*, 95–104.

(56) Ramaniah, L. M.; Bernasconi, M.; Parrinello, M. Ab initio molecular-dynamics simulation of K^+ solvation in water. *J. Chem. Phys.* **1999**, *111*, 1587–1591.

(57) White, J. A.; Schwegler, E.; Galli, G.; Gygi, F. The solvation of Na^+ in water: First-principles simulations. *J. Chem. Phys.* **2000**, *113*, 4668–4673.

(58) Ikeda, T.; Boero, M.; Terakura, K. Hydration of alkali ions from first principles molecular dynamics revisited. *J. Chem. Phys.* **2007**, *126*, 034501/1–9.

(59) Sulpizi, M.; Carloni, P. Cation-π versus OH- π Interactions in Proteins: A Density Functional Study. *J. Phys. Chem. B* **2000**, *104*, 10087–10091.

(60) Wu, P.; O'Malley, P. J. Environmental effects on phenoxyl free radical spin densities and hyperfine couplings. *J. Mol. Struct. THEOCHEM* **2005**, *730*, 251–254.

(61) Allen, M. P.; Tildesley, D. J. How to analyse the results. In *Computer Simulation of Liquids*; Oxford University Press Inc.: New York, 1987; Section 6.4, pp 191−195.

(62) Zhu, D.; Herbert, B. E.; Schlautman, M. A.; Carraway, E. R. Characterization of Cation-π Interactions in Aqueous Solution Using Deuterium NMR Spectroscopy. *J. Environ. Qual.* **2004**, *33*, 276–284.

(63) DeFelippis, M. R.; Murthy, C. P.; Faraggi, M.; Klapper, M. H. Pulse Radiolytic Measurement of Redox Potentials: The Tyrosine and Tryptophan Radicals. *Biochemistry* **1989**, *28*, 4847–4853.

CT8000415

J. Chem. Theory Comput. **2008,** *4,* 1057–1064

JCTC Journal of Chemical Theory and Computation

Electronic Structure Calculation of MgO (001) Surface with Aggregated Oxygen Vacancies

Motoi Tobita* and Shirun Ho

Hitachi, Ltd., Advanced Research Laboratory (ARL), 1-280 Higashi-Koigakubo,
Kokubunji-shi, Tokyo 185-8601, Japan

Received November 8, 2007

Abstract: The electronic structure of an MgO (001) surface with aggregated oxygen vacancies at the surface was investigated by using a density-functional embedded-cluster method using $Mg_{45}O_x$ ($x = 39-45$) clusters and the B3LYP/CEP-121G level of calculation. The investigation found that the distribution of energy levels corresponding to oxygen-vacancy states strongly depends on the number and configuration of oxygen vacancies. Molecular orbitals of the oxygen vacancy states consist of orbitals at the vacancy sites and their neighboring magnesium ions. These orbitals can simply be modeled by orbitals formed from interacting pseudoatoms. A pseudoatom is an electronically neutral atom representing an oxygen vacancy and electrons of magnesium ions near the vacancy. With this model, an N-atom system having n vacancies can be modeled by n pseudoatoms. Interactions among oxygen vacancies in MgO were qualitatively reproduced by the interacting pseudoatoms model.

Introduction

Magnesium oxide (MgO) is an important material in many industrial applications. A perfect crystal of MgO has a NaCl-like structure and features hardness, transparency, and good insulating properties. In MgO materials, as in other oxides, vacancies are often present. Vacancies alter optical and electronic properties significantly from those observed in the perfect crystal. MgO can therefore be engineered in order to meet a specific need by controlling the position, charged state, and density of vacancies. The secondary-electron-emission coefficient is an example of a property that strongly depends on the position, charged state, and density of oxygen vacancies.[1] This coefficient can be approximately calculated as a function of several parameters including valence-bandwidth, energy levels of the oxygen-vacancy states, band gap, and electron affinity.[2]

Some of the above parameters have been obtained in prior experimental[1,3–11] and theoretical[11–22] investigations. Optical-absorption measurements[1,3,4] showed that the band gap of bulk MgO is 7.4−7.8 eV. Systems with an oxygen monovacancy (called F-center or F^+-center) with a trapping charge of −2 or −1, respectively, have been studied, and the vacancy energy level has been identified by luminescence measurements.[5–10] *Ab initio* calculations of oxygen vacancies as well as structural and energetic investigations have been done for bulk and surface MgO.[13–23]

On the other hand, much less is known about MgO with aggregated oxygen vacancies, which is the subject of the current paper. Experimentally, aggregated vacancies have been observed in many materials by surface reconstruction driven by annealing (see ref 18 and references therein). Using semiempirical complete neglect of differential overlap calculations, Castanier and Noguera predicted that aggregated oxygen vacancies are energetically more favored over periodically spread oxygen vacancies.[15] Finocchi et al. theoretically investigated interaction between oxygen vacancies using the local density approximation (LDA) functional under the periodic boundary condition (PBC).[18] Their calculations used pseudopotential for core electrons. The wave function of valence electrons was expanded by using a plane-wave basis set. They calculated vacancy-formation energy as a function of concentration and configuration of vacancies and derived a two-body model potential accounting for interaction between vacancies.

The LDA functional performed well in predicting equilibrium geometry and vibrational frequencies of corundum (α-Al_2O_3).[24] However, LDA is not the best method for

---

* Corresponding author e-mail: motoi.tobita.mb@hitachi.com.

10.1021/ct700303v CCC: $40.75

predicting band gaps. Therefore, the previous energy levels calculated by LDA are of only qualitative importance.

In this study, density-functional-theory calculations were performed on finite-size clusters of MgO embedded in the crystalline electrostatic potential. In the calculation, core electrons are described by the pseudopotential, as done in the previous LDA study. The wave function of valence electrons is expanded by a linear combination of atomic orbitals. The exchange-correlation term is evaluated by using the B3LYP functional,[25,26] which is known to predict band gaps of a variety of materials, typically within an error of a few tenths of an electronvolt when a large enough basis set is used. Examples of band gap calculations include polypeptides,[27] polyynes,[28] molecular crystals,[29] and metal oxides.[30] The success of the B3LYP functional is due to the correction of the self-interaction error inherent in the LDA and in the generalized gradient approximation functional by a mixing of Hartree-Fock exchange in B3 exchange functional form.[25] The B3LYP functional is thus expected to better predict energy levels of MgO with various oxygen-vacancy configurations. The embedded-cluster model defines an MgO structure with locally aggregated oxygen vacancies. The bulk-size effect, which is important in calculations of ionic materials such as MgO, is approximately included in the model by using embedding-electrostatic potentials.

In this study, orbital-energy-related quantities such as valence-bandwidth, energy levels of oxygen-vacancy states, band gap, and electron affinity of an MgO (001) surface with aggregated oxygen vacancies were calculated. For comparison, calculations were also performed for bulk MgO, an MgO perfect surface, and a surface F-center. It was found that, when oxygen vacancies aggregate, vacancy energy levels split, and the amount of split depends on the number of oxygen-vacancy pairs separated by the nearest-neighbor oxygen distance. The main components of molecular orbitals corresponding to the oxygen-vacancy states are located at oxygen-vacancy sites. An oxygen-vacancy site can thus be viewed as a pseudoatom. Indeed, a system where each oxygen vacancy site is represented by a helium atom reproduces the shape and energy-ordering of molecular orbitals corresponding to the oxygen-vacancy states of MgO with aggregated oxygen vacancies.

In summary, MgO with aggregated oxygen vacancies can exhibit significantly different optical and electronic properties as compared with MgO with monovacancies. Thus, secondary-electron-emission phenomena should largely be influenced by the size and density of surface oxygen vacancies of a material.

Methods

A. Modeling. *1. Embedded-Cluster Model.* The MgO crystal structure has a NaCl-type cubic lattice with a lattice constant of 4.2112 Å. In our calculations, MgO was modeled by finite-size rectangular parallelepiped clusters, where (001) is exposed at each face of the rectangular parallelepiped. The clusters were embedded in a box of point charges that mimicked the Madelung potential made by an ionic crystal lattice. The effect of point charges on the quantum-chemically treated cluster was included in the Hamiltonian through one-electron integrals.

Bulk MgO was also modeled, as well as the MgO surface, to clarify the effect on the surface. In the bulk-MgO model, the cluster was embedded in the center of the point-charge box; however, in the MgO-surface model, the cluster was embedded in the center of the top surface of the point-charge box. The positions of the point charges were fixed at magnesium and oxygen sites of the MgO perfect crystal. The effects of two different types of point-charge quantities, that is, formal charge and consistent charge,[31] were compared. The formal-charge quantity is $+2.0$ for point charges at magnesium sites and -2.0 for those at oxygen sites. The consistent charge was determined so that the point-charge quantity at each magnesium site is equal to the average of Mulliken-charge quantities on the Mg atoms of the cluster. The size of the point-charge box was fixed at $15 \times 16 \times 13$ for the bulk MgO and $15 \times 16 \times 8$ for the MgO surface. With these box sizes, the calculated orbital-energy-related quantities converged to 0.01 eV.

2. Cluster Structure. Convergence of the orbital-energy-related quantities with respect to the cluster size was investigated using six different sizes of rectangular parallelepiped clusters: $1 \times 2 \times 1$ (MgO), $3 \times 4 \times 2$ ($Mg_{12}O_{12}$), $5 \times 6 \times 3$ ($Mg_{45}O_{45}$), $5 \times 6 \times 4$ ($Mg_{60}O_{60}$), $7 \times 8 \times 3$ ($Mg_{84}O_{84}$), and $7 \times 8 \times 4$ ($Mg_{112}O_{112}$). In this model, the third parameter z of "$x \times y \times z$" represents layer thickness. These cluster sizes satisfy the conditions recommended by Lü et al.; namely, in cluster calculations of metal oxides, the cluster should be charge-neutral and stoichiometric.[32]

Effects of geometry relaxation on the orbital-energy-related quantities were investigated using clusters possessing one, three, or five oxygen vacancies, where each of them traps a charge of -2. These cluster structures were constructed by removing oxygen atoms from the charge-neutral and stoichiometric $7 \times 8 \times 3$ cluster surface. The considered structures are depicted in Figure 1a. In the force calculation during geometry relaxation, interactions between an atom and a point charge were described by the electrostatic potential made by the point charge with an additional electron-repulsion potential of the exponential form

$$A \exp(-r/\rho)$$

where ρ is an nonbonded interaction parameter, r is the distance between a point charge and an atom, and A is a proportional parameter. The values of ρ and A that are used are 0.252 $Å^{-1}$ and 18065.8 kcal/mol for point charges at magnesium sites and 0.26 $Å^{-1}$ and 58298.9 kcal/mol for those at oxygen sites, respectively.[33]

Clusters possessing two to six oxygen vacancies, where each of them traps a charge of -2, were constructed by removing two to six oxygen atoms from the charge-neutral and stoichiometric $5 \times 6 \times 3$ cluster surface. The geometry of the remaining atoms was fixed at the perfect crystal structure, and the point-charge quantity was fixed at the formal charge values. The considered structures, including the perfect surface and the surface F-center, are depicted in Figure 1b. The number included in each structure label shows the number of oxygen vacancies in the cluster.

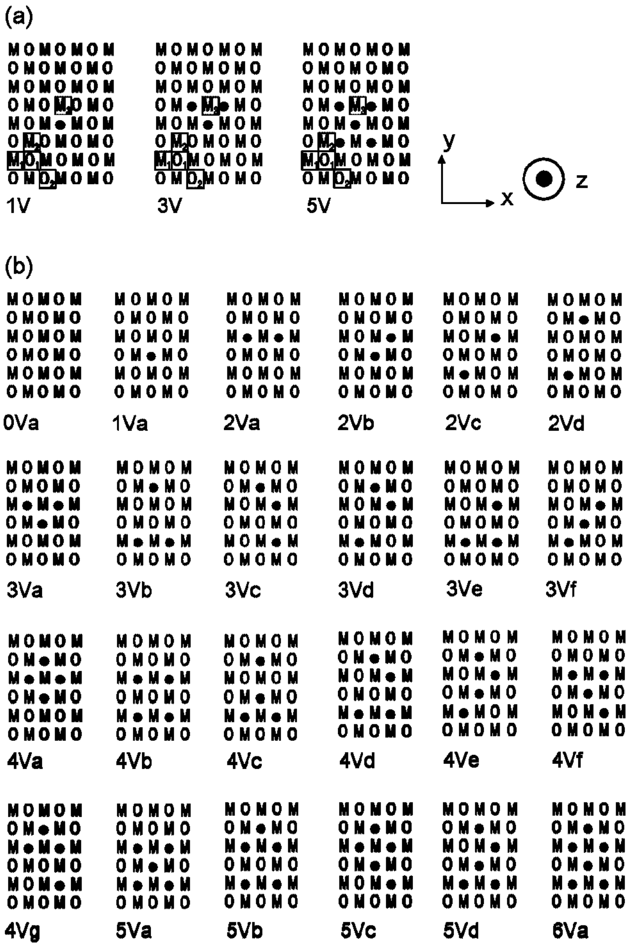


Figure 1. Structures of MgO clusters investigated. (a) 7 × 8 × 3 clusters with one, three, or five oxygen vacancies. (b) 5 × 6 × 3 clusters with zero to six oxygen vacancies. Only the top layer is shown. The second and third layers include no oxygen vacancies; M, magnesium atom; O, oxygen atom; closed circle, oxygen vacancy with a trapping charge of −2.

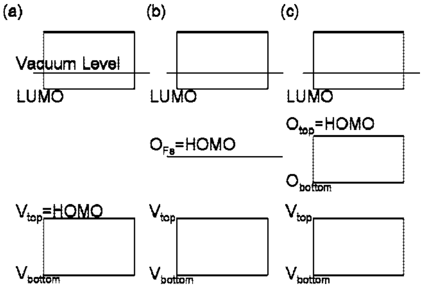


Figure 2. Distribution of orbital energies for (a) perfect MgO, (b) MgO with surface F-center, and (c) MgO with aggregated oxygen vacancies at the surface.

B. Calculation of Orbital-Energy-Related Quantities. An initial investigation of cluster size and the geometry relaxation effect on the orbital-energy-related quantities was performed at the B3LYP[25,26]/CEP-4G[34] level of calculation. Further analyses were performed using a larger basis set at the B3LYP/CEP-121G[34] level. The B3LYP functional is a Hartree-Fock/pure-DFT hybrid functional. Its exchange term contains a 20% contribution from the Hartree-Fock exchange term. The CEP-4G and CEP-121G basis sets are an effective-core-potential basis set. Oxygen 1s electrons and magnesium 1s through 2p electrons are described by the effective core potential. In CEP-4G, valence electrons are expanded by one s-type function and a set of p-type functions. Whereas in CEP-121G, three s-type functions and three sets of p-type functions are used in the expansion. Although the B3LYP functional predicts the band gap of materials with good accuracy when a large enough basis set is used, the CEP-121G basis set is not large enough. To estimate orbital energies with the basis set limit, occupied energy levels (E_{OCC}) and the lowest unoccupied molecular orbital (LUMO) energy (E_{LUMO}) were corrected by using the following empirical relations derived for the highest occupied molecular orbital (HOMO) and LUMO energy correction:[35]

$$E_{\mathrm{OCC}}=(E_{\mathrm{OCC}}{}^{*} + 0.050)/1.001 \tag{1}$$

$$E_{\mathrm{LUMO}} = 0.6091 E_{\mathrm{LUMO}}{}^{*} - 0.475 \tag{2}$$

where energy terms with an asterisk are calculated quantities and terms without an asterisk are corrected quantities for the basis set limit. This kind of empirical correction is valid because using an incomplete basis set introduces a systematic error in orbital energies.[35–37] Note that the coefficients in the above relations are not optimized for B3LYP/CEP-121G, but for B3LYP/6-31G(d). We assume that these coefficients are still qualitatively "good" because the used functional is the same and because both basis sets are similar in terms of the number of functions per atom.

The calculated orbital energies were used to calculate the valence bandwidth as the energy difference between the top and bottom of the valence orbitals. Band gap was calculated as the energy difference between HOMO and LUMO. Electron affinity was calculated as the energy difference between the vacuum level and LUMO.

All of the calculations were performed using Gaussian03.[38]

Results and Discussion

Before the calculated results are discussed, the orbital-energy distribution of MgO is briefly explained, using Figure 2. The energy distribution for a perfect MgO surface, MgO with a surface F-center, and MgO with aggregated oxygen vacancies at the surface is shown in Figure 2a−c, respectively. In all of the cases, LUMO lies below the vacuum level. Energy levels of oxygen-vacancy states lie in between the LUMO and the top of valence orbitals, V_{top}, as seen in Figure 2b,c. As a consequence, the top of the oxygen-vacancy levels becomes the HOMO when vacancies exist in the MgO.

A. Initial Investigation of Cluster Size and Geometry Relaxation Effects. The orbital-energy-related quantities of the MgO surface calculated with different sizes of cluster are given in Figure 3. Calculated orbital energy differences LUMO−HOMO, HOMO−V_{bottom}, and −LUMO were 5.85, 4.92, and −1.14 eV, respectively, when a 7 × 8 × 4 cluster was used. The energy differences obtained with a 5 × 6 × 3 cluster differed by 0.03 eV for LUMO−HOMO and −LUMO and by 0.3 eV for HOMO−V_{bottom} as compared with the values obtained with the 7 × 8 × 4 cluster. From these data, we judged it suitable to perform further investiga-

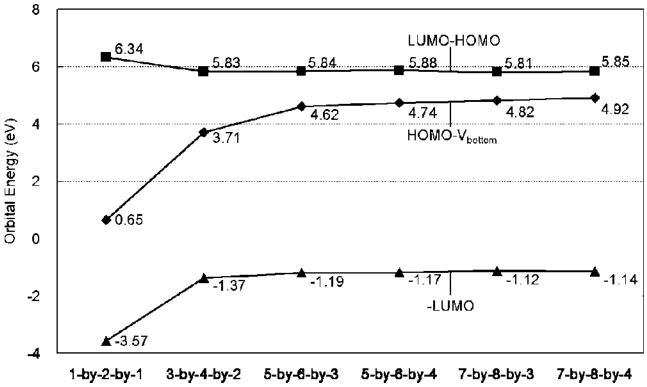


Figure 3. Convergence of orbital-energy-related quantities with respect to cluster size calculated at the B3LYP/CEP-4G level.

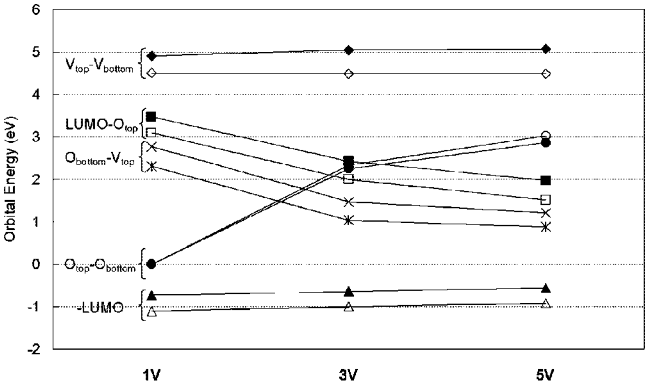


Figure 4. Orbital energies of clusters having one, three, or five oxygen vacancies with and without geometry relaxation effects calculated at the B3LYP/CEP-4G level. Stars and closed diamonds, squares, circles, and triangles show values without geometry relaxation. Crosses and open diamonds, squares, circles, and triangles show values with geometry relaxation.

tions using the 5 × 6 × 3 cluster. A prior investigation also supports the use of at least a three-layer model for a quantitative description of the surface F-center.[17]

The orbital-energy-related quantities of the MgO surface having one, three, or five oxygen vacancies with and without geometry relaxation effects are given in Figure 4. The amounts of shift in the energy differences are almost constant regardless of the number of oxygen vacancies. Those values between the top and bottom of the valence orbital ($V_{top}-V_{bottom}$), between the LUMO and the top of the oxygen vacancy orbital (LUMO$-O_{top}$), between the bottom of the oxygen vacancy orbital and the top of the valence orbital ($O_{bottom}-V_{top}$), and between the vacuum level and the LUMO ($-$LUMO) are all around 0.4$-$0.5 eV. The energy difference between the top and bottom of the oxygen vacancy orbital ($O_{bottom}-O_{top}$) is much less affected by the geometry relaxation effects. These results demonstrate that the contribution of the geometry relaxation effects to the orbital energy difference is small. Thus, further investigations (subsection B and later) using a larger basis set were performed using the perfect crystal structure. Geometries of systems with vacancies are constructed by simply deleting coordinates of atoms from those of the perfect crystal.

Table 1. The Amount of Displacement from Crystalline Lattice for the Selected Atoms in Particular Direction after Geometry Relaxation (Å)[a]

		structure		
atom	direction	1V	3V	5V
M_1	*x*	0.026	0.027	0.023
M_1	*z*	0.028	0.027	0.027
M_2	*x*	−0.019	−0.010	−0.063
M_2	*z*	0.024	0.023	0.016
M_3	*y*	−0.016	−0.012	−0.040
M_3	*z*	0.023	0.011	0.009
O_1	*x*	0.006	0.014	0.011
O_1	*z*	−0.026	−0.028	−0.020
O_2	*x*	0.014	0.011	0.003
O_2	*z*	−0.034	−0.036	−0.034

[a] The definitions of the atoms and direction are defined in Figure 1a. The results are obtained with the B3LYP/CEP-4G level of calculation.

To see the geometrical change after the geometry relaxation, the amount of shift in the atomic position from the lattice of the MgO crystal is calculated. Table 1 shows the result for the selected atoms, defined in Figure 1a, in particular directions. Five selected atoms, M_1, M_2, M_3, O_1, and O_2, are classified into two groups. The atoms in the first group, M_1 and O_2, are at the periphery of the cluster and thus are adjacent to point charges. Those in the second group, M_2, M_3, and O_1, are not adjacent to point charges and are closer to oxygen vacancy sites. From Table 1, the largest displacement is observed for the second-group atoms that are adjacent to oxygen vacancy sites, such as atom M_2 in structure 5V for the *x* direction. This result coincides with a prior work by D'Ercole and Pisani. They reported that only the first-neighbor atoms of vacancies moves significantly after the geometry relaxation.[20] Here, we showed that even the largest amount of shift of 0.063 Å corresponds to only 3% of a change in atomic position with respect to the Mg$-$O distance in the crystal, 2.1056 Å. Also, the amount of shift in the atoms of the first group is even smaller. Therefore, the geometrical errors caused by the use of point charges are considered to be negligible.

B. Perfect Crystal and Surface. Calculated consistent charges and orbital-energy differences of the bulk MgO and MgO surface, modeled with the 5 × 6 × 3 cluster and calculated with B3LYP/CEP-121G, are given in Table 2.

The values of consistent point charge are 1.50 for the bulk and 1.45 for the surface. Although MgO is considered as a nearly perfect ionic crystal, the calculated values are significantly smaller than the value expected for the perfectly ionic crystal, 2.0. Gerson and Bredow obtained a consistent point charge of ±1.11 from 8 × 8 × 4 and 8 × 8 × 5 clusters using a semiempirical method with a Löwdin population analysis.[21] When the charge is defined according to a Löwdin- or Mulliken-type population analysis, the consistent charge seems to become significantly smaller than the formal charge value. However, a presumably better description of charge, such as a charge determined by natural-population analysis,[39] gave a charge close to ±1.9.[22] The present consistent charge quantities could be an artifact of using a Mulliken-type charge for an ionic system.[40] Therefore,

Table 2. Calculated Energy Differences between HOMO and the Bottom of Valence Orbitals (HOMO−V_{bottom}), between LUMO and HOMO (LUMO−HOMO), between the Vacuum Level and LUMO (−LUMO), and between the Vacuum Level and HOMO (−HOMO) for Bulk MgO and MgO Surface Modeled by a 5 × 6 × 3 Cluster[a]

	point-charge type	point-charge quantity (|*e*|)	HOMO−V_{bottom} (eV)	LUMO−HOMO (eV)	−LUMO (eV)	−HOMO (eV)
bulk	formal	2.00	3.87	6.39 (7.99)	2.89 (1.29)	9.28 (9.32)
	consistent	1.50	3.91	6.24 (7.41)	1.79 (0.62)	8.03 (8.07)
	exp.[1,3,4]		4.0∼8.5[b]	7.4∼7.8[c]	0.85∼1.0[d]	
surface	formal	2.00	3.90	6.01 (7.48)	2.52 (1.06)	8.54 (8.58)
	consistent	1.45	4.25	5.32 (6.52)	1.85 (0.65)	7.17 (7.21)
	exp. [11]					6.7 ± 0.4

[a] The formal charge and consistent point charge were used for the energy-difference calculations. Numbers in parentheses are the empirically corrected values using eqs 1 and 2. [b] Valence-bandwidth. [c] Band gap. [d] Electron affinity.

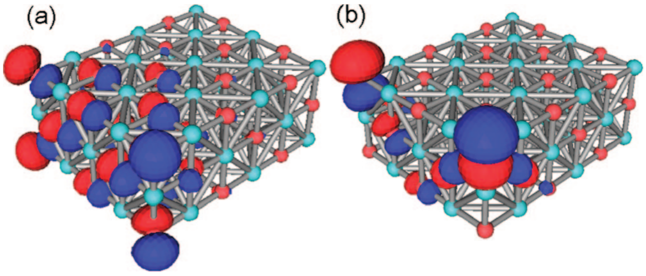


Figure 5. (Color online) HOMO of 5 × 6 × 3 MgO clusters modeling (a) bulk MgO and (b) MgO surface; sky-blue (lighter) sphere, magnesium atom; red (darker) sphere, oxygen atom; blue (darker) and red (lighter) shapes, isosurface (level = 0.03) of the molecular orbital.

calculations for systems with aggregated oxygen vacancies (subsection C) were performed using the formal charge.

The value of −HOMO in the bulk is larger than that in the surface by about 0.7−0.9 eV, as seen from Table 2. However, the difference in the other orbital-energy-related quantities between the bulk and surface is smaller. Thus, we looked at the HOMO more closely. Figure 5 shows HOMOs of the bulk MgO and the MgO surface. In both cases, the HOMO consists of oxygen 2p orbitals. In the bulk MgO (Figure 5a), the majority of oxygen 2p orbitals lie along the surface. As a consequence, interaction between neighboring oxygen atoms in the same layer has an antibonding character, while the one in the neighboring layer has a bonding character. As a sum, HOMO has an almost nonbonding character. In the MgO surface (Figure 5b), oxygen 2p orbitals stand perpendicular to the surface. The main components of the HOMO separate at the edges of the cluster, and interaction between orbitals at the edges has an antibonding character. The HOMO of the bulk MgO is therefore more stable than that of the MgO surface. The difference in HOMOs is caused by the presence/absence of point charges above the top surface of the cluster.

The empirically corrected orbital energies and experimentally observed energies[1,3,4,11] compare well. Although the calculated values are slightly underestimated in the bulk and overestimated in the surface, the degree of error from the experimental values is within a few tenths of an electronvolt.

C. Aggregated Oxygen Vacancies at the Surface. Figure 6 shows the distributions of orbital energies of the 24 structures shown in Figure 1b. Formal charge was used as point charge, and geometry-relaxation effects were excluded. V_{bottom}*, V_{top}*, and LUMO* are the orbital energy of the bottom of valence orbitals, that of the top of valence orbitals, and the LUMO of structure 0Va (see Figure 1b). Orbital energies shown between V_{top}* and LUMO* are oxygen-vacancy levels, of which the highest and lowest energies are denoted as O_{top} and O_{bottom}, respectively, in Figure 6. Structures are sorted according to the number of oxygen vacancies. Among the structures with the same number of oxygen vacancies, structures are sorted according to the energy of O_{bottom}. The energy of O_{top} increases, while the energy of O_{bottom} decreases when the structure contains a larger number of oxygen vacancies.

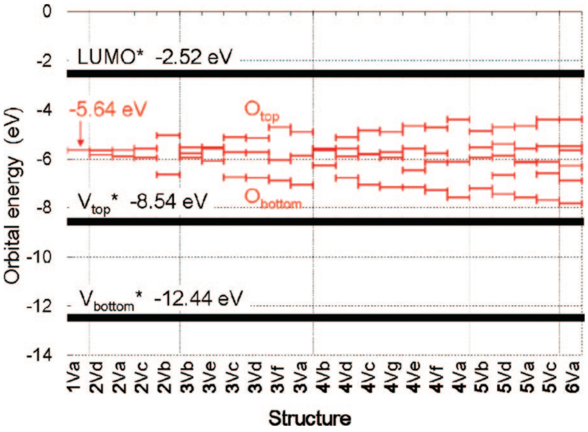


Figure 6. (Color online) Distribution of orbital energies for the 24 structures defined in Figure 1b. V_{bottom}*, orbital energy of the bottom of valence orbitals of structure 0V; V_{top}*, orbital energy of the top of valence orbitals of structure 0V; LUMO*, LUMO of structure 0V; O_{top}, the top of oxygen-vacancy energy levels; O_{bottom}, the bottom of oxygen-vacancy energy levels.

In structure 1Va, the model for the surface F-center, the value of the oxygen vacancy level, O_{Fs}, is −5.64 eV. Experimentally, the absorption energy of the F-center is known as 5.01 eV.[5] This energy corresponds to the energy difference between the LUMO and the oxygen vacancy level. Applying eqs 1 and 2 to the O_{Fs} and LUMO, respectively, leads to an energy difference of 4.51 eV, as shown in Table 3. Although the inclusion of the geometry relaxation effects would increase the current difference between the calculation and experiment by 0.4−0.5 eV as expected from Figure 4, the error of within 1 eV is still a satisfactory agreement with the experiment.

Coming back to Figure 6, among the structures with the same number of oxygen vacancies, large deviations in the energy difference between O_{top} and O_{bottom} (O_{top}−O_{bottom}) were observed. The O_{top}−O_{bottom} values are relatively large for structures 2Vb, 3Vc, 3Vd, and so forth, as shown in the

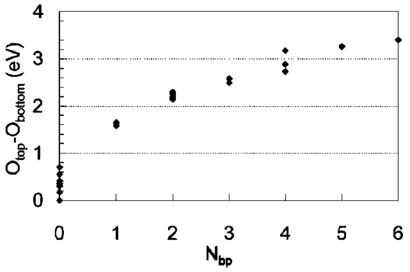


Figure 7. The relationship between the number of oxygen-vacancy pairs separated by the nearest-neighbor oxygen distance and the energy difference between the top and bottom of oxygen-vacancy levels ($O_{top}-O_{bottom}$).

Table 3. Orbital Energies of the LUMO and the Oxygen Vacancy Level (O_{Fs}) and Their Difference as well as Experimentally Measured Absorption Energy of the F Center[a]

	LUMO	O_{Fs}	LUMO-O_{Fs}
uncorrected	−2.52	−5.64	3.12
corrected by eqs 1 and 2	−1.06	−5.57	4.51
experiment			5.01[5]

[a] A B3LYP/CEP-121G calculation was performed for a MgO surface with one oxygen vacancy that traps a charge of −2 (F_s-center) modeled by a $Mg_{45}O_{44}$ (structure 1Va in Fig. 1b) cluster embedded in a 15 × 16 × 8 formal point-charge-box. No geometry relaxation effects were included.

gaps between red lines in Figure 6. The values of $O_{top}-O_{bottom}$ seem to be related to the number of oxygen-vacancy pairs separated by the nearest-neighbor oxygen distance, hereafter denoted as N_{bp}. For example, in structure 2Vd, the two oxygen vacancies are separated by more than the nearest-neighbor distance between two oxygen atoms, thus giving $N_{bp} = 0$, and the value of $O_{top}-O_{bottom}$ is about 0.2 eV. On the other hand, in structure 2Vb, the two oxygen vacancies are separated by the nearest-neighbor oxygen distance ($N_{bp} = 1$), and the value of $O_{top}-O_{bottom}$ is about 1.6 eV. To clarify this point further, the relationship between N_{bp} and $O_{top}-O_{bottom}$ for all of the considered structures is plotted in Figure 7. As N_{bp} increases, the $O_{top}-O_{bottom}$ value also tends to increase. Therefore, N_{bp} is a convenient index to understand the orbital energy of oxygen-vacancy levels.

Figure 8 shows molecular orbitals of O_{bottom} and O_{top} levels for structures 2Vd and 2Vb. In structure 2Vd, the two oxygen vacancies shown by crosses are relatively far apart, and both molecular orbitals corresponding to O_{bottom} (Figure 8a) and to O_{top} (Figure 8b) localize at the oxygen-vacancy sites. However, in structure 2Vb, molecular orbitals corresponding to O_{bottom} (Figure 8c) and to O_{top} (Figure 8d) delocalize along the two oxygen vacancies. In both structures, 2Vd and 2Vb, the molecular orbital corresponding to O_{bottom} shows bonding character, and that corresponding to O_{top} shows antibonding character. Since those observed molecular orbitals are in the vicinity of the oxygen vacancy sites, it is possible to view these orbitals as those made by pseudoatoms. Under this viewpoint, molecular orbitals corresponding to O_{bottom} and O_{top} are interpreted as bonding and antibonding molecular orbitals made by two interacting pseudoatoms.

The concept of the pseudoatom was applied to the system with six oxygen vacancies, structure 6V. A pseudoatom is an electronically neutral atom representing an oxygen vacancy and electrons of magnesium ions near the vacancy. The system was modeled by a cluster composed of six helium atoms. The helium atom is chosen as a representation of the pseudoatom because both a helium atom and an oxygen vacancy with a trapping charge of −2 have two electrons. In Figure 9, molecular orbitals of oxygen-vacancy levels for structure 6Va and those of the helium cluster are shown in the order of increasing orbital energy. Molecular orbitals of structure 6Va and the helium cluster have one-to-one correspondence in the correct energy ordering. The degree of delocalization differs, however, because in structure 6Va, 3s orbitals of surface magnesium atoms are dominant, while in the helium cluster, composing atomic orbitals are much more confined 1s orbitals.

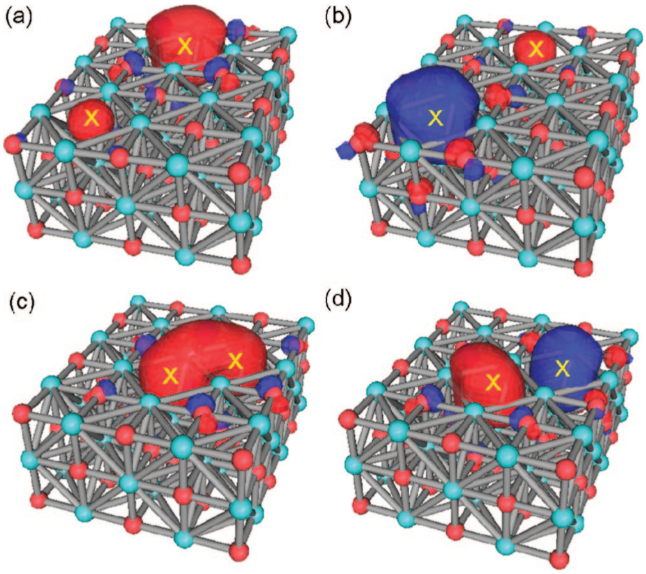


Figure 8. (Color online) Molecular orbitals of the top and bottom of the oxygen-vacancy levels for structures 2Vd and 2Vb. (a) Structure 2Vd, O_{top}; (b) structure 2Vd, O_{bottom}; (c) structure 2Vb, O_{top}; (d) structure 2Vb, O_{bottom}. Sky-blue (lighter) sphere, magnesium atom; red (darker) sphere, oxygen atom; cross, oxygen-vacancy site; blue (darker) and red (lighter) shapes, isosurface (level = 0.03) of the molecular orbital.

It is informative to compare obtained orbital energies with those obtained using the PBC calculated by Finocchi et al.[18] They defined a unit cell of 16 atoms (4 × 4) at the surface, of which eight are oxygen sites. Their structures containing two oxygen vacancies at the surface, either homogeneously or in a close-packed manner (0.25(h) and 0.25(cp), see Figure 1 of ref 18), are compared to our structures 2Vc and 2Vb, respectively. Similarly, their structures containing four oxygen vacancies at the surface (0.5(h) and 0.5(cp)) are compared to our structures 4Vb and 4Va, respectively.

When one or more pairs of oxygen vacancies are located at the nearest-neighbor oxygen distance, as in structures 2Vb and 4Va, orbital energies agree with each other within 0.5 eV. For these systems, adjacent oxygen vacancies form extended molecular orbitals that correspond to O_{top} and O_{bottom}. Therefore, interaction between more distant oxygen vacancies introduced by the use of the PBC does not significantly alter the oxygen-vacancy energy levels. On the other hand, when no oxygen-vacancy pairs are separated by the nearest-neighbor oxygen distance, as in structures 2Vc

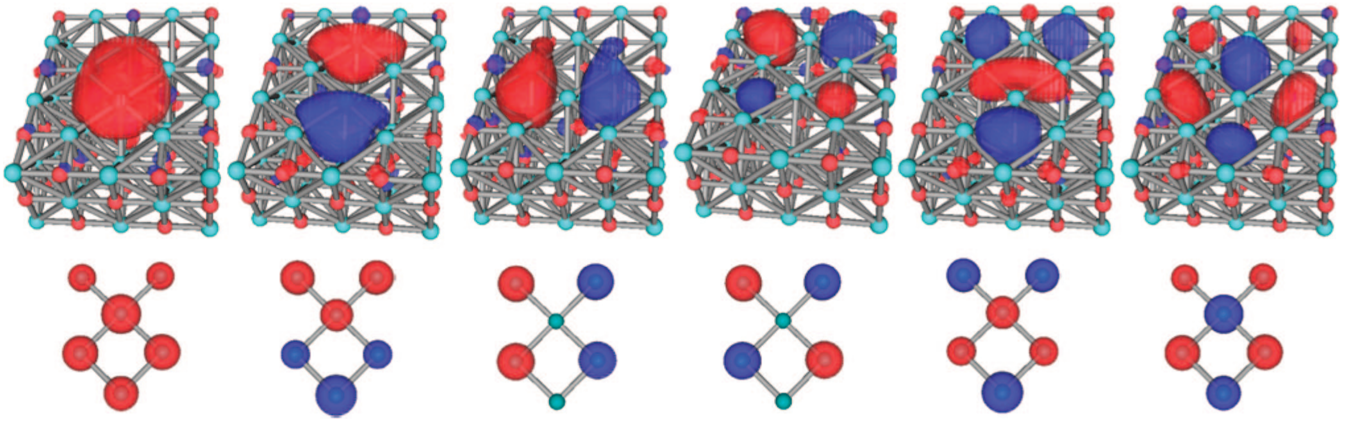


Figure 9. (Color online) Molecular orbitals of oxygen-vacancy levels for structure 6V (upper row) and those of the helium cluster (lower row). In the upper row: sky-blue (lighter) sphere, magnesium atom; red (darker) sphere, oxygen atom. In the lower row: a sphere represents a helium atom; blue (darker) and red (lighter) shapes, isosurface (level = 0.03) of the molecular orbital.

and 4Vb, the difference in orbital energies enlarges to more than 0.8 eV. In this case, the sum of long-range interaction between oxygen vacancies, which is present only in the PBC model, is no longer negligible.

The PBC models a structure with oxygen vacancies periodically spread over MgO materials. This structure may be seen as annealed MgO, where oxygen vacancies evenly spread over materials. On the other hand, in the cluster model, aggregated oxygen vacancies are isolated. The considered structures may thus be seen as a locally damaged MgO surface by ion bombardment.

Conclusions

Electronic structures of perfect bulk MgO, perfect surface MgO, an MgO surface with a surface F-center, and an MgO surface with aggregated oxygen vacancies at the surface were calculated using an embedded-cluster approach. The orbital-energy-related quantities converge within 0.3 eV when a 5 × 6 × 3 cluster is used. The inclusion of geometry-relaxation effects shifts the orbital energies. However, the amount of the shift is almost irrelevant to the number of oxygen vacancies. In the calculation of perfect bulk MgO, perfect MgO surface, and an MgO surface with one oxygen vacancy, measured valence-bandwidth, band gap, and electron affinity were reproduced within a few tenths of an electronvolt of error. To achieve this accuracy, an empirical correction, accounting for the incompleteness of a basis set, on orbital energies was applied. For an MgO surface with aggregated oxygen vacancies at the surface, the distribution of energy levels corresponding to oxygen-vacancy states depends strongly on the number and configuration of oxygen vacancies. When oxygen vacancies are separated by the nearest-neighbor oxygen distance, the energy difference between the top and bottom of the oxygen-vacancy levels ($O_{top}-O_{bottom}$) becomes larger. As the number of oxygen-vacancy pairs separated by the nearest-neighbor oxygen distance increases, the value of $O_{top}-O_{bottom}$ increases. Molecular orbitals of oxygen-vacancy states largely localize in the vicinity of the vacancy sites. An oxygen-vacancy site can therefore be viewed as a pseudoatom. Molecular orbitals of the oxygen-vacancy states can be interpreted as orbitals formed by interacting pseudoatoms. Oxygen-vacancy energy levels for isolated vacancies modeled by cluster calculation and those for evenly spread vacancies modeled by PBC are similar when two vacancies are adjacent. On the other hand, when no vacancy pairs are separated by the nearest-neighbor oxygen distance, the difference in the vacancy energy levels becomes larger. It is concluded that MgO with aggregated oxygen vacancies can exhibit significantly different optical and electronic properties as compared with MgO with monovacancies or periodically distributed vacancies.

References

(1) Aboelfotoh, M. O.; Lorenzen, J. A. Influence of Secondary-electron Emission from MgO Surfaces on Coltage-breakdown Curves in Penning Mixutures for Insulated-electrode Discharges. *J. Appl. Phys.* **1977**, *48*, 4754–4759.

(2) Motoyama, Y.; Sato, F. Calculation of Secondary Electron Emission Yield γ from MgO Surface. *IEEE Trans. Plasma Sci.* **2006**, *34*, 336–342.

(3) Johnson, P. D. Some Optical Properties of MgO in the Vacuum Ultraviolet. *Phys. Rev.* **1954**, *94*, 845–846.

(4) Roessler, D. M.; Walker, W. C. Electronic Spectrum and Ultraviolet Optical Properties of Crystalline MgO. *Phys. Rev.* **1967**, *159*, 733–738.

(5) Chen, Y.; Kolopus, J. L.; Sibley, W. A. Luminescence of the F^+ Center in MgO. *Phys. Rev.* **1969**, *186*, 865–870.

(6) Williams, R. T.; Williams, J. W.; Turner, T. J.; Lee, K. H. Kinetics of Radiative Recombination in Magnesium Oxide. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1979**, *20*, 1687–1699.

(7) Summers, G. P.; Wilson, T. M.; Jeffries, B. T.; Tohver, H. T.; Chen, Y.; Abraham, M. M. Luminescence from Oxygen Vacancies in MgO Crystals Thermochemically Reduced at High Temperatures. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1983**, *27*, 1283–1291.

(8) Rosenblatt, G. H.; Rowe, M. W.; Williams, G. P., Jr.; Williams, R. T.; Chen, Y. Luminescence of F and F^+ centers in Magnesium Oxide. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1989**, *39*, 10309–10318.

(9) Chaudhri, M. M.; Sands, H. S. Photoluminescence from Indented MgO Crystals Using a Near Ultraviolet/Visible Raman Microscope. *J. Appl. Phys.* **1997**, *82*, 785–791.

(10) González, R.; Monge, M. A.; Munos Santiuste, J. E.; Pareja, J. E.; Chen, Y.; Kotomin, E.; Kukla, M. M.; Popov, A. I. Photoconversion of F-type Centers in Thermochemically

Reduced MgO Single Crystals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 4786–4790.

(11) Kantorovich, L. N.; Shluger, A. L.; Sushko, P. V.; Günster, J.; Stracke, P.; Goodman, D. W.; Kempter, V. Mg Clusters on MgO surfaces: Study of the Nucleation Mechanism with MIES and Ab Initio Calculations. *Faraday Discuss.* **1999**, *114*, 173–194.

(12) Wang, Q. S.; Holzwarth, N. A. W. Electronic Structure of Vacancy Defects in MgO Crystals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1990**, *41*, 3211–3225.

(13) Gibson, A.; Haydock, R.; LaFemina, J. P. The Electronic Structure of Neutral and Charged Surface Vacancy Defects in Periclase. *Appl. Surf. Sci.* **1993**, *72*, 285–293.

(14) Gibson, A.; Haydock, R.; LaFemina, J. P. Stability of Vacancy Defects in MgO: The Role of Charge Neutrality. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 2582–2592.

(15) Castanier, E.; Noguera, C. Non-stoichiometric Reconstructions on MgO(100). *Surf. Sci.* **1996**, *364*, 17–29.

(16) Scorza, E.; Birkenheuer, U.; Pisani, C. The Oxygen Vacancy at the Surface and in Bulk MgO: An Embedded-Cluster study. *J. Chem. Phys.* **1997**, *107*, 9645–9657.

(17) Orlando, R.; Millini, R.; Perego, G.; Dovesi, R. Catalytic Properties of F-centres at the Magnesium Oxide Surface: Hydrogen Abstraction from Methane. *J. Mol. Catal.* **1997**, *119*, 253–262.

(18) Finocchi, F.; Goniakowski, J.; Noguera, C. Interaction between Oxygen Vacancies on MgO(100). *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 5178–5188.

(19) D'Ercole, A.; Giamello, E.; Pisani, C.; Ojämae, L. Embedded-Cluster Study of Hydrogen Interaction with an Oxygen Vacancy at the Magnesium Oxide Surface. *J. Phys. Chem. B* **1999**, *103*, 3872–3876.

(20) D'Ercole, A.; Pisani, C. Ab Initio Study of Hydrogen Dissociation at a Surface Divancacy on the (001) MgO Surface. *J. Chem. Phys.* **1999**, *111*, 9743–9753.

(21) Gerson, A. R.; Bredow, T. MgO(100) Surface Relaxation and Vacancy Defects: a Semi-Empirical Quantum-Chemical Study. *Phys. Chem. Chem. Phys.* **1999**, *1*, 4889–4896.

(22) Sushko, P. V.; Shluger, A. L.; Catlow, C. R. A. Relative Energies of Surface and Defect States: Ab Initio Calculations for the MgO (001) Surface. *Surf. Sci.* **2000**, *450*, 153–170.

(23) Sushko, P. V.; Gavartin, J. L.; Shluger, A. L. Electronic Properties of Structural Defects at the MgO (001) Surface. *J. Phys. Chem. B* **2002**, *106*, 2269–2276.

(24) Montanari, B.; Civalleri, B.; Zicovich-Wilson, C. M.; Dovesi, R. Properties, Dynamics, and Electronic Structure of Condensed Systems and Clusters Influence of the Exchange-Correlation functional in All-Electron Calculations of Vibrational Frequencies of Corundum (Al_2O_3). *Int. J. Quantum Chem.* **2006**, *106*, 1703–1714.

(25) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(26) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.

(27) Ladik, J.; Bogar, F.; Penke, B. Comparison of HF, HF+MP2, LDA, BLYP and B3LYP Band Structures of the Homopeptides. *Int. J. Quantum Chem.* **2004**, *98*, 522–527.

(28) Yang, S.; Kertesz, M. Bond Length Alternation and Energy Band Gap of Polyyne. *J. Phys. Chem. A* **2006**, *110*, 9771–9774.

(29) Perger, W. F. Calculation of Band Gaps in Molecular Crystals Using Hybrid Functional Theory. *Chem. Phys. Lett.* **2003**, *368*, 319–323.

(30) Muscat, J.; Wander, A.; Harrison, N. M. On the Prediction of Band Gaps from Hybrid Functional Theory. *Chem. Phys. Lett.* **2001**, *342*, 397–401.

(31) Xu, X.; Nakatsuji, H.; Ehara, M.; Lü, X.; Wang, N. Q.; Zhang, Q. E. Cluster Modeling of Metal Oxides: the Influence of the Surrounding Point Charges on the Embedded Cluster. *Chem. Phys. Lett.* **1998**, *292*, 282–288.

(32) Lü, X.; Xu, X.; Wang, N.; Zhang, Q.; Ehara, M.; Nakatsuji, H. Cluster Modeling of Metal Oxides: How to Cut out a Cluster. *Chem. Phys. Lett.* **1998**, *291*, 445–452.

(33) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(34) Stevens, W. J.; Basch, H.; Krauss, M. Compact Effective Potentials and Efficient Shared-Exponent Basis Sets for the First- and Second-row Atoms. *J. Chem. Phys.* **1984**, *81*, 6026–6033.

(35) Zhan, C.-G.; Nichols, J. A.; Dixon, D. A. Ionization Potential, Electron Affinity, Electronegativity, Hardness, and Electron Excitation Energy: Molecular Properties from Density Functional Theory Orbital Energies. *J. Phys. Chem. A* **2003**, *107*, 4184–4195.

(36) Salzner, U.; Lagowski, J. B.; Pickup, P. G.; Poirier, R. A. Design of Low Band Gap Polymers Employing Density Functional Theory — Hybrid Functionals Ameliorate Band Gap Problem. *J. Comput. Chem.* **1998**, *18*, 1943–1953.

(37) Takahata, Y.; Chong, D. P. Density-Functional Calculations of Molecular Electron Affinities. *J. Braz. Chem. Soc.* **1999**, *10*, 354–358.

(38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision D.02; Gaussian, Inc., Wallingford, CT, 2004.

(39) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *J. Chem. Phys.* **1985**, *83*, 735–746.

(40) Collins, J. B.; Streitwieser, A. Integrated Spatial Electron Populations in Molecules: Application to Simple Molecules. *J. Comput. Chem.* **1980**, *1*, 81–87.

CT700303V

J. Chem. Theory Comput. **2008,** *4,* 1065–1072 **1065**

JCTC Journal of Chemical Theory and Computation

Calculation of Fukui Functions Without Differentiating to the Number of Electrons. 3. Local Fukui Function and Dual Descriptor

Tim Fievez,† Nick Sablon,† Frank De Proft,† Paul W. Ayers,‡ and Paul Geerlings*,†

Eenheid Algemene Chemie (ALGC), Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussel, Belgium, and Department of Chemistry, McMaster University, Hamilton, Ontario L8S 4M1, Canada

Received January 25, 2008

Abstract: An alternative approach for the calculation of DFT-based reactivity descriptors involving derivatives of the energy with respect to the number of electrons and the external potential is further evaluated. Using functional derivatives with respect to the external potential, the finite difference approximation was avoided for the local calculation of the Fukui functions and the dual descriptor. A relevant set of molecules has been calculated after the optimization of computational parameters. It is shown that the new approach correctly predicts the nucleophilic attack on CH_2O, the formation of CO metal complexes, the regioselectivity of monosubstituted benzenes, and the softest nucleophilic site in some ambident nucleophiles.

1. Introduction

Density functional theory (DFT), with the electron density ($\rho(\bar{r})$) as the central quantity, provides an extraordinary tool for the theoretical study of chemical systems. On the basis of the Hohenberg and Kohn theorems[1] and the subsequent work of Kohn and Sham,[2] DFT[3,4] became highly popular as an ab initio method during the past two decades, mainly due to the development of accurate exchange-correlation functionals[5,6] and the implementation of the formalism in mainstream ab initio programs.

The advantages of DFT mainly focus on its superior accuracy per unit computational cost and the possibility of defining chemical concepts in a precise way using the so-called conceptual DFT.[7–10] This flourishing DFT domain, founded by Parr and co-workers, creates a theoretical framework for various chemical reactivity concepts, allowing their qualitative and quantitative use in reactivity studies. For example, quantities such as electronegativity,[11] hardness,[12] and softness[13] received a strict definition in DFT. Insight into the different descriptors showed that most reactivity descriptors are defined as derivatives of the electronic energy with respect to the number of electrons or the external potential.[14] (For an isolated system, the external potential is the potential caused by the atomic nuclei.) The Fukui function $f(\bar{r})$,[15] for example, is defined as the (mixed) variational derivative of the energy with respect to the number of electrons N and the external potential $v(\bar{r})$:

$$f(\bar{r}) = \frac{\partial^2 E}{\partial N \delta\, v(\bar{r})} \tag{1}$$

A simpler form of the Fukui function can be formulated using the electron density, $\rho(\bar{r})$:

$$f(\bar{r}) = \left(\frac{\partial\, \rho(\bar{r})}{\partial N}\right)_{v(\bar{r})} \tag{2}$$

Formulas 1 and 2 are problematic since the "derivative discontinuity" in the energy (and the density) with respect to the number of electrons[16–21] makes $f(\bar{r})$ undefined for an integer number of electrons and hence for any isolated molecule. This can be overcome by introducing derivatives from the right, f^+, and from the left, f^-. In line with Perdew et al.'s[16,17] conclusions, these derivatives can be computed exactly with the electron densities of the anionic ($\rho_{N+1}(\bar{r})$), cationic ($\rho_{N-1}(\bar{r})$), and neutral systems ($\rho_N(\bar{r})$):[19,21]

$$f_N^+(\bar{r}) = \left(\frac{\partial\, \rho_N(\bar{r})}{\partial N}\right)^+_{v(\bar{r})} = \rho_{N+1}(\bar{r}) - \rho_N(\bar{r}) \tag{3}$$

$$f_N^-(\bar{r}) = \left(\frac{\partial\, \rho_N(\bar{r})}{\partial N}\right)^-_{v(\bar{r})} = \rho_N(\bar{r}) - \rho_{N-1}(\bar{r}) \tag{4}$$

---

* Corresponding author e-mail: pgeerlin@vub.ac.be.

† Vrije Universiteit Brussel.

‡ McMaster University.

10.1021/ct800027e CCC: $40.75

The Fukui function from above, f^+, represents the best way to change the electron density in response to an increase in the number of electrons; consequently, it can be employed to predict the preferred site for a nucleophilic attack. The Fukui function from below, f^-, represents the best way to change the electron density in response to a decrease in the number of electrons; consequently, it can be employed to predict the preferred site for an electrophilic attack.

Since practical DFT calculations generally use approximate exchange-correlation functionals with a N-electron self-interaction error,[22] eqs 3 and 4 are not exact anymore. Although the finite difference expressions are only approximate[19,20,23,24] for those situations, they are conventionally used as the reference method for the calculation of Fukui functions. This is justified by the fact that approximate exchange-correlation functionals are usually more accurate for systems with integer numbers of electrons than they are for systems with fractional electron numbers. For example, the ionization potentials computed by the finite difference approximation are more accurate than those computed from the highest occupied molecular orbital (HOMO) energy.[25]

Examination of the higher-order energy derivatives can lead to new descriptors of chemical importance.[26] Morell et al. introduced the dual descriptor[27–30] this way:

$$f^{(2)}(\bar{r}) = \frac{\partial^3 E}{\partial^2 N \delta\, v(\bar{r})} \tag{5}$$

The actual significance of this new descriptor becomes clear by incorporating the definition of the Fukui function:

$$f^{(2)}(\bar{r}) = \left(\frac{\partial f(\bar{r})}{\partial N}\right)_{v(\bar{r})} \cong f^+(\bar{r}) - f^-(\bar{r}) \tag{6}$$

Within this finite difference approximation, the dual descriptor can be interpreted as the difference between the Fukui functions for nucleophilic and electrophilic attacks. The dual descriptor thus gives a combination of both Fukui functions: it is positive for locations where a nucleophilic attack is more probable than an electrophilic attack and negative where electrophilic attack is more probable.

The theoretical basis for the dual descriptor is explained in ref 28, as is its application to various molecules. Its usefulness has been shown by the determination of the regioselectivities in the electrophilic aromatic substitutions on monosubstituted benzenes[27] and other organic reaction types.[29–31]

In this article, a recent approach for the calculation of Fukui functions, presented by some of the authors,[32,33] is further evaluated and extended for the dual descriptor. The Fukui function is calculated as the functional derivative of the electronic chemical potential to the external potential:

$$f(\bar{r}) = \left(\frac{\delta\mu}{\delta\, v(\bar{r})}\right)_N \tag{7}$$

This alternative definition can easily be obtained from eq 2 using a Maxwell relation. The derivative can then be calculated on the basis of the responses of the electronic chemical potential (μ) to a large number of perturbations in the external potential modeled by randomly placed point charges.

As the dual descriptor can be expressed in terms of the functional derivative of the hardness to the external potential, an analogous scheme can be developed for its calculation.

2. Methodology

2.1. Functional Derivatives. When the derivative discontinuity is taken into account, the following definitions for the Fukui functions can be rewritten as

$$f_N^+ = \left(\frac{\delta\mu_N^+}{\delta\, v(\bar{r})}\right)_N \tag{8}$$

$$f_N^- = \left(\frac{\delta\mu_N^-}{\delta\, v(\bar{r})}\right)_N \tag{9}$$

Similar to formulas 3 and 4, the finite difference approximation seems inevitable for the evaluation of μ^+ and μ^- if an approximate exchange-correlation functional is used. In this case, the proposed formulas 8 and 9 would give the same result as eqs 3 and 4. However, because the exchange-correlation potentials associated with the local density approximation (LDA) and generalized gradient approximations (GGAs)[34,35] do not have a "derivative discontinuity", the orbital energies are exactly equal to the chemical potentials in those approximations.[36–38]

$$\mu^+ = \varepsilon_{\mathrm{LUMO}} \tag{10}$$

$$\mu^- = \varepsilon_{\mathrm{HOMO}} \tag{11}$$

Here, $\varepsilon_{\mathrm{LUMO}}$ denotes the lowest unoccupied molecular orbital (LUMO) energy and $\varepsilon_{\mathrm{HOMO}}$ the HOMO energy.

Incorporation of these conclusions yields the following expressions for the Fukui functions:

$$f_N^+(\bar{r}) = \left(\frac{\delta\, \varepsilon_{\mathrm{LUMO}}}{\delta\, v(\bar{r})}\right)_N \tag{12}$$

$$f_N^-(\bar{r}) = \left(\frac{\delta\, \varepsilon_{\mathrm{HOMO}}}{\delta\, v(\bar{r})}\right)_N \tag{13}$$

These new formulas are exact only for LDA- and GGA-type functionals.

Extending this for the dual descriptor, one finds that

$$f^{(2)}(\bar{r}) = 2\left(\frac{\delta\eta}{\delta\, v(\bar{r})}\right) \cong \left(\frac{\delta(\mu^+ - \mu^-)}{\delta\, v(\bar{r})}\right)_N = \left\{\frac{\delta(\,\varepsilon_{\mathrm{LUMO}} - \varepsilon_{\mathrm{HOMO}})}{\delta\, v(\bar{r})}\right\}_N \tag{14}$$

with η being the hardness, defined as the second derivative of the energy to the number of electrons.

2.2. Theoretical Method. The functional derivatives were calculated using the scheme described in ref 32. This technique can be used for evaluating the functional derivative of any property Q with respect to the external potential:

$$\left(\frac{\delta\, Q[v(r)]}{\delta\, v(\bar{r})}\right)_N \tag{15}$$

When the definition of a functional derivative is used, the responses of $Q[v(\bar{r})]$ to P perturbations $\{w_i(\bar{r})\}_{i=1}^{P}$ in the external potential can be written as

$$Q[v+w_i]-Q[v]=\int \left(\frac{\delta\, Q[v]}{\delta\, v(\bar{r})}\right)_N w_i(\bar{r})\, \mathrm{d}\bar{r} \quad \text{with } i\in \{1,\ 2,\ ...,\ P\} \tag{16}$$

This equation is only valid in the linear response regime, that is, only for small perturbations.

Expansion of the functional derivative in a basis set

$$\left(\frac{\delta\, Q[v(\bar{r})]}{\delta\, v(\bar{r})}\right)_N=\sum_{j=1}^{K} q_j\, \beta_j(\bar{r}) \tag{17}$$

gives

$$Q[v+w_i]-Q[v]=\sum_{j=1}^{K} q_j \int w_i(\bar{r})\, \beta_j(\bar{r})\ \mathrm{d}\bar{r} \quad \text{with } i\in \{1,\ 2,\ ...,P\} \tag{18}$$

This set of simultaneous linear equations for the coefficients of expansion can be rewritten as

$$\mathbf{d}=\mathbf{B}\mathbf{q} \tag{19}$$

where **B** is a $K \times P$ matrix with elements $B_{ij} = \int w_i(\bar{r})\, \beta_j(\bar{r})\, \mathrm{d}\bar{r}$, **d** is a P-dimensional column matrix with elements $d_i = Q[v + w_i] - Q[v]$, and **q** is a K-dimensional column matrix with elements q_i.

Equation 19 can be solved once $P > K$ using a linear least-squares fitting. In practice, P needs to be much larger than K to ensure convergence. If desired, the normalization of the calculated functional derivative can be imposed by adding an extra equation to the set.

2.3. Implementation. The external potential is perturbed using point charges which are randomly distributed around the molecule. The response of the HOMO and LUMO energy determines the matrix **d**. **B** is analytically computed from the basis set and the point charges. Equation 19 is then solved for **q**. Finally, substitution of the resulting expansion coefficients into eq 17 yields the desired functional derivative. In the case of the calculation of the Fukui functions, Q should be taken as $\varepsilon_{\mathrm{LUMO}}$ or $\varepsilon_{\mathrm{HOMO}}$ (see eqs 12 and 13).

The calculation of the dual descriptor occurs in a similar way, though the responses of the hardness with respect to variations in the external potential are needed. Using the "finite difference" approximation for the explicit calculation of the hardness, one finds

$$f^{(2)}(\bar{r})=2\left(\frac{\delta\eta}{\delta\, v(\bar{r})}\right)_N \cong \left(\frac{\delta\,(\mu^+-\mu^-)}{\delta\, v(\bar{r})}\right)_N \tag{20}$$

For LDA and GGA functionals, this is equivalent to

$$\begin{aligned} f^{(2)}(\bar{r}) &= \left\{\frac{\delta(\,\varepsilon_{\mathrm{LUMO}}-\varepsilon_{\mathrm{HOMO}})}{\delta\, v(\bar{r})}\right\}_N \\ &= \left\{\left[\frac{\delta\, \varepsilon_{\mathrm{LUMO}}}{\delta\, v(\bar{r})}\right]_N-\left[\frac{\delta\, \varepsilon_{\mathrm{HOMO}}}{\delta\, v(\bar{r})}\right]_N\right\} \\ &= f^{+}(\bar{r})-f^{-}(\bar{r}) \end{aligned} \tag{21}$$

The proposed calculation scheme for $f^{(2)}(\bar{r})$ is thus equal to the one in eq 6, except that the used Fukui functions are obtained using our alternative method.

3. Computational Details

All molecular structures were optimized at the B3LYP/6-311+G(d,f)[39] level of theory. Calculations of the energy and orbital energies were carried out using the Perdew−Becke−Enrzerhof (PBE) exchange-correlation functional and the 6-311++G(d) basis set.

All calculations were performed with Gaussian 03[40] in combination with the program described in refs 32 and 33. Contour plots were generated using Mathematica,[41] while isosurfaces are constructed with Gaussview.[42]

4. Results and Discussion

In the following section, we examine the reactivity of different molecules using local Fukui functions and the local dual descriptor computed with the perturbational method described above. We will first test its implementation, concentrating on local reactivity indices for the molecules CH_2O, SCN^-, and CO. Having established the reliability of the method, an elaborate set of more complex molecules will be investigated, focusing mainly on the results for the local dual descriptor.

4.1. Determination of Computational Parameters. An important issue when using the aforementioned perturbational technique for the calculation of Fukui functions is the determination of the optimal working parameters of the program, like the distribution, the size and the number of the point charges, and the option to enforce the normalization of the functions.

Prior to the determination of these parameters, we had to decide which basis set to use for the expansion of the functional derivatives, compare eq 17. In order to get an optimal accuracy, this basis set should be as flexible as possible and should also incorporate the decay of the density and its derivatives, which is twice as fast as the one for wave functions. These arguments lead to the choice of the Ahlrichs Coulomb Fitting auxiliary basis set,[43,44] which is a density-based basis set, typically used for the accelerated calculation of Coulomb integrals in DFT calculations.

For our purposes, the incorporation of only s and p functions is sufficient. The contraction coefficients are neglected, which makes the set more flexible.

The first parameter to be considered is the positioning of the point charges. Each of the point charges is randomly distributed around an atom in a volume defined by two concentric spheres with radii $R_{\min}$ and $R_{\max}$, which are expressed in units of the van der Waals radius of the atom under consideration. As the functional derivative can only be accurate in regions where sampling has occurred, a large sampling region is required for the local calculation of functional derivatives. Consequently, we chose to place the point charges between 0.01 and 2.0 units of van der Waals radii of the composing atoms. This region, for example, corresponds to a sampling between distances of 0.02 and 4.43 Å from the nucleus for the hydrogen atom.

The magnitude of the point charges determines the size of the change of the orbital energies. Because we are restricted to the linear response region for the change in orbital energy, the determination of this parameter follows from a careful consideration between producing significant changes and remaining in the linear response region.

A comparison of a series of plots for the formaldehyde molecule for different acceptable point charge sizes showed

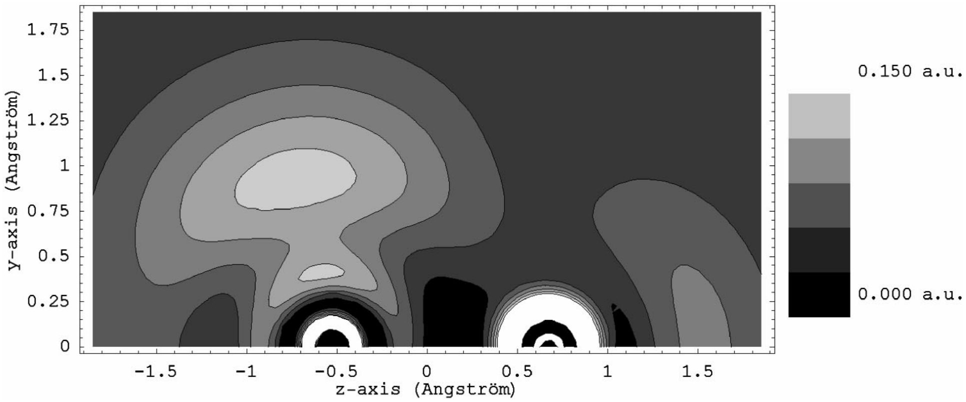


Figure 1. $f(+)$ contourplot of the carbonyl group in formaldehyde in the plane perpendicular to the molecular plane. Coordinates: C at −0.52 Å, O at 0.67 Å on the *z* axis.

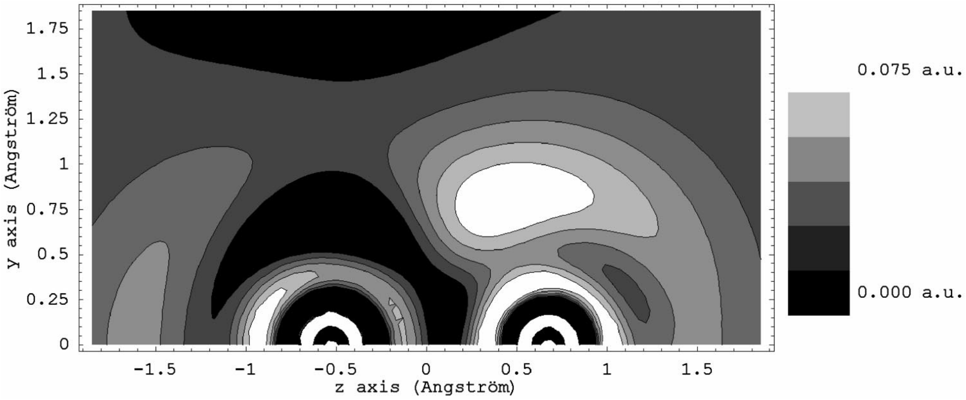


Figure 2. $f(-)$ contourplot of the carbonyl group in formaldehyde in the plane perpendicular to the molecular plane. Coordinates: C at −0.52 Å, O at 0.67 Å on the *z* axis.

that the most reliable results were generated with point charges of 0.005 *e* or 0.010 *e*. Because of this, we chose to use fixed point charges of 0.005 *e*, which are automatically doubled if the orbital energy response is too small to be significant.

Finally, the only remaining parameter to be determined is the number of point charges we use. Since each placed point charge implies a single-point energy calculation, it is necessary to balance between an acceptable CPU time and a reliable convergence of the least-squares fitting. On the basis of the inspection of the molecular symmetry in the $f^+(r)$ contour plots for the molecular plane of CH_2O (i.e., does the local function show the correct molecular symmetry), we decided to use 400 point charges per atom. With the chosen set of working parameters and by adding a normalization equation with a weight of 1, normalizations turn out to be good and usually only differ from the exact values by a factor of 10^{-2} or even 10^{-3}.

4.2. Fukui Function Calculations of Small Molecules. Before turning to the dual descriptor, we will perform an analysis of CH_2O, SCN^-, and CO using Fukui functions with contour and isosurface plots, computed as described above.

Formaldehyde is a typical reference molecule for testing Fukui functions because of its reactive carbonyl group: the carbon atom is known to undergo nucleophilic attack, while the oxygen atom is the preferred site for an electrophilic attack. When this is taken into account, $f^+(\bar{r})$ should have large values around the carbon atom, while $f^-(\bar{r})$ should be large close to the oxygen atom. Figures 1 and 2, obtained with the perturbational method, depict contour plots of these functions for the carbonyl group in the plane perpendicular to the molecular plane.

As can be seen, the contour plots are marked by high Fukui function values in the neighborhood of the carbon atom for $f^+(\bar{r})$ and high $f^-(\bar{r})$ values for the oxygen atom, in perfect agreement with experimental reactivity and finite difference results.[24] Both plots reveal spurious fluctuations of the Fukui functions near the nucleus. This effect is a consequence of the limited sampling in the direct surroundings of the nuclei. Elaborate sampling in this region solves the problem but requires a significant increase of CPU time without generating any reactivity-related information.

Passing to carbon monoxide (CO), as an alternative to contour plots, isosurfaces are used. In view of the extensiveness of the isosurface in the direct environment of the carbon atom, this atom is predicted to be the most reactive for nucleophilic and electrophilic attacks. Nonetheless, both surfaces in Figure 3 have a distinct shape, indicating that the orientation of an incoming nucleophile or electrophile at the initial stage of the reaction might be different.

As a third case, we investigated the reactivity of the ambident nucleophile SCN^-, which possesses two reactive sites for an electrophilic attack.[45] In this situation, the hardness of the reactive site plays—via the HSAB principle—an important role in the selectivity.

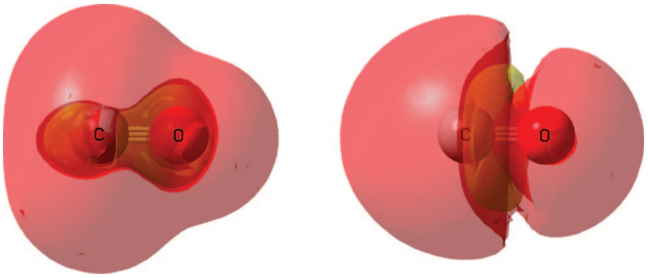


Figure 3. Isosurface (0.01 au) of *f*(+) and *f*(−) for CO.

It is well-known that, for anions, GGA functionals yield positive or very small HOMO energies, which exhibit considerable basis set dependence.[46] For that reason, μ^- was estimated using vertical ionization energies (eq 22), implying an extra single-point calculation of the $N - 1$ system for each perturbation.

$$\mu^- = \left(\frac{\partial E}{\partial N}\right)^-_{v(\bar{r})} = E_N - E_{N-1} \qquad (22)$$

In Figure 4, we show the contourplot of $f^-(\bar{r})$ for the molecular plane of SCN^-. The results seem similar to the ones obtained in ref 47: the Fukui function clearly demonstrates that the softest reaction site is the sulfur atom. Hard electrophiles will prefer to interact with the nitrogen atom. These results are in agreement with experimental data,[45] as are the ones for two other typical ambident nucleophiles, CH_2CHO^- and NO_2^-, for which data can be found in the Supporting Information.

4.3. Application of Dual Descriptor to Small Molecules. The excellent quality of the Fukui functions using the perturbational approach gives the necessary confidence for the application of the new methodology in the computation of local descriptors. Hence, the local dual descriptor will be calculated for the molecules CH_2O and CO. The ambident nucleophile, SCN^-, will be left out of consideration, because its reactivity is almost entirely determined by describing its behavior toward electrophilic attacks, which is best characterized by f^- (cf. previous section).

Results of $f^{(2)}$ for formaldehyde are displayed in Figure 5: areas with $f^{(2)}(\bar{r}) > 0$ are susceptible to undergoing a nucleophilic attack, and areas with $f^{(2)}(\bar{r}) < 0$ are more likely to accept an electrophilic attack. As the Fukui functions (Figures 1 and 2), $f^{(2)}(\bar{r})$ generates the correct reactivity for the carbonyl of formaldehyde with positive values near the carbon atom and negative values close to the oxygen atom.

An interesting advantage of the dual descriptor is its ability to smoothen both Fukui functions involved. This effect also diminishes the spurious oscillations around the nuclear positions.

For CO, $f^{(2)}(\bar{r})$ sketches an image analogous to the Fukui functions (Figure 3), as the negative values of the descriptor (electrophilic attack) are mainly situated at the carbon terminus of the CO bond and positive values (nucleophilic attack) are located near the carbon atom in a tire-shaped way, positioned perpendicular to the bonding axis.

The preceding results can be seen as a generalization of the classical explanation for the bonding patterns observed in transition-metal−carbonyl compounds. In these systems, an interaction occurs between the (electrophilic) metal ion and the electrons in the CO HOMO orbital, which is mainly localized on the carbon atom, resulting in a bonding σ orbital. Further stabilization is then obtained through a π interaction between the electrons in the d orbitals of the metal (the nucleophile) and the LUMO of CO.[48]

4.4. Regioselectivity in Electrophilic Aromatic Substitution. Because of the remarkable selectivity difference between ortho/para and meta directors in electrophilic aromatic substitution (EAS) reactions,[49] which has already been studied in previous conceptual DFT work,[50,51] this reaction provides interesting test cases for the new perturbational method.

Since $f^{(2)}(\bar{r})$ is a combination of the Fukui functions for electrophilic and nucleophilic attack, it should give enough relevant information to elucidate the reactivity. In order to facilitate the interpretation of the ring reactivity, no fitting basis functions were placed on the hydrogen atoms.

First, phenol and aniline were considered. Their OH and NH_2 groups are electron donors that promote substitutions on ortho and para positions in an EAS.

Figure 7 depicts the results for $f^{(2)}(\bar{r})$; the yellow regions correspond to zones with a negative dual descriptor which are susceptible to an electrophilic attack. These regions are the most pronounced for the para positions, whereas the reactivity of the ortho positions, which are expected to be susceptible to undergoing electrophilic attacks too, is less clear. Therefore, the same isosurfaces are shown from another angle in Figure 8. In these figures, one can see small nested zones at the ortho positions which are sensible for an electrophilic attack.

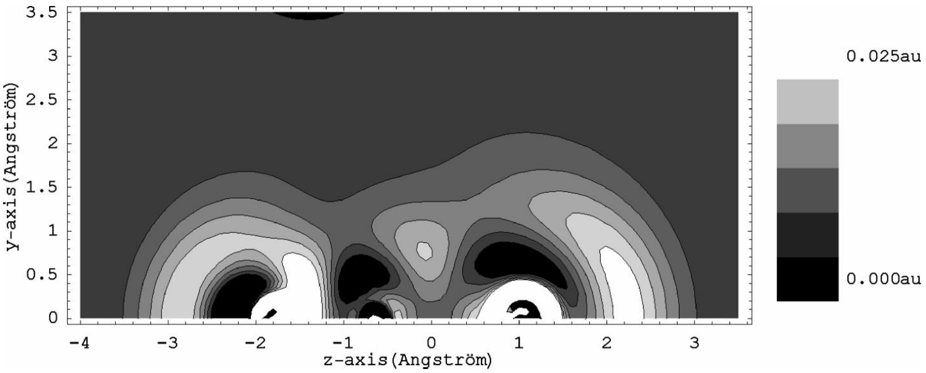


Figure 4. *f*(−) contourplot of SCN^-. Coordinates of the atoms in Å: N at −1.81, C at −0.63, and S at 1.03 on the *z*-axis.

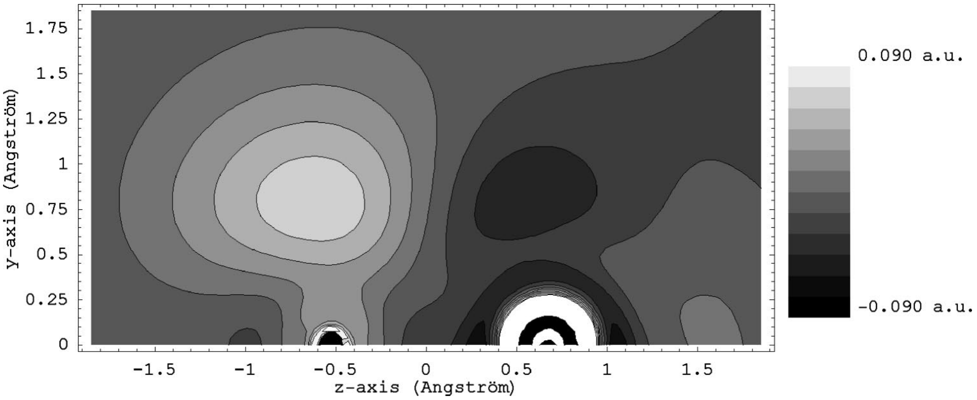


Figure 5. $f^{(2)}$ contourplot of the carbonyl group in formaldehyde in the plane perpendicular to the molecular plane. Coordinates: C at −0.52 Å, O at 0.67 Å on the *z* axis.

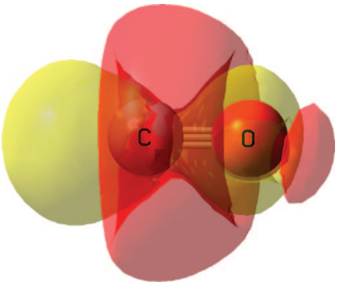


Figure 6. Isosurface of $f^{(2)}$ (0.02 au, positive regions are red, negative regions are yellow) for CO.

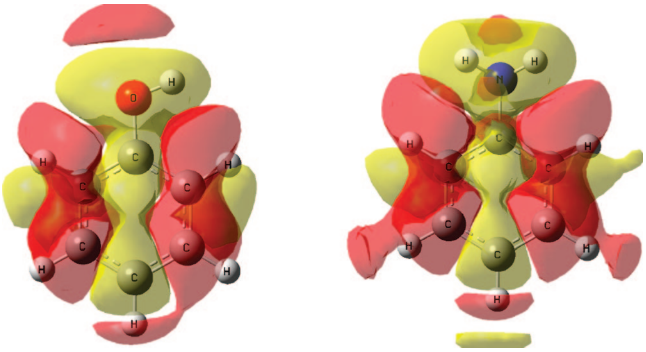


Figure 7. View of the isosurface of $f^{(2)}$ (0.002 au, positive regions are red, negative regions are yellow) for the molecular plane of phenol and aniline.

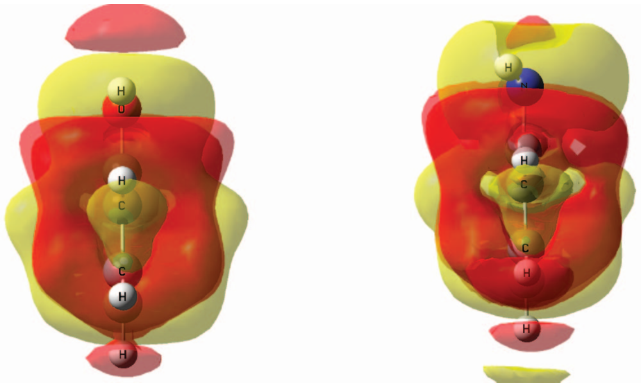


Figure 8. View of the isosurface of $f^{(2)}$ (0.002 au, positive regions are red, negative regions are yellow) for the plane perpendicular to the molecular plane of phenol and aniline.

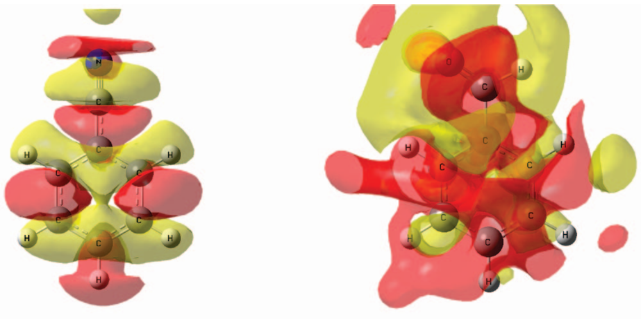


Figure 9. View of the isosurface of $f^{(2)}$ (0.002 au, positive red, negative yellow) for the molecular plane of cyanobenzene and benzaldehyde.

Next, two deactivating, meta-directing cases, benzaldehyde and cyanobenzene, were calculated.

As can be seen in Figure 9, an important reactivity switch occurred in comparison with the ortho/para directors. For cyanobenzene, the para position is clearly deactivated for an electrophilic attack, while the meta position now exhibits a yellow region. Quite surprisingly, the bond between the ipso and ortho positions is surrounded by a yellow surface. An inspection of the finite difference $f^{(2)}(\bar{r})$ at the same level of theory (PBE/6-311++G(d,f)) reveals good agreement with the results of the perturbational method. Deviations of our results from the previously obtained finite difference results[27] at the Hartree−Fock (HF/6-311G(d,f)) level, which do not predict an ipso reactivity, could thus be ascribed to the use of the PBE functional in the DFT calculations.[52,53]

Results for benzaldehyde pose another difficulty because no specific reactivity pattern can be recognized in the isosurface of Figure 9, whereas the finite difference dual descriptor (PBE/6-311++G(d)) is able to predict a more realistic reactivity. In search of possible explanations, the composing Fukui functions were examined. Primarily, $f^-(\bar{r})$ exhibited strange behavior with a reactivity concentrated on the substituent, leaving the ring basically inactive for an EAS. As previously explained, $f^-(\bar{r})$ is constructed from the responses of the HOMO energy. A plot of the HOMO density reveals that this orbital is predominantly located on the substituent and is thus unable to predict the correct ring reactivity. Since the two lower-lying orbitals are located on the aromatic ring, we used these orbitals together with the HOMO in order to obtain an adjusted version of $f^-(\bar{r})$:

$$f^{-}(\bar{r}) = \left(\frac{\delta\,(\varepsilon_{\mathrm{HOMO}} + \varepsilon_{\mathrm{HOMO-1}} + \varepsilon_{\mathrm{HOMO-2}})/3}{\delta\, v(\bar{r})} \right)_N \qquad (23)$$

Figure 10 shows that the adjusted $f^{-}(\bar{r})$ (eq 23) improves the reactivity picture. Meta positions exhibit yellow regions; nonetheless, the rest of the ring, mainly the ortho−ipso bond, presents similar behavior. The improved results obtained with formula 23 suggest the occurrence of orbital mixing during the different perturbations: when several orbitals are close in energy, different perturbations will access different orbitals.[54–56] Because the inclusion of the three highest orbitals does not provide a perfect prediction of the chemical reactivity yet, it may be necessary to include additional orbitals in eq 23. It is also possible that the present method—which models an infinitesimal change in the number of electrons—does not adequately predict the orbital relaxation contribution to the Fukui function.[50,57–60] (It is known, for example, that DFT functionals are more accurate for an integer electron number than they are for a fractional electron number.[61]) So, it is also possible that a more-accurate model of orbital relaxation effects would bring the picture in Figure 10 into closer agreement with the experimentally observed reactivity preferences.

In addition, the Supporting Information contains a number of other molecules like propene, propenal, ethanal, propanon, and chlorobenzene that were tested with the perturbational method. In all of the cases, the dominant reactivity for electrophilic addition, nucleophilic addition, and electrophilic aromatic substitution was recovered with regard to experimental results.

5. Conclusions

We have given a validation of and extension to the perturbational method[32,33] for the calculation of DFT-based reactivity descriptors by computing local Fukui functions and the dual descriptor. The method is based on a Maxwell relation to circumvent the need to take the derivatives with respect to the number of electrons and uses functional derivatives with respect to the external potential instead.

The results confirm the usefulness of the dual descriptor as a new reactivity index for electrophilic and nucleophilic attacks. Moreover, the convincing results for both the Fukui functions ($(\delta\mu/\delta v(\bar{r}))_N$) and the dual descriptor ($(\delta\eta/\delta v(\bar{r}))_N$) also indicate the general validity of the perturbational approach, making it an interesting methodology for the calculation of functional derivatives to the external potential of any global property.

As such, the long-lasting problem concerning DFT-based reactivity descriptors involving derivatives with respect to the number of electrons can successfully be circumvented by switching to expressions with functional derivatives to the external potential, obtained using Maxwell relations.

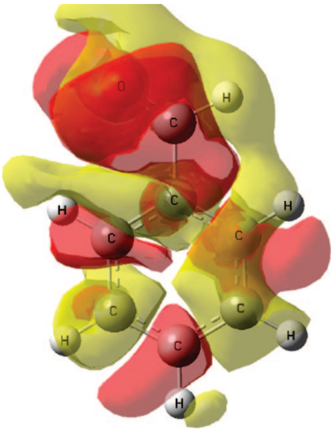


Figure 10. View of isosurface of $f^{(2)}$ (0.002 au, positive regions are red, negative regions are yellow) for the molecular plane of benzaldehyde using adjusted formula 23.

Acknowledgment. T.F. and N.S. thank the Fund for Scientific Research Flanders (FWO) for a predoctoral (Aspirant) fellowship. P.G and F.D.F thank the Fund for Scientific Research Flanders (FWO) and the VUB for their continuous support. P.W.A. would like to thank NSERC, the Canada Research Chairs, and SHARCNET for research support.

Supporting Information Available: A PDF file containing contourplots of $f^{-}(\bar{r})$ for ambident nucleophiles NO_2^- and CH_2CHO^- and isosurfaces of $f^{(2)}(\bar{r})$ for propene, propanon, ethanal, and chlorobenzene is included. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hohenberg, P.; Kohn, W. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1964**, *136*, 864–871.

(2) Kohn, W.; Sham, L. J. *Phys. Rev. A: At., Mol., Opt. Phys.* **1965**, *140*, 1133–1138.

(3) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974–12980.

(4) Kohn, W. *Rev. Mod. Phys.* **1999**, *71*, 1253–1266.

(5) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(6) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. *J. Chem. Phys.* **2005**, *123*, 062201.

(7) Chermette, H. *J. Comput. Chem.* **1999**, *20*, 129–154.

(8) Geerlings, P.; De Proft, F.; Langenaeker, W. *Chem. Rev.* **2003**, *103*, 1793–1874.

(9) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989; p 333.

(10) Ayers, P. W.; Anderson, J. S. M.; Bartolotti, L. J. *Int. J. Quantum Chem.* **2005**, *101*, 520–534.

(11) Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. *J. Chem. Phys.* **1978**, *68*, 3801–3807.

(12) Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516.

(13) Yang, W.; Parr, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6723–6726.

(14) Nalewajski, R. F.; Parr, R. G. *J. Chem. Phys.* **1982**, *77*, 399–407.

(15) Parr, R. G.; Yang, W. *J. Am. Chem. Soc.* **1984**, *106*, 4049–4050.

(16) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L. *Phys. Rev. Lett.* **1982**, *49*, 1691–1694.

(17) Zhang, Y.; Yang, W. *Theor. Chem. Acc.* **2000**, *103*, 346–348.

(18) Yang, W.; Zhang, Y.; Ayers, P. W. *Phys. Rev. Lett.* **2000**, *4*, 5172–5175.

(19) Ayers, P. W.; Levy, M. *Theor. Chem. Acc.* **2000**, *103*, 353–360.

(20) Ayers, P. W.; Parr, R. G. *J. Am. Chem. Soc.* **2000**, *122*, 2010–2018.

(21) Ayers, P. W. *J. Math. Chem.* **2008**, *43*, 285–303.

(22) Mori-Sanchez, P.; Cohen, A. J.; Yang, W. T. *J. Chem. Phys.* **2006**, *125*, 201102.

(23) Michalak, A.; De Proft, F.; Geerlings, P.; Nalewajski, R. F. *J. Phys. Chem. A* **1999**, *103*, 762–771.

(24) Geerlings, P.; De Proft, F.; Langenaeker, W. *Chem. Rev.* **2003**, *103*, 1808.

(25) Tozer, D. J.; Handy, C. H. *Mol. Phys.* **2003**, *101*, 2669–2675.

(26) Geerlings, P.; De Proft, F. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3028–3042.

(27) Morell, C.; Grand, A.; Toro-Labbé, A. *J. Phys. Chem. A* **2005**, *109*, 205–212.

(28) Morell, C.; Grand, A.; Toro-Labbé, A. *Chem. Phys. Lett.* **2006**, *425*, 342–346.

(29) Morale, C. Un nouveau descripteur de la reactivité locale, Ph.D. Thesis, Université de Grenoble, Grenoble, France, 2006; p 133.

(30) Morell, C.; Grand, A.; Gutiérrez-Oliva, S.; Toro-Labbé, A. Using the reactivity-selectivity descriptor $\Delta f(\bar{r})$ in organic chemistry. In *Theoretical Aspects of Chemical Reactivity*; Toro-Labbé, A., Ed.; Elsevier: Amsterdam, The Netherlands, 2006; pp 101−118.

(31) Ayers, P. W.; Morell, C.; De Proft, F.; Geerlings, P. *Chem.*−*Eur. J.* **2007**, *13*, 8240–8247.

(32) Ayers, P. W.; De Proft, F.; Borgoo, A.; Geerlings, P. *J. Chem. Phys.* **2007**, *126*, 224107.

(33) Sablon, N.; Ayers, P. W.; De Proft, F.; Geerlings, P. *J. Chem. Phys.* **2007**, *126*, 224108.

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

(35) Perdew, J. P.; Burke, M.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

(36) Perdew, J. P.; Levy, M. *Phys. Rev. Lett.* **1983**, *51*, 1884–1887.

(37) Sham, L. J.; Schluter, M. *Phys. Rev. Lett.* **1983**, *51*, 1888–1891.

(38) Perdew, J. P.; Levy, M. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *56*, 16021–16028.

(39) Krishnan, R.; Frisch, M. J.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 4244–4245.

(40) Frisch, M. J. ; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision D.01; Gaussian, Inc.: Wallingford, CT, 2004.

(41) *Mathematica*, Version 5.2; Wolfram Research, Inc.: Champaign, IL, 2005.

(42) *GaussView*, Version 3.09; Dennington, R., II, Keith, T., Millam, J., Eppinnett, K., Hovell, W. L., Gilliland, R., Eds.; Semichem, Inc.: Shawnee Mission, KS, 2003.

(43) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283–289.

(44) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119–124.

(45) Hudson, R. F. Nucleophilic Reactivity. In *Chemical Reactivity and Reaction Paths*; Klopman, G., Ed.; Wiley: New York, 1974; pp 167−252.

(46) Simons, J.; Jordan, K. D. *Chem. Rev.* **1987**, *87*, 535–555.

(47) Langenaeker, W.; De Proft, F.; Geerlings, P. *THEOCHEM* **1996**, *362*, 175–179.

(48) Atkins, P.; Overton, T.; Rourke, J.; Weller, M.; Amstrong, F. *Inorganic Chemistry*, 4th ed.; Oxford University Press: Oxford, United Kingdom, 2006; pp. 59.

(49) Clayden, J.; Greeves, N.; Warren, S.; Wothers, P. *Organic Chemistry*; Oxford University Press: Oxford, United Kingdom, 2001; pp 547−579.

(50) Langenaeker, W.; De Proft, F.; Geerlings, P. *J. Phys. Chem.* **1995**, *99*, 6424–6431.

(51) De Proft, F.; Van Alsenoy, C.; Peeters, A.; Langenaeker, W.; Geerlings, P. *J. Comput. Chem.* **2002**, *23*, 1198–1209.

(52) Langenaeker, W.; Demel, K.; Geerlings, P. *THEOCHEM* **1991**, *80*, 329–342.

(53) Anderson, J. S. M.; Melin, J.; Ayers, P. W. *J. Chem. Theory Comput.* **2007**, *3*, 375–389.

(54) da Silva, R. R.; Ramalho, T. C.; Santos, J. M.; Figueroa-Villar, J. D. *J. Phys. Chem. A* **2006**, *110*, 1031–1040.

(55) Maksic, Z. B.; Vianello, R. *J. Phys. Chem. A* **2006**, *110*, 10651–10652.

(56) da Silva, R. R.; Ramalho, T. C.; Santos, J. M.; Figueroa-Villar, J. D. *J. Phys. Chem. A* **2006**, *110*, 10653–10654.

(57) Yang, W.; Parr, R. G.; Pucci, R. *J. Chem. Phys.* **1984**, *81*, 2862–2863.

(58) Bartolotti, L. J.; Ayers, P. W. *J. Phys. Chem. A* **2005**, *109*, 1146–1151.

(59) Ayers, P. W. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3387–3390.

(60) Melin, J.; Ayers, P. W.; Ortiz, J. V. *J. Phys. Chem. A* **2007**, *111*, 10017–10019.

(61) Zhang, Y.; Yang, W. *J. Chem. Phys.* **1998**, *109*, 2604–2608.

CT800027E

JCTC Journal of Chemical Theory and Computation

Computational Studies on the Mechanisms and Dynamics of OH Reactions with $CHF_2CHFOCF_3$ and $CHF_2CH_2OCF_3$

Lei Yang, Jing-yao Liu,* and Ze-sheng Li*

Institute of Theoretical Chemistry, State Key Laboratory of Theoretical and Computational Chemistry, Jilin University, Changchun 130023, P. R. China

Received January 30, 2008

Abstract: The dual-level direct dynamics method has been employed to investigate the multichannel hydrogen abstraction reactions of $CHF_2CHFOCF_3 + OH$ (R1) and $CHF_2CH_2OCF_3 + OH$ (R2) theoretically. The optimized geometries and frequencies of the stationary points and the minimum energy path are calculated at the B3LYP/6-311G(d,p) level, and the energetic information is further refined by the MC-QCISD method. There are two hydrogen abstraction channels for reaction R1 and three hydrogen abstraction channels for reaction R2. And four products, $CF_2CHFOCF_3$, CHF_2CFOCF_3, $CF_2CH_2OCF_3$, and CHF_2CHOCF_3, are produced during these two reactions. The enthalpies of formation for the species involved in the two reactions ($CHF_2CHFOCF_3$, $CF_2CHFOCF_3$, CHF_2CFOCF_3, $CHF_2CH_2OCF_3$, $CF_2CH_2OCF_3$, and CHF_2CHOCF_3) are calculated via isodesmic reaction. Furthermore, the reaction mechanisms of subsequent reactions of product radicals ($CF_2CHFOCF_3$, CHF_2CFOCF_3, $CF_2CH_2OCF_3$, and CHF_2CHOCF_3) with the OH radical are studied at the same level. The rate constants of reactions R1 and R2 are evaluated by means of canonical variational transition-state theory including the small-curvature tunneling correction over a wide range of temperatures, from 250 to 1200 K. Our results show that the tunneling correction plays an important role in the rate constant calculation in the low-temperature range. Agreement between the calculated and experimental data available at 298 K is good. It is found that the fluorine substitution decreases the reactivity of the C−H bond, and as a result, reaction R2 may proceed much easier than reaction R1. The Arrhenius expressions over a wide temperature range are obtained, and the kinetic isotope effects for reactions $CHF_2CHFOCF_3/CDF_2CDFOCF_3 + OH$ and $CHF_2CH_2OCF_3/CDF_2CD_2OCF_3 + OH$ are estimated so as to provide theoretical estimation for future laboratory investigation.

Introduction

Since chlorofluorocarbons (CFCs) cause the depletion of the stratospheric ozone layer as well as global warming, a large number of replacement compounds such as hydrochlorofluorocarbons (HCFCs) and hydrofluorocarbons (HFCs) have been selected for industrial applications. Unfortunately, some of them have considerable global warming potentials. Furthermore, HCFCs contain Cl atoms and thus contribute to ozone depletion. Finding the suitable industrial replacement of CFCs remains a challenge, because of the complex considerations of performance, safety, and environmental properties required. Recently, hydrofluoroethers (HFEs) have been suggested as substitutes for CFCs and have been used in various industrial applications,[1,2] since they contain neither chlorine nor bromine atoms and have no ozone depletion potential (ODP).[3] Moreover, the introduction of ether linkage −O− may lead to an even greater reactivity in the troposphere, and thus HFEs are considered to be promising substitutes of CFCs. Prior to its large-scale industrial use, it is necessary to assess its environmental impact by determining its atmospheric lifetime.

---

* Corresponding author. Fax: +86-431-88498026. E-mail: ljy121@mail.jlu.edu.cn (J.-y. L.); zeshengli@mail.jlu.edu.cn (Z.-s. L.).

10.1021/ct800032e CCC: $40.75

In general, the degradation of HFEs is initiated by the oxidation reaction with OH radicals in the atmosphere. Therefore, study of the reactivity of HFEs against OH radicals is important for the evaluation of their atmospheric lifetime. In the present work, our attention will focus on the reactions of molecules $CHF_2CHFOCF_3$ and $CHF_2CH_2OCF_3$ with OH radicals. As Ravishankara et al.[3] mentioned, chemical reactions involving CF_3O and CF_3OO lead to negligibly small ODPs, which means that the molecules containing CF_3 groups may be no more harmful to the stratospheric ozone layer. For the reaction of $CHF_2CHFOCF_3$ with OH, the hydrogen atom can be abstracted from α- and β-carbon atoms, so two H-abstraction channels are feasible.

$$CHF_2CHFOCF_3 + OH \rightarrow CF_2CHFOCF_3 + H_2O \tag{R1a}$$

$$CHF_2CHFOCF_3 + OH \rightarrow CHF_2CFOCF_3 + H_2O \tag{R1b}$$

Oyaro et al.[4] studied this reaction using the relative rate method, and the rate constant at 298 K was given as $(6.8 \pm 1.1) \times 10^{-15}$ cm^3 molecule^{-1} s^{-1}. However, for the reaction of $CHF_2CH_2OCF_3$ with OH, no experimental data are reported. Only Urata et al.[5] evaluated the rate constant by using the empirical estimation method to give a value of 3.27×10^{-14} cm^3 molecule^{-1} s^{-1} at 298 K. Since $CHF_2CH_2OCF_3$ has C_1 symmetry, the two hydrogen atoms in the CH_2 group are not equivalent, and as a result, three distinct channels are found, that is,

$$CHF_2CH_2OCF_3 + OH \rightarrow CF_2CH_2OCF_3 + H_2O \tag{R2a}$$

$$CHF_2CH_2OCF_3 + OH \rightarrow CF_2CHOCF_3 + H_2O \tag*{(R2b,R2b′)}$$

To the best of our knowledge, no other experimental or theoretical study is carried out for the two reactions. Furthermore, the kinetic isotope effects (KIEs) may play an important role in investigating the kinetic nature of the reactions, but there has been no experimental determination for the two reactions. To ascertain the environmental impact of these two molecules, theoretical studies on the mechanisms of reactions $CHF_2CHFOCF_3$/$CHF_2CH_2OCF_3$ + OH, the temperature dependence of the rate constants, and KIEs for reactions $CHF_2CHFOCF_3$/$CDF_2CDFOCF_3$ + OH and $CHF_2CH_2OCF_3$/$CDF_2CD_2OCF_3$ + OH are very desirable.

Here, a dual-level (X//Y) direct dynamics[6–8] method is employed to study the kinetic nature of reactions R1 and R2. In this approach, dynamics calculations are based directly on the results of electronic structure calculations. As usual, X//Y refers to optimization and frequencies at a low level of theory with single-point energies from a high level of theory. The rate constants for each reaction channel are evaluated using the variational transition state theory (VTST)[9–11] with an interpolated single-point energies (ISPE) approach.[12] The centrifugal-dominant small-curvature semiclassical adiabatic ground-state tunneling (CD-SCSAG) approximation[13,14] is included. Then, the total rate constants within 250−1200 K, the product branching ratios, and KIEs are obtained. The comparison of theoretical and literature results is made.

The radicals $CF_2CHFOCF_3$, CHF_2CFOCF_3, $CF_2CH_2OCF_3$, and CHF_2CHOCF_3 are the major products produced in reactions R1 and R2, and they may further degrade through their reactions with the OH radical. However, there is no report on the properties of these radicals or their further degradation processes in the atmosphere, which are important to assess their environmental impact. Thus, the other goal of the present work is to provide deep insight into the reaction mechanism of these product radicals with OH.

$$CF_2CHFOCF_3 + OH \rightarrow \text{products} \tag{R3}$$

$$CF_2CH_2OCF_3 + OH \rightarrow \text{products} \tag{R4}$$

$$CHF_2CFOCF_3 + OH \rightarrow \text{products} \tag{R5}$$

$$CHF_2CHOCF_3 + OH \rightarrow \text{products} \tag{R6}$$

Calculation Methods

In the present study, all of the electronic structure calculations are carried out with the Gaussian 03 program.[15] The optimized geometries and harmonic vibrational frequencies of all of the stationary points (reactants, transition states, complexes, and products) involved in reactions R1–R6 are calculated by using the B3LYP method (Becke's three-parameter nonlocal-exchange functional[16] with the nonlocal correlation functional of Lee, Yang, and Parr[17]) with the 6-311G(d,p) basis set. To quantify the level of trust in the method, the geometries and harmonic vibrational frequencies of reactants, products, and transition states for reaction R1 are also performed at the MP2 level of theory (restricted or unrestricted second-order Møller−Plesset perturbation theory) with the 6-311G(d,p) basis set as well as using the BB1K method (the Becke88−Becke95 one-parameter model[18]) using the 6-31+G(d,p) basis set.[19] At the B3LYP/6-311G(d,p) and BB1K/6-31+G(d,p) levels, the minimum energy paths (MEPs) are calculated by the intrinsic reaction coordinate (IRC) theory. Also, the energy derivatives, including gradients and Hessians at geometries along the MEP, are obtained. Single-point energy calculations for the stationary points and the extra energies along the MEP are carried out using the multicoefficient correlation method based on quadratic configuration interaction with single and double excitations (MC-QCISD).[20] Then, the energy profile is refined at the MC-QCISD//B3LYP/6-311G(d,p) and MC-QCISD//BB1K/6-31+G(d,p) levels.

The initial information on the potential energy surface is put into the POLYRATE 8.4.1 program,[21] and then the dynamic calculations for reactions R1 and R2 are calculated by using VTST[9–11] with the ISPE[12] method. The ISPE method is a dual-level direct dynamics scheme, in which a few extra energies calculated at a high level of theory along the low-level MEP are used to correct the classical energetic profile of the reaction. The rate constants are calculated using canonical variational transition state theory (CVT)[22] with the CD-SCSAG approximation proposed by Truhlar and coworkers.[13,14] Most of the vibrational modes are treated as quantum-mechanical separable harmonic oscillators. The hindered-rotor approximation of Truhlar and Chuang is used for calculating the partition functions of the lowest modes associated with the torsion. In the calculation of the electronic partition functions, the excited state of the OH radical, with a 140 cm^{-1} splitting, is included. The total rate constant of each reaction is obtained as the sum of the individual rate constant of each H-abstraction channel.

The enthalpies of formation of species are estimated by using isodesmic reactions.[23] Isodesmic reactions are those in which the number and type of bonds are conserved. Because of the electronic similarity between reactants and products, these reactions will cancel the systematic errors in the calculations and lead to accurate results. Here, we predict the enthalpies of formation for species $CHF_2CHFOCF_3$, $CF_2CHFOCF_3$, CHF_2CFOCF_3, $CHF_2CH_2OCF_3$, $CF_2CH_2OCF_3$, and CHF_2CHOCF_3 using the following isodesmic reactions:

$$CHF_2CHFOCF_3 + CH_4 \rightarrow CH_3OCH_3 + CF_3CF_3 \tag{R7}$$

$$CF_2CHFOCF_3 + CH_4 + CH_4 \rightarrow CH_3OCH_3 + CF_3 + CF_3CH_3 \tag{R8}$$

$$CHF_2CFOCF_3 + CH_4 + CH_4 \rightarrow CH_3OCH_3 + CF_3 + CF_3CH_3 \tag{R9}$$

$$CHF_2CH_2OCF_3 + CH_4 + CH_4 \rightarrow CH_3OCH_3 + CHF_2CH_3 + CHF_3 \tag{R10}$$

$$CF_2CH_2OCF_3 + CH_4 + CH_4 \rightarrow CH_3OCH_3 + CHF_2CH_3 + CF_3 \tag{R11}$$

$$CHF_2CHOCF_3 + CH_4 + CH_4 + CH_4 \rightarrow CH_3OCH_3 + CH_2F_2 + CHF_3 + CH_2CH_3 \tag{R12}$$

Results and Discussion

Electronic Structure Calculations. The optimized structural parameters of the stationary points involved in R1 and R2 calculated at the B3LYP/6-311G(d,p), BB1K/6-31+G(d,p), and MP2/6-311G(d,p) levels are displayed in Figure 1, along with the limited experimental data.[24,25] In addition, others involved in R3–R6 are given in the Supporting Information (Figure S1). As shown in Figure 1, when comparison is available, the optimized parameters at the B3LYP/6-311G(d,p), BB1K/6-31+G(d,p), and MP2/6-311G(d,p) levels are in reasonable agreement with each other. With respect to reaction R1, the reactant complexes are stabilized by the hydrogen-bond attractive interaction between O′ and H″ and between F′ and H′ (Figure 1), due to the high electronegativities of the fluorine and oxygen atoms. Also, the product complexes CP1a and CP1b with energies less than the products are located at the exits of channels R1a and R1b. For reaction R2, two reactant complexes (CR2a and CR2b) and three product complexes (CP2a, CP2b, and CP2b′) are identified at the reactant and product sides of the three channels, respectively. These indicate that each reaction channel may proceed via an indirect mechanism. With respect to the transition states TS1a, TS1b, TS2a, TS2b, and TS2b′, the breaking C−H bonds are stretched by 11.4, 12.0, 9.8, 10.5, and 11.4%, respectively, compared to the corresponding reactants, and the forming O−H bonds are longer than the equilibrium bond lengths in H_2O by 35.7, 35.1, 39.4, 34.5, and 32.7%, respectively, at the B3LYP/6-311G(d,p) level. The elongations of the forming bonds are greater than those of the breaking bonds, indicating that these transition states are all reactant-like, so H-abstraction reactions R1 and R2 may proceed via "early" transition states. The similar results can be obtained for reaction R1 at the BB1K/6-31+G(d,p) level.

The harmonic vibrational frequencies of the stationary points calculated at the B3LYP/6-311G(d,p), BB1K/6-31+G(d,p), and MP2/6-311G(d,p) levels as well as the available experimental data[24,26] are presented in Table S1 (Supporting Information). The vibrational frequencies are calculated to characterize the nature of each critical point and make zero-point energy (ZPE) correction. The number of imaginary frequencies (0 or 1) indicates whether a minimum or a transition state has been located. "Zero" imaginary frequencies correspond to a local minimum, while "one" imaginary frequency corresponds to transition structure. At the B3LYP/6-311G(d,p) level, the imaginary frequencies of TS1a, TS1b, TS2a, TS2b, and TS2b′ are 940*i*, 1009*i*, 610*i*, 790*i*, and 954*i* cm^{-1}, respectively.

Energetics. An accurate knowledge of the enthalpies of formation ($\Delta H^{\circ}_{f,298}$) for species is required for a thorough understanding of the kinetics and mechanisms of their reactions, particularly in atmospheric modeling. The $\Delta H^{\circ}_{f,298}$ values of $CHF_2CHFOCF_3$, $CF_2CHFOCF_3$, CHF_2CFOCF_3, $CHF_2CH_2OCF_3$, $CF_2CH_2OCF_3$, and CHF_2CHOCF_3 are obtained by using the calculated reaction enthalpies of the isodesmic reactions R7–R12 and the experimental $\Delta H^{\circ}_{f,298}$ of the other species involved in the reactions (CH_4, −17.89 kcal/mol;[27] CF_3CF_3, −321.2 kcal/mol;[27] CHF_3, −166.60 kcal/mol;[27] CF_3, −112.4 kcal/mol;[27] CH_2F_2, −107.71 kcal/mol;[27] CH_3OCH_3, −43.99 ± 0.12 kcal/mol;[28] CHF_2CH_3, −118.78 ± 0.95 kcal/mol;[29] CF_3CH_3, −178.94 kcal/mol;[30] CH_2CH_3, 28.4 ± 0.5 kcal/mol[31]). The enthalpies of formation calculated at the MC-QCISD//B3LYP/6-311G(d,p) level are −362.89 ± 0.12, −338.21 ± 0.12, −338.10 ± 0.12, −312.43 ± 1.07, −289.82 ± 1.07, and −265.36 ± 0.65 kcal/mol for $CHF_2CHFOCF_3$, $CF_2CHFOCF_3$, CHF_2CFOCF_3, $CHF_2CH_2OCF_3$, $CF_2CH_2OCF_3$, and CHF_2CHOCF_3, respectively, which are presented in Table 1. Note that the error limits are calculated by adding the maximum uncertainties of the $\Delta H^{\circ}_{f,298}$ values of reference species.

The reaction enthalpies (ΔH°_{298}) and potential barriers (ΔE) calculated at the MC-QCISD//B3LYP/6-311G(d,p), MC-QCISD//BB1K/6-31+G(d,p), and MC-QCISD//MP2/6-311G(d,p) levels are displayed in Table 2. As shown in Table 2, the ΔH°_{298} and ΔE values obtained at the MC-QCISD//B3LYP, MC-QCISD//BB1K, and MC-QCISD//MP2 levels show good mutual agreement. It is known that MP2 is expensive for such systems, so we employed B3LYP/6-311G(d,p) as the low level to get the rate constants of the two systems. To further verify the reliability of B3LYP, we also performed the dynamic calculations for reaction R1 at the MC-QCISD//BB1K/6-31+G(d,p) level.

Schematic potential energy surfaces obtained at the MC-QCISD//B3LYP/6-311G(d,p) level for reactions $CHF_2CHFOCF_3$ + OH (R1) and $CHF_2CH_2OCF_3$ + OH (R2) are

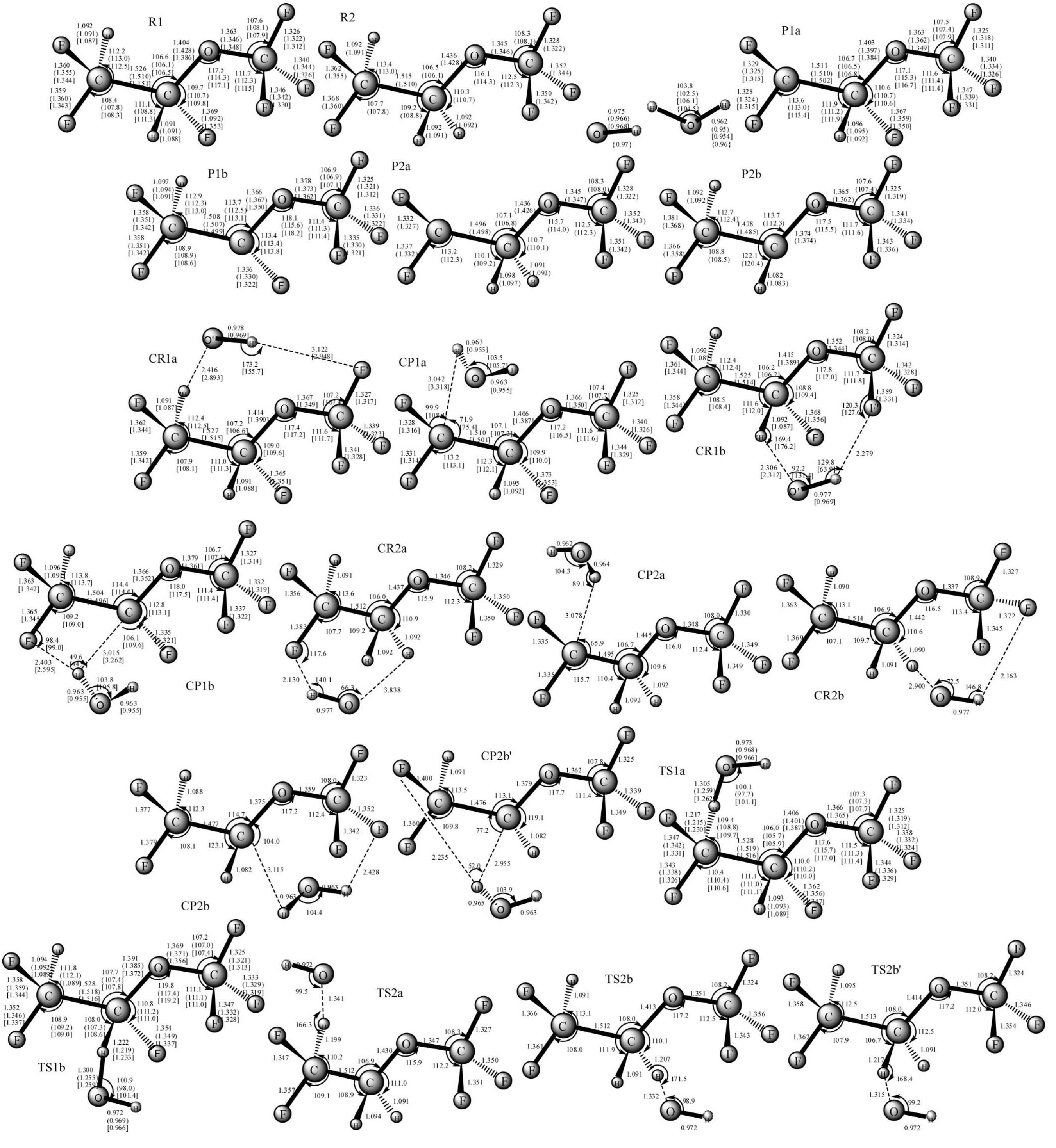


Figure 1. Optimized geometries of reactants, products, complexes, and transition states at the B3LYP/6-311G(d,p), MP2/6-311G(d,p) (in parentheses), and BB1K/6-31+G(d,p) (in square brackets) levels of reactions $CHF_2CH_2OCF_3$/$CHF_2CHFOCF_3$ + OH. The numbers in curly brackets are the experimental values.[24,25] Bond lengths are in angstroms, and angles are in degrees.

plotted in Figure 2a and b. The energy of reactant is set to zero for reference. Seen from Figure 2a, we can find that two hydrogen-bond complexes, CR1a (−2.54 kcal/mol) and CR1b (−0.57 kcal/mol), are located at the reactant side of channels R1a and R1b, respectively. And two complexes, CP1a and CP1b, with the energies being 0.52 and 0.57 kcal/mol lower than the products, are located at the product sides of R1a and R1b, respectively. The barrier height of channel R1a is 4.61 kcal/mol, which is only about 0.8 kcal/mol lower than that of channel R1b. Furthermore, channels R1a and R1b are exothermic at about −16.81 and −16.77 kcal/mol, which is almost the same. Thus, H abstractions from $-CHF_2$ and $-CH_2-$ groups may be kinetically and thermodynamically competitive channels. Furthermore, the relative energies of the transition states, complexes, and products obtained at the MC-QCISD//BB1K level are also displayed in Figure

Table 1. Enthalpies of Formation (kcal/mol) of the Species $CHF_2CHFOCF_3$, $CF_2CHFOCF_3$, CHF_2CFOCF_3, $CHF_2CH_2OCF_3$, $CF_2CH_2OCF_3$, and CHF_2CHOCF_3 Calculated at the MC-QCISD//B3LYP/6-311G(d,p) Level

species	$\Delta H^{\circ}_{f,298}$	species	$\Delta H^{\circ}_{f,298}$
$CHF_2CHFOCF_3$	−362.89 ± 0.12	$CHF_2CH_2OCF_3$	−312.43 ± 1.07
$CF_2CHFOCF_3$	−338.21 ± 0.12	$CF_2CH_2OCF_3$	−289.82 ± 1.07
CHF_2CFOCF_3	−338.10 ± 0.12	CHF_2CHOCF_3	−265.36 ± 0.62

2a. It is seen that the barrier heights obtained at the MC-QCISD//BB1K level are in good agreement with those obtained at the MC-QCISD//B3LYP level. At the B3LYP/6-311G(d,p) level, two hydrogen-bond complexes (CR2a and CR2b) are located at the reactant side of the channels (R2a and R2b) of reaction R2; the stabilization energies are −1.45 and −1.91 kcal/mol for CR2a and CR2b, respectively. Also, three complexes, CP2a, CP2b, and CP2b′, are located at the exits of channels R2a, R2b, and R2b′ with the energies lower than the corresponding products. With respect to the transition states of R2, the relative energy of TS2b, 4.10 kcal/mol, is the lowest one, indicating that OH attack on the H atoms in the $-CH_2-$ group is most favored kinetically. In addition, TS2b′ (H′ in the $-CH_2-$ group is abstracted by OH) lies about 0.8 kcal/mol higher than TS2b. By analysis, we find that, for channel R2b, the abstracted hydrogen (H) lies between H and F, whereas for channel R2b′, the abstracted hydrogen (H′) lies between F and F, so the steric hindrance for channel R2b′ is a little larger than the one for channel R2b, which leads to the higher potential energy barrier of R2b. Although the barrier height of channel R2a is about 0.4 kcal/mol lower than that of R2b′, the latter is more exothermic than the former by about 2.6 kcal/mol. So, channels R2a and R2b′ are expected to be competitive, while channel R2b is the dominant route kinetically and thermodynamically. This view is further supported in the following rate constant study. In addition, the energies of transition state TS1b (5.43 kcal/mol, relative to the reactants) is higher than those of TS2b (4.10 kcal/mol) and TS2b′ (4.93 kcal/mol) at the MC-QCISD//B3LYP/6-311G(d,p) level, indicating that the substitution of H by F decreases the reactivity of the C−H bond toward hydrogen abstraction. A similar conclusion can be obtained from the reactions OH with CH_3OCH_3 and CF_3OCH_3.[32]

Now, our focus will be fixed on the subsequent reactions of product radicals with OH: $CF_2CHFOCF_3$ + OH (R3), $CF_2CH_2OCF_3$ + OH (R4), CHF_2CFOCF_3 + OH (R5), and CHF_2CHOCF_3 + OH (R6). The potential energy surfaces (PESs) of these reactions obtained at the B3LYP/6-311G(d,p) level are shown in Figure 3a and b. From Figure 3a and b, we can find that all of these reactions take the addition−elimination mechanisms. Without any association barriers, the OH radical initially attacks the product radicals on the free radical site via electrostatic interaction between the reactant species. These are barrierless association processes, and four intermediates, M1, M2, M3, and M4, are produced with the energies 111.36, 109.33, 106.93, and 99.77 kcal/mol, respectively, lower than the reactants. Starting from M1, it can take the concerted 1,3-H-shift and C−H-bond-rupture processes to form product complex CP1 (−75.13 kcal/mol). Subsequently, CP1 dissociates to the final products CF_2CFOCF_3 and H_2O, that is,

$$CF_2CHFOCF_3 + OH \rightarrow C(OH)F_2CHFOCF_3\ (M1) \rightarrow CP1 \rightarrow CF_2CFOCF_3 + H_2O \quad (R3)$$

A similar case can be found in reactions $CF_2CH_2OCF_3$ + OH (R4) and CHF_2CFOCF_3 + OH (R5). However, the case is different for reaction CHF_2CHOCF_3 + OH (R6). The intermediate M4 has two H atoms which are not equivalent. With the large amount of heat released from the addition processes, M4 can take a concerted 1,3-H shift and cleavage of a neighbor C−H bond via transition state TS4, forming CF_2CHOCF_3 and H_2O, which is similar to the situation discussed above. Also, M4 can take a concerted 1,2-H shift and adjacent C−H cleavage to form CP5 via transition state TS5 and then produce products CHF_2COCF_3 and H_2O. These reactions can be summarized as follows:

$$CF_2CH_2OCF_3 + OH \rightarrow C(OH)F_2CH_2OCF_3\ (M2) \rightarrow CP2 \rightarrow CF_2CHOCF_3 + H_2O \quad (R4)$$

$$CHF_2CFOCF_3 + OH \rightarrow CHF_2C(OH)FOCF_3\ (M3) \rightarrow CP3 \rightarrow CF_2CFOCF_3 + H_2O \quad (R5)$$

$$CHF_2CHOCF_3 + OH \rightarrow CHF_2CH(OH)OCF_3\ (M4) \rightarrow CP4 \rightarrow CF_2CHOCF_3 + H_2O \quad (R6a)$$

$$CHF_2CHOCF_3 + OH \rightarrow CHF_2CH(OH)OCF_3\ (M4) \rightarrow CP5 \rightarrow CHF_2COCF_3 + H_2O \quad (R6b)$$

From Figure 3a and b, we can see that all of the transition states and products lie below the respective reactants, which means that the addition−elimination processes are all thermodynamically and kinetically accessible, and the products CF_2CFOCF_3, CF_2CHOCF_3, and CHF_2COCF_3 may be easily produced during these processes. So, these products may be expected to be observed experimentally. We hope that the present theoretical prediction may assist in future laboratory investigations.

Dynamic Calculations. For the H-abstraction reactions R1 and R2, direct dynamic calculations are performed on the basis of the electronic structure information above. The reactant and product complexes for the reaction channel (if it exists) are taken into account in the dynamic calculations performed by the Polyrate program. Similar cases have been discussed by Galano et al.[33] The MEPs of reactions R1 and R2 are calculated by IRC theory at the B3LYP/6-311G(d,p) level, and the potential energy profile is further improved with the ISPE method at the MC-QCISD//B3LYP level. As pointed out by Truhlar et al.,[34] four single-point calculations are important to correct the lower-level reaction path for the ISPE method: two points are close to the 300 K turning points and are useful primarily to estimate the width of the barrier; two other points are close to the lower-level saddle point and are useful primarily to locate the dual-level saddle point. So the single-point energies evaluated at four extra points (denoted as VTST−ISPE-4) are used for the spline fit. For the purpose of comparison, the dynamics calculations of reaction R1 are also carried out with the VTST method at the MC-QCISD//BB1K level.

Table 2. The Enthalpies (ΔH°_{298}) and Potential Barriers (ΔE) (kcal/mol) with ZPE Correction at the MC-QCISD//B3LYP, MC-QCISD//BB1K, and MC-QCISD//MP2 Levels

levels	MC-QCISD//B3LYP		MC-QCISD//BB1K		MC-QCISD//MP2	
	ΔH°_{298}	ΔE	ΔH°_{298}	ΔE	ΔH°_{298}	ΔE
$CHF_2CHFOCF_3$ + OH → $CF_2CHFOCF_3$ + H_2O (R1a)	−16.51	4.61	−16.33	4.32	−16.45	4.67
$CHF_2CHFOCF_3$ + OH → CHF_2CFOCF_3 + H_2O (R1b)	−16.39	5.43	−16.18	5.36	−16.19	5.36
$CHF_2CH_2OCF_3$ + OH → $CF_2CH_2OCF_3$ + H_2O (R2a)	−17.55	4.46			−17.55	
$CHF_2CH_2OCF_3$ + OH → CHF_2CHOCF_3 + H_2O (R2b)	−20.07	4.10			−19.95	
$CHF_2CH_2OCF_3$ + OH → CHF_2CHOCF_3 + H_2O (R2b′)	−20.07	4.93			−19.95	

Figure 4 shows the plot of the classical potential energy curve (V_{MEP}), ground-state vibrationally adiabatic energy curve ($V_{\mathrm{a}}^{\mathrm{G}}$), and zero-point energy curve (ZPE) as functions of s (amu)$^{1/2}$·bohr at the MC-QCISD//B3LYP level for channel R1a, where $V_{\mathrm{a}}^{\mathrm{G}} = V_{\mathrm{MEP}} + \mathrm{ZPE}$. Due to the similarity, the plots for R1b, R2a, R2b, and R2b′ are depicted in Figure S4 (Supporting Information). From Figure 4, we can find that the ZPE is practically constant as s varies, with only a gentle drop near the saddle point. The V_{MEP} and $V_{\mathrm{a}}^{\mathrm{G}}$ curves are similar in shape, and the maxima are also located at the same position, implying that the variational effect will be small. The following study will further support this view.

The rate constants of each reaction channel are calculated by using TST, CVT, and CVT with small-curvature tunneling (SCT) correction over the temperature range 250−1200 K at the MC-QCISD//BLYP/6-311G(d,p) level. The plots of

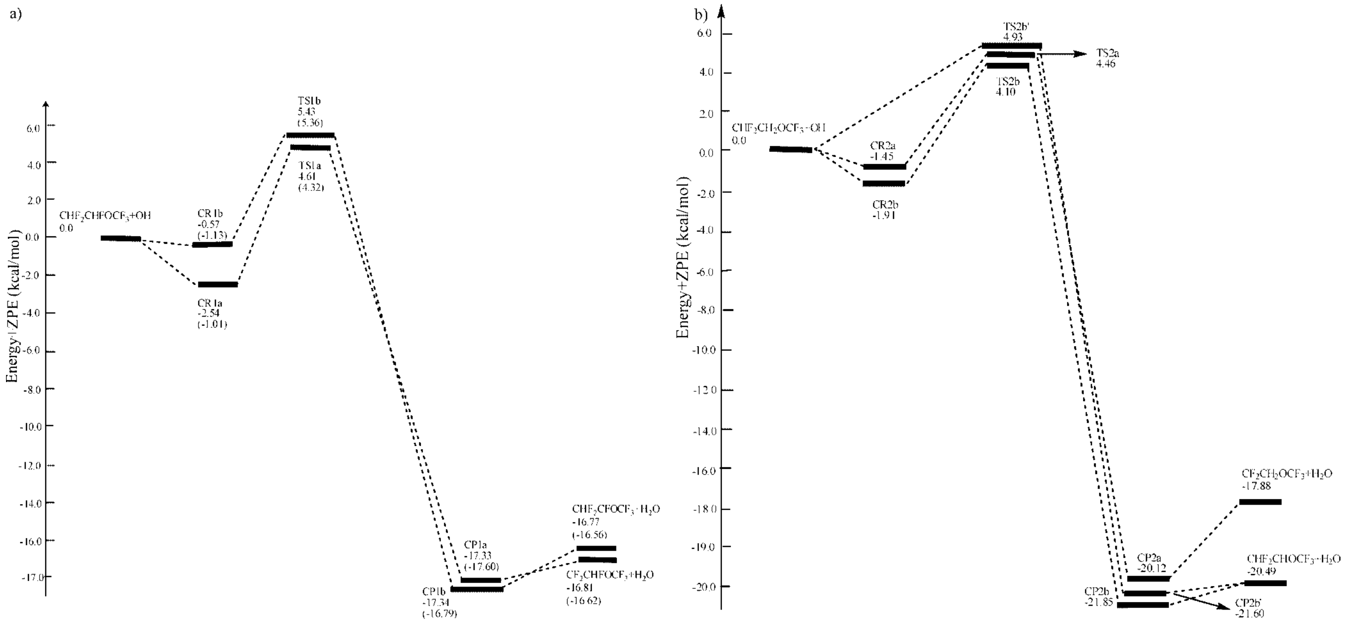


Figure 2. (a) Schematic potential energy surfaces for reaction $CHF_2CHFOCF_3$ + OH. Relative energies (in kcal/mol) are calculated at the MC-QCISD//B3LYP+ZPE and MC-QCISD//BB1K+ZPE levels. (b) Schematic potential energy surfaces for reaction $CHF_2CH_2OCF_3$ + OH. Relative energies (in kcal/mol) are calculated at the MC-QCISD//B3LYP+ZPE level.

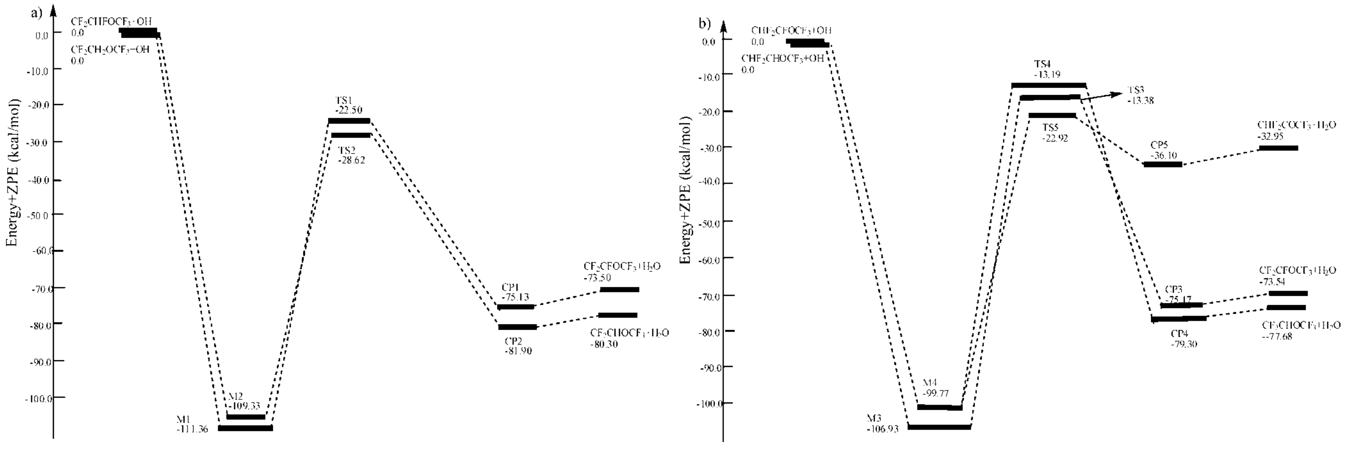


Figure 3. Schematic potential energy surfaces for reactions $CF_2CHFOCF_3$/$CF_2CH_2OCF_3$ + OH. Relative energies (in kcal/mol) are calculated at the MC-QCISD//B3LYP/6-311G(d, p)+ZPE level. (b) Same as part a except for reactions CHF_2CFOCF_3/CHF_2CHOCF_3 + OH.

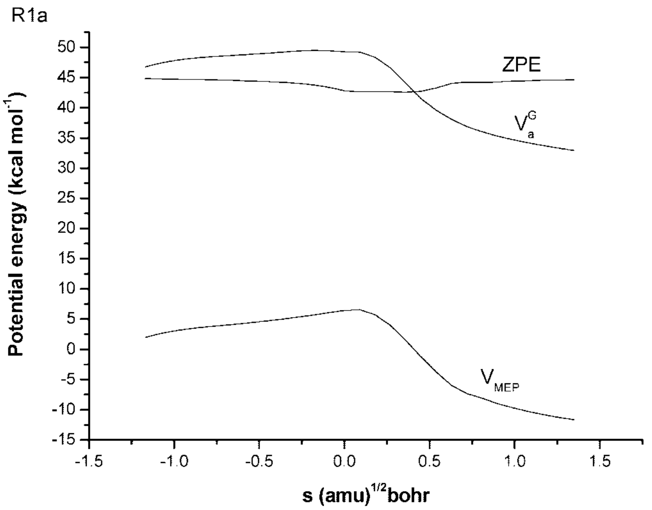


Figure 4. Classical potential energy curve (V_{MEP}), ground-state vibrationally adiabatic energy curve (V^G_a), and zero-point energy curve (ZPE) as functions of s (amu)$^{1/2}$ bohr at the MC-QCISD//B3LYP/6-311G(d,p) level for channel $CHF_2CHFOCF_3 + OH \rightarrow CF_2CHFOCF_3 + H_2O$ (R1a).

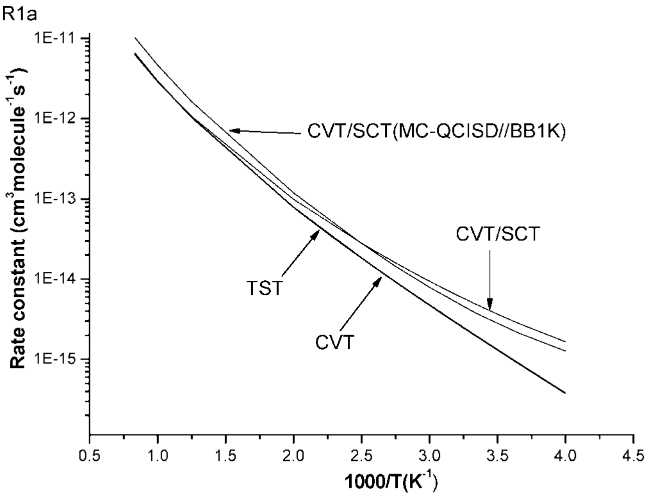


Figure 5. The TST, CVT, and CVT/SCT rate constants calculated at the MC-QCISD//B3LYP level and the CVT/SCT rate constants calculated at the MC-QCISD//BB1K level as a function of 10^3/T for the reaction channel R1a.

TST, CVT, and CVT/SCT rate constants of channel R1a as functions of the reciprocal of the temperature are present in Figure 5, and the corresponding plots of R1b, R2a, R2b, and R2b′ are shown in Figure S5 (Supporting Information). Also, the calculated CVT/SCT rate constants at the MC-QCISD//BB1K/6-31+G(d,p) level for channels R1a (k'_{1a}) and R2a (k'_{1b}) are displayed in Figures 5 and Figure S5 (Supporting Information), respectively. Seen from the figures, the calculated CVT/SCT rate constants at the MC-QCISD//B3LYP level agree well with those obtained at the MC-QCISD//BB1K level. At the MC-QCISD//B3LYP level, the TST and CVT rate constants for H abstraction from $-CHF_2$ are very close over the whole temperature range, which indicates that the variational effect is small. These results are in accord with the above analysis. Moreover, by contrasting the CVT and CVT/SCT rate constants, it is found that SCT correction plays an important role for each channel at the lower temperatures. For example, the ratios of $k_{CVT/SCT}/k_{CVT}$ at 250 K are 4.4, 1.1, 4.6, 9.5, and 11.4 for R1a, R1b, R2a, R2b, and R2b′, respectively.

The total rate constants of R1 and R2 are obtained as the sum of the individual rate constants associated with the corresponding channels. Figure 6 presents the individual CVT/SCT rate constants of each reaction channel of R1 and R2 (k_{1a}, k_{1b}, k_{2a}, k_{2b}, and $k_{2b'}$) and the total rate constants k_1 and k_2 ($k_1 = k_{1a} + k_{1b}$ and $k_2 = k_{2a} + k_{2b} + k_{2b'}$) obtained at the MC-QCISD//B3LYP level, along with the available literature values.[4,5] Also, the total rate constants of channel R1 calculated at the MC-QCISD//BB1K level ($k'_1 = k'_{1a} + k'_{1b}$) are plotted in Figure 6 for comparison. In addition, the temperature dependence of the k_{1a}/k_1, k_{1b}/k_1, k_{2a}/k_2, k_{2b}/k_2, and $k_{2b'}/k_2$ branching ratios is plotted in Figure 7. For reaction R1, H abstraction from the $-CHF_2$ group (R1a) is a more competitive channel over the whole temperature range, whereas the contribution of channel R1b should be considered at the lower temperature range. For example, k_{1a}/k_1 branching ratios are 71.2, 88.7, and 96.1% at 298, 500, and 1200 K, respectively. From Figure 6, we can see that the total rate constants calculated at the two levels show good agreement in the whole temperature range, within a factor of 0.6−1.2. The values of k_1 and k'_1 are 6.64×10^{-15} and 5.33×10^{-15} cm^3 molecule^{-1} s^{-1} at 298 K, respectively, which are consistent with the only experimental value [$(6.80 \pm 1.1) \times 10^{-15}$ cm^3 molecule^{-1} s^{-1}] obtained by Oyaro et al.[4] The good agreements between theoretical and experimental values and between the results at the two levels imply that the B3LYP method is reliable in the location of the transition state and the reaction path for such systems. So, for reaction R2, the dynamic calculations are just performed at the MC-QCISD//B3LYP level. For reaction R2, seen in Figure 7, the dominant reaction channel is channel R2b, that is, H abstraction from the $-CH_2-$ group, leading to the product CHF_2CHOCF_3, and the channels R2a and R2b′ are competitive channels. The k_{2a}/k_2, k_{2b}/k_2, and $k_{2b'}/k_2$ branching ratios are 16.4, 73.4, and 10.2% at 298 K and 28.3, 42.0, and 29.7% at 1200 K. Thus, all of these channels (R2a, R2b, and R2b′) should be taken into account over the whole temperature range. Our calculated rate constant at 298 K (2.74×10^{-14} cm^3 molecule^{-1} s^{-1}) is slightly lower than the available literature value, 3.27×10^{-14} cm^3 molecule^{-1} s^{-1}, estimated by using the empirical estimation method.[5] In addition, the rate constant of reaction R2 (k_2) is about 1 order of magnitude larger than that of reaction R1 (k_1); that is to say, reaction R2 may proceed much easier than reaction R1, and the fluorine substitution decreases the reactivity of the C−H bond. This is because F is strongly electron-withdrawing; the electron density on the carbon atom of $-CHF$ is more reduced in $CHF_2CHFOCF_3$ than that of $-CH_2-$ in $CHF_2CH_2OCF_3$. Consequently, the smaller rate constants are found for reaction R1.

The rate constants exhibit strong non-Arrhenius behavior in the temperature range of 250−1200 K. Since there is little experimental data for these two reactions, to provide useful guides for future experimental study, the three-parameter fits based on the CVT/SCT rate constants within 250−1200 K give the expressions as follows (in units of cm^3 molecule^{-1} s^{-1}): $k_1 = [2.49 \times 10^{-24}]T^{4.10} \exp(-488/T)$, $k_2 = [4.26 \times 10^{-23}]T^{3.80} \exp(-404/T)$.

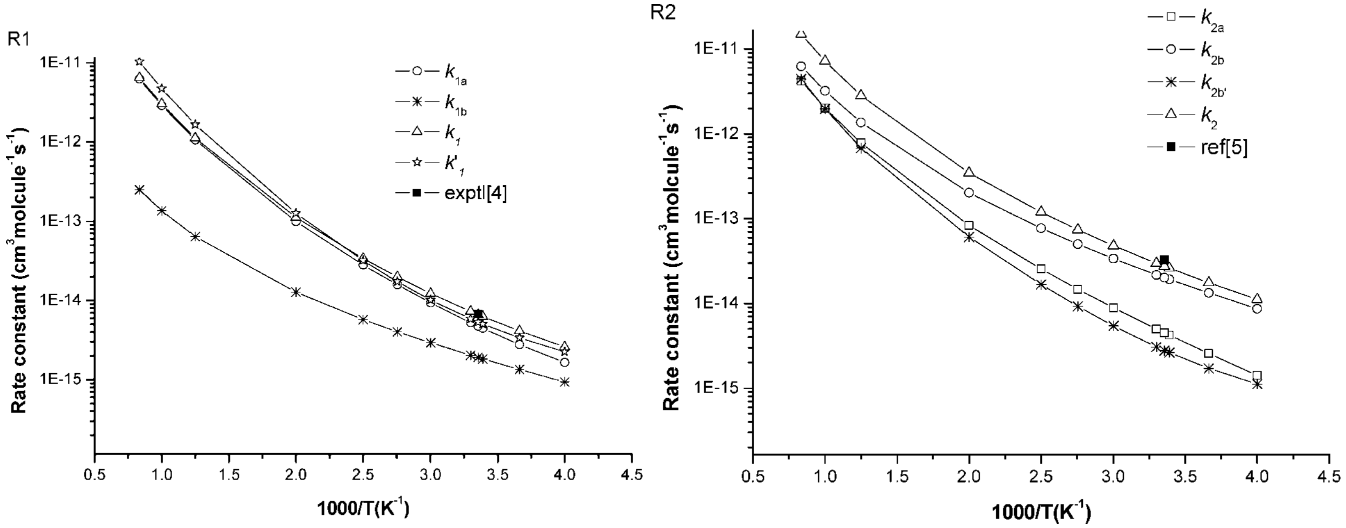


Figure 6. The rate constants k_{1a} and k_{1b} for the reaction channels of $CHF_2CHFOCF_3$ + OH and k_{2a}, k_{2b}, and $k_{2b'}$ for $CHF_2CH_2OCF_3$ + OH; the total rate constants k_1 ($k_1 = k_{1a} + k_{1b}$) and k_2 ($k_2 = k_{2a} + k_{2b} + k_{2b'}$) calculated at the MC-QCISD//B3LYP/6-311G(d, p) level; and the total rate constants k'_1 ($k'_1 = k'_{1a} + k'_{1b}$) calculated at the MC-QCISD//BB1K/6-31+G(d,p) level along with the literature values[4,5] as a function of $10^3/T$.

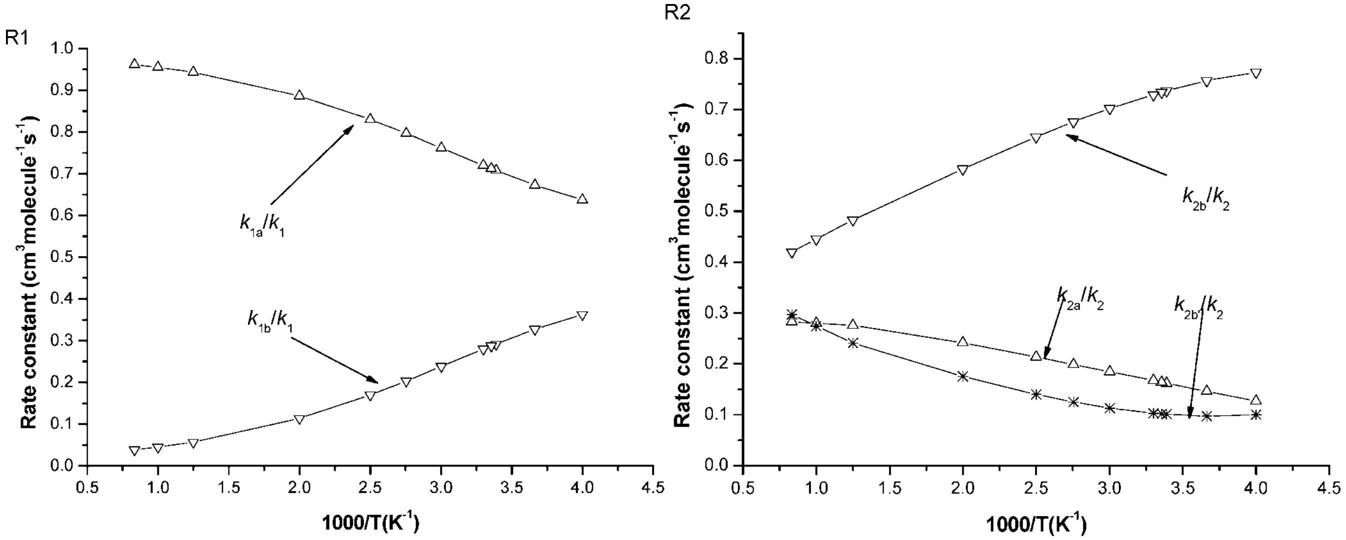


Figure 7. Calculated branching ratio for reactions $CHF_2CHFOCF_3$ + OH and $CHF_2CH_2OCF_3$ + OH as a function of $10^3/T$ at the MC-QCISD//B3LYP/6-311G(d, p) level.

Furthermore, the kinetic isotope effects for the reactions $CHF_2CHFOCF_3$ + OH (R1)/$CDF_2CDFOCF_3$ + OH (R1*) and $CHF_2CH_2OCF_3$ + OH (R2)/$CDF_2CD_2OCF_3$ + OH (R2*) are calculated over the whole temperature range. The calculated KIEs for the total reactions R1 (k_1/k_1*) and R2 (k_2/k_2*) are shown in Figure 8, and the KIEs for individual channels R1a, R1b, R2a, R2b, and R2b′ are displayed in the Supporting Information (Figure S8). As shown in Figure 8, the two reactions present normal KIEs, that is, greater than 1. The KIE values are 12.2 and 2.0 at 298 K for reactions R1 and R2, respectively, and the KIEs decrease with the temperature increase. Since there is no experimental value to make a comparison, we hope the present predictions for KIEs may be helpful for future experimental investigations.

Conclusions

In this paper, the dual-level direct dynamic method is employed to investigate theoretically the hydrogen abstraction reaction systems $CHF_2CHFOCF_3$ + OH (R1) and $CHF_2CH_2OCF_3$ + OH (R2). The PES information is obtained at the B3LYP/6-311G(d,p) level, and the higher-level energies for the stationary points and extra points along the minimum energy path are refined at the MC-QCISD theory. Two and three hydrogen abstraction channels are found for reactions R1 and R2, respectively, and four products, $CF_2CHF_2OCF_3$, $CHF_2CF_2OCF_3$, $CF_2CH_2OCF_3$, and CHF_2CHOCF_3, are produced. Using group balance isodesmic reactions as working reactions, the $\Delta H^{\circ}_{f,298}$ values are −362.89 ± 0.12, −338.21 ± 0.12, −338.10 ± 0.12, −312.43 ± 1.07, −289.82 ± 1.07, and −265.36 ± 0.65 kcal/mol for $CHF_2CHFOCF_3$, $CF_2CHFOCF_3$, CHF_2CFOCF_3, $CHF_2CH_2OCF_3$, $CF_2CH_2OCF_3$, and CHF_2CHOCF_3, respectively, at the MC-QCISD//B3LYP/6-311G(d,p) level. Furthermore, the subsequent reaction mechanisms of the reactions of the above four product radicals with the OH radical are

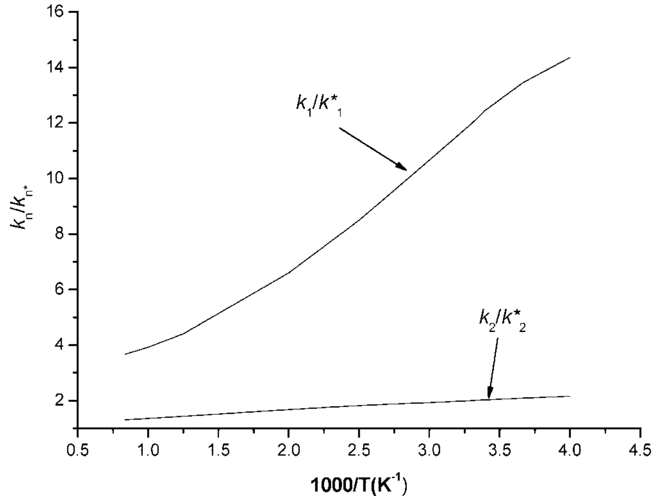


Figure 8. Plot of the calculated ratios k_1/k^*_1 and k_2/k^*_2 versus 1000/T at the temperature range 250−1200 K.

investigated theoretically at the same level in this article. We found that all of these reactions take the addition−elimination mechanisms. These processes are all thermodynamically and kinetically accessible, and the products CF_2CFOCF_3, CF_2CHOCF_3, CHF_2COCF_3, and H_2O are expected to be observed. The rate constants of the H-abstraction reaction channels are calculated by using CVT with the SCT correction. For each H-abstraction reaction channel, the variational effect is small over the whole temperature region, and the SCT correction plays an important role in the rate constant calculation in the lower-temperature range. For reaction R1, H abstraction from the $-CHF_2$ group (R1a) is the more competitive channel over the whole temperature range, whereas the contribution of channel R1b should be considered at the lower temperatures. For reaction R2, the preferred reaction channel is channel R2b. However, the contribution of channels R2a and R2b′ should be also taken into account over the whole temperature range. In addition, the rate constant of reaction R2 (k_2) is about 1 order of magnitude larger than that of reaction R1 (k_1). To provide good estimation for future laboratory investigations, the fitted three-parameter expressions within 250−1200 K for these two reactions are given, $k_1 = [2.49 \times 10^{-24}]T^{4.10} \exp(-488/T)$, $k_2 = [4.26 \times 10^{-23}]T^{3.80} \exp(-404/T)$ cm^3 molecule^{-1} s^{-1}.

Acknowledgment. We thank Professor Donald G. Truhlar for providing the POLYRATE 8.4.1 program. This work is supported by the National Natural Science Foundation of China (Grants 20333050, 20073014, and 20303007), Doctor Foundation by the Ministry of Education Foundation for University Key Teacher by the Ministry of Education, Key Subject of Science and Technology by the Ministry of Education of China, and the Innovation Foundation by Jilin University.

Supporting Information Available: Calculated frequencies, potential energy curves, computed rate constant curves, and a plot of calculated ratios. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Sekiya, A.; Misaki, S. *J. Fluorine Chem.* **2000**, *101*, 215.

(2) Sekiya, A.; Misaki, S. *CHEMTECH* **1996**, *26*, 44.

(3) Ravishankara, R. A.; Turnipseed, A. A.; Jensen, N. R.; Barone, S.; Mills, M.; Howark, C. J.; Solomon, S. *Science* **1994**, *263*, 71.

(4) Oyaro, N.; Sellevag, S. R.; Nielsen, C. J. *J. Phys. Chem. A* **2005**, *109*, 337.

(5) Urata, S.; Takada, A.; Uchimaru, T.; Chandra, A. K. *Chem. Phys. Lett.* **2003**, *368*, 215.

(6) Truhlar, D. G. In *The Reaction Path in Chemistry: Current Approaches and Perspectives*; Heidrich, D. , Ed.; Kluwer: Dordrecht, The Netherlands, 1995; pp 229.

(7) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771.

(8) Hu, W. P.; Truhlar, D. G. *J. Am. Chem. Soc.* **1996**, *118*, 860.

(9) Truhlar, D. G.; Garrett, B. C. *Acc. Chem. Res.* **1980**, *13*, 440.

(10) Truhlar, D. G.; Isaacson, A. D.; Garrett, B. C. In *The Theory of Chemical Reaction Dynamics*; Baer, M., Ed.; CRC Press: Boca Raton, FL, 1985; pp 65.

(11) Truhlar, D. G.; Garrett, B. C. *Annu. Rev. Phys. Chem.* **1984**, *35*, 159.

(12) Chuang, Y. Y.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 1140.

(13) Lu, D. H.; Truong, T. N.; Melissas, V. S.; Lynch, G. C.; liu, Y. P.; Garrett, B. C.; Steckler, R.; Issacson, A. D.; Rai, S. N.; Hancock, G. C.; Lauderdale, J. G.; Joseph, T.; Truhlar, D. G. *Comput. Phys. Commun.* **1992**, *71*, 235.

(14) Liu, Y.-P.; Lynch, G. C.; Truong, T. N.; Lu, D.-H.; Truhlar, D. G.; Garrett, B. C. *J. Am. Chem. Soc.* **1993**, *115*, 2408.

(15) Frisch, M. J.; Truck, 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.; Boboul, A. G.; Stefnov, B. B.; Liu, G.; Liaschenko, A.; Piskorz, P.; Komaromi, L.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; AlLaham, 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 A.1 ed.; Guassian, Inc.: Pittsburgh, PA, 2003.

(16) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(17) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(18) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2715.

(19) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.

(20) Fast, P. L.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 6111.

(21) Chuang, Y.-Y.; Corchado, J. C.; Fast, P. L.; Villa, J.; Hu, W.-P.; Liu, Y.-P.; Lynch, G. C.; Jackels, C. F.; Nguyen, K. A.; Gu, M. Z.; Rossi, I.; Coitino, E. L.; Clayton, S.; Melissas, V. S.; Lynch, B. J.; Steckler, R.; Garrett, B. C.; Isaacson,

A. D.; Truhlar, D. G.; *Polyrate*, version 8.4.1 ed.; University of Minnnesota: Minneapolis, MN, 2000.

(22) Garrett, B. C.; Truhlar, D. G. *J. Chem. Phys.* **1979**, *70*, 1593.

(23) Good, D. A.; Francisco, J. S. *J. Phys. Chem. A* **1998**, *102*, 7143.

(24) Huber, K. P.; Herzberg, G. In *Molecular Spectra and Molecular Structure. IV. Constants of Diatomic Molecules*; Van Nostrand Reinhold Co.: New York, 1979.

(25) Hoy, A. R.; Bunker, P. R. *J. Mol. Struct.* **1979**, *74*, 1.

(26) Shimanouchi, T. *Tables of Molecular Vibrational Frequencies, Consolidated 1, NSRDS NBS-39*; U.S. Department of Commerce: Washington, DC.

(27) Chase, M. W., Jr *J. Phys. Chem. Ref. Data, Monograph* **1998**, *9*, 1.

(28) Pilcher, G.; Pell, A. S.; Coleman, D. J. *Trans. Faraday Soc.* **1964**, *60*, 499.

(29) Kolesov, V. P.; Shtekher, S. N.; Martynov, A. M.; Skuratov, S. M. *Russ. J. Phys. Chem. (Engl. Transl.)* **1968**, *42*, 975.

(30) Wu, E.; Rodgers, A. S. *J. Phys. Chem.* **1974**, *78*, 2315.

(31) Tsang, W. Heats of Formation of Organic Free Radicals by Kinetic Methods. In *Energetics of Organic Free Radicals*; Martinho Simoes, J. A., Greenberg, A., Liebman, J. F., Eds.; Blackie Academic and Professional: London, 1996; pp 22.

(32) Wu, J. Y.; Liu, J. Y.; Li, Z. S.; Sun, C. C. *J. Chem. Phys.* **2003**, *118*, 10986.

(33) Galano, A.; Alvarez-ldaboy, J. R.; Ruiz-Santoyo, M. E.; Vivier-Bunge, A. *ChemPhysChem* **2004**, *5*, 1379.

(34) Chuang, Y. Y.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 1140.

CT800032E

J. Chem. Theory Comput. **2008,** *4,* 1083–1093

JCTC Journal of Chemical Theory and Computation

Atomistic Simulation of Adiabatic Reactive Processes Based on Multi-State Potential Energy Surfaces

Jonas Danielsson and Markus Meuwly*

Chemistry Department, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland

Received February 28, 2008

Abstract: The adiabatic reactive molecular dynamics (ARMD) method provides a framework to study chemical reactions using molecular dynamics simulations with minimal computational overhead. Here, ARMD is generalized to an arbitrary reactive process between two states in which reactants and products can be treated by an atomistic force field. The implementation is described, and the method is applied to two systems: the kinetics of NO rebinding to myoglobin (Mb) as a validation system and the conformational transition in neuroglobin (Ngb) which explores the full functionality of ARMD. For MbNO, the nonexponential kinetics observed both in experiment and earlier ARMD studies is reproduced. Furthermore, the sensitivity of the results with respect to the asymptotic separation between the two potential energy surfaces (NO bound and unbound) is studied.

1. Introduction

Describing the dynamics of chemical reactions and extracting meaningful information that can be compared with experimental data is a major challenge in computational chemistry. For most chemically and biologically relevant processes the inherent difficulty lies in the large span of time- and length scales involved. Ideally, the computational treatment of chemical reactions in condensed phases, on surfaces, or in macromolecular systems requires methods that are capable of describing motions and time scales in the sub-Å/fs range (electrons, atomic rearrangements) to the tens-of-Å/ms range in a realistic fashion. The computational cost of accurate electronic structure calculations (ab initio calculations including electron correlation) prohibits the direct use of such methods including the time dependence of the process for large systems. On the other hand, the simplified form of the interaction potential used in standard force fields is not designed to describe changes in the electronic structure. The quantum mechanical/molecular mechanical (QM/MM) method[1–4] is a popular way to combine the accuracy of electronic structure calculations for the reacting fragments with the efficiency of classical force fields. This method has been used to locate critical points on the PES, which can provide useful information.[5–7] Typically, studies including dynamics and configurational heterogeneity are carried out at the semiempirical level.[8,9] Exceptions to this are, e.g., proton transfer reactions in small water- and ammonia-containing systems.[10–12] The Empirical Valence Bond (EVB)[13,14] is a method in which the PES is described by two or more valence states, each described parametrized as a force field, which are mixed according to an empirical coupling term. EVB is comparable in efficiency to conventional force field calculations which allows dynamics simulations. In its original version, an EVB representation is calibrated by forcing the calculated ground-state energy at certain points to reproduce their experimentally determined free energies.[13] In more recent applications EVB potentials have also been calibrated in view of data from electronic structure calculations. If the reaction involves a change of electronic state, then surface hopping methods[15,16] where the system can switch between several surfaces in a stochastic process provides a useful way to treat the dynamics. However, computational efficiency is still an issue since the switching probability has to be calculated from nonadiabatic coupling elements, which requires detailed knowledge of the electronic structure. The latter is particularly challenging for systems with a complicated electronic structure, such as metal-containing proteins, where strong multireference character of the metal-center is common.

---

* Corresponding author e-mail: m.meuwly@unibas.ch.

10.1021/ct800066q CCC: $40.75

Recently an approach was presented which is applicable to reactions involving two potential energy surfaces where quantum effects (e.g., tunneling or coherence) are less important because the transitions are to a good approximation adiabatic (e.g., due to strong coupling between the two states). Adiabatic Reactive Molecular Dynamics (ARMD) was successfully used to quantitatively study the nonexponential rebinding kinetics of nitric oxide (NO) to myoglobin (Mb).[17] Photolysis and subsequent rebinding of small molecules to the heme group of globular proteins such as myoglobin and hemoglobin has become one of the paradigm processes in biophysics and has provided a substantial insight into the interplay between structure, function, and dynamics of protein−ligand complexes.[18,19] Earlier experiments of myoglobin rebinding kinetics focused on the binding of CO vs O_2.[20–22] Given the important biological role of NO as a messenger molecule involved in platelet aggregation and immune response,[23,24] the interest in the reactive dynamics between NO and Mb has considerably increased.[25–29] Finally, MbNO is also an interesting system for ultrafast spectroscopy given its rapid rebinding kinetics with a fast time scale in the picosecond range[28,30] and for molecular dynamics simulations,[17,27] since many interesting processes occur on a time scale suitable for computational work.

Neuroglobin is a recently discovered member of the globin protein family expressed in the nerve cells of vertebrates.[31] The function of this protein is unknown, although it seems to be upregulated in cases of hypoxia (oxygen deficiency), suggesting a protective role.[32] The crystal structure of unligated neuroglobin was recently solved and revealed a number of interesting differences compared to other globins.[33] Unlike myoglobin, the unligated protein is hexacoordinated, with active-site histidines acting as axial ligands. Analysis of the X-ray experiments found that the measured electron density in the active site could not be well fitted with a single conformation. Using a second conformation with 30% occupancy and a different heme orientation the quality of the fitting was considerably improved. The biological significance for these almost degenerate conformations is unknown, although NMR experiments indicate that binding of CN^- is significantly slower in one of the conformations.[34]

In the present work the general formulation of ARMD to treat reactions involving two potential energy surfaces is presented and applied to two system of biological interest. First the Mb-NO rebinding problem[27,28] is reconsidered to validate the generalized implementation, and second we study the equilibrium and switching between different conformational states in neuroglobin[33] to demonstrate the generality of the code and also show how one can parametrize the method from experimental data. The current work is structured as follows. First, details about the ARMD algorithm and its implementation are provided. Next, the computational setup for the two test systems is described, and the results from extensive ARMD simulations are presented. Finally, conclusions summarize the work.

2. Computational Methods.

2.1. The ARMD Algorithm. In the following, the potential energy surfaces (PESs) considered have the standard CHARMM[35] form

$$\begin{aligned} V(\vec{x}) = {} & \sum_{bonds} \frac{k_i}{2}(r_i - r_i^0)^2 + \sum_{angles} \frac{k_j}{2}(\theta_j - \theta_j^0)^2 + \\ & \sum_{dihedrals} \sum_{m=1}^{mmax} k_k^m \cos(m\varphi_k - \gamma_k) + \\ & \sum_{impropers} \frac{k_l}{2}(\phi_l - \phi_l^0)^2 + \\ & \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon 0 r_{ij}} + 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \end{aligned} \tag{1}$$

where r_i, θ_j, ϕ_k, and φ_l are bond lengths, covalent angles, dihedrals, and improper dihedrals, respectively, the k_x parameters are force constants, γ_k is a dihedral offset, q_i are atomic partial charges, and σ_{ij}, ε_{ij} are pairwise Lennard-Jones radius and well-depths. In addition to these terms, Morse potentials are used for anharmonic bonds

$$V_{Morse} = D_e(1 - \exp(-\beta(r - r_0)))^2 \tag{2}$$

where r_0 is the equilibrium bond length, D_e is the dissociation energy, and β is the anharmonicity parameter. V_{Morse} is particularly useful in studying the breaking and formation of bonds.

The force field description (eq 1) of the reactants (V_R) and the products (V_P) differs by a number of force field parameters. In a macromolecular system the number of energy terms by which V_R and V_P differ is usually small compared to the total number of energy terms. The latter is of order N for bonded and of order N^2 for nonbonded terms, where N is the total number of atoms. Thus, it is only necessary to provide a small number of additional parameters compared to a standard simulation to describe the differences between R and P.

For the general case where V_R and V_P represent different electronic states, the two PESs are separated by an asymptotic energy difference that does not appear in a classical force field. In standard force fields only energy differences have physical significance, whereas the absolute energy scale is unimportant. For a meaningful comparison of the energies of the two force fields defining V_R and V_P, a constant Δ (the "asymptotic separation") is introduced so that $V_R+\Delta$ and V_P share a common 'zero' of energy. In principle, the energy difference between two electronic states at a chosen geometry can be estimated from a single quantum chemical calculation, which allows to set Δ. In practice, however, it is preferable to calibrate Δ to reproduce experimental data.

The dynamics of the system is initiated and propagated on the ground state $V_0(\vec{x}) = \min[V_R+\Delta, V_P]$. At each time step the energy difference $\Delta V = V_R - V_P + \Delta$ is calculated and compared with the previous step. If ΔV switches sign at time t_{cross}, a crossing between the PESs has been detected. Since prior to the crossing point the dynamics takes place on one diabatic surface, the mixing of surfaces due to coupling has to be introduced a posteriori to create the adiabatic surface. This is achieved by reconstructing the system at the time point $t_{cross} - T_{mix}/2$. The crossing between the two diabatic surfaces is then carried out by gradually switching the PES from $V_R + \Delta$ to V_P by using a time dependent switching function $f(t)$ according to

$$V = f(t)(V_R + \Delta) + (1 - f(t))V_P \tag{3}$$

$f(t)$ is chosen so that at the crossing time t_{cross} the surfaces are mixed with equal contributions, i.e $f(t_{cross}) = 0.5$. In this way the procedure is time reversible in the sense that a backward trajectory will follow the same path. Furthermore $0 \leq f(t) \leq 1$ and $f(t)$ increases monotonically with t. Among the possible functions that fulfill these conditions the following was deemed suitable

$$f(t) = \frac{1}{2}[\tanh(\alpha(t - t_{cross})/(T_{mix} - \delta t)) + 1] \tag{4}$$

here written for a switch from V_P to V_R. The constant α is chosen so that $f(t)$ is close to 0 and 1 for $0 < t < t_{cross} - T_{mix}/2$ and $t > t_{cross} + T_{mix}/2$, respectively. In the present simulations $\alpha = 4.8$ was used. In the opposite switch process, $f(t)$ is replaced with $1 - f(t)$. After the switch the dynamics continues on V_P. During the switch and for a time interval $t = T_{mix}/2$ after a crossing, no new crossing can take place. This avoids that two crossings overlap in time. For $t > T_{mix}/2$ the algorithm is reset, ΔV is monitored, and configurations are stored for a next surface crossing.

2.2. Implementation and User Interface. The strategy outlined above was implemented in CHARMM together with a user interface designed to make the method as general and flexible as possible. For details concerning conventions used in CHARMM the original reference should be consulted.[35] The input data required are the parameters Δ, T_{mix}, and α and all force field parameters that differ between V_R and V_P. If the atom connectivities for the system in V_R and V_P differ (i.e., a chemical reaction is investigated) the exclusion lists controlling the nonbonded interactions between atoms separated by two or three covalent bonds in the two states have to be modified. Such modifications are automatically generated from the protein structure file and the additional input concerning the changes in topology between the two surfaces.

Energy terms that differ between V_R and V_P are modified according to $f(t)(V_R{+}\Delta) + (1{-}f(t))V_P$ term by term in eq 1 during a surface crossing. However, for the electrostatic interaction, instead of scaling the Coulombic energy term, the charges q_i which change between V_R and V_P are scaled: $q_i(t) = f(t)q_i^{(R)} + (1{-}f(t))q_i^{(P)}$, and then the electrostatic energy and forces can be calculated by any available method. This simplifies the use of different techniques such as shifted cutoff, reaction fields,[36] or Ewald summation[37] for treating the long-range electrostatics. Note that this leads to a quadratic scaling of electrostatic interaction terms between pairs of atoms where both charges are subject to change during the switch.

2.3. Simulation Setup and Protocols. All simulations were carried out using CHARMM[35] together with the CROSS module described above. Except for the parameters used for the reacting regions (see further below) all parameters are from the CHARMM22[38] all-atom force field. For water molecules the TIP3P model was used.[39]

For both systems (rebinding of NO to myoglobin and conformational substates in Ngb) stochastic boundary conditions[40] were used. The water sphere (radius of 25 Å) was centered around the iron atom of the heme group, and the radius of the inner region where free Newtonian dynamics takes place was 20 Å. The systems were solvated by overlaying the protein with a pre-equilibrated water sphere and removing all waters in close contact (<2.8 Å) with any atom of the protein. The water sphere was rotated, and the overlay was repeated twice. For all simulations the SHAKE[41] algorithm was applied to all covalently bound hydrogen atoms which allows a time step of 1 fs.

Table 1. Parameters of the Two Surfaces Used in the Neuroglobin Simulations

parameter	surface 1 (70%)	surface 2 (30%)
r_0/Å (His64(N_ε)-Fe)	2.40	1.85
k_b/kcal·mol^{-1}·Å^{-2} (His64(N_ε)-Fe)	50.0	50.0
r_0/Å (His96(N_ε -Fe))	1.90	2.00
k_b/kcal·mol^{-1}·Å^{-2} (His96(N_ε)-Fe)	75.0	75.0
θ_0/° (His64(N_ε)-Fe-N_{Heme})	87.00	80.00
k_a/kcal·mol^{-1}·rad^{-2} (His64(N_ε)-Fe-N_{Heme})	50.00	40.00
θ_0/° (His96(N_ε)-Fe-N_{Heme})	93.00	100.00
k_a/kcal·mol^{-1}·rad^{-2} (His96(N_ε)-Fe-N_{Heme})	50.00	40.00
θ_0/° (His64(N_ε)-Fe-His96(N_ε))	177.00	150.00
k_a/kcal·mol^{-1}·rad^{-2} (His64(N_ε)-Fe-His96(N_ε))	50.00	60.00
ϕ_0/° (His64/96(C_δ)-His64/96(N_ε)-Fe-N_{Heme})	0.00	0.00
k_d/kcal·mol^{-1}·rad^{-2} (His64/96(C_δ)-His64/96(N_ε)-Fe-N_{Heme})	0.05	0.05
m (His64/96(C_δ)-His64/96(N_ε)-Fe-N_{Heme})	4	4

Rebinding in MbNO. The initial coordinates for MbNO were taken from the previous study of Nutt and Meuwly,[17] to allow direct comparison. These initial conditions represent a state where the ligand has been photodissociated from the heme group after equilibration in the bound state. The velocities are then randomized by sampling from the appropriate Maxwell–Boltzmann distribution. Subsequently, statistics on the rebinding events is collected as described previously.[17] The experimentally accessible observable is the fraction $N(t)$ of unbound ligand survived. In ARMD this corresponds to the fraction of trajectories still on the unbound PES at time t after photodissociation. The force field parameters for the 5- and 6-coordinated heme group were taken from Meuwly et al.[27] The NO was represented by a simple two-center charge model with charges $-qO = qN = 0.063e$ and a harmonic oscillator as internal potential energy function. In the previous ARMD study of this system[17] a two-dimensional PES from extensive QM calculations was used with the ligand center-of-mass (CoM) to iron distance and the Fe–CoM–O angle as the coordinates.[29] For the current validation simulations the force field described in previous work was used for the bound state, namely a Morse oscillator for the ligand-heme coordination bond, and a harmonic term for the Fe–N–O angle.[27] The simpler PES for the Fe-NO interaction was used to test ARMD under common conditions where no specialized interaction potentials for specific coordinates are available. Also, it is of interest to see whether the salient features of the rebinding reaction can be understood without a more sophisticated PES.

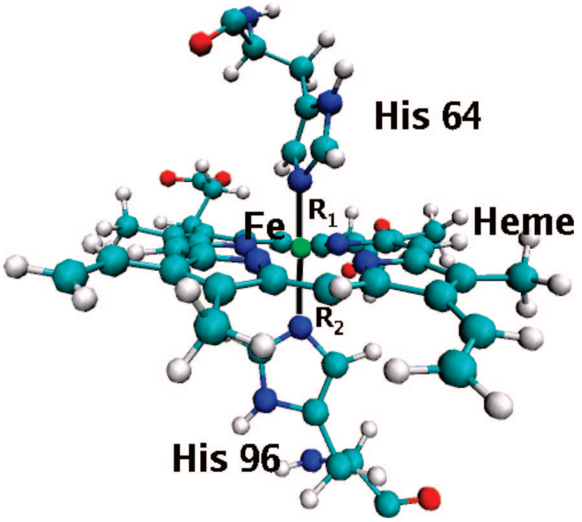


Figure 1. Active site of neuroglobin with the two coordinating histidines (64 and 96), the heme group with the central Fe atom (green), and the two Fe−N distances R_1 and R_2 used to define the reaction coordinate R_u in the umbrella sampling calculations.

Conformational Substates in Ngb. In the experimental structure of neuroglobin two conformations with relative populations 70/30 were observed.[33] The heme group remains hexacoordinated in both cases, but the orientation and bond lengths between the ferric heme iron and the two coordinating histidine groups differ. Switching between these two states can therefore be regarded as a bound-to-bound transition. The force field parameters (see Table 1) for the two observed conformations of neuroglobin were derived from the experimentally observed geometries in combination with standard force constants adapted from similar systems.

The initial structure for Ngb was the wild-type, ligand-free structure (PDB accession code 1Q1F). Minimization and equilibration of the protein were carried out individually on both surfaces by adjusting the Δ parameter to large positive or negative values which avoids crossings between the two states. The resulting average potential energies can be used to estimate Δ such that a certain relative energy between the two stable forms are obtained. Longer simulations of 500 ps each were used to explore the relative population of the two surfaces for different separations Δ between the two PESs. From these preliminary simulations, an approximate progression coordinate can be constructed which is a useful guide for more detailed sampling methods, such as umbrella sampling. This procedure was used in neuroglobin along the coordinate R_u defined by the ratio R_1/R_2 where R_1 and R_2 are shown in Figure 1. A single window with a harmonic umbrella potential was sufficient to obtain good sampling along the entire relevant range of R_u ε (0.8−1.2). The umbrella potential was centered at $R_u = 0.93$, which corresponds approximately to the transition state, and a force constant of 40 kcal/mol was used. From these simulations, reliable estimates of the free energy profile and relative population of the two states can be calculated.

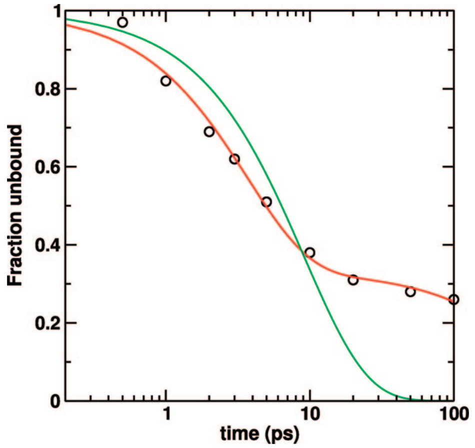


Figure 2. Fraction of unbound trajectories in photodissociated MbNO as a function of time (logarithmic scale) after photodissociation (black circles) for Δ = 65 kcal/mol, together with a double exponential (red) and the best single exponential fit (green).

3. Results and Discussion

3.1. Validation of the Implementation: MbNO as a Test Case. Experimental work on the rebinding kinetics has shown that the rebinding reaction after photodissociation of MbNO is nonexponential in time.[25,28] For a geminate rebinding process, $N(t)$ follows an exponential decay with a single time constant τ and a corresponding rate constant $\lambda = \tau^{-1}$ according to

$$N(t) = e^{-t/\tau} \tag{5}$$

For cases where such a decay law does not reproduce the data, either a time-dependent rate constant $\tau(t)$ (which gives rise to a so-called stretched exponential decay)[42] is used or an ansatz as a sum of i exponentials

$$N(t) = \sum_i w_i e^{-t/\tau_i} \quad \sum_i w_i = 1 \tag{6}$$

with weights w_i is assumed. The generalization of this expression for continuous rate distributions $g(\lambda)$ leads to

$$N(t) = \int g(\lambda) e^{-\lambda t} d\lambda \tag{7}$$

From a number N_{traj} of ARMD trajectories with initial conditions that represent the ensemble of ligands after photodissociation, $N(t)$ can then be approximated by the expression

$$N(t) = \frac{1}{N_{traj}} \sum_{j=1}^{N_{traj}} H(t - t_{cross}^{(j)}) \tag{8}$$

where $H(t)$ is the Heaviside function, and $t_{cross}^{(j)}$ is the observed crossing time for trajectory j. For the present validation $N_{traj} = 150$, which is considerably less than in the original study of Nutt and Meuwly where $N_{traj} = 6000$. However, it is still sufficient to distinguish between exponential and nonexponential rebinding kinetics.

The function $N(t)$ from the present simulations is shown in Figure 2. The survival fraction has a nonexponential time dependence. A single exponential decay (green) does not provide a meaningful fit, particularly not for the long time

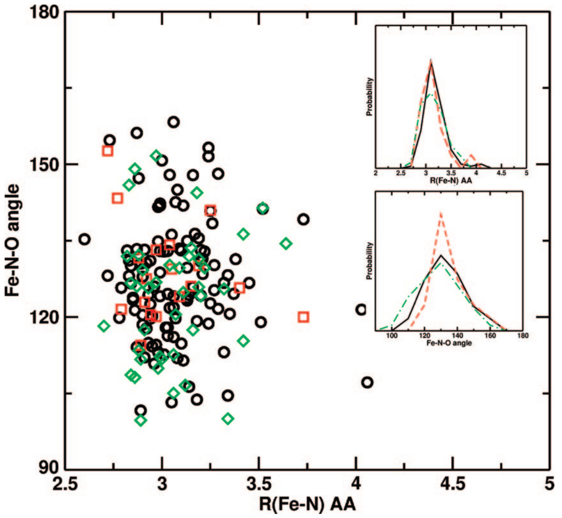


Figure 3. The crossing geometries of NO rebinding to Mb with 3 different values of Δ (red squares Δ = 60, black circles Δ = 65, and green diamonds Δ = 70 kcal/mol) projected onto the distance between iron and the nitrogen of the ligand R_{FeN} and the angle between iron and the NO ligand θ_{FeNO}. The insets show the distributions of crossing points along the R_{FeN} (upper) and θ_{FeNO} (lower) coordinates. For the insets, broken, solid, and dash-dotted lines denote Δ = 60, 65, and 70 kcal/mol, respectively.

behavior. If a double exponential is used, then the data can be fit over the entire time range. The two time constants are $\tau_1 = 3.6$ ps and $\tau_2 = 373$ ps, although τ_2 is not well defined and the uncertainty associated with it is quite large, given that only 8 recrossing events are observed with $t_{cross} > 20$ ps. The fast and slow components have weights of $w_1 = 0.66$ and $w_2 = 0.34$, respectively. Experimentally, ultrafast IR spectroscopy results could be fitted to a double exponential with time constants 5.3 and 133 ps and weights of 0.54 and 0.46 for the disappearance of the free NO stretch mode after photolysis.[25,28] The earlier computational study gave time constants $\tau_1 = 3.1$ ps and $\tau_2 = 18.9$ ps, but there a considerably more complex PES was used for both, the unbound ligand (a 3 point charge model and a Morse potential to describe the NO stretch) and the bound state (a two-dimensional PES based on extensive DFT calculations), that allowed for rebinding both in the Fe-NO and Fe-ON orientation.[17,29]

To visualize the crossing seam all observed crossing geometries are projected onto a plane containing the iron nitrogen distance $RFe\text{-}N$ and the angle formed by the iron and the ligand $\theta Fe\text{-}NO$ (see Figure 3, black circles). As in the previous study the most probable iron-ligand separation $RFe\text{-}N$ lies between 3.0 Å and 3.1 Å, but values as large as 4.05 Å occur. The distribution of $\theta Fe\text{-}NO$ is centered around 127° with a width of ≈50°. Figure 3 establishes that the transition between the bound and the unbound state is not a single point in phase space but rather an ensemble of structures. Furthermore, it is clear that $RFe\text{-}N$ and $\theta Fe\text{-}NO$ alone will not provide a good description of the crossing manifold but that the crossing geometry is also modulated by additional environmental degrees of freedom.

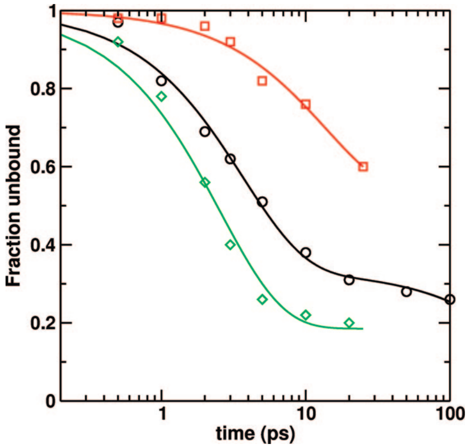


Figure 4. Rebinding curves for the MbNO system calculated as in Figure 2 with three different values of Δ (red squares Δ = 60, black circles Δ = 65, and green diamonds Δ = 70 kcal/mol). The solid lines in respective color represent double exponential fits to the simulated data.

The influence of the asymptotic separation Δ on the observed kinetics is summarized in Figure 4 where $N(t)$ is reported for values of Δ 5 kcal/mol higher and lower than the one chosen above (65 kcal/mol). For these calculations a smaller number $N_{traj} = 50$ of reactive trajectories was used, and the maximum trajectory length was 25 ps. As found previously,[17] the choice of Δ has a marked influence on the observed time scales. However, the existence of two rebinding time scales is independent of the value for Δ. Because the time scale of the present simulations does not fully describe the longer decay constant only the fast time scale τ_1 is analyzed in more detail. The constants τ_1 are 2.6, 3.6, and 13.8 ps for Δ = 70, 65, and 60 kcal/mol, respectively. This reflects an increase of the rebinding barrier (the so-called "inner barrier") with decreasing Δ. It is also instructive to consider the weights w_1 as Δ decreases. They decrease from 0.81 to 0.48. Again, this is a manifestation of the importance of the inner barrier which determines the fast rebinding rate.

The value of Δ not only influences the barrier of the reaction but also potentially changes the ensemble that determines the crossing seam. To investigate the sensitivity of the crossing geometries on Δ the crossing geometries were projected as described above for the different values of Δ (Figure 3). For the current system, the crossing manifold is found to be largely insensitive to the value of Δ which, however, may be different for other applications. For the distributions or $RFe\text{-}N$ no effect can be seen, whereas a weak trend toward a larger average and smaller spread of $\theta_{Fe\text{-}NO}$ can be seen (124° with Δ = 70 kcal/mol to 129° with Δ = 60 kcal/mol).

It is also of interest to consider the dependence on the mixing time T_{mix}. To this end, constant-energy (*NVE*) simulations have been performed, and the total energy was followed as a function of T_{mix}. For very short T_{mix} (≤5 fs), an irregular change of the total energy is observed. This limit represents an abrupt change of the PES during the crossing. The T_{mix} that minimizes the total energy change (≤1 kcal/mol, i.e. "chemical accuracy") for the system studied here is around 10 fs. With increasing T_{mix}, the error in total energy increases during the crossing which is due to the longer time

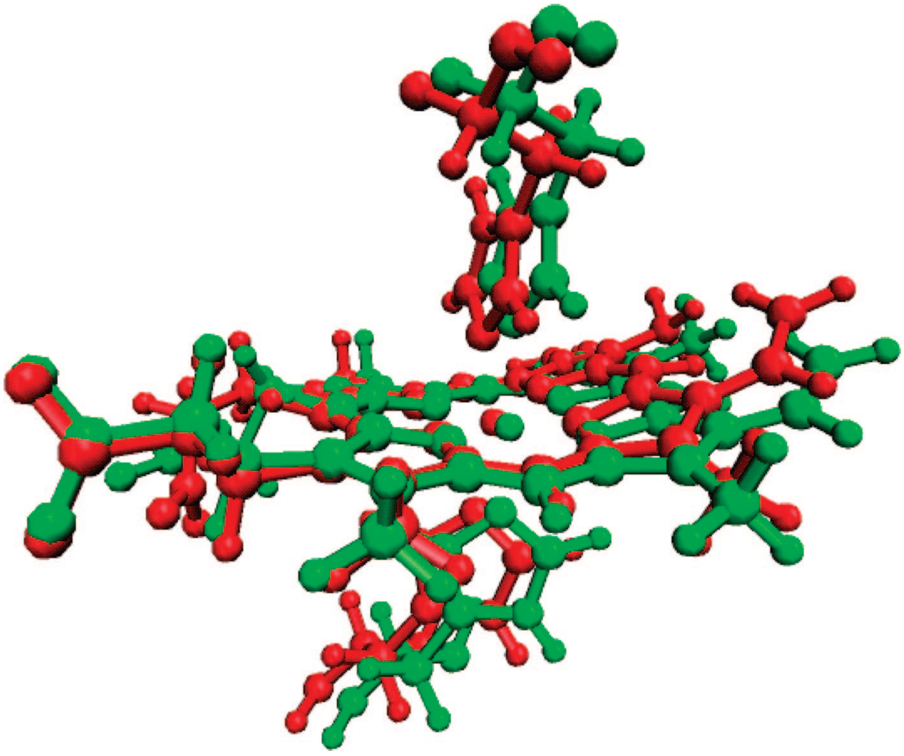


Figure 5. Comparison of the active site structure in neuroglobin for the lower populated state p_2 (red) and the state with higher population p_1 (green). The picture shows the heme group and the two coordinating histidines His64 (above the heme) and His96 (below the heme).

the system spends in the crossing region during the switching. However, even with switching times up to 30 fs, the error is less than 2 kcal/mol which is much smaller than the fluctuations for simulations in the canonical ensemble, which is the usual mode in which ARMD simulations are run. The optimal mixing times found here are comparable to those reported for nonadiabatic reactions using the Coherent Switching with Decay of Mixing (CSDM) which was developed for small molecular systems.[43,44]

Recently, a multidimensional PES for CO-bound and free Mb was presented which includes the effect of spin−orbit interactions.[45] The 6-dimensional PES was fitted to ground-state ab initio energies and the gaps between the ground and the electronically excited states. However, until now, the PESs were not used in dynamics simulations, and, thus, their validity in view of the experimental results (energetics, infrared spectroscopy, rebinding dynamics) for either of the states (bound and unbound) cannot be judged.

3.2. Conformational Equilibrium in Ngb. A more general application of ARMD is the investigation of the equilibrium between two conformational states in Ngb (see Figure 5). In this case at least one of all common bonded force field terms (bonds, angles, and dihedral angles) change between the two states in question. Experimentally, the relative populations and the geometrical features of the conformations involved have been characterized.[33] Furthermore, the two conformational substates are likely to be of biological relevance (allosteric control), which further motivates its understanding with the generalized ARMD procedure. To study the conformational equilibrium, the two states were treated with separate potential energy surfaces. Table 1 summarizes the parameters used for these simulations. This is mainly a convenient way to represent a conformational transition between two well-defined states on a complex potential energy surface. However, the conformational change itself does not involve an electronic transition between two states. This example also serves as an illustration of how Δ can be calibrated from experimental data.

The experimental information available about these two conformational substates is their structures and their relative population (70% and 30%).[33] Consequently, Δ is chosen such as to reproduce the correct occupation of the two states: substate 1 (characterized by PES 1) with 70% population and substate 2 (characterized by PES 2) with 30% population. These two PESs only differ in the force field parameters to correctly describe the experimentally determined structures of substates 1 and 2.

Exploratory simulations showed that the barrier for the conformational change is sufficiently high to make it a “rare” event: typically several hundreds of picoseconds pass between transitions. Thus, the direct sampling of the conformational change and determination of Δ is computationally demanding. To better characterize the transitions, umbrella sampling[46] was chosen as a method to guide the system toward the crossing seam. The progression coordinate, $R_u = R_1/R_2$, was chosen such as to separate the two PESs and still be restricted to a finite interval so that a single harmonic umbrella potential could be used to increase sampling over this interval. Figure 6 shows that $R_u(t)$ switches between two stable positions at 0.85 and 1.01. The switches between these are almost perfectly correlated in time with the crossings as seen from the time series of the energy difference between the two conformational states. It can also

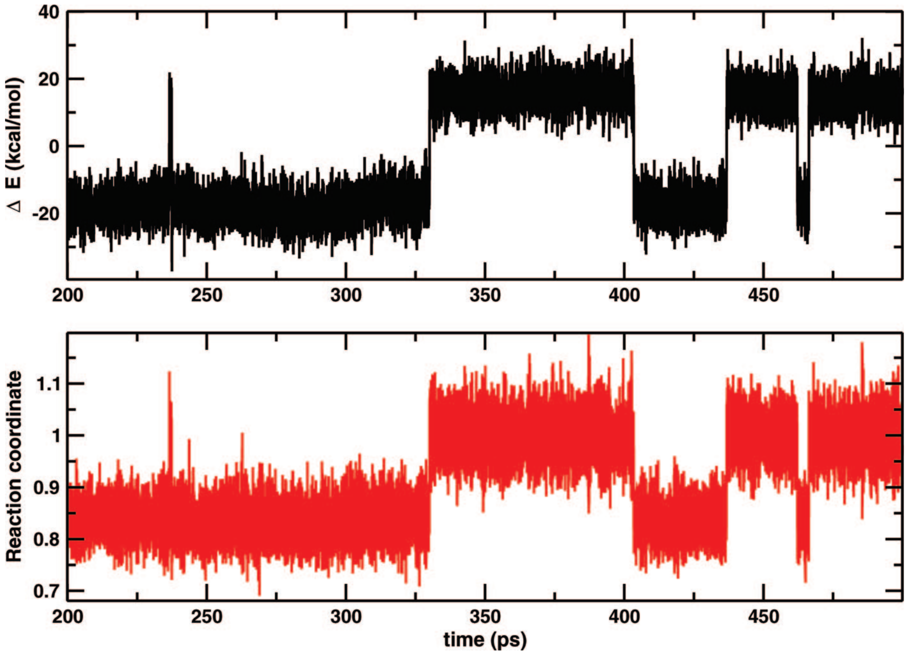


Figure 6. Comparison of the time series for the energy difference $\Delta E = V_1 - V_2$ (above) and the umbrella coordinate R_u (below) defined in the text for the conformational transition in neuroglobin. The correlation between ΔE and the switching of R_u is evident.

be seen in Figure 6 that although R_u clearly can be used to separate the two states, there is a small range around $R_u \approx 0.95$ which is sampled regardless of which state the system is in. Thus, no simple single coordinate can fully describe the transition between the two manifolds, since the energy difference ΔE is also modulated by collective coordinates of the environment, leading to e.g fluctuations of the electrostatic field.

A suitable center for the umbrella potential was determined from a simulation which sampled the conformational transition (see Figure 6) and provided $R_u = 0.93$. Applying the umbrella potential increases the number of transitions from which meaningful statistics for the populations p_1 and p_2 can be estimated. From experiment, the ratio $p_1/p_2 = 70/30$ was determined.[33] The following two-state model is useful to describe the population p_1 as a function of the free energy difference between states 1 and 2. If the free energy difference between the two surfaces at $\Delta = 0$ is $\Delta G(\Delta{=}0) \equiv G = G_1 - G_2$ and the zero of the free energy is assumed for PES 2 ($G_2{=}0$), then the following expression describes the population p_1 on surface 1

$$p_1 = \frac{\exp(-\beta(G+\Delta))}{1+\exp(-\beta(G+\Delta))} = \frac{1}{1+\exp(\beta(G+\Delta))} \tag{9}$$

Thus, a relative population of 70/30 corresponds to a free energy difference of −1.4 kcal/mol between the two states at $T = 300$ K.

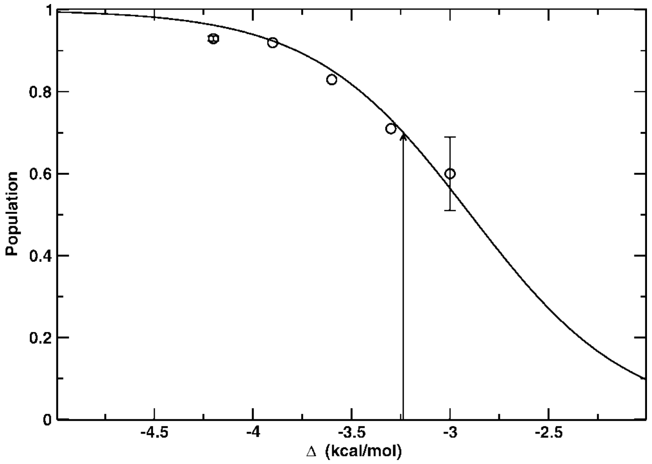


Figure 7. Calculated population of V_1 from umbrella sampling simulations as a function of Δ (circles) and a fit based on the two-state statistical model (solid line). The error bars for $\Delta = -3.0$ and -4.2 kcal/mol are based on the standard deviation from 5 independent simulations. The experimentally observed population $p_1 = 0.7$ is marked by an arrow.

The value of Δ was calibrated as follows. Energy minimizations on the two surfaces V_1 and V_2 give an estimate for the (conformationally unaveraged) energy difference $V_1 - V_2 = 4.8$ kcal/mol, which has the wrong sign compared to experiment. To match $V_1 - V_2 + \Delta$ with the estimated free energy difference from experiment $\Delta = -6.2$ kcal/mol was chosen. Next, short umbrella sampling runs starting from each surface were run which suggested that $\Delta = -6.2$ kcal/mol was too low (trajectories initiated on V_2 cross in less than 1 ps and never cross back). To find a range where an equilibrium could be simulated, Δ values between −3.0 and −4.5 kcal/mol were chosen. Five 500 ps simulations with Δ values of −3.0, −3.3, −3.6, −3.9, and −4.2 kcal/mol were run, each with a harmonic umbrella centered at $R_u = 0.93$ and a force constant of 40 kcal/mol. The population of p_1 for each Δ is plotted in Figure 7 and can be represented by eq 9. However, the data from the simulations contain statistical uncertainty, because only ≈10 crossings occur during each 500 ps trajectory. To better characterize the

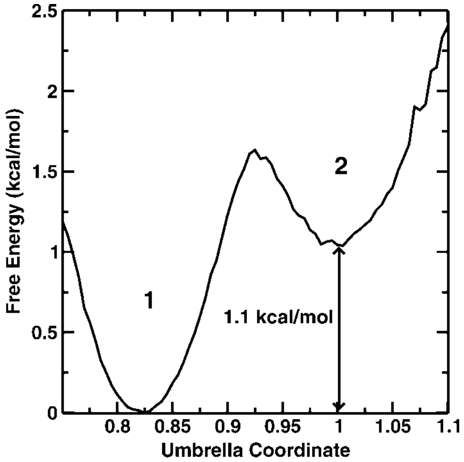


Figure 8. Free energy profile along R_u in neuroglobin from umbrella sampling calculations with $\Delta = -3.3$ kcal/mol. The higher energy state 2 is destabilized by 1.1 kcal/mol relative to the more stable state 1. The forward barrier 1→2 is 1.6 kcal/mol compared to a reverse barrier 2 → 1 of 0.5 kcal/mol.

populations, error bars for $\Delta = -3.0$ and -4.2 kcal/mol were determined from 5 independent simulations of 500 ps each (see Figure 7).

Fitting the simulated data to eq 9 at $T = 300$ K provides an analytical expression for $p_1(\Delta)$. The only free parameter in eq 9 is the free energy difference G at $\Delta = 0$ kcal/mol, which gives $G = 2.9$ kcal/mol. From the fitted curve $\Delta \approx -3.3$ kcal/mol reproduces the experimentally observed population $p_1 = 0.7$. To improve the statistics, the trajectory with $\Delta = -3.3$ kcal/mol was continued for another 500 ps leading to a total of 22 crossings, and the data were used to calculate a free energy profile $G(R_u)$, shown in Figure 8. The free energy difference $\Delta G = 1.1 \pm 0.25$ kcal/mol between the two minima is in quite good agreement with that derived from experiment ($\Delta G = 1.4$ kcal/mol). The error in ΔG is estimated from the error bar of $p(1)$ for $\Delta = 3.0$ kcal/mol (Figure 7). The estimated asymptotic separation $\Delta = 3.3$ kcal/mol could be used in further studies of Ngb to investigate e.g different mutants and their influence on the equilibrium between these two states.

4. Conclusions

The present work has generalized and presented in detail Adiabatic Reactive Molecular Dynamics for studying reactive processes and transitions between potential energy surfaces in large systems where quantum effects are not dominant (adiabatic reactions) or even absent (conformational transition where conformers can be described by different force fields). The implementation was applied to bond breaking/bond formation processes and conformational transitions, and a user interface handles the necessary input. Two parameters (Δ and T_{mix}) set by the user define the method, and the influence of these was analyzed. By applying the method to two realistic and different test cases (myoglobin and neuroglobin) the generality was demonstrated and possible procedures to calibrate the parameters were discussed. Furthermore, ARMD interfaces with most sampling enhancing methods such as targeted MD,[47] metadynamics,[48] replica exchange MD,[49] or umbrella sampling.[46] The latter was used here to investigate the transitions between two conformational substates in neuroglobin, where the barrier between the two states was significantly above the thermal energy.

As a validation for the general ARMD method the rebinding of NO to Mb was considered. The results from the simulations of Mb-NO show that the nonexponential kinetics of ligand rebinding after photodissociation is also observed using the generalized version of ARMD, as was the case with the dedicated implementation.[17] It was found that the fast phase kinetics is controlled by the parameter Δ which determines the inner barrier. The crossing seam is quite similar as the one observed in the earlier study[17] and does not change appreciably for different values of Δ.

The application of ARMD to the conformational equilibrium in neuroglobin establishes that the method can be successfully applied to general reactions where an arbitrary number of force field terms (see eq 1) can be created, deleted, or modified. It was demonstrated how a computational scan of the parameter Δ in combination with a simple two-state model can be used to calibrate ARMD in view of experimental data. This opens the possibility to characterize the reaction at an atomistic level and provides insight which is complementary to experimental investigations.

It is of interest to discuss ARMD in the light of other methods proposed to study reactive processes. An approach related to the one presented here was applied to the rebinding of NO in myoglobin.[50] This work considered a model where the reaction progresses along the Fe−N*NO* separation ρ. To describe the transition between the bound and the unbound state, a switching function depending on ρ was employed. With this approach the correct trends for the picosecond recombination in different Mb mutants at position 29 could be reproduced. A disadvantage is that the approach relies on a geometrical definition of the progression coordinate. It is known that only in the simplest systems (e.g., some proton transfer reactions) a meaningful progression coordinate can be guessed. In more complex systems the relevant coordinates are less obvious, and different choices can lead to different results, as was shown for the isomerization of a tyrosine ring in the bovine pancreatic trypsin inhibitor.[51] Instead of the explicit definition of a (potentially unsuitable) progression coordinate, ARMD performs the switching in time. This opens the possibility to use ARMD to determine suitable progression coordinates a posteriori from the motions between reactants and products.

An alternative procedure employs a Hamiltonian which interpolates between the bound and the unbound system.[52] From a long MD simulation transition state configurations are approximately located, further relaxed (quenched) by minimization, and either end up in the product or the reactant well. From several hundred simulations of unbound Mb-CO the distribution of barrier heights at different temperatures was determined, and results for the width and peak position of the distribution were found to be in qualitative agreement with experiment. Because the approach is primarily based on relaxing the transition state structures it is not clear how to calculate dynamical properties from it. Also, the effect of the surrounding solvent is difficult to account for explicitly with this approach. In fact, the simulations were carried out with implicit solvent.

It is also of interest to briefly discuss similarities and differences between ARMD and EVB/EVB-like approaches. Both methods try to combine the speed of molecular dynamics simulations with added functionality in a force field description of the total energy of a system. Both methods are parametrized and parameter optimization may be a lengthy and arduous process, and their determination is preferably done with respect to experimental data. The starting point for ARMD is that the interactions and dynamics in the reactant (R) and product (P) states of a molecular system can be characterized in a meaningful way using force fields and atomistic simulations. For this, observables (which may be different for R and P) for both end points of the reactions are calculated and compared to experiment. If adjustments to the intermolecular interactions (i.e., the force field parameters) are required, then they can be carried out for R and P independently. The force fields may include additional functionality (such as multipolar representations of particular fragments) for one state which are not deemed important for the other state. In a second step, ARMD is used to find adiabatic transitions between the two states. To obtain a meaningful overall energetics the constant Δ is calibrated in view of experimental data. EVB is based on a superposition of ionic and covalent resonance forms from which a—typically—small secular matrix is constructed.[13] As such, EVB is rooted in the well established concept of valence bond theory. The resonance forms are mixed through off diagonal (coupling) matrix elements, and the resulting matrix is diagonalized for every conformation. In its original form, the various parametrized terms were fitted to experimental data.[13,53] First, the Hamiltonian is parametrized for a system in the gas phase. By leaving the covalent and off-diagonal terms unchanged and only modifying the ionic terms, a Hamiltonian for the system in a solvent environment is set up. The off-diagonal term couples different states, depends on the solute coordinates, and is typically represented as an exponential or a Gaussian function. The parametrization assumes that the off-diagonal term does not change between the reaction in the gas phase and in solution.[54] In summary, while the two methods are related through the aim to optimally explore the speed force fields to follow reactions in complex environments, the underlying procedures to arrive at such a representation are quite different. It should be noted that in ARMD the R and P side of the reaction are parametrized individually, whereas they are coupled in EVB.

Limitations of ARMD as presented here include processes where the quantum nature (e.g., tunneling, nonadiabaticity, coherence) of the reaction dominates. This is most obviously the case in small molecular systems for which other and potentially more suitable approaches have been developed.[15,16,55] An example from the condensed phase is bond-breaking in MbNO where coherent motion arises from the coupling of the electronic degrees of freedom to the nuclear dynamics during the spin-state change.[56] Experimentally, distinct oscillations with periods of 430 and 150 fs were observed which were related to heme doming and the iron-histidine motion. On the other hand, isomerization reactions or ligand rebinding are examples where coherence is expected to be unimportant. An improvement that will allow to more realistically investigate reactions from ARMD is to introduce a coordinate-dependent asymptotic shift $\Delta(\vec{q})$. Here, $\vec{q}$ describes directions (or progression coordinates) which are found to be important from simulations with a coordinate-independent Δ. Once $\vec{q}$ is known, $\Delta(\vec{q})$ can be obtained from electronic structure calculations.

Other interesting problems that could be treated with ARMD include, for example, (1) adiabatic electron transfer, (2) adiabatic atom transfer, or (3) solid state phase transformations. Concerning (1), the charge transfer band of organic analogs of symmetrical transition-metal-centered intervalence compounds have been investigated experimentally.[57] It was found that for such strongly adiabatic systems classical theory without tunneling correction reproduces the observed rate constants. Mixed-valence isomers for which strong coupling has been found by using infrared spectroscopy[58] and class III intervalence compounds[59,60] are additional candidates for which ARMD may be useful. Other interesting systems in this class include electron transport in hematite[61] (where the interaction between the reactant and product electronic state is 0.20 eV, which is consistent with adiabatic electron transfer) or photoinduced adiabatic electron transfer in strongly coupled dye/semiconductor colloidal systems.[62] Regarding (2), ARMD may be used to study the dynamics and energetics of adiabatic atom transfer reactions such as concerted proton—electron transfer in the oxidation of hydrogen-bonded phenols which have been found to be not highly nonadiabatic.[63] Also, it might be of interest to investigate the dynamics of the decomposition reaction $H_2SO_4 \rightarrow SO_3 + H_2O$. The mechanism involved is high vibrational excitation of the OH stretching motion in the electronic ground state.[64] For solid state phase transition (example (3) above) one typically assumes that transformations follow adiabatic pathways. For such processes, ARMD could be applied in much the same way as for the conformational equilibrium in neuroglobin. In this context it is interesting to note that very recently the role of nonadiabaticity in the iron bcc to hcp phase transformation was investigated. These calculations used a nudged elastic band approach to connect the initial and final states with a predescribed reaction coordinate.[65] More detailed investigations including free nuclear dynamics could give additional insight in the reorganization mechanism and the role of nonadiabaticity for such problems.

ARMD in its current formulation assumes that once the crossing between the two potential energy surfaces is reached, the reaction takes place with probability one. However, there are also cases such as long-range electron transfer[66] and nonadiabatic photochemistry[67] where a more detailed knowledge of the crossing region is necessary, or where (nuclear) quantum effects become important. Such extensions are possible by using a probability

$p < 1$ for switching the surface when a crossing is found or by allowing hops between the potential energy surfaces away from the crossing seam. Their development will provide a natural extension of ARMD and open up new applications for studying reactive events using validated atomistic force fields.

Acknowledgment. We thank Profs. M. Karplus and J. D. Doll and Dr. S. Mishra for discussions. This work is supported by the Schweizerischer Nationalfonds Project 200021-117810.

References

(1) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227.

(2) Singh, U.; Kollman, P. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the CH_3Cl+Cl^- exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.* **1986**, *7*, 718.

(3) Field, M.; Bash, P.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700.

(4) Bash, P.; Field, M.; Karplus, M. Free energy perturbation method for chemical reactions in the condensed phase: a dynamical approach based on a combined quantum mechanical and molecular mechanical potential. *J. Am. Chem. Soc.* **1987**, *109*, 8092.

(5) Marti, S.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J.; Field, M. A hybrid potential reaction path and free energy study of the chorismate mutase reaction. *J. Am. Chem. Soc.* **2001**, *123*, 1709.

(6) Shaik, S.; Kumar, D.; Visser, S.; Altun, A.; Thiel, W. Theoretical Perspective on the Structure and Mechanism of Cytochrome P450 Enzymes. *Chem. Rev.* **2005**, *105*, 2279.

(7) Field, M. Simulating enzyme reactions: Challenges and perspectives. *J. Comput. Chem.* **2002**, *23*, 48.

(8) Stewart, J. Semiempirical Molecular Orbital Theory. In *Reviews in Computational Chemistry;* Lipkowitz, K., Boyd, D., Eds.; John Wiley & Sons, Inc.:1990; *Vol 1.*

(9) Elstner, M. The SCC-DFTB method and its application to biological systems. *Theor. Chem. Acc.* **2006**, *116*, 316.

(10) Mei, H. S.; Tuckerman, M. E.; Sagnella, D. E.; Klein, M. L. Quantum nuclear ab initio molecular dynamics study of water wires. *J. Phys. Chem. B* **1998**, *102*, 10446.

(11) Meuwly, M.; Karplus, M. Simulation of proton transfer along ammonia wires: An “ab initio” and semiempirical density functional comparison of potentials and classical molecular dynamics. *J. Chem. Phys.* **2002**, *116*, 2572.

(12) Sauer, J.; Döbler, J. Gas-phase infrared spectrum of the protonated water dimer: Molecular dynamics simulation and accuracy of the potential energy surface. *Comput. Phys. Commun.* **2005**, *6*, 1706.

(13) Warshel, A.; Weiss, R. An Empirical Valence Bond Approach For Comparing Reactions In Solution and In Enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218.

(14) Åqvist, J.; Warshel, A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.* **1993**, *93*, 2523.

(15) Tully, J.; Preston, R. Trajectory surface hopping approach to nonadiabatic molecular collisions: the reaction of H^+ with D_2. *J. Chem. Phys.* **1971**, *55*, 562.

(16) Tully, J. Molecular Dynamics with Electronic Transitions. *J. Chem. Phys.* **1990**, *93*, 1061.

(17) Nutt, D. R.; Meuwly, M. Studying reactive processes with classical dynamics: Rebinding dynamics in MbNO. *Biophys. J.* **2006**, *90*, 1191.

(18) Frauenfelder, H.; McMahon, B.; Austin, R.; Chu, K.; Groves, J. The role of structure, energy landscape, dynamics and allostery in the enzymatic function of myoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2370.

(19) Frauenfelder, H.; McMahon, B. Hydration, slaving and protein function. *Biophys. Chem.* **2002**, *98*, 35.

(20) Gibson, Q. H.; Smith, M. H. *J. Physiol.* **1957**, *136*, P27.

(21) Keyes, M.; Falley, M.; Lumry, R. Studies of Heme Proteins. II. Preparation and Thermodynamic Properties of Sperm Whale Myoglobin. *J. Am. Chem. Soc.* **1971**, *93*, 2035.

(22) Straub, J. E.; Karplus, M. Molecular-Dynamics Study of the Photodissociation of Carbon-Monoxide from Myoglobin - Ligand Dynamics in the 1st 10 ps. *Chem. Phys.* **1991**, *158*, 221.

(23) Traylor, T.; Sharma, V. Why NO? *Biochemistry* **1992**, *31*, 2847.

(24) Snyder, S.; Bredt, D. Biological roles of nitric oxide. *Sci. Am.* **1992**, *266*, 68.

(25) Petrich, J.; Lambry, J.-C.; Kuczera, K.; Karplus, M.; Poyart, C.; Martin, J.-L. Ligand binding and protein relaxation in heme proteins: a room temperature analysis of NO geminate recombination. *Biochemistry* **1991**, *30*, 3975.

(26) Zhu, L.; Sage, J.; Champion, P. Observation Of Coherent Reaction Dynamics In Heme-Proteins. *Science* **1994**, *266*, 629.

(27) Meuwly, M.; Becker, O. M.; Stote, R.; Karplus, M. NO rebinding to myoglobin: a reactive molecular dynamics study. *Biophys. Chem.* **2002**, *98*, 183.

(28) Kim, S.; Jin, G.; Lim, M. Dynamics Of Geminate Recombination of NO with Myoglobin in Aqueous Solution Probed by Femtosecond Mid-IR Spectroscopy. *J. Phys. Chem. B* **2004**, *108*, 20336.

(29) Nutt, D. R.; Karplus, M.; Meuwly, M. Potential energy surface and molecular dynamics of MbNO: Existence of an unsuspected FeON minimum. *J. Phys. Chem. B* **2005**, *109*, 21118.

(30) Ionascu, D.; Gruia, F.; Ye, X.; Yu, A.; Rosca, F.; Beck, C.; Demidov, A.; Olson, J.; Champion, P. Temperature-dependent studies of NO recombination to heme and heme proteins. *J. Am. Chem. Soc.* **2005**, *127*, 16921.

(31) Burmester, T.; Weich, B.; Reinhardt, S.; Hankeln, T. A vertebrate globin expressed in the brain. *Nature* **2000**, *407*, 520.

(32) Sun, Y.; Jin, K.; Peel, A.; Mao, X. O.; Xie, L.; Greenberg, D. Neuroglobin protects the brain from experimental stroke in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3497.

(33) Vallone, B.; Nienhaus, K.; Brunori, M.; Nienhaus, G. The Structure of Murine Neuroglobin: Novel Pathways for Ligand Migration and Binding. *Proteins* **2004**, *56*, 85.

(34) Du, W.; Syvitski, R.; Dewilde, S.; Moens, L.; La Mar, G. Solution ^{1}H NMR Characterization of Equilibrium Heme Orientational Disorder with Functional Consequences in Mouse Neuroglobin. *J. Am. Chem. Soc.* **2003**, *125*, 8080.

(35) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187.

(36) Lee, S.; Warshel, A. A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. *J. Chem. Phys.* **1992**, *97*, 3100.

(37) Ewald, P. 'Die Berechnung optischer and elektrostatischer Gitterpotentiale′. *Ann. Phys.* **1921**, *369*, 253.

(38) 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., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J. E.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586.

(39) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functionsfor simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.

(40) Brooks, C., III; Karplus, M. Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.* **1983**, *79*, 6312.

(41) Van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.

(42) Young, R.; Frauenfelder, H.; Johnson, J.; Lamb, D.; Nienhaus, G.; Phillipp, R.; Scholl, R. Time- and temperature dependence of large-scale conformational transitions in myoglobin. *Chem. Phys.* **1991**, *158*, 315.

(43) Zhu, C.; Nangia, S.; Jasper, A.; Truhlar, D. G. Coherent switching with decay of mixing: An Improved treatment of electronic coherence for non-Born-Oppenheimer trajectories. *J. Chem. Phys.* **2004**, *121*, 7658.

(44) Zhu, C.; Jasper, A.; Truhlar, D. G. Non-Born-Oppenheimer trajectories with self-consisting decay of mixing. *J. Chem. Phys.* **2004**, *120*, 5543.

(45) Margulis, C.; Guallar, V.; Sim, E.; Friesner, R.; Berne, B. A New Semiempirical Approach to Study Ground and Excited States of Metal Complexes in Biological Systems. *J. Phys. Chem. B* **2002**, *106*, 8038.

(46) Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187.

(47) Schlitter, J.; Engels, M.; Krüger, A. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graph.* **1994**, *12*, 84.

(48) Laio, A.; Parinello; M. Escaping free-energy minima. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12562.

(49) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141.

(50) Li, H.; Elber, R.; Straub, J. E. Molecular dynamics simulation of NO recombination to myoglobin mutants. *J. Biol. Chem.* **1993**, *268*, 17908.

(51) Northrup, S. H.; Pear, M. R.; Lee, C.-Y.; McCammon, J. A.; Karplus, M. Dynamical theory of activated processes in globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4035.

(52) Panchenko, A. R.; Wang, J.; Nienhaus, G. U.; Wolynes, P. G. Analysis of ligand binding to heme proteins using a fluctuating path description. *J. Phys. Chem.* **1995**, *99*, 9278.

(53) Coulson, C. A.; Danielsson, U. Ionic and covalent contributions to the hydrogen bond. 1. *Ark. Fys.* **1954**, *8*, 239.

(54) Hong, G. Y.; Rosta, E.; Warshel, A. Using the constrained DFT approach in generating diabatic surfaces and off diagonal empirical valence bond terms for modeling reactions in condensed phases. *J. Phys. Chem. B* **2006**, *110*, 19570.

(55) Miller, W. H.; George, T. F. Semiclassical Theory of Electronic Transtions in Low-Energy Atomic and Molecular Collisions Involving Several Nuclear Degrees of Freedom. *J. Chem. Phys.* **1972**, *56*, 5637.

(56) Zhu, L.; Sage, J. T.; Champion, P. M. *Science* **1994**, *266*, 629.

(57) Nelsen, S. F.; Ismagilov, R. F.; Trieber, D. A., II. Adiabatic electron transfer: comparison of modified theory with experiment. *Nature* **1997**, *278*, 846.

(58) Londergan, C. H.; Salsman, J. C.; Lear, B. J.; Kubiak, C. P. Observation and dynamics of "mixed-valence isomers" and a thermodynamic estimate of electronic coupling parameters. *Chem. Phys.* **2006**, *324*, 57.

(59) Nielsen, S. F.; Weaver, M. N.; Luo, Y.; Lockard, J. V.; Zink, J. I. Use of the neighboring orbital model for analysis of electronic coupling in Class III intervalence compounds. *Chem. Phys.* **2006**, *324*, 195.

(60) D'Alessandro, D. M.; Keene, F. R. Current trends and future challenges in the experimental, theoretical and computational analysis of intervalence charge transfer (IVCT) transitions. **2006**, *35*, 424.

(61) Rosso, K. M.; Smith, D. M. A.; Dupuis, M. An ab initio model of electron transport in hematite (α-Fe_2O_3 basal planes). *J. Chem. Phys.* **2003**, *118*, 6455.

(62) Huber, R.; Moser, J.-E.; Gratzel, M.; Wachtveitl, J. Real-time observation of photoinduced adiabatic electron transfer in strongly coupled dye/semiconductor colloidal systems with a 6 fs time constant. *J. Phys. Chem. B* **2002**, *106*, 6494.

(63) Rhile, I. J.; Markle, T. F.; Nagao, H.; DiPasquale, A. G.; Lam, O. P.; Lockwood, M. A.; Rotter, K.; Mayer, J. M. Concerted proton-electron transfer in the oxidation of hydrogen-bonded phenols. *J. Am. Chem. Soc.* **2006**, *128*, 6075.

(64) Vaida, V.; Kjaergaard, H. G.; Hintze, P. E.; Donaldson, D. J. Photolysis of sulfuric acid vapor by visible solar radiation. *Science* **2003**, *299*, 1566.

(65) Johnson, D. F.; Carter, E. A. Nonadiabaticity in the iron bcc to hcp phase transformation. *J. Chem. Phys.* **2008**, *128*, 104703.

(66) Marcus, R. Theory of Oxidation-Reduction Reactions Involving Electron Transfer 0.1. *J. Chem. Phys.* **1956**, *24*, 966.

(67) Yarkony, D. Current Issues In Nonadiabatic Chemistry. *J. Phys. Chem.* **1996**, *100*, 18612.

CT800066Q

J. Chem. Theory Comput. **2008,** *4,* 1094–1106

JCTC Journal of Chemical Theory and Computation

Calculation of One-Photon and Two-Photon Absorption Spectra of Porphyrins Using Time-Dependent Density Functional Theory

Paul N. Day,*[†,‡] Kiet A. Nguyen,[†,§] and Ruth Pachter*[†]

Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio 45433, General Dynamics Information Technology, Inc., Dayton, Ohio 45431, and UES, Inc., Dayton, Ohio 45432

Received March 6, 2008

Abstract: Time-dependent density functional theory has been used to calculate the one-photon and two-photon absorption spectra of free-base porphyrin, a substituted zinc porphyrin, and a zinc porphyrin dimer, in order to assess the validity of the method to reproduce the large increase in the two-photon absorption (TPA) cross-section for the dimer. Three hybrid functionals with varying amounts of exact exchange were tested, and the calculated one-photon absorption spectra for each of the molecular systems were shown to be in qualitative agreement with the measured spectra. All three functionals predict a large enhancement in the TPA cross-section for the dimer relative to the monomer, in agreement with experimental results. However, because of the sensitivity of the resonance enhancement factor to small differences in the relevant state energies, quantitative prediction of the TPA cross-section by this method is still a challenge.

I. Introduction

Porphyrin systems are important in a variety of biological functions; for example, because of their combination of biocompatibility and nonlinear optical properties, they have been used in photodynamic therapy.[1–4] Photofrin,[5] a porphyrin oligomer, has been approved for the treatment of some forms of cancer. Porphyrins have long been studied also for nonlinear absorption, such as reverse-saturable absorption[6–8] due to excited-state absorption, but were not found useful for two-photon absorption (TPA). However, a remarkable enhancement in the TPA cross-section in porphyrin dimers linked by ethyne or butadiyne has been observed in recent studies.[9–11]

Drobizhev et al.[11] measured the one-photon absorption (OPA) and TPA spectra of a porphyrin monomer, [5,15-bis(3,5-bitert-butylphenyl)-10,20-bis(trihexylsilylethynyl)porphyrinato]zinc, labeled yPy, and several porphyrin dimers, the simplest of which is 5,5′-(ethyne)bis[[10,20-bis(3,5-bitert-butylphenyl)-porphyrinato]zinc], labeled PyP (see Figure 1). The maximum TPA cross-section measured for PyP was about 350 times the maximum TPA cross-section of yPy. In this work, we aim to assess time-dependent density functional theory (TDDFT) for the prediction of OPA and TPA spectra for the porphyrin monomer and dimer, in particular, the enhancement of the TPA cross-section for PyP as compared to yPy.

For comparison, we also discuss the simplest porphyrin, free-base porphin (FBP, see Figure 1), for which numerous excited-state studies have been carried, in part due to the availability of experimental gas-phase UV−vis spectra.[12] In addition, results for the Q band are available that were measured in a supersonic expansion,[13] and they are in good agreement with previous measurements in the gas phase.[12] The theoretical methods that have been used for FBP include complete-active-space with second-order perturbation corrections (CASPT2),[14–16] symmetry-adapted-cluster configuration interaction (SAC-CI),[15,17–20] similarity-transformed equation-of-motion coupled-cluster (STEOM-CC),[21] quantum Monte Carlo (QMC),[22] the semiempirical CNDO-π-SCF-MO-PSDCI method,[23] and TDDFT.[15,24–29] In this paper, the body of spectral data on FBP will be reviewed

---

* Corresponding authors. E-mail: Paul.Day@wpafb.af.mil (P.N.D.); Ruth.Pachter@wpafb.af.mil (R.P.).

[†] Air Force Research Laboratory

[‡] General Dynamics Information Technology, Inc.

[§] UES, Inc.

10.1021/ct800080w CCC: $40.75

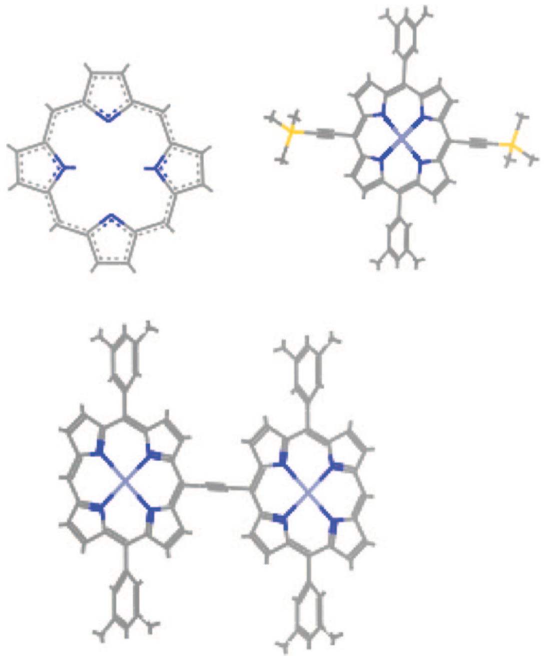


Figure 1. Structure of FBP, yPy, and PyP.

and expanded upon, with the goal of determining the appropriate exchange-correlation (x-c) density functional and basis set to be used to provide reliable spectra for larger porphyrins.

II. Theory

The TPA cross-section is calculated by the expression given previously,[30–33]

$$\delta_{f0}(E_1+E_2) = \frac{8\pi^4}{(ch)^2} E_1 E_2 g(E_1+E_2) |S_{f0}(u_1, u_2)|^2 \tag{1}$$

where g is a line width function, $|S_{f0}(u_1,u_2)|^2$ is the two-photon probability corresponding to a transition from the ground state (0) to a final state (f),[23,34–41]

$$|S_{f0}(u_1, u_2)|^2 = \left|\sum_i^N \left[\frac{(u_1 \cdot \mu_{i0})(\mu_{fi} \cdot u_2)}{E_i - E_1 + i\Gamma_i} + \frac{(u_2 \cdot \mu_{i0})(\mu_{fi} \cdot u_1)}{E_i - E_2 + i\Gamma_i}\right]\right|^2 \tag{2}$$

and E_1 and $E_2$are the energies of the two photons with unit polarization vectors u_1 and u_2, respectively. The transition dipole moments are given by μ_{ij}, and the state energies and line widths are given by E_i and Γ_i, respectively. As we noted previously,[32,33] electronic spectra can be obtained from linear response (LR) TDDFT, while TPA cross-sections require the application of quadratic response theory. The two-photon probabilities can be evaluated by the above sum-over-state (SOS) expression, or directly from the single residue of the quadratic response (SRQR)[42–44] in TDDFT. To carry out the SOS, the ground-to-excited-state transition dipole moments needed are obtained from LR TDDFT, while the excited-state to excited-state moments are obtained from the double residue of the quadratic response (DRQR).

The line shape can be represented by a Gaussian function:

$$g^{G}(E_1+E_2) = \left(\frac{4h^2 \ln 2}{\pi E_{\mathrm{FWHM}}^2}\right)^{\frac{1}{2}} \exp\left[\frac{-4 \ln 2}{E_{\mathrm{FWHM}}^2}(E_1+E_2-E_f)^2\right] \tag{3}$$

or by a Lorentzian:

$$g^{L}(E_1+E_2) = \frac{h}{\pi}\frac{E_{\mathrm{FWHM}}/2}{(E_1+E_2-E_f)+(E_{\mathrm{FWHM}}/2)^2} \tag{4}$$

where E_{FWHM} is the full-width at half-maximum for the final state f, and E_f is the energy of state f relative to the ground state. Note that, when a Gaussian function is used, the calculated peak TPA cross-section is 1.48 times larger than for a Lorentzian, for a given line width.

The TPA probability is dependent on the orientation of the molecule relative to the incident photons, so orientationally averaged expressions have been derived for various beam orientations and polarizations.[39] To correspond with experimental results, the expression used here is for photons that are linearly polarized with parallel polarization.

TPA processes can be divided into type I and type II.[32,45,46] In centrosymmetric molecules, the ground state has gerade symmetry, and only states with ungerade symmetry are one-photon-allowed, while only states of gerade symmetry are two-photon-allowed. The TPA in centrosymmetric molecules is of type I, where the state i in eq 2 must be of ungerade symmetry in order to have a nonzero transition dipole moment with the gerade ground state, and the final state f is of gerade symmetry and thus has a nonzero transition dipole moment with state i. The type I TPA cross-section can be estimated by the three-state approximation, which assumes that a single term dominates the sum in eq 2, and in the case where the two photons have the same energy (E_λ), the cross-section is given by

$$\delta^{\mathrm{I}}_{f0} = \frac{32\pi^4 g_{\max}}{15(ch)^2}\frac{E_\lambda^2}{(E_i - E_\lambda)^2}|\mu_{0i}|^2|\mu_{if}|^2(2\cos^2\Theta_{\mu\mu} + 1) \tag{5}$$

where $g_{\max}$, the maximum in the line width function, is obtained by setting $E_f = E_1 + E_2 = 2E_\lambda$ in eq 3 or 4, and $\Theta_{\mu\mu}$ is the angle between the two transition dipole moment vectors. While the three-state approximation is often not adequate for quantitative results, it can be useful in analyzing the origin of the TPA intensity.

Type II TPA occurs in noncentrosymmetric molecules where there is a change in the dipole moment between the ground state and the final state. In this case, the sum may be dominated by the terms from these two states, and a two-state approximation to the TPA cross-section may be used:

$$\delta^{\mathrm{II}}_{f0} = \frac{32\pi^4 g_{\max}}{15(ch)^2}|\mu_{0f}|^2|\Delta\mu_{0f}|^2(2\cos^2\Theta_{\mu\Delta\mu} + 1) \tag{6}$$

where $\Delta\mu_{0f} = \mu_{ff} - \mu_{00}$, and $\Theta_{\mu\Delta\mu}$ is now the angle between the transition dipole moment and the dipole difference vectors. Although not all of the systems in this study are strictly centrosymmetric, such as the yPy(1p) molecular system, they have enough symmetry that type I TPA is the dominant process.

III. Computational Methods

The geometry of each molecular system was optimized using the B3LYP functional with the 6-31G** basis set. Geometry optimizations and some LR TDDFT calculations were carried out with Gaussian 03.[47] LR, SRQR, and DRQR TDDFT calculations were carried out with the Dalton program.[48] The DRQR calculations were used to investigate the minimal-states approximations and also to check the TPA cross-sections obtained by the SRQR method, and good convergence of the SOS in the DRQR method was found with less than 20 excited states. All of the calculated TPA cross-sections reported here use the more efficient SRQR method. Gaussian line-width functions were used to compute the TPA cross-sections, with the line widths obtained from the experimental TPA spectra for yPy and PyP.[11] For FBP, no experimental line width is available, so we used a line width of 0.26 eV to be consistent with the previous theoretical study[23] and because it is a reasonable approximation to the TPA line width observed in the experimental studies on FBTPP.[1,49] Generalized gradient approximation (GGA) functionals such as BLYP[50,51] were tested for the TDDFT calculations but were found to be unsatisfactory due to the number of so-called "ghost" states produced near the position of the strong B band. These "ghost" states, which are clearly artifacts of the functional as based on results from both experiments and other levels of theory, have been previously noted with the use of GGA functionals for FBP.[29]

The functionals used for the TDDFT calculations were B3LYP,[50–54] CAMB3LYP,[55–58] and mCAMB3LYP.[33,59] The B3LYP functional is the most widely used "hybrid" functional, where a fraction of exact exchange (0.2 in this case) is combined with a GGA exchange functional to improve its behavior in the asymptotic region. However, to further improve the prediction accuracy of charge-transfer states and Rydberg states, a long-range corrected (LC) functional was recently introduced[60] to increase the fraction of exact exchange in the asymptotic region. This functional was further improved by modification of it into a three-parameter model called the coulomb-attenuated model (CAM).[55] This model used the same Becke exchange[50] and Lee, Yang, and Parr correlation (LYP)[51] used in the B3LYP functional to get the CAMB3LYP[55] functional, which has a fraction of exact exchange that varies from 0.19 (similar to the value in B3LYP) at small separations to 0.65 ($\alpha + \beta$) at large separations (using the default parameter values $\alpha = 0.19$, $\beta = 0.46$, and $\mu = 0.33$). The parameter μ, which scales the separation in the argument to the error function, was kept fixed at 0.33, which had been optimized by Tawada et al.[60] for the first to third row atoms. Other modified forms have been introduced, such as the LC-ωPBE functional of Vydrov and Scuseria,[61] but in a recent assessment,[62] none were found that proved to be superior to CAMB3LYP. However, for many molecules with intramolecular charge transfer, carrying out TDDFT with the standard CAMB3LYP functional (where $\beta = 0.46$) yields excitation spectra in poorer agreement with experimental results than when the B3LYP functional is used, so a modified version of this functional was introduced and tested (mCAMB3LYP)[33,59] with $\beta = \alpha = 0.19$, making it a two-parameter LC model with the maximum fraction of exact exchange equal to 0.38.

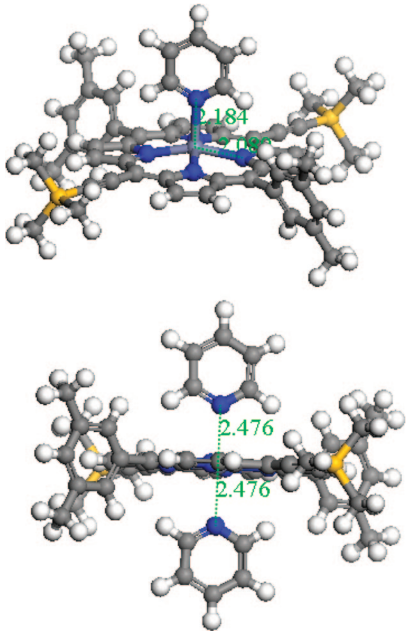


Figure 2. Structure of yPy(1p) and yPy(2p).

Two continuum solvation models were tested, namely, the self-consistent reaction field model with a spherical cavity (SCRF-S) by Mikkelson et al.,[63,64] which has been implemented in Dalton 2.0,[48] and the polarizable continuum model (PCM),[65,66] which was used previously[67] with TDDFT to improve the agreement with measured spectra of substituted porphyrins in chloroform. The PCM was used in both Gaussian 03[47] and in a developmental version of Dalton.[68] The internally stored parameters for the experimental solvent of interest, dichloromethane (aka methylene chloride), were used and are consistent between the two programs. The value used for the static dielectric constant was 8.93, and for the optical-frequency dielectric constant, a value of 2.02 was used. In our reported calculations, model compounds were used with methyl groups substituted for the hexyl and *tert*-butyl groups on yPy and PyP. Note that the solvent in the experimentally measured OPA and TPA spectra for the zinc porphyrin systems was dichloromethane with 1% pyridine,[11] presumably to prevent aggregation by coordinating with the zinc atoms in the porphyrin. In order to simulate the effect of the pyridine, the geometry of yPy was reoptimized with one and two pyridine molecules coordinated to the zinc atom to form the structures yPy(1p) and yPy(2p), respectively (see Figure 2). At the given level of theory (B3LYP/6-31G**), both of these structures are about 15 kcal/mol more stable than the noncoordinated molecules. However, the bond between the pyridine and the Zn atom in yPy(1p) is shorter (2.1 Å) and stronger than the corresponding bonds (2.5 Å) in yPy(2p). Results will be reported for the bare yPy molecule and for yPy(1p), to assess the effects of a coordinating ligand.

If the isolated PyP molecule is held flat, enforcing D_{2h} symmetry, the resulting stationary point has several imaginary frequencies, and thus it is not a minimum, but is 3.1

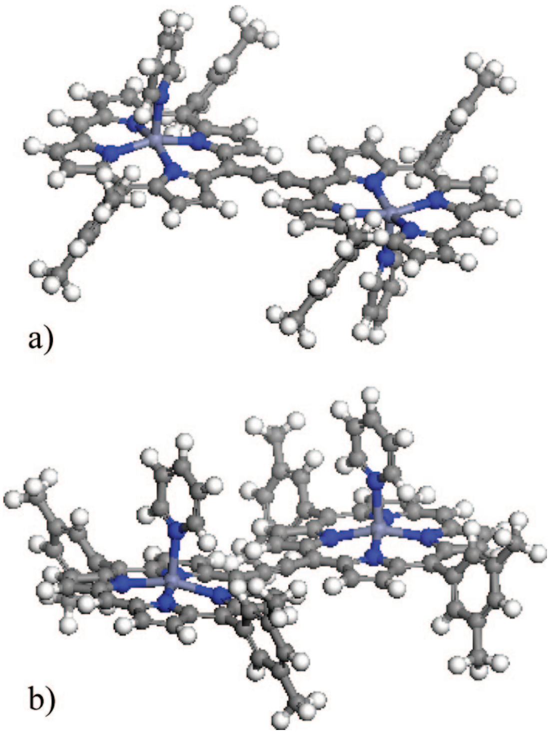


Figure 3. Structure of PyP(2p) with symmetry (a) C_{2h} and (b) C_{2v}.

kcal/mol higher in energy than the minimum energy structure. In the minimum energy structure, the two porphyrin planes are twisted 36° from each other, as was found previously at the AM1[11,69] and B3LYP[70] levels for this molecule. Structures with two pyridine molecules coordinated with PyP, one to each zinc atom, as is shown in Figure 3, have been optimized in C_{2v} symmetry (with the two pyridine molecules on the same side) and in C_{2h} symmetry (with the pyridine molecules on opposite sides). The C_{2h} structure was found to be 1.5 kcal/mol lower in energy than the C_{2v} structure, so it was used in further calculations reported here. A TDDFT calculation was carried out also for the C_{2v} structure using the B3LYP functional, and the OPA spectra are nearly identical to the spectra obtained using the C_{2h} geometry. However, the C_{2h} geometry has several small imaginary frequencies and is still about 0.8 kcal/mol higher in energy than the minimum energy geometry, which has the two porphyrin planes twisted by 31°. Results for both the C_{2h} geometry and the minimum-energy C_1 geometry will be presented here. While the results from the C_1 geometry are expected to be in better agreement with experimental results, the results from the C_{2h} geometry should give the maximum nonlinear absorption to be expected for this molecule, as the planar structure maximizes conjugation and π-orbital delocalization. This may be closer to what would be seen in a solid crystal. Also, the calculations in the higher symmetry aid in identifying and classifying the excited states. Coordinates for all structures are given in the Supporting Information.

The largest calculations carried our here, the quadratic TDDFT on PyP(2p) with no symmetry, took 96 h on 32 processors on an SGI Altix. The linear TDDFT calculations could be carried out at approximately one-fourth the cost of the quadratic TDDFT. Also, calculations on the porphyrin monomer system yPy(1p) required only about 20% of the computational effort of the porphyrin dimer system.

IV. Results and Discussion

A. Free-Base Porphyrin (FBP). *1. OPA.* The calculated excitation energies and oscillator strengths for free-base porphyrin (FBP), also known as porphin, are listed in Table 1 and compared to experimental results and previous calculations. The results for the B3LYP and CAMB3LYP functionals agree with those reported previously[15,26,29] using the 6-31G(d) basis set (although the assignment of the B and N states with B3LYP was reversed by Cai et al.[15]). The results with the 6-311G(d,p) basis show that the basis set effects are small, as was previously shown for this molecule using TDDFT with nonhybrid density functionals.[25,29]

The spectra of porphyrins is typically interpreted through the Gouterman four-orbital model (GFOM).[12,71] In FBP, which has D_{2h} symmetry, the two highest occupied molecular orbitals, the HOMO ($5b_{1u}$) and HOMO-1 ($2a_u$), and the two lowest unoccupied molecular orbitals, LUMO ($4b_{3g}$) and LUMO+1 ($4b_{2g}$) are significantly separated in energy from the other orbitals and play a major role in explaining the spectral features. The HOMO and HOMO-1 are nearly degenerate as are the LUMO and LUMO-1, and transitions between these four orbitals are used to explain the Q and B bands in porphyrin spectra. For FBP, each band consists of "*x*" (excitation parallel to the inner hydrogen−hydrogen axis with B_{3u} symmetry, see Figure 1) and "*y*" (perpendicular to the inner hydrogen−hydrogen axis with B_{2u} symmetry) components. At the lower energy, the transitions interfere destructively, and thus the Q band is very weak, while the B band is strong due to constructive interference.

The vapor-phase excitation energies predicted using the B3LYP functional are in excellent agreement with experimental results for both the Q and B bands, consistent with previous calculations,[15,24,26,29] while in the CAMB3LYP results, the excitation energy of the B band is slightly overestimated, similar to that obtained in the high-level ab initio methods SAC-CI,[15] CASPT2,[15] and STEOM-CC.[21] The measured intensity of the Q band is several orders of magnitude larger than calculated, likely due to vibronic coupling, an effect not included in the calculations. The total oscillator strength calculated for the B band using B3LYP (1.04) is in good agreement with the measured value (1.15), while CAMB3LYP overestimates the oscillator strength by about 70%, and CASPT2,[15] STEOM-CC,[21] and CAS-CI[15] overestimate it by 50%, 65%, and 110%, respectively. The mCAM results are intermediate to B3LYP and CAMB3LYP, overestimating the B-band oscillator strength by about 40%.

In the experiment, the N band is a small shoulder on the B band with an estimated maximum at 3.65 eV, while in the calculations, the N band is a distinct peak of intensity comparable to the B band (depending on the functional used). The N band is still close in energy to the B band, and it appears there could be some mixing of the orbitals involved in these transitions. The major contributions to each of these excited states from each orbital transition for the three

Table 1. Excitation Energies (ΔE in eV) and Oscillator Strengths (f) for Porphin Calculated by TDDFT and Compared to Other Levels of Theory and Experiment

			1^1B_{3u}	1^1B_{2u}	2^1B_{3u}	2^1B_{2u}	3^1B_{2u}	3^1B_{3u}
	basis		Q_x	Q_y	B_x	B_y	N_y	N_x
CNDO-CI[23]		ΔE	1.85	2.53	3.23	3.49	3.98	3.83
		f	0.00	0.06	0.32	0.73	0.88	1.30
B3LYP[24]	SVP	ΔE	2.27	2.43	3.32	3.49		
B3LYP[29]	SVP	ΔE	2.24	2.39	3.27	3.45	3.70	3.79
		f	0.00	0.00	0.40	0.61	0.55	0.82
B3LYP[15,26]	6-31G*	ΔE	2.28	2.44	3.34	3.51	3.77	3.87
		f	0.00	0.00	0.42	0.63	0.51	0.78
B3LYP[a]	6-31G*	ΔE	2.28	2.44	3.33	3.51	3.77	3.86
		f	10^{-6}	10^{-5}	0.41	0.63	0.51	0.78
B3LYP[a]	6-311G**	ΔE	2.27	2.42	3.31	3.48	3.74	3.83
		f	10^{-6}	10^{-5}	0.44	0.67	0.48	0.76
B3LYP/PCM[a,b]	6-311G**	ΔE	2.29	2.43	3.23	3.33	3.74	3.73
		f	10^{-4}	10^{-4}	0.94	1.27	0.20	0.61
CAMB3LYP[15]	6-31G*	ΔE	2.19	2.42	3.55	3.69	4.50	4.28
		f	0.00	0.00	0.81	1.16	0.07	0.59
CAMB3LYP[a]	6-31G*	ΔE	2.19	2.42	3.55	3.69	4.50	4.27
		f	10^{-3}	10^{-3}	0.80	1.16	0.07	0.59
CAMB3LYP[a]	6-311G**	ΔE	2.17	2.39	3.52	3.65	4.46	4.24
		f	10^{-3}	10^{-3}	0.83	1.17	0.07	0.57
CAM/PCM[a,b]	6-311G**	ΔE	2.19	2.38	3.37	3.44	4.49	4.19
		f	10^{-3}	10^{-3}	1.30	1.49	0.06	0.43
mCAM[a]	6-31G*	ΔE	2.27	2.45	3.44	3.61	4.05	4.03
		f	10^{-4}	10^{-4}	0.60	1.00	0.19	0.69
mCAM[a]	6-311G**	ΔE	2.25	2.43	3.41	3.57	4.02	4.00
		f	10^{-4}	10^{-4}	0.63	1.02	0.18	0.67
mCAM/PCM[a,b]	6-311G**	ΔE	2.27	2.43	3.29	3.38	4.05	3.92
		f	10^{-7}	10^{-3}	1.13	1.40	0.11	0.50
SAC-CI[15]	6-31G*	ΔE	1.71	2.10	3.46	3.61	4.19	4.09
		f	0.00	0.00	1.00	1.49	0.36	0.89
SAC-CI[15]	6-31G	ΔE	1.80	2.29	3.62	3.77	4.39	4.22
		f	0.00	0.00	1.10	1.64	0.27	0.85
CASPT2[15]	6-31G	ΔE	1.84	2.24	3.56	3.12	3.95	3.72
		f	0.00	0.01	0.98	0.76	0.51	0.03
CASPT2[14]	SV(d)	ΔE	1.63	2.11	3.12	3.08	3.42	3.53
		f	0.00	0.00	0.70	0.91	0.46	0.83
SAC[17]	SV(d)	ΔE	1.77	2.01	3.47	3.73	4.38	4.20
		f	0.00	0.01	0.77	1.62	0.34	1.32
STEOM-CC[21]	DZP	ΔE	1.75	2.40	3.47	3.62	4.35	4.06
		f	0.00	0.01	0.69	1.20	0.42	0.93
QMC[22]		ΔE (eV)		2.46				
		f						
measured[12]		ΔE (eV)	1.98	2.42	3.33		3.65	
		f	0.01	0.06	1.15		0.10	
measured[13]		ΔE (eV)	2.02	2.47				
Measured (in CH_2Cl_2)[12]		ΔE (eV)	2.02	2.39	3.15			

[a] Current work. [b] The PCM solvation model for CH_2Cl_2 was included.

functionals are given in Table 2. The two GFOM transitions are listed first for each state. For all three functionals, the first two excited states have only GFOM transitions as significant contributers, thus clearly defining the Q band.

In a recent study,[15] when the B3LYP functional was used, the third and fourth excited states were assigned to the N band, while the fifth and sixth excited states were assigned to the B band, which is the opposite of the assignments in three earlier B3LYP studies.[24,26,29] The assignments here are consistent with the earlier studies.[24,26,29] When the B3LYP functional is used, a transition from a non-Gouterman orbital (HOMO-2) has the largest contribution to the third excited state, but the two GFOM transitions have comparable contributions, and this is identified as the B_x state. The situation is similar for the fourth excited state, where the two GFOM transitions and a transition from HOMO-2 all have significant and comparable contributions, and this state is identified as B_y. The next two states have smaller contributions from the GFOM transitions, with larger contributions from the previously mentioned transitions from HOMO-2, so these states are labeled N_y and N_x. When the CAMB3LYP functional is used, the contributions from the GFOM transitions for the third and fourth excited states are clearly dominant, leaving no doubt that this is the B band, and there is much less mixing of the GFOM transitions into the N band. The results using the mCAM functional are intermediate relative to the other two functionals, but the GFOM transitions are still the largest contributers to the third and fourth excited states, verifying their assignment to the B band, and there is less mixing of the GFOM into the N-band states than when B3LYP was used. The effect of an applied external field on the calculated states was also tested, in a manner similar to that carried out previously,[15] and while the results showed the B3LYP states to be affected more by

Table 2. Weight Percent Contribution of Each Orbital Transition for Each Excited State of Porphin[a]

		CASPT2[14]	B3LYP[b]	CAMB3LYP[b]	mCAM[b]
1^1B_{3u}	**Q_x**	**1.630**	**2.281**	**2.186**	**2.268**
H → L + 1	$5b_{1u} \rightarrow 4b_{2g}$	0.370	0.293	0.270	0.283
H − 1 → L	$2a_u \rightarrow 4b_{3g}$	0.340	0.204	0.231	0.214
1^1B_{2u}	**Q_y**	**2.110**	**2.442**	**2.417**	**2.454**
H → L	$5b_{1u} \rightarrow 4b_{3g}$	0.410	0.272	0.252	0.264
H − 1 → L + 1	$2a_u \rightarrow 4b_{2g}$	0.390	0.225	0.246	0.233
2^1B_{3u}	**B_x**	**3.120**	**3.335**	**3.553**	**3.436**
H → L + 1	$5b_{1u} \rightarrow 4b_{2g}$	0.130	0.111	0.180	0.144
H − 1 → L	$2a_u \rightarrow 4b_{3g}$	0.280	0.191	0.238	0.221
H − 2 → L + 1	$4b_{1u} \rightarrow 4b_{2g}$	0.270	0.294	0.084	0.136
2^1B_{2u}	**B_y**	**3.080**	**3.511**	**3.693**	**3.607**
H → L	$5b_{1u} \rightarrow 4b_{3g}$	0.210	0.177	0.253	0.232
H − 1 → L + 1	$2a_u \rightarrow 4b_{2g}$	0.300	0.195	0.256	0.253
H − 2 → L	$4b_{1u} \rightarrow 4b_{3g}$	0.170	0.136	0.007	0.028
3^1B_{2u}	**N_y**	**3.420**	**3.770**	**4.495**	**4.045**
H → L	$5b_{1u} \rightarrow 4b_{3g}$		0.053		0.008
H − 1 → L + 1	$2a_u \rightarrow 4b_{2g}$	0.070	0.083	0.005	0.018
H − 2 → L	$4b_{1u} \rightarrow 4b_{3g}$	0.320	0.357	0.452	0.459
H − 7 → L	$3b_{1u} \rightarrow 4b_{3g}$	0.150	0.012	0.033	0.013
3^1B_{3u}	**N_x**	**3.530**	**3.865**	**4.274**	**4.031**
H → L + 1	$5b_{1u} \rightarrow 4b_{2g}$	0.220	0.101	0.062	0.081
H − 1 → L	$2a_u \rightarrow 4b_{3g}$	0.090	0.108	0.041	0.071
H − 2 → L + 1	$4b_{1u} \rightarrow 4b_{2g}$	0.290	0.294	0.377	0.347

[a] Excited state energies (in eV) are given in bold. [b] Current work.

the applied field than the CAMB3LYP states, the results did not motivate a change in the assignment of the calculated states. Note, however, that, while the B3LYP functional gives the most accurate excitation energy and intensity for the B band, it also yields a more intense N band, which is qualitatively incorrect. The CAMB3LYP functional overestimates the excitation energy and intensity of the B band more than the other two functionals, and thus the mCAM functional could be a compromise for the prediction of spectra in porphyrins.

The effect of a solvent on the spectra was investigated using the PCM with each of the three density functionals, and comparing the results to the spectra measured for FBP in dichloromethane.[12] The PCM successfully predicts the very small blue shift of the Q_x state, the very small red-shift of the Q_y state, and the red-shift of nearly 0.2 eV for the B band. When the PCM is included in the calculation, the strength of the B band is significantly larger than that of the N band, which is in better qualitative agreement with experimental results.

Table 3. Calculated TPA Resonance Transition Energies (ΔE, in eV) and the Corresponding TPA Cross-Sections (σ, in GM) for FBP, Using the 6-31G* Basis, with Comparison to Measured TPA Peak Cross-Sections for a Similar Molecule, FBTPP[a]

		$1A_g$	$2A_g$	$3A_g$	$4A_g$	$5A_g$
measured[1,49]	ΔE	3.25				
(FBTPP)	σ	25				
CNDO-CI[23]	ΔE	3.26	3.46	3.82		
reported	σ	13.76	1.05	585.71		
corrected	σ	6.88	0.53	292.85		
B3LYP	ΔE	3.61	4.27	4.44	4.85	5.19
	σ	0.28	17.28	1.53	0.45	13.94
CAMB3LYP	ΔE	4.38	4.87	5.28	5.42	5.92
	σ	157451	137.67	4.13	3.00	3.57
mCAMB3LYP	ΔE	3.93	4.54	4.79	5.08	5.47
	σ	0.55	136.52	3.82	0.08	7.07

		$1B_{1g}$	$2B_{1g}$	$3B_{1g}$	$4B_{1g}$	$5B_{1g}$
CNDO-CI[23]	ΔE	3.13	3.49	3.84	3.91	
reported	σ	1.58	0.61	23.27	50.33	
corrected	σ	0.79	0.31	11.63	25.16	
B3LYP	ΔE	3.42	4.05	4.25	4.68	5.23
	σ	0.00	11.44	0.15	0.03	9.82
CAMB3LYP	ΔE	3.88	4.77	4.87	5.29	6.09
	σ	0.02	2.82	222.52	2.60	9.83
mCAMB3LYP	ΔE	3.61	4.38	4.49	4.93	5.62
	σ	0.01	13.57	0.66	13.41	32.55

[a] A Gaussian lineshape was used with FWHM = 0.26 eV. The CNDO-CI[23] cross-sections have been corrected for consistency with our definition (see text).

2. *TPA*. The maximum TPA cross-sections for FBP calculated using the B3LYP, CAMB3LYP, and mCAMB3LYP functionals are given in Table 3, along with the calculated results from a previous study[23] using CNDO-π-SCF-MO-PSDCI and SOS and the measured peak TPA cross-sections for a similar molecule, free-base meso-tetraphenylporphin (FBTPP).[1,49] The CNDO study[23] used an alternative definition of the TPA cross-section, which gives values twice as large as with our convention, as we have mentioned previously;[32,33] thus, their reported values have been corrected for this factor for direct comparison to our results. The calculated TPA cross-section using B3LYP of 17.3 GM for the A_g state at 4.27 eV is similar in magnitude to the peak TPA cross-section measured for FBTPP.[1,49] However, when the CAMB3LYP and mCAMB3LYP functionals are used, resonance effects dominate the TPA cross-section. When CAMB3LYP is used, the computed A_g TPA state at 4.38 eV is nearly exactly resonant with the computed Q_x state at 2.19 eV. Similarly for mCAMB3LYP, a TPA state with A_g symmetry is computed at 4.54 eV, which is nearly perfectly resonant with the computed Q_x state at 2.27 eV. The calculation of such large, unphysical TPA cross-sections for these states illustrates the importance of accurately

Table 4. Calculated Excitation Energies (ΔE in eV) and Oscillator Strengths (f) for yPy Compared to Experiments[a]

solvent[b]		basis		Q_x	Q_y	B_y	B_x	
none	CV-INDO-CI[70]		ΔE	2.07		3.15	3.19	
			f	0.03		1.73	2.09	
none	B3LYP	6-31G**	ΔE	2.17	2.19	3.07	3.16	
			f	0.19	0.00	1.22	1.47	
1p	B3LYP	6-31G**	ΔE	2.09	2.14	2.99	3.11	
			f	0.21	0.00	0.99	1.14	
1p + PCM	B3LYP	6-31G**	ΔE	2.05	2.14	2.87	2.99	
			f	0.31	0.00	1.31	1.41	
none	CAMB3LYP	6-31G**	ΔE	2.18	2.20	3.33	3.34	
			f	0.14	0.01	1.37	1.77	
1p	CAMB3LYP	6-31G**	ΔE	2.07	2.14	3.19	3.26	
			f	0.17	0.00	1.24	1.47	
1p + PCM	CAMB3LYP	6-31G**	ΔE	2.04	2.13	3.06	3.13	
			f	0.26	0.00	1.52	1.69	
none	mCAMB3LYP	6-31G**	ΔE	2.21	2.23	3.20	3.25	
			f	0.17	0.00	1.24	1.64	
1p	mCAMB3LYP	6-31G**	ΔE	2.10	2.16	3.06	3.19	3.13
			f	0.20	0.00	0.85	1.19	0.45
1p + PCM	mCAMB3LYP	6-31G**	ΔE	2.07	2.16	2.95	3.05	
			f	0.30	0.00	1.41	1.55	
CH_2Cl_2	measured[11]		ΔE	1.92	2.07	2.73	2.80	
			f	0.15	0.00	0.66	0.94	

[a] The isolated yPy molecule has C_{2h} symmetry, and all the excited states reported here have B_u symmetry. When a pyridine molecule is coordinated to yPy, the symmetry is reduced to C_2 and the symmetry of the excited states is B. [b] 1p denotes that one pyridine molecule is coordinated with the Zn atom, as shown in Figure 2. PCM denotes the polarizable continuum model with the parameters for CH_2Cl_2 as the solvent.

obtaining the state energies. Further improvements are needed in the (x-c) functional to be applied for the calculation of TPA cross-sections in porphyrins, as we also demonstrate for yPy and PyP.

B. yPy. *1. OPA.* The addition of the ethynyl groups to the basic zinc porphyrin reduces the symmetry from D_{4h} to D_{2h}, red-shifts both the Q and B bands, and increases the Q-band intensity. While the Q band is still weak compared to the B band, the quasidegeneracy of the Gouterman four orbitals is substantially perturbed by this additional conjugation to remove the near-perfect cancelation of the transition dipoles in the Q band. The role the intensity of the Q band plays in the strength of the TPA can be seen in examining the three-state approximation (eq 5); one of the Q-band states is the primary intermediate state for the lowest TPA bands. Table 4 lists the calculated excitation energies and oscillator strengths for the Q and B states in comparison to experimental results. As was noted by Zhu et al.,[70] the x and y labels in the experimental spectra[11] of yPy were reversed, as the Q_x line is the lower-energy and more intense of the Q states, where the molecule has been aligned such that the x axis is the long molecular axis. The isolated yPy molecule has C_{2h} symmetry, and all of the reported OPA states are of B_u symmetry, while in yPy(1p), the symmetry is reduced to C_2 and the OPA states have B symmetry. As with porphin, the computed excitation energies and intensities are larger than the experimental ones.

The inclusion of a single pyridine molecule coordinated to the zinc atom in yPy produces a significant red-shift. The SCRF-S model[63] in Dalton 2.0 has little effect on the calculated excitation energies, but the PCM model produces a significant red-shift, improving the agreement with experimental results. Note that the inclusion of either of the continuum solvation models increases the calculated intensity. Without including PCM, the calculated B band using the CAMB3LYP functional for yPy(1p) is blue-shifted nearly 0.5 eV from experimental results, compared to about 0.3 eV using the B3LYP functional, with the mCAM result about midway between the two. Previously reported excitation energies calculated using CV-INDO-MRDCI[70] for the isolated yPy are quite close to our results using CAMB3LYP on yPy(1p), also overestimating the B-band energies by about 0.5 eV. The PCM reduces our error to about 0.3 eV for CAMB3LYP and less than 0.2 eV for B3LYP. The calculated Q-band excitation energy is closer to experimental results and less dependent on the functional, as all three functionals predict this excitation energy to be nearly 0.2 eV higher than measured with good agreement with the experimental intensity. When the PCM is included, the error in this energy is reduced to about 0.1 eV, but the predicted intensity is increased to nearly double that found experimentally. In the recently reported CV-INDO-CI calculations, the intensity of the Q band was severely underestimated,[70] while the intensity of the B band was substantially overestimated.

2. TPA. The calculated transition energies for the TPA states and the corresponding resonant TPA cross-sections for yPy and yPy(1p) are listed in Table 5, and the calculated TPA cross-sections for yPy(1p) fitted to Gaussian line shapes are plotted in comparison to experimental results in Figure 4. The isolated yPy molecule has C_{2h} symmetry, and the listed TPA states have A_g symmetry, while yPy(1p) has C_2 symmetry and the TPA states have A symmetry. While the B_g states in C_{2h} symmetry and the B states in C_2 symmetry are not strictly TPA-forbidden, the calculated TPA cross-sections for these states are much smaller than those reported in Table 5.

The recently reported TPA cross-sections calculated using CV-INDO-MRDCI[70] used a Lorentzian line-shape function with a FWHM of 0.2 eV, while we used a Gaussian line-shape function with a FWHM of 0.45 to correspond with experimental results. In order to directly compare the calculated[70] TPA cross-sections with ours, the values have been converted to our line shape by multiplying by the factor 1.48 × 0.2/0.45. Both the reported and corrected values are reported in Table 5, while Figure 4 shows only the corrected values. After correction, the first TPA peak at 3.04[70] is in good agreement with experimental results, which is surprising considering the underestimation of the Q-band intensity by this method and its importance in the three-state approximation of the TPA cross-section. A much larger TPA cross-section at 3.78 eV was also reported.

While the reported measured maximum TPA cross-section is about 20 GM at a transition energy of about 2.9 eV (corresponding to degenerate photons at 1.45 eV), the QRSR-TDDFT results predict a larger cross-section at a higher transition energy. Each of the three functionals predicts a peak between 3.8 and 3.9 eV with a cross-section between 100 and 200 GM, as well as a much larger peak in the 4.0−4.5 eV range. The calculated TPA states between 4.09 and 4.31 eV have very large cross-sections due to resonance enhancement from the Q_x state located at about 2.1 eV.

Table 5. Calculated TPA Resonance Transition Energies (ΔE, in eV) and the Corresponding TPA Cross-Sections (σ, in GM) for yPy[a]

solvent[b]								
none	CV-INDO-MRDCI[70]	ΔE	3.04	3.78				
	reported	σ	46	406				
	corrected (see text)	σ	30	266				
none	B3LYP/6-31G**	ΔE	3.34	3.43	3.47	3.62	3.68	
		σ	28.3	0.02	27.5	0.23	107.8	
SCRF	B3LYP/6-31G**	ΔE	3.34	3.43	3.45	3.62	3.68	
		σ	44.4	0.04	35.1	0.34	167.8	
1p	B3LYP/6-31G**	ΔE	3.28	3.32	3.38	3.50	3.60	3.61
		σ	0.26	14.1	4.7	66.0	1.4	4.1
		ΔE	3.61	3.70	3.78	3.82	3.89	3.90
		σ	0.71	50.3	7.6	3.5	15.7	34.9
		ΔE	3.96	3.98	4.14	4.16	4.23	4.31
		σ	1.5	58.73	4.2×10^5	2.7×10^6	1.7×10^4	1082
1p	mCAM/6-31G**	ΔE	3.55	3.61	3.70	3.82	3.92	3.98
		σ	1.8	4.5	0.05	101	8.4	4.6
		ΔE	4.02	4.09	4.11	4.17	4.22	4.24
		σ	2.0	5360	8330	1.3×10^5	1.2×10^9	3.2×10^6
1p	CAMB3LYP/6-31G**	ΔE	3.83	3.90	4.28	4.29	4.31	
		σ	76.6	74.6	5.9×10^7	1.0×10^5	5.4×10^4	
CH_2Cl_2	measured[11]	ΔE	2.93					
		σ	23.7					

[a] The isolated yPy molecules have C_{2h} symmetry, and the symmetry of the listed TPA states is A_g. When the yPy is coordinated with a pyridine molecule, the symmetry of the system is reduced to C_2, and that of the TPA states is reduced to A symmetry. We used a Gaussian lineshape with FWHM = 0.45 eV, except for the states above 4.0 eV, where we used FWHM = 0.03 eV to minimize resonance effects on the rest of the spectrum. [b] See footnote to Table 4.

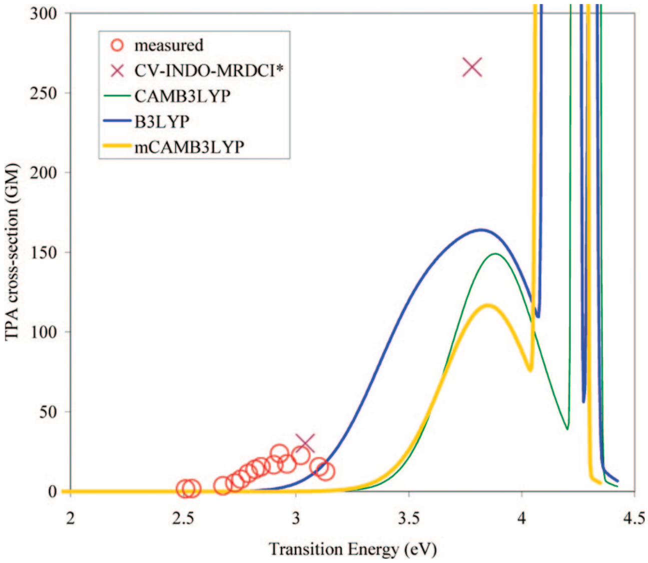


Figure 4. TPA for yPy. The TDDFT calculations have a single pyridine molecule coordinated to the zinc atom. The basis set is 6-31G**, and FWHM = 0.45 eV, except for states above 4.0 eV, where FWHM = 0.03 was used to prevent these states from obscuring the spectra from the lower energy states. The CV-INDO-CI[70] results have been corrected to our line shape function as described in the text.

Experimental TPA is not reported for this range, as OPA will dominate in this energy range. While the experimental FWHM of 0.45 eV has been used for the lower-energy TPA states, a FWHM of 0.03 eV was used for the states between 4.09 and 4.31 eV to prevent them from completely obscuring the lower-energy TPA states. As can be seen in Figure 4, the use of B3LYP results in the highest peak cross-section (164 GM) between 3 and 4 eV, and also the broadest peak. This is due to its being the sum of four major TPA states at 3.50 eV (66 GM), 3.70 eV (50 GM), 3.90 eV (35 GM), and 3.98 eV (59 GM), as well as several minor TPA states. By contrast, the mCAMB3LYP peak of 117 GM is primarily from one TPA state at 3.82 eV (101 GM), and the CAMB3LYP peak of 149 GM is from just two TPA states at 3.83 eV (77 GM) and 3.90 eV (75 GM). This illustrates a major difference in the results from the three functionals: B3LYP yields 14 TPA states below 4.05 eV, mCAMB3LYP yields seven TPA states in this range, and CAMB3LYP gives just two.

The calculated peak TPA cross-section is 5−7 times larger than the measured one and is blue-shifted about 0.8 eV. This overprediction of the TPA cross-section can be attributed to the overprediction of the energy of the two-photon state, thus resulting in a large resonance enhancement factor in eq 5. With some rearrangement, the resonance enhancement factor can be written as

$$\eta = \left(\frac{E_i}{E_f} - 0.5\right)^{-2} \tag{7}$$

where E_i is the energy of the Q_x state, and E_f is the energy of the TPA state. This factor calculated from experimental energies is only 42, but for the four major TPA states calculated using B3LYP, it is 109, 249, 842, and 1659, while for the major TPA state calculated using mCAMB3LYP, it is 396. The CAMB3LYP results illustrate how sensitive to the state energies the calculated TPA cross-section can be due to this resonance enhancement factor, as its value is 576 for the TPA state at 3.83 eV, but for the nearby TPA state at 3.90 eV, it is 1020.

While using the SCRF-S solvation model does not significantly change the calculated state energies, the more accurate PCM can be used to give more accurate state energies. All three functionals, including the PCM, cause a

Table 6. OPA Spectra for PyP Calculated with TDDFT and the 6-31G Basis Set[a]

solvent	functional/sym.		Q_x	Q_x'	B_x	B_x'	N_x	N_x'	M_x	M_x'
none	CV-INDO-CI[70]	ΔE	2.03		2.81	3.60				
		f	0.18		3.24	0.76				
none	B3LYP/D_{2h}	ΔE	1.93		2.80	3.31	2.89		3.47	
		f	0.93		1.01	0.64	0.00		0.14	
2p	B3LYP/C_{2h}	ΔE	1.77	2.46	2.73	3.14	2.86	3.18	3.22	3.36
		f	0.94	0.00	0.84	0.39	0.00	0.07	0.13	0.02
2p	B3LYP/C_1	ΔE	1.84	2.52	2.66	3.12	2.86	3.17	3.19	3.43
		f	0.82	0.00	0.46	0.63	0.00	0.00	0.49	0.00
2p + PCM	B3LYP/C_{2h}	ΔE	1.69	3.05	2.70	3.12	2.89		3.14	3.37
		f	1.20	0.00	0.94	0.49	0.00		0.30	0.01
2p	CAMB3LYP/C_{2h}	ΔE	1.87	3.50	3.00	3.80	4.03	3.70	4.04	4.23
		f	0.71	0.00	1.92	0.31	0.01	0.00	0.03	0.06
2p	CAMB3LYP/C_1	ΔE	1.95		3.01	3.68	3.72			
		f	0.59		1.83	0.24	0.25			
2p + PCM	CAMB3LYP/C_{2h}	ΔE	1.81	3.67	2.94	3.71	4.06	4.04	4.08	
		f	0.94	0.01	2.17	0.28	0.00	0.01	0.01	
2p	mCAMB3LYP/C_{2h}	ΔE	1.83	2.88	2.86	3.46	3.32	3.39	3.70	3.74
		f	0.87	0.02	1.31	0.47	0.00	0.01	0.12	0.00
2p	mCAMB3LYP/C_1	ΔE	1.91		2.83	3.36	3.33	3.39	3.41	
		f	0.75		1.02	0.50	0.01	0.32	0.20	
2p + PCM	mCAMB3LYP/C_{2h}	ΔE	1.76	3.47	2.81	3.40	3.35	3.35	3.68	
		f	1.13	0.03	1.50	0.50	0.02	0.02	0.04	
CH_2Cl_2	experiment	ΔE	1.74	2.16	2.57	2.78				
		f	0.32	0.08	0.76	0.39				

solvent	functional/sym.		Q_y	Q_y'	B_y	B_y'	B_y''	N_y	N_y'	M_y
none	B3LYP/D_{2h}	ΔE	2.15	2.88	2.60	3.42	3.45			
		f	0.01	0.00	0.22	0.08	1.85			
2p	B3LYP/C_{2h}	ΔE	2.01	2.79	2.51	2.95	3.32	3.40	3.01	3.11
		f	0.06	0.17	0.13	0.00	1.60	0.13	0.00	0.04
2p	B3LYP/C_1	ΔE	2.06	2.76	2.51	2.98	3.31		2.52	3.11
		f	0.06	0.16	0.05	0.00	1.58		0.01	0.06
2p + PCM	B3LYP/C_{2h}	ΔE	2.02	2.77	2.49	2.89	3.21	3.40		3.45
		f	0.09	0.29	0.27	0.00	2.11	0.02		0.00
2p	CAMB3LYP/C_{2h}	ΔE	2.16	3.62	3.15	3.55	3.86	4.00	4.10	4.15
		f	0.05	0.47	1.38	0.00	0.70	0.06	0.00	0.02
2p	CAMB3LYP/C_1	ΔE	2.19	3.61	3.21	3.56	3.84	3.59		
		f	0.04	0.33	1.38	0.00	0.52	0.16		
2p + PCM	CAMB3LYP/C_{2h}	ΔE	2.16	3.61	3.04	3.49	3.83	4.15		4.01
		f	0.06	0.50	2.08	0.00	0.60	0.15		0.03
2p	mCAMB3LYP/C_{2h}	ΔE	2.12	3.12	2.80	3.21	3.47	3.68	3.44	3.54
		f	0.06	0.30	0.41	0.00	0.95	0.10	0.01	0.61
2p	mCAMB3LYP/C_1	ΔE	2.16	3.10	2.83	3.23			2.93	
		f	0.05	0.36	0.26	0.00			0.00	
2p + PCM	mCAMB3LYP/C_{2h}	ΔE	2.12	3.10	2.76	3.15	3.41	3.68	3.77	
		f	0.09	0.46	0.81	0.00	1.61	0.04	0.05	
CH_2Cl_2	experiment	ΔE	2.01	2.24	2.64	2.83	3.01			
		f	0.03	0.11	0.38	0.19	1.16			

[a] The excitation energies (ΔE) are in eV. For the cases where the solvent includes two pyridine molecules (denoted "2p"), they are coordinated on opposite sides of the molecule, as shown in Figure 3a.

red-shift of about 0.12 eV for the B band and a red-shift of about 0.04 eV for Q_x. However, the energy shift is much smaller for the TPA states; when the CAMB3LYP is used, each of the first five TPA states is red-shifted by 0.02 eV, while with B3LYP and mCAMB3LYP, we observe similarly small shifts, some in each direction. We conclude that both continuum solvation models predict a significantly larger peak TPA cross-section at about the same transition energy.

C. PyP. *1. OPA.* The calculated and experimental excitation energies and corresponding oscillator strengths for PyP are reported in Table 6. Extinction coefficients are plotted in Figure 5 for the B3LYP functional. Drobizhev et al.[11] reported four excited electronic states with "*x*" polarization (corresponding to the B_u states in the C_{2h} system PyP(2p)) and five excited electronic states with "*y*" polarization (corresponding to the A_u states). The reported combined oscillator strengths[11] of 0.35 for the Q band and 2.91 for the B band were resolved by Gaussian fits into the oscillator strengths reported in Table 6. An oscillator strength of 0.11 for the $Q_x(0-1)$ state at 1.83 eV was combined with an oscillator strength of 0.21 for the $Q_x(0-0)$ state at 1.74 eV.

The B3LYP results with the two different basis sets, 6-31G and 6-31G**, show a small basis set effect, with no change in the position of the Q band and just a small red-shift of the most intense line in the B band with the addition of the polarization functions. There is also a small decrease in the intensity of the Q band and a small increase in the intensity of the B band.

As with yPy, the calculated spectra tend to be blue-shifted and more intense than the measured data. However, the

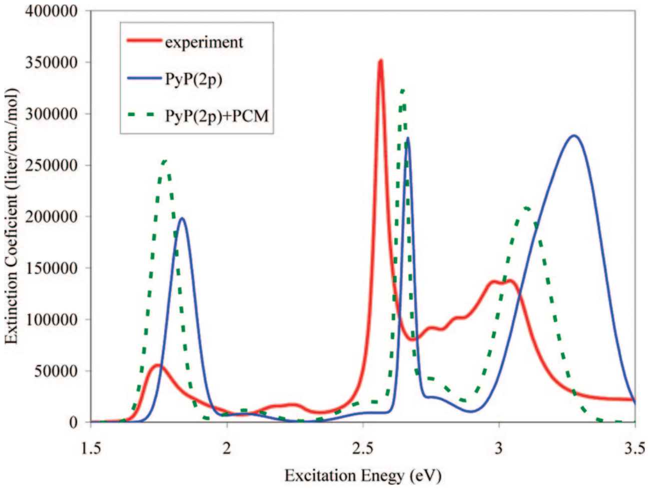


Figure 5. Calculated OPA of PyP(2p) with the B3LYP functional and 6-31G basis set in C_1 symmetry both with and without the PCM solvation model, with comparison to experimental results.

excitation energy calculated for the Q_x state of PyP(2p) with the B3LYP functional using C_{2h} symmetry (either with or without PCM) is in excellent agreement with experimental results, although the calculated oscillator strength (0.9 in the gas phase, 1.1 with PCM) is about 3 times the experimental value (0.32). Allowing the system to relax to the C_1 minimum-energy geometry causes a small blue-shift of 0.1 eV and a small decrease in the oscillator strength to 0.8. Since the amount of charge transfer involved in the Q band is small, the effect of the twisting out of plane of the two porphyrin units and consequent loss in conjugation is also small. Using CAMB3LYP blue-shifts the Q band by about 0.1 eV and reduces the oscillator strength by about 25%. The mCAMB3LYP results fall in between those obtained with B3LYP and CAMB3LYP, with all three functionals in good agreement with the measured Q-band excitation energy while overestimating the oscillator strength. In the previously reported CV-INDO-CI results,[70] the Q_x band position was overestimated by 0.3 eV and the oscillator strength was underestimated by 50%.

The B_x line calculated using B3LYP is in good agreement with experimental results, while the CAMB3LYP functional overestimates the excitation energy and oscillator strength for this transition. Relaxing the symmetry constraint to C_1 substantially decreases the oscillator strength for this line when B3LYP is used, and this intensity appears to be transferred to the B_x' state. Relaxing the symmetry constraint has little effect when the functional is CAMB3LYP; the B3LYP functional predicts more charge transfer in the B-band states than does the CAMB3LYP functional, and thus the loss of conjugation in the C_1 structure has a bigger effect on the B3LYP results. The increased long-range exact exchange in the CAMB3LYP functional decreases charge transfer between the porphyrin units; however, the B3LYP results are in better agreement with the experimental absorption spectra. The results from mCAMB3LYP are again intermediate relative to the other two functionals. In the recent CV-INDO-CI study,[70] the excitation energies calculated for the B_x and B_x' states are higher than experimental results by 0.2 and 0.8 eV, respectively, and the oscillator strengths are overestimated by factors of 4 and 2.

The assignment of the "*y*" polarization states in the B band may seem surprising since the most intense state calculated using B3LYP/PCM (at 3.15 eV) and the most intense state calculated using CAMB3LYP/PCM (at 2.99 eV) are assigned to different states. However, each of the calculated states assigned to the B_y'' line is dominated by the transition from the highest-occupied B_g orbital to the lowest unoccupied B_u orbital. By comparing the results obtained with the mCAMB3LYP functional, we can see how the introduction of the additional asymptotic exchange in the CAM functional induces a transfer of intensity from the B_y'' state to the B_y state. While the states calculated using different functionals can be matched using the calculated molecular orbital transitions, the assignment of the calculated states to experiments is carried out using just the symmetry group, the transition energy, and the oscillator strength. While there are significant differences between the B3LYP and CAMB3LYP for some states, the assignments in Table 6 seem to be the most likely solution.

2. TPA. The TPA resonant cross-sections for PyP are given in Table 7, and fitted TPA cross-sections have been plotted in Figure 6. The CV-INDO-CI results[70] have again been corrected for the different line-shape functions used in that study and ours, this time by multiplying by 1.48 × 0.20/0.22. Calculated results are given for both the bare PyP (in D_{2h} symmetry) and the solvated PyP(2p) (in both C_{2h} and C_1 symmetries) using the B3LYP, CAMB3LYP, and mCAMB3LYP functionals. Experimentally, a strong TPA peak is observed reaching 8500 GM near a transition energy of 3.0 eV. The CV-INDO-CI study[70] reported a peak near this energy, but with only about one-fourth the intensity. They also predicted a peak with an intensity similar to that found experimentally, but with a transition energy of 3.7 eV. Using B3LYP for the bare PyP, two smaller TPA peaks are calculated, one about 0.5 eV lower in energy than experimental results and the other 0.5 eV higher than experimental results. Including the two pyridines in the calculation induces a red-shift in the lower TPA peak, and while the position of the second peak does not change, the red-shift in the Q-band state (Table 6) brings it into resonance with this state, thus causing an unphysically large result for this TPA cross-section. As can be seen in Figure 6, there is little difference in the peak at 2.25 eV when the C_1 geometry is used, and the 20 TPA states calculated in C_1 geometry extend to only 3.11 eV, thus missing the large resonance near 3.49 eV. When CAMB3LYP is used to model the bare molecule, the first TPA peak is blue-shifted relative to experimental results, but when the two pyridine molecules are included, excellent agreement with experimental results is obtained. A peak TPA cross-section of about 15 000 GM is predicted near a transition energy of 3.06 eV for the C_{2h} geometry, but the experimental results did not report TPA above 3.0 eV, and the TPA cross-section appears to still be increasing at that energy. When the C_1 geometry is used, there is little change in the transition energy, but the peak TPA cross-section is reduced to about 9000 GM, which also appears to agree well with the available experimental data. Much of this reduction

Table 7. TPA Resonances of PyP with Cross-Sections Greater than 100 GM[a]

solvent									
none		CV-INDO-CI[70]	ΔE	2.96	3.70				
		reported	σ	1755	8022				
		corrected, see text	σ	2354	10762				
none	D_{2h}	B3LYP	ΔE	2.39	2.44	3.51			
			σ	604	1869	561			
SCRF	D_{2h}	B3LYP	ΔE	2.39	2.44	3.49			
			σ	1082	3333	1235			
none	D_{2h}	CAMB3LYP	ΔE	3.25	3.77	3.71	4.05	4.07	4.13
			σ	11234	270	373	41505	1.1E+06	4.3E+05
2p	C_{2h}	B3LYP	ΔE	2.25	2.29	3.36	3.49		
			σ	1904	431	168	1.5E+06		
2p	C_{2h}	CAMB3LYP	ΔE	3.06	3.48	3.57	3.60	3.71	
			σ	15052	174	1693	506	1.0E+06	
					4.00	4.03	4.06	4.33	
					143	102	270	18135	
2p	C_{2h}	mCAMB3LYP	ΔE	2.58	3.46	3.75			
			σ	4330	156	52267			
2p	C_1	B3LYP	ΔE	2.25	2.28				
			σ	1935	289				
2p	C_1	CAMB3LYP	ΔE	3.08	3.59	3.59			
			σ	9087	257	142			
2p	C_1	mCAMB3LYP	ΔE	2.59					
			σ	3565					
CH_2Cl_2	experiment		ΔE	3.02					
			σ	8510					

[a] A Gaussian lineshape was used with FWHM = 0.22 eV, and the 6-31G basis set was used. The second column lists the symmetry used in the calculation.

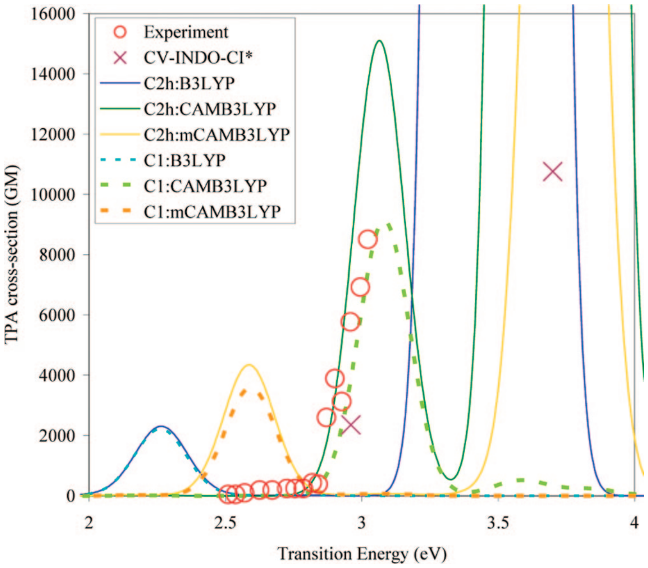


Figure 6. TPA of PyP(2p) calculated using the 6-31G basis and FWHM = 0.22 eV. The experimental data do not extend to transition energies above 3.02 eV. Also, the C_1 calculations do not include enough excited states to reach the highly resonant region of the spectrum. The CV-INDO-CI[70] TPA cross-sections have been corrected to be consistent with our lineshape function (see text).

in the calculated cross-section can be attributed to the decrease in the resonance enhancement factor η (eq 6) of 33%, due to the 0.1 eV blue-shift of the Q_x state (see Table 6). A similar blue-shift in Q_x in the calculation using B3LYP decreases η by 18%, but other factors must compensate for this (perhaps beyond the three-state approximation), since there is no decrease in the calculated TPA cross-section. As the OPA calculated with CAMB3LYP is not in such good agreement with experimental results, the excellent agreement with experimental results of the TPA spectra calculated with CAMB3LYP is still under investigation.

While the B3LYP and mCAMB3LYP functionals underestimate the peak TPA cross-section by factors of 4 and 2, respectively, the main source of the discrepancy appears to be the resonance enhancement factor, as it was in the overestimation of the TPA cross-section of yPy. Most of the enhancement in the TPA cross-section in changing from the monomer to the dimer is in the increase in the magnitudes of the most relevant transition dipole moments. The three-state approximation, eq 5, can be used to understand the source of this enhancement by analyzing the term in the sum with the major contribution to the cross-section. The excited-state transition dipole moment, μ_{if}, can be obtained from a QRDR calculation, while the transition dipole moments from the ground state, μ_{0i}, are obtained from the LR-TDDFT calculations, as previously mentioned. In calculations for the isolated yPy and PyP molecules using the B3LYP functional with the 6-31G basis and the Q_x state as the intermediate state, the two transition dipole moments are 5.1 and 2.3 D in yPy (corresponding to OPA and TPA states at 2.22 and 3.80 eV, respectively), while they are 11.3 and 12.2 D for PyP (for OPA and TPA states at 1.93 and 2.44 eV). When the product of the squares of these transition dipole moments is evaluated, we observe that this factor predicts a TPA cross-section enhancement factor of about 140 in comparing the monomer to the dimer. The actual three-state approximations in these two cases are 66 GM for yPy (compared to 82 GM for this state in Table 5) and 1565 GM for PyP (compared 1870 GM in Table 7). Thus, TDDFT with the B3LYP functional can qualitatively predict, using the three-state model, the enhancement of the TPA that has been observed experimentally; however, once again, the resonance enhancement factor is highly sensitive to the exact state energies, and none of the functionals tested has consistently yielded the accuracy needed for these state energies.

IV. Conclusions

Theoretical OPA and TPA spectra have been calculated with TDDFT for FBP, yPy, and PyP and compared to experimental results. The B3LYP functional was most effective at predicting the OPA, as the CAMB3LYP functional generally overestimated the excitation energy. The overestimation of excited-state energies may be attributed to the additional stabilization of the occupied orbitals that is a result of the inclusion of more exact exchange. It is possible that in CAMB3LYP the exchange and correlation have not been properly balanced. Thus, an improved and rigorous parametrization, not yet carried out, is needed and is currently in progress.

The large enhancement of the TPA cross-section for the porphyrin dimer as compared to the porphyrin monomer, observed experimentally, was predicted by the calculations. The prediction of quantitatively correct TPA cross-sections seems to be limited by the accuracy with which the most relevant state energies can be predicted. The prediction of TDDFT/B3LYP that the transition energy for the relevant TPA state in yPy is about 0.5 eV larger than observed experimentally led to a 5-fold overprediction of the TPA cross-section for this molecule. For PyP, the CAMB3LYP functional accurately predicts the TPA cross-section due to the accuracy of the calculated transition energy, while the underestimation of this transition energy by 0.5 eV with mCAMB3LYP and by 0.8 eV with B3LYP leads to an underestimation of the TPA cross-section by factors of 2 and 4, respectively.

Finally, although the results show the need for improved x-c functionals, at the same time, these calculations qualitatively agree with experimental results by predicting the large TPA enhancement for the porphyrin dimer. Because of the sensitivity of the TPA cross-section to small changes in state energies, a method which will accurately predict state energies is necessary for quantitative prediction of TPA spectra.

Acknowledgment. Computational resources for this research have been provided by the Air Force Research Laboratory High Performance Computing Major Shared Resource Center. We would like thank the Agren group for providing a modified version of the Dalton program which included the CAM-B3LYP (x-c) functional and the PCM solvation model.

Supporting Information Available: The Cartesian coordinates of all of the optimized geometries used in this study are given. OPA and TPA spectra calculated for the yPy and PyP systems with additional basis sets are also listed. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Karotki, A.; Kruk, M.; Drobizhev, M.; Rebane, A.; Nickel, E.; Spangler, C. W. *IEEE J. Selected Top. Quantum Electron.* **2001**, *7*, 971.

(2) Spangler, C. W.; Starkey, J. R.; Meng, F.; Gong, A.; Drobizhev, M.; Rebane, A.; Moss, B. *Optical Methods for Tumor Treatment and Detection: Mechanisms and Techniques in Photodynamic Therapy XIV*; SPIE: Bellingham, WA, 2005; Vol. 5689, p 141.

(3) Bhawalkar, J. D.; Kumar, N. D.; Zhao, C.-F.; Prasad, P. N. *J. Clin. Lasers Med. Surg.* **1997**, *15*, 201.

(4) Aijun, G.; Fanqing, M.; Starkey, J. R.; Moss, B. L.; Rebane, A.; Drobizhev, M.; Spangler, C. W. *ACS Polymeric Materials: Science and Engineering*; American Chemical Society: Washington, DC, 2006; Vol. 95, p 289.

(5) Photofrin Web site. www.photofrin.com (accessed May 2008).

(6) Mansour, K., Jr.; Perry, K. J.; Choong, I.; Marder, S. R.; Perry, J. W. *Proc. SPIE* **1993**, *1853*, 132.

(7) Perry, J. W.; Mansour, K.; Lee, I.-Y. S.; Wu, X.-L.; Bedworth, P. V.; Chen, C.-T.; Ng, D.; Marder, S. R.; Miles, P.; Wada, T.; Tian, M.; Sasabe, H. *Science* **1996**, *273*, 1533.

(8) Su, W.; Cooper, T. M.; Brant, M. C. *Chem. Mater.* **1998**, *10*, 1212.

(9) Karotki, A.; Drobizhev, M.; Dzenis, Y.; Taylor, P. N.; Anderson, H. L.; Rebane, A. *Phys. Chem. Chem. Phys.* **2004**, *6*, 7.

(10) Drobizhev, M.; Stepanenko, Y.; Dzenis, Y.; Karotki, A.; Rebane, A.; Taylor, P. N.; Anderson, H. L. *J. Am. Chem. Soc.* **2004**, *126*, 15352.

(11) Drobizhev, M.; Stepanenko, Y.; Dzenis, Y.; Karotki, A.; Rebane, A.; Taylor, P. N.; Anderson, H. L. *J. Phys. Chem. B* **2005**, *109*, 7223.

(12) Edwards, L.; Dolphin, D. H.; Gouterman, M.; Adler, A. D. *J. Mol. Spectrosc.* **1971**, *38*, 16.

(13) Even, U.; Jortner, J. *J. Chem. Phys.* **1982**, *77*, 4391.

(14) Serrano-Andres, L.; Merchan, M.; Rubio, M.; Roos, B. O. *Chem. Phys. Lett.* **1998**, *295*, 195.

(15) Cai, Z.-L.; Crossley, M. J.; Reimers, J. R.; Kobayashi, R.; Amos, R. D. *J. Phys. Chem. B* **2006**, *110*, 15624.

(16) Cortina, H.; Senent, M. L.; Smeyers, Y. G. *J. Phys. Chem. A* **2003**, *107*, 8968.

(17) Tokita, Y.; Hasegawa, J.; Nakatsuji, H. *J. Phys. Chem. A* **1998**, *102*, 1843.

(18) Yamamoto, S.; Tatewaki, H.; Kitao, O.; Diercksen, G. H. F. *Theor. Chem. Acc.* **2001**, *106*, 287.

(19) Kitao, O.; Ushiyama, H.; Miura, N. *J. Chem. Phys.* **1999**, *110*, 2936.

(20) Hasegawa, J.-y.; Takata, K.; Miyahara, T.; Neya, S.; Frisch, M. J.; Nakatsuji, H. *J. Phys. Chem. A* **2005**, *109*, 3187.

(21) Gwaltney, S. R.; Bartlett, R. J. *J. Chem. Phys.* **1998**, *108*, 6790.

(22) Aspuru-Guzik, A.; Akramine, O. E.; Grossman, J. C.; Lester, W. A. J. *J. Chem. Phys.* **2004**, *120*, 3049.

(23) Masthay, M. B.; Findsen, L. A.; Pierce, B. M.; Bocian, D. F.; Lindsey, J. S.; Birge, R. R. *J. Chem. Phys.* **1986**, *84*, 3901.

(24) Bauernschmitt, R.; Ahlrichs, R. *Chem. Phys. Lett.* **1996**, *256*, 454.

(25) van Gisbergen, S. J. A.; Rosa, A.; Ricciardi, G.; Baerends, E. J. *J. Chem. Phys.* **1999**, *111*, 2499.

(26) Nguyen, K. A.; Day, P. N.; Pachter, R. *J. Phys. Chem. A* **2000**, *104*, 4748.

(27) Šeda, J.; Burda, J. V.; Brázdová, V.; Kapsa, V. *Int. J. Mol. Sci.* **2004**, *5*, 196.

(28) Fabian, J.; Diaz, L. A.; Seifert, G.; Niehaus, T. *THEOCHEM* **2002**, *594*, 41.

(29) Sundholm, D. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2275.

(30) Rumi, M.; Ehrlich, J. E.; Heikal, A. A.; Perry, J. W.; Barlow, S.; Hu, Z.; McCord-Maughon, D.; Parker, T. C.; Rockel, H.; Thayumanavan, S.; Marder, S. R.; Beljonne, D.; Bredas, J.-L. *J. Am. Chem. Soc.* **2000**, *122*, 9500.

(31) Sutherland, R. L. *Handbook of Nonlinear Optics*; Marcel Dekker, Inc.: New York, 1996.

(32) Day, P. N.; Nguyen, K. A.; Pachter, R. *J. Phys. Chem. B* **2005**, *109*, 1803.

(33) Day, P. N.; Nguyen, K. A.; Pachter, R. *J. Chem. Phys.* **2006**, *125*, 094103.

(34) Orr, B. J.; Ward, J. F. *Mol. Phys.* **1971**, *20*, 513.

(35) Birge, R. R.; Pierce, B. M. *J. Chem. Phys.* **1979**, *70*, 165.

(36) Birge, R. R.; Bennett, J. A.; Hubbard, L. M.; Fang, H. L.; Pierce, B. M.; Kliger, D. S.; Leroi, G. E. *J. Am. Chem. Soc.* **1982**, *104*, 2519.

(37) Peticolas, W. L. *Annu. Rev. Phys. Chem.* **1967**, *18*, 233.

(38) McClain, W. M. *Acc. Chem. Res.* **1974**, *7*, 129.

(39) Monson, P. R.; McClain, W. M. *J. Chem. Phys.* **1970**, *53*, 29.

(40) Gold, A. *Proceedings of the International School of Physics*; Academic: New York, 1969.

(41) McClain, W. M. *J. Chem. Phys.* **1971**, *55*, 2789.

(42) Olsen, J.; Jorgensen, P. *J. Chem. Phys.* **1985**, *82*, 3235.

(43) Furche, F. *J. Chem. Phys.* **2001**, *114*, 5982.

(44) Salek, P.; Vahtras, O.; Guo, J.; Luo, Y.; Helgaker, T.; Agren, H. *Chem. Phys. Lett.* **2003**, *374*, 446.

(45) Birge, R. R.; Zhang, C.-F. *J. Chem. Phys.* **1990**, *92*, 7178.

(46) Dick, B.; Hohlneicher, G. *J. Chem. Phys.* **1982**, *76*, 5755.

(47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; 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*, A.11.4 ed.; Gaussian, Inc.: Pittsburgh, PA, 2003.

(48) Dalton, a molecular electronic structure program, release 2.0. http://www.kjemi.uio.no/software/dalton/dalton.html (accessed May 2008).

(49) Kruk, M.; Karotki, A.; Drobizhev, M.; Kuzmitsky, V.; Gael, V.; Rebane, A. *J. Lumin.* **2003**, *105*, 45.

(50) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

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

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

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

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

(55) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51.

(56) Peach, M. J. G.; Cohen, A. J.; Tozer, D. J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4543.

(57) Peach, M. J. G.; Helgaker, T.; Salek, P.; Keal, T. W.; Lutnæs, O. B.; Tozer, D. J.; Handy, N. C. *Phys. Chem. Chem. Phys.* **2006**, *8*, 558.

(58) Rudberg, E.; Sałleka, P.; Helgaker, T.; Ågren, H. *J. Chem. Phys.* **2005**, *123*, 184108.

(59) Nguyen, K. A.; Day, P. N.; Pachter, R. *J. Chem. Phys.* **2007**, *126*, 094303.

(60) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2004**, *120*, 8425.

(61) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 234109.

(62) Jacquemin, D.; Perpète, E. A.; Vydrov, O. A.; Scuseria, G. E.; Adamo, C. *J. Chem. Phys.* **2007**, *127*, 094102.

(63) Mikkelsen, K. V.; Dalgaard, E.; Swanstrøm, P. *J. Phys. Chem.* **1987**, *91*, 3081.

(64) Mikkelsen, K. V.; Agren, H.; Jensen, H. J. A.; Helgaker, T. *J. Chem. Phys.* **1988**, *89*, 3086.

(65) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.

(66) Frediani, L.; Rinkevicius, Z.; Agren, H. *J. Chem. Phys.* **2005**, *122*, 244104.

(67) Zhu, Y.; Zhou, S.; Kan, Y.; Su, Z. *Int. J. Quantum Chem.* **2007**, *107*, 1614.

(68) Agren, H. Personal communication.

(69) Rubtsov, I. V.; Susumu, K.; Rubtsov, G. I.; Therien, M. J. *J. Am. Chem. Soc.* **2003**, *125*, 2687.

(70) Zhu, L.; Yi, Y.; Shuai, Z.; Schmidt, K.; Zojer, E. *J. Phys. Chem. A* **2007**, *111*, 8509.

(71) Gouterman, M. *J. Mol. Spectrosc.* **1961**, *6*, 138.

CT800080W

J. Chem. Theory Comput. **2008,** *4,* 1107–1118

JCTC Journal of Chemical Theory and Computation

Computing the Heat of Adsorption using Molecular Simulations: The Effect of Strong Coulombic Interactions

T. J. H. Vlugt,*,† E. García-Pérez,‡ D. Dubbeldam,§ S. Ban,‖ and S. Calero‡

Process & Energy Laboratory, Delft University of Technology, Leeghwaterstraat 44, 2628CA Delft, The Netherlands, Department of Physical, Chemical, and Natural Systems, University Pablo de Olavide, Ctra. Utrera km 1, 41013 Sevilla, Spain, Chemical and Biological Engineering Department, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, and Condensed Matter and Interfaces, Utrecht University, P.O. Box 80000, 08 TA Utrecht, The Netherlands

Received December 18, 2007

Abstract: Molecular simulations are an important tool for the study of adsorption of hydrocarbons in nanoporous materials such as zeolites. The heat of adsorption is an important thermodynamic quantity that can be measured both in experiments and molecular simulations, and therefore it is often used to investigate the quality of a force field for a certain guest–host (*g* - *h*) system. In molecular simulations, the heat of adsorption in zeolites is often computed using either of the following methods: (1) using the Clausius-Clapeyron equation, which requires the partial derivative of the pressure with respect to temperature at constant loading, (2) using the energy difference between the host with and without a single guest molecule present, and (3) from energy/particle fluctuations in the grand-canonical ensemble. To calculate the heat of adsorption from experiments (besides direct calorimetry), only the first method is usually applicable. Although the computation of the heat of adsorption is straightforward for all-silica zeolites, severe difficulties arise when applying the conventional methods to systems with nonframework cations present. The reason for this is that these nonframework cations have very strong Coulombic interactions with the zeolite. We will present an alternative method based on biased interactions of guest molecules that suffers less from these difficulties. This method requires only a single simulation of the host structure. In addition, we will review some of the other important issues concerning the handling of these strong Coulombic interactions in simulating the adsorption of guest molecules. It turns out that the recently proposed Wolf method (*J. Chem. Phys.* **1999**, *110*, 8254) performs poorly for zeolites as a large cutoff radius is needed for convergence.

1. Introduction

Zeolites are microporous crystalline materials with pores of about the same size of a small molecule like water or n-hexane. The structure of a zeolite consists of covalently bonded TO_4 units, in which the T-atom is usually a silicon (Si) or an aluminum (Al) atom. To obey charge neutrality, the substitution of a silicon atom by an aluminum atom requires the presence of a nonframework cation (usually Na^+ or K^+) or a proton (H^+). There are approximately 170 different zeolite framework types that have been synthesized.[1] Zeolitic materials are widely used as water softener, selective adsorber, and catalyst for hydrocarbon conversions (catalytic cracking and isomerization).

---

* Corresponding author e-mail: t.j.h.vlugt@tudelft.nl.

† Delft University of Technology.

‡ University Pablo de Olavide.

§ Northwestern University.

‖ Utrecht University.

10.1021/ct700342k CCC: $40.75

As molecular simulations can provide a fundamental understanding of processes and properties at the molecular scale, in the past few years this type of simulations has become an important tool for investigating the adsorption properties of small guest (*g*) molecules in zeolite hosts (*h*). As guest−host interactions are often dominated by the dispersive interactions of the oxygen atoms with the guest,[2] classical force fields based on Lennard-Jones (LJ) interactions have become very popular in this field of research.[3–10] Recently, it has been shown that an optimal parametrization can lead to united atom force fields for alkanes that are transferable to many all-silica zeolite framework types.[11–13] This transferability is often tested by a comparison between simulations and experimental adsorption data that had not been used to calibrate the force field parameters.[13]

When nonframework cations are present in the framework, strong electrostatic interactions between the nonframework cations and the framework atoms (Si, Al, O) have to be included for a correct description of the system. Although the Ewald summation with tin foil boundary conditions is often used,[14] the so-called Wolf method has been introduced recently as a pairwise alternative.[15,16] The strong interaction between the nonframework cations and hydrocarbons lead to an enormous effect on adsorption properties.[17–20]

In this work, we will consider the effect of strong electrostatic interactions between the framework and the nonframework cations or adsorbates on the computation of the heat of adsorption. This quantity describes the change in enthalpy when a molecule is transferred from the gas phase into the pores of a zeolite. In experiments, the heat of adsorption is usually computed using the Clausius-Clapeyron equation[21] or direct calorimetry experiments[22] while in molecular simulations this quantity is usually calculated directly from the total energy of the simulated system.[4,23,8] However, we will show that a direct computation using energy differences results in very inaccurate results for zeolites with strongly bound nonframework cations.

This paper is organized as follows. In section 2 we briefly review the various schemes to compute the heat of adsorption in molecular simulations, and we will explain the advantages and disadvantages of each method. In particular, we will show that a direct calculation of the heat of adsorption using energy differences may lead to problems for systems with strongly interacting nonframework cations. We will introduce an alternative method based on biased insertions that does not suffer from these difficulties. Results of the various methods for a typical system (described in section 3) are presented in section 4. In section 5 we discuss some of the other important issues for handing the Coulombic interactions in these systems and show why the Ewald summation is here preferable over the recently proposed Wolf method. Our findings are summarized in section 6.

2. Calculating the Heat of Adsorption

In the remainder of this paper, we will denote the zeolite and the included nonframework cations as the "host" (*h*). The scaled positions of all the atoms that belong to the host are denoted by the vector **h**. The adsorbate is denoted as "guest" (*g*), and its conformation is denoted by **g**. In most of the derivations in this work, the heat of adsorption is calculated in molecular units, i.e. in J (or kJ) per molecule instead of the often used J per mol.

Following Wood, Panagiotopoulos, and Rowlinson,[21] the heat of adsorption q (or enthalpy of adsorption - ΔH) at loading θ is defined using the famous Clausius-Clapeyron equation

$$-q = \Delta H = k_{\mathrm{B}}\left(\frac{\partial \ln[P/P_0]}{\partial T^{-1}}\right)_{\theta=\mathrm{constant}} = \left(\frac{\partial \ln[P/P_0]}{\partial \beta}\right)_{\theta=\mathrm{constant}} \tag{1}$$

where P is the pressure, P_0 is an arbitrary reference pressure, T is the absolute temperature, θ is the loading of guest molecules, $k_{\mathrm{B}} = R/N_{\mathrm{av}}$ is the Boltzmann constant, R is the gas constant, N_{av} is Avogadro's number, and $\beta = 1/(k_{\mathrm{B}}T)$. For an in-depth review of the thermodynamic definition of the heat of adsorption we refer the reader to ref 24. To compute the heat of adsorption at loading θ directly from this equation, one needs the adsorption isotherm $\theta(P)$ for various temperatures T. At sufficiently low loadings, the gas phase will behave as an ideal gas, and the adsorption isotherm will become a linear function

$$K_{\mathrm{H}} = \frac{\theta}{VP} \tag{2}$$

in which V is the volume of the host, and K_{H} is the so-called Henry coefficient (in units of molecules per unit of host volume per unit of pressure). The heat of adsorption at low loading then becomes

$$-q = \Delta H = -\frac{\partial \ln[K_{\mathrm{H}}/K_{\mathrm{H0}}]}{\partial \beta} \tag{3}$$

in which K_{H0} is an arbitrary constant (that has the same units as K_{H}). In molecular simulations, the most convenient way to calculate the Henry coefficient is using Widom's test particle method [14,25–27]

$$K_{\mathrm{H}} = \beta \times \exp[-\beta \mu_{\mathrm{ex}}] = \beta \times \frac{\langle \exp[-\beta u^{+}] \rangle_{\mathrm{H}}}{\langle \exp[-\beta u_{\mathrm{IG}}^{+}] \rangle_{\mathrm{EB}}} \tag{4}$$

in which μ_{ex} is the excess chemical potential of the guest in the zeolite, u^{+} is the energy of a test (guest) molecule inserted at a random position in the zeolite, and u_{IG}^{+} is the energy of a test (guest) molecule inserted at a random position in an empty box without the presence of the host (often referred to as an isolated chain). The brackets $\langle \cdots \rangle_{\mathrm{H}}$ denote an average in the NVT ensemble over all conformations of the host (and positions of the test particle), and the brackets $\langle \cdots \rangle_{\mathrm{EB}}$ denote an ensemble average for a test chain in an empty simulation box (ideal gas phase). For chain molecules like alkanes, it is well-known that insertion of a test chain at a random position in the zeolite nearly always results in overlaps with zeolite atoms, and therefore the sampling statistics of the average $\langle \exp[-\beta u^{+}]\rangle_{\mathrm{H}}$ will be extremely poor.[23] For chains that are not too long (<50 monomers) it is convenient to use the Configurational-bias Monte Carlo (CBMC) method[28–31,14] to insert test chains. In this case, the Henry coefficient is computed from[23]

$$K_{\mathrm{H}} = \beta \times \frac{\langle W\rangle_{\mathrm{H}}}{\langle W_{\mathrm{IG}}\rangle_{\mathrm{EB}}} \tag{5}$$

in which $\langle W\rangle_{\mathrm{H}}$ is the average Rosenbluth weight of a test chain in the host, and $\langle W_{\mathrm{IG}}\rangle_{\mathrm{EB}}$ is the average Rosenbluth weight of an isolated test chain in an empty box.

Calculating the heat of adsorption at low loading directly using either eq 1 or eqs 3 and 4 requires several simulations or adsorption experiments at different temperatures, and the final answer must be computed by a numerical differentiation with respect to $1/T$. As the accuracy of such a numerical differentiation strongly depends on the accuracy of the individual simulations,[32] many long simulations are required to obtain an accurate value for the heat of adsorption. Therefore, two alternative methods to compute the heat of adsorption are often used in molecular simulations.

(1) From energy differences computed in the canonical (NVT) ensemble[21,23]

$$-q = \Delta H = \langle U_1\rangle_1 - \langle U_0\rangle_0 - \langle U_{\mathrm{g}}\rangle - \frac{1}{\beta} \tag{6}$$

in which U_N is the total energy of a host with N guest molecules present, $\langle\cdots\rangle_X$ refers to an ensemble average at constant V, T, and X guest molecules, and $\langle U_{\mathrm{g}}\rangle$ is the average energy of an isolated guest molecule (without the host present). The average $\langle U_{\mathrm{g}}\rangle$ for a certain guest molecule only depends on temperature and needs to be calculated only once. Later, we will see that for zeolites with nonframework cations, the difference $|\langle U_1\rangle_1 - \langle U_0\rangle_0|$ is very small compared to either $\langle U_0\rangle_0$ or $\langle U_1\rangle_1$. Therefore, a direct computation of both $\langle U_0\rangle_0$ and $\langle U_1\rangle_1$ can lead to a very inaccurate estimate of ΔH. Note that eq 6 only applies to the heat of adsorption at zero coverage and that it assumes ideal gas behavior.

(2) From energy/particle fluctuations in the grand-canonical (μVT) ensemble,[24] we can approximate the change in potential energy upon adsorption of a single guest molecule[24]

$$\langle U_{N+1}\rangle_{N+1} - \langle U_N\rangle_N \approx \left(\frac{\partial\langle U\rangle_\mu}{\partial\langle N\rangle_\mu}\right)_\beta = \left(\frac{2\langle U\rangle_\mu}{2\langle N\rangle_\mu}\right)_\beta = \frac{\langle U\times N\rangle_\mu - \langle U\rangle_\mu\langle N\rangle_\mu}{\langle N^2\rangle_\mu - \langle N\rangle_\mu\langle N\rangle_\mu} \tag{7}$$

where the brackets $\langle\cdots\rangle_\mu$ denote an average in the grand-canonical ensemble, N is the number of guest molecules, and μ is the chemical potential of the guest molecules. This leads to[24]

$$-q = \Delta H = \frac{\langle U\times N\rangle_\mu - \langle U\rangle_\mu\langle N\rangle_\mu}{\langle N^2\rangle_\mu - \langle N\rangle_\mu\langle N\rangle_\mu} - \langle U_g\rangle - \frac{1}{\beta} \tag{8}$$

ΔH in eq 8 is usually defined as the *isosteric heat of adsorption*, and it is often applied at nonzero loading.[24] Again, it is assumed here that the gas phase is ideal. The disadvantage of using this method in practice is that it relies on many particle insertions and removals in the grand-canonical ensemble. The fraction of accepted trial moves in grand-canonical simulations however can be quite low, even if advanced insertion techniques are used.[33–35] In practice, long simulations in the grand-canonical ensemble are needed to obtain accurate statistics for the averages in eq 7, especially in the limit of low chemical potential where the number of guest molecules is either 0 or 1. Therefore, we will not consider the direct use of eq 8 in the remainder of this manuscript.

It is instructive to show that eqs 3 and 6 should give the same value for ΔH at low loading. Our starting point is the observation that the Henry coefficient is related to the free energy of a guest molecule inside the host (eq 4). We will make use of the fact that partial derivatives of free energies can always be expressed as ensemble averages.[14] Consider the partial derivative

$$\frac{\partial \ln\langle \exp[-\beta u^{+}]\rangle_{\mathrm{H}}}{\partial\beta} \tag{9}$$

and denote the scaled coordinates of the host (nonframework cations included) by **h** and the scaled coordinates of the guest molecules by **g**. $U_0(\mathbf{h})$ is the energy of the host and $u^{+}(\mathbf{g},\mathbf{h}) = U_1(\mathbf{g},\mathbf{h}) - U_0(\mathbf{h})$ is the energy of a single guest molecule adsorbed in the host. For the canonical partition function of the host (without the guest) we can write

$$Q_0 = Q_{\text{zeolite without guest}} = \int d\mathbf{h}\ \exp - [\beta U_0(\mathbf{h})] \tag{10}$$

in which we have omitted the constant prefactor of the partition function.[14] For the zeolite with a single guest molecule we can write

$$\begin{aligned} Q_1 &= Q_{\text{zeolite with single guest molecule}} \\ &= \int d\mathbf{h} \int d\mathbf{g}\, \exp[-\beta(U_0(\mathbf{h}) + u^{+}(\mathbf{g},\mathbf{h}))] \end{aligned} \tag{11}$$

In this equation we separated the total energy of a zeolite with a single guest molecule into two contributions: the zeolite and the test particle. The ensemble average $\langle \exp[-\beta u^{+}]\rangle_{\mathrm{H}}$ is therefore

$$\begin{aligned} \langle \exp[-\beta u^{+}]\rangle_{\mathrm{H}} &= \frac{\int d\mathbf{h} \int d\mathbf{g}\ \exp[-\beta(U_0(\mathbf{h}) + u^{+}(\mathbf{g},\mathbf{h}))]}{\int d\mathbf{h}\ \exp[-\beta U_0(\mathbf{h})]} \\ &= \frac{\int d\mathbf{h}[\exp[-\beta U_0(\mathbf{h})] \int d\mathbf{g}\ \exp[-\beta u^{+}(\mathbf{g},\mathbf{h})]]}{\int d\mathbf{h}\ \exp[-\beta U_0(\mathbf{h})]} \end{aligned} \tag{12}$$

Taking the partial derivative of its logarithm with respect to β we get

$$\begin{aligned}
&\frac{\partial \ln\langle \exp[-\beta u^{+}]\rangle_{\mathrm{H}}}{\partial \beta}\\
&=\frac{\partial}{\partial \beta}\left[\ln\left(\frac{\int d\mathbf{h}[\exp[-\beta U_0(\mathbf{h})]\int d\mathbf{g}\ \exp[-\beta u^{+}(\mathbf{g},\mathbf{h})]]}{\int d\mathbf{h}\exp[-\beta U_0(\mathbf{h})]}\right)\right]\\
&=\frac{\int d\mathbf{h}\ \ \exp[\beta U_0(\mathbf{h})]}{\int d\mathbf{h}[\exp[-\beta U_0(\mathbf{h})]\int d\mathbf{g}\ \exp[-\beta u^{+}(\mathbf{g},\mathbf{h})]]}\times\\
&\frac{\int d\mathbf{h}\ \ \exp[-\beta U_0(\mathbf{h})]\times\int d\mathbf{h}\int d\mathbf{g}\ \exp[-\beta(U_0(\mathbf{h})+u^{+}(\mathbf{g},\mathbf{h}))](-U_0(\mathbf{h})-u^{+}(\mathbf{g},\mathbf{h}))}{(\int d\mathbf{h}\ \ \exp[-\beta U_0(\mathbf{h})])^2}\\
&-\frac{\int d\mathbf{h}\ \ \exp[-\beta U_0(\mathbf{h})]}{\int d\mathbf{h}[\exp[-\beta U_0(\mathbf{h})]\int d\mathbf{g}\ \exp[-\beta u^{+}(\mathbf{g},\mathbf{h})]]}\times\\
&\frac{\int d\mathbf{h}\ \exp[-\beta U_0(\mathbf{h})](-U_0(\mathbf{h}))\times\int d\mathbf{h}\int d\mathbf{g}\ \exp[-\beta(U_0(\mathbf{h})+u^{+}(\mathbf{g},\mathbf{h}))]}{(\int d\mathbf{h}\ \exp[-\beta U_0(\mathbf{h})])^2}\\
&=-\frac{\int d\mathbf{h}\int d\mathbf{g}\ \exp[-\beta(U_0(\mathbf{h})+u^{+}(\mathbf{g},\mathbf{h}))](U_0(\mathbf{h})+u^{+}(\mathbf{g},\mathbf{h}))}{\int d\mathbf{h}\int d\mathbf{g}\ \exp[-\beta(U_0(\mathbf{h})+u^{+}(\mathbf{g},\mathbf{h}))]}\\
&+\frac{\int d\mathbf{h}\ \exp[-\beta U_0(\mathbf{h})]U_0(\mathbf{h})}{\int d\mathbf{h}\ \ \exp[-\beta U_0(\mathbf{h})]}\\
&=\langle U_0\rangle_0-\langle U_0+u^{+}\rangle_1\\
&=\langle U_0\rangle_0-\langle U_1\rangle_1
\end{aligned}\tag{13}$$

The same can be derived for the other term in eq 6 (corresponding to the isolated chain)

$$\begin{aligned}
&\frac{\partial \ln\langle \exp[-\beta u_{\mathrm{IG}}^{+}]\rangle_{\mathrm{EB}}}{\partial \beta}\\
&=\langle U_{\mathrm{IG}}\rangle_{\text{empty box no chain}}-\langle U_{\mathrm{IG}}+u_{\mathrm{IG}}^{+}\rangle_{\text{empty box with single chain}}\\
&=-\langle u_{\mathrm{IG}}^{+}\rangle_{\text{empty box with single chain}}\\
&=-\langle U_{\mathrm{g}}\rangle
\end{aligned}\tag{14}$$

in which U_{IG} is the energy of an empty box (zero) and u_{IG}^{+} is the energy of an isolated chain in this empty box. From eqs 3 and 4, we obtain

$$\begin{aligned}
\Delta H=-\frac{\partial \ln K_{\mathrm{H}}}{\partial \beta}=\\
-\frac{\partial}{\partial \beta}[\ln\beta+\ln\langle\exp[-\beta u^{+}]\rangle_{\mathrm{H}}-\ln\langle\exp[-\beta u_{\mathrm{IG}}^{+}]\rangle_{\mathrm{EB}}]
\end{aligned}\tag{15}$$

and by substituting eqs 13 and 14 we end up with eq 6. This means that eqs 3 and 6 should give identical result for ΔH at low loading.

In this work, we propose to use eq 8 in such a way that $\langle U_{N+1}\rangle_{N+1}$ and $\langle U_N\rangle_N$ are computed in a single simulation in the NVT ensemble. This means that for the calculation of ΔH at a certain loading (N guest molecules present in the system), only a single simulation of the host structure is needed. This simulation is performed in the ensemble with N guest molecules present (denoted as $\langle\cdots\rangle_N$). Our starting point is

$$\begin{aligned}
&\langle U_{N+1}\rangle_{N+1}-\langle U_N\rangle_N\\
&=\langle U_N+u^{+}\rangle_{N+1}-\langle U_N\rangle_N\\
&=\frac{\int d\mathbf{h}\int d\mathbf{g}\ \exp[-\beta(U_N(\mathbf{g},\mathbf{h})+u^{+}(\mathbf{g},\mathbf{h}))](U_N(\mathbf{g},\mathbf{h})+u^{+}(\mathbf{g},\mathbf{h}))}{\int d\mathbf{h}\int d\mathbf{g}\ \exp[-\beta(U_N(\mathbf{g},h)+u^{+}(\mathbf{g},\mathbf{h}))]}-\langle U_N\rangle_N\\
&=\frac{\dfrac{\int d\mathbf{h}\int d\mathbf{g}\ \exp[-\beta(U_N(\mathbf{g},\mathbf{h})+u^{+}(\mathbf{g},\mathbf{h}))](U_N(\mathbf{g},\mathbf{h})+u^{+}(\mathrm{g},\mathbf{h}))}{\int d\mathbf{h}\int d\mathbf{g}\ \exp[-\beta U_N(\mathbf{g},\mathbf{h})]}}{\dfrac{\int d\mathbf{h}\int d\mathbf{g}\ \exp[-\beta(U_N(\mathbf{g},\mathbf{h})+u^{+}(\mathbf{g},\mathbf{h}))]}{\int d\mathbf{h}\int d\mathbf{g}\exp[-\beta U_N(\mathbf{g},\mathbf{h})]}}-\langle U_N\rangle_N\\
&=\frac{\langle(U_N+u^{+})\exp[-\beta u^{+}]\rangle_N}{\langle\exp[-\beta u^{+}]\rangle_N}-\langle U_N\rangle_N
\end{aligned}\tag{16}$$

in which $u^{+}=U_{N+1}-U_N$ is the energy of a test (guest) chain. Note that the coordinates of the test chain are included in $\mathbf{h}$. For chain molecules, it is much more convenient to use the Rosenbluth algorithm[36] to generate a conformation of a test chain. In this case, it is trivial to show that

$$\langle\exp[-\beta u^{+}]\rangle_N=\langle W\rangle_N \tag{17}$$

$$\langle(U_N+u^{+})\times\exp[-\beta u^{+}]\rangle_N=\langle(U_N+u^{+})\times W\rangle_N \tag{18}$$

and therefore

$$\langle U_{N-1}\rangle_{N-1}-\langle U_N\rangle_N=\frac{\langle(U_N+u^{+})\times W\rangle_N}{\langle W\rangle_N}-\langle U_N\rangle_N \tag{19}$$

In the Rosenbluth algorithm, it is often convenient to use only part of the nonbonded energy to select trial segments leading to a modified Rosenbluth weight $\langle W^* \rangle$. In the Appendix, it is derived that

$$\langle W \rangle_N = \langle W^* \exp[-\beta\delta] \rangle_N \tag{20}$$

$$\langle (U_N + u^+) \times W \rangle_N = \langle (U_N + u^+) \times W^* \exp[-\beta\delta] \rangle_N \tag{21}$$

in which δ is the energy difference between the total nonbonded part of the potential and the nonbonded part of the potential that was used to select trial segments.[27,37] The final expression then becomes

$$-q = \Delta H = \frac{\langle (U_N + u^+) \times W^* \exp[-\beta\delta] \rangle_N}{\langle W^* \exp[-\beta\delta] \rangle_N} - \langle U_N \rangle_N - \langle U_g \rangle - \frac{1}{\beta} \tag{22}$$

The ensemble averages $\langle \cdots \rangle_N$ can be computed from the same simulation. Calculating $\langle U_g \rangle$ requires a simulation of an isolated guest molecule at the same temperature. It is important to note that eqs 20 and 8 lead to identical results and both can be applied at nonzero loading (in contrast to eqs 3, 4, and 5). At low loading ($N = 0$), eq 22 becomes identical to eqs 3 and 6 and not only the heat of adsorption but also the Henry coefficient can be computed using a single simulation of the host structure (this in contrast to the conventional test particle method).

In principle, one could also perform simulations in the ensemble $\langle \cdots \rangle_N$ to compute averages in the ensemble $\langle \cdots \rangle_{N-1}$. This leads to

$$\langle U_{N-1} \rangle_{N-1} = \frac{\langle U_{N-1} \times \exp[\beta u^+] \rangle_N}{\langle \exp[\beta u^+] \rangle_N} \tag{23}$$

with $u^+ = U_N - U_{N-1}$. This would correspond to a real-particle version of eq 16. However, as this is equivalent to computing the chemical potential using a real particle formulation, this approach will be extremely inefficient.[14]

3. Model and Simulation Details

To test the different methods to compute the heat of adsorption, we studied the adsorption of n-alkanes in zeolite LTA5A using the force field of ref 18. We simulated a single unit cell with dimensions of $a = b = c = 24.555$ Å ($\alpha = \beta = \gamma = 90°$) and with composition $Na_{32}Ca_{32}Al_{96}Si_{96}O_{384}$. Periodic boundary conditions are used. The unit cell of LTA consists of 8 cages that are interconnected by small windows, see Figure 1. The crystallographic positions of the framework atoms (Al, Si, O) as well as the initial positions of the nonframework cations (Na, Ca) are taken from the work of Pluth et al.[38] Electrostatic interactions are handled using the Ewald summation with parameters $\alpha = 0.3$ Å^{-1} and $k = 9$ wave vectors in each reciprocal direction.[14]

Simulations were performed in the NVT ensemble at 500 K. In each Monte Carlo cycle, it is chosen at random with a fixed probability to displace, rotate, or regrow a hydrocarbon chain (if present in the system) at a random position.

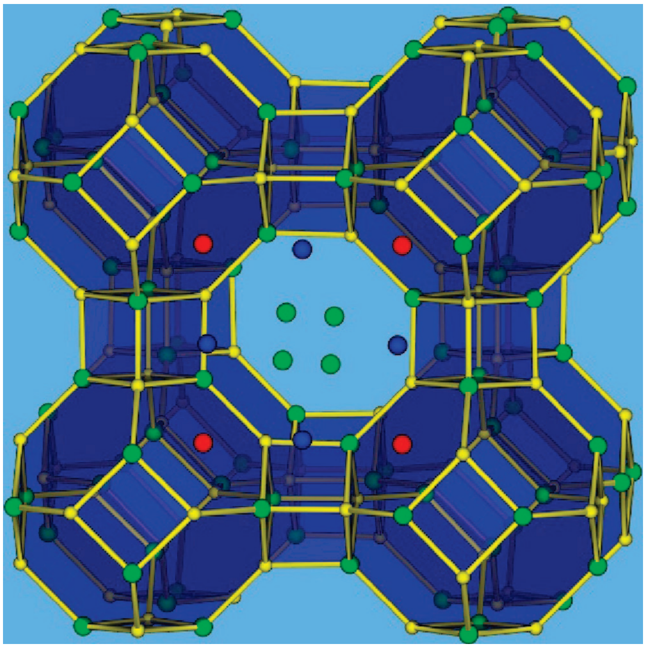


Figure 1. Schematic representation of framework type LTA.

As it is well-known that fixing the positions of the nonframework cations (Na, Ca) may lead to erroneous results,[17] trial moves to displace the nonframework cations are included. The number of trial moves in each cycle is equal to the number of nonframework cations. The positions of the framework atoms (Si, Al, O) are kept fixed. During the simulations, test chains are grown using the Rosenbluth algorithm from which data on the average Rosenbluth weight of the sorbate are collected; this is used to compute the heat of adsorption using eq 22. The number of test insertions per cycle is equal to the number of nonframework cations. In this way, the amount of CPU time per cycle is approximately equal for the methods to compute the heat of adsorption (eqs 6 and 22).

4. Results

In Figure 2, we have plotted the heat of adsorption of various n-alkanes at zero loading computed using various methods. Clearly, the new method (eq 22) produces exactly the same result as the method using the temperature derivative of the Henry coefficient (eqs 3 and 4) as well as with the method based on energy differences (eq 6) within the accuracy of the simulations.

In Figure 3 we have plotted the running average of the heat of adsorption of n-hexane as a function of the number of MC cycles when small displacements of the nonframework cations are allowed. Here, the initial positions of the nonframework cations were taken from the crystallographic positions. The method based on energy differences requires two simulations of the zeolite, one with a single n-hexane chain and one without, while our new method requires only a single simulation of the host system. It turns out that each MC cycle takes approximately the same amount of CPU time for each method, and therefore it becomes clear from Figure 3 that our new method converges much more quickly to the final answer. As the difference $|\langle U_1 \rangle_1 - \langle U_0 \rangle_0|$ is typically

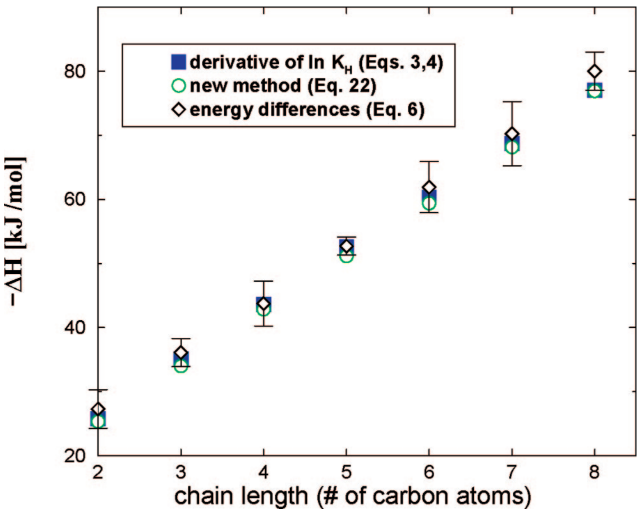


Figure 2. Heat of adsorption at zero loading for various n-alkanes in zeolite LTA5A computed using various methods. $T = 500$ K. We have included the error bars for the method based on energy differences. For the other two methods, we checked that the error bars are always smaller than the symbol size.

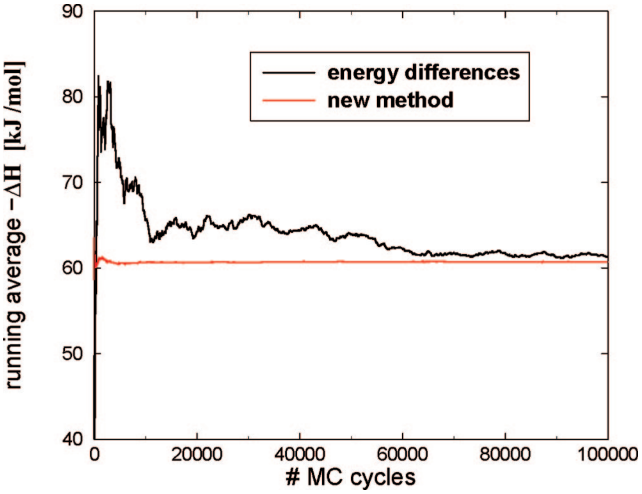


Figure 3. Running average of the heat of adsorption of n-hexane at zero loading as a function of the number of MC cycles for the method based on energy differences (eq 6) and our new method (eq 20). $T = 500$ K. Small displacements of the nonframework cations are allowed. Initial positions of the nonframework cations were taken from the crystallographic positions.[38]

10^4 times smaller than either $\langle U_1\rangle_1$ or $\langle U_0\rangle_0$, and slightly smaller than the typical fluctuations of either $\langle U_1\rangle_1$ or $\langle U_0\rangle_0$, it is not surprising that the method based on energy differences performs poorly.

In all previous simulations (Figures 2 and 3), the initial positions of the nonframework cations were taken from the known crystallographic positions of LTA5A.[38] For this zeolite, also the positions of the framework Al atoms are well-known. Although the nonframework cations are allowed to move during the simulations, as expected their positions during the simulations do not significantly deviate from the crystallographic positions. However, in most cases, for a given zeolite the crystallographic positions of the nonframework cations or the framework Al atoms are not known from experiments. However, one could still be interested in computing the adsorption properties in this case. One should keep in mind that for some zeolites the adsorption properties strongly depend on the positions of the framework Al atoms, while for other zeolites this dependency is not present.[19,20]

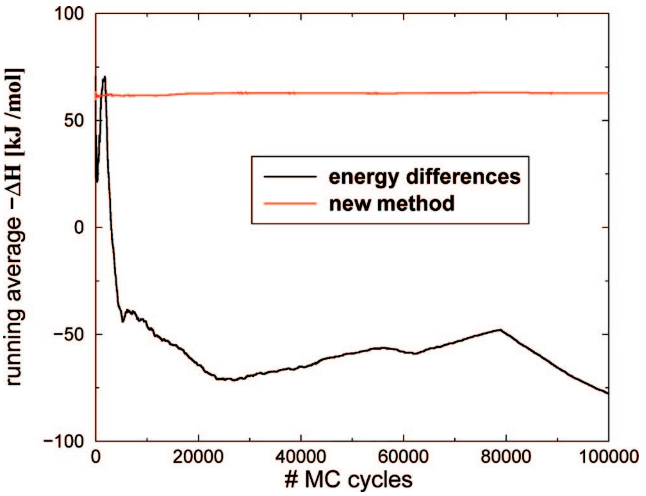


Figure 4. Running average of the heat of adsorption of n-hexane at zero loading as a function of the number of MC cycles for the method based on energy differences (eq 6) and our new method (eq 20), not starting from the crystallographic positions of the nonframework cations. $T = 500$ K. A simulation starting from random initial nonframework cation positions was conducted, and a long equilibration period was used. The final configuration of this simulation was used here as the starting point to compute the heat of adsorption. Trial moves to put nonframework cations at random positions in the zeolite are included to accelerate equilibration of the positions of the nonframework cations.

It is instructive to study this usual senario further, i.e. the positions of the nonframework cations are not known. In this case, the only choice is to start from random initial positions of the nonframework cations and to use a long equilibration period of the system. Trial moves to put a nonframework cation at a random position in the zeolite will then significantly accelerate equilibration. It is important to note that this approach relies on the assumption that during the simulation the nonframework cations are able to find their equilibrium positions. This is a major cause for concern, as the interactions between the nonframework cations and the framework atoms are dominated by very strong electrostatic interactions, especially for multivalent ions like Ca.

To investigate the effect of possible rearrangements of the nonframework cations during the simulation, we conducted a simulation starting from random initial positions for the nonframework cations that was equilibrated for 10^5 cycles. Trial moves to put the nonframework cations at a random position in the zeolite were used to accelerate equilibration. The final configuration of the nonframework cations was then used as a starting point to compute the heat of adsorption using eqs 6 and 22, see Figure 4. Clearly, the method based on energy differences is extremely inaccurate as during the simulations the positions of the nonframework cations in the simulation with the n-hexane chain apparently deviate from those in the simulation without the n-hexane chain, even though these simulations were started from the same initial nonframework cation positions. This difference in nonframework cation positions will have a huge effect on the total energies $\langle U_1\rangle_1$ and $\langle U_0\rangle_0$, and therefore it will dramatically affect the efficiency of any method to compute the heat of adsorption that requires more than one simulation of the host structure. In principle, the huge difference of Figure 4 would eventually disappear if one would simulate much longer, so that all rearrangements of nonframework cations would be visited according to their statistical weight. Unfortunately, due to the very strong interactions of the nonframework cations with the framework this will take extremely long simulations, even at high temperature (500 K).

From Figures 3 and 4 we can conclude that the method based on energy differences is very sensitive to the displacements of the nonframework cations and far less sensitive to the adsorption energy of a hydrocarbon. Therefore, the use of this method is not recommended for zeolites with strongly interacting nonframework cations. Note that the heat of adsorption of n-hexane calculated using our new method is approximately 60.7 kJ/mol for the simulations of Figure 3 and 62.7 kJ/mol for the simulations of Figure 4. As this difference it very small, one can conclude that the positions of the nonframework cations only have a minor influence on the heat of adsorption for LTA5A. However, performing the simulations with (arbitrary) fixed nonframework cation positions is not a sensible option as this may introduce other artifacts (especially when the temperature or loading is varied).[17]

5. Handling of Coulombic Interactions in Zeolites

In the previous section we have analyzed the difficulties introduced by strong electrostatic interactions in zeolites. Electrostatic interactions are very long-ranged due to their r^{-1} behavior. It is well-known that simple truncation of these interactions may lead to incorrect simulation results.[14,39,40] Also, a direct summation of Coulombic pair interactions is conditionally convergent, i.e. the final result depends on the order of the summation. It is for this reason why the well-known Ewald summation[41,26,14] or equivalent method is currently the generally accepted method in molecular simulations. For the Ewald summation, the total Coulombic energy of the system can be written as the sum of three contributions (assuming tin foil boundary conditions):

(1) The real-space part, which is proportional to

$$E_{\text{real}} = \sum_{i<j} \frac{q_i q_j \operatorname{erfc}(\alpha r_{ij})}{r_{ij}} \tag{24}$$

in which erfc(x) is the complementary error function, r_{ij} is the distance between particles i and j, q_i and q_j are the charges on particles i and j, respectively, and α is a constant damping factor. The summation $\sum_{i<j}$ is in principle a summation over all periodic images of i and j. However, it is convenient to choose α in such a way that only the nearest images have to be considered (as erfc(x) is close to zero for large x). This requires a value of α that is not too small. The real-space part of the Ewald summation and the conventional Lennard-Jones interactions can be calculated simultaneously using the same cutoff radius.

(2) The self-energy

$$E_{\text{self}} = -\frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 \tag{25}$$

which does not depend on the positions of the ions in the system.

(3) The Fourier part of the energy, which is proportional to

$$E_{\text{Fourier}} = \frac{1}{2V} \sum_{\mathbf{k}\neq 0}^{\infty} \frac{\exp[-k^2/4\alpha^2]}{k^2} \left| \sum_{j=1}^{N} q_j \exp[-i\mathbf{k} \bullet \mathbf{r}_j] \right|^2 \tag{26}$$

in which V is the volume of the system and the vectors $\mathbf{k}$ are linear combinations of the reciprocal basis vectors of the system. For larger systems (>10000 charges), it is more efficient to replace eq 24 by a grid-based computation like Particle Mesh Ewald.[14]

The convergence of the total energy is controlled by the real-space cutoff r_{cut}, the value of α, and the number of vectors in each reciprocal direction (k). Provided that only nearest images are used to calculate the real-space energy (eq 22), the total energy as a function of the number of k-vectors converges to the same value for a large number of k-vectors, independent of the value of α (provided that α is sufficiently large).

The Fourier part of the system clearly depends on the positions $\mathbf{r}_i$ (and charges q_i) of all N particles in the system. This may suggest that in a Monte Carlo simulation, in every trial move the positions of all particles in the system have to be considered for the calculation of the energy difference between the new and old configuration. Fortunately, it is possible to rewrite the Ewald summation in such a way that only the atoms with a different position have to be considered. The structure of eq 26 can be expressed as

$$E_{\text{Fourier}} \propto \sum_{\mathbf{k}\neq 0} |x(\mathbf{k}, \mathbf{r}^N, \mathbf{q}^N)|^2 = \sum_{\mathbf{k}\neq 0} [\boldsymbol{R}(x(\mathbf{k}))^2 + \boldsymbol{I}(x(\mathbf{k}))^2] \tag{27}$$

in which the complex numbers x depends on the positions (r^N) and charges (q^N) of all the atoms in the system. The real and imaginary component of this vector can be expressed as

$$\boldsymbol{R}(x(\mathbf{k})) = \sum_{i=1}^{N} q_i \cos(\mathbf{k} \bullet \mathbf{r}_i) \tag{28}$$

and

$$\boldsymbol{I}(x(\mathbf{k})) = \sum_{i=1}^{N} q_i \sin(\mathbf{k} \bullet \mathbf{r}_i) \tag{29}$$

As $\mathbf{R}(x(\mathbf{k}))$ and $\mathbf{I}(x(\mathbf{k}))$ can be expressed as a summation over all particles in the system, it is convenient to store $\mathbf{R}(x(\mathbf{k}))$ and $\mathbf{I}(x(\mathbf{k}))$ in the memory of the computer (one complex number for each $\mathbf{k}$). For particle displacements, rotations, regrows, (test) insertions, deletions, and identity

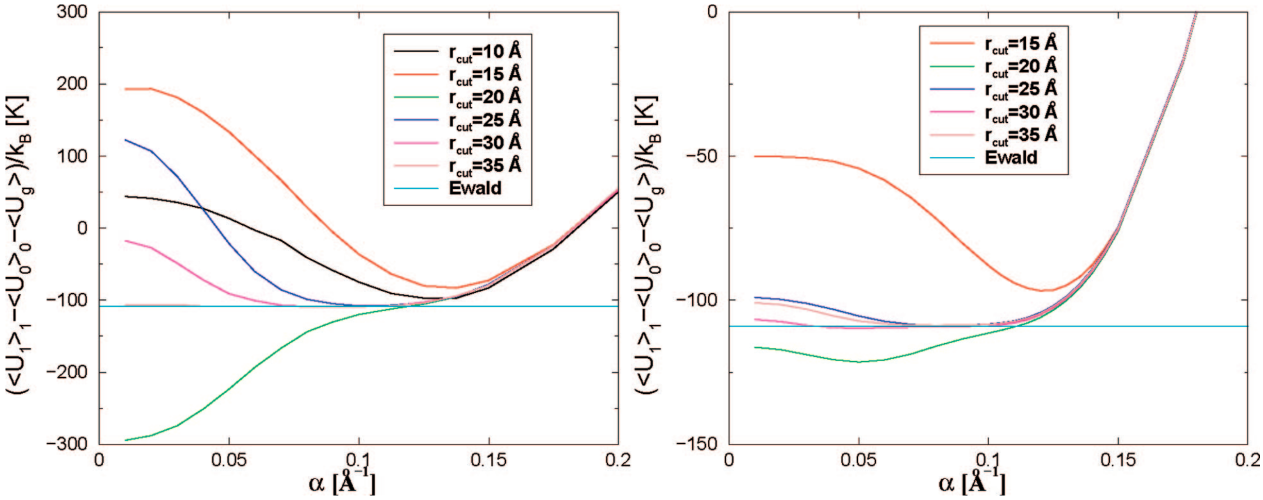


Figure 5. Average energy difference $\langle U_1\rangle_1 - \langle U_0\rangle_0 - \langle U_g\rangle$ for CO_2 in all silica MFI-type zeolite computed using the Wolf method (left) or using the modification by Fennell and Gezelter (right). The (unweighted) average is taken over all possible configurations of CO_2 that do not result in an overlap with zeolite atoms. The zeolite consists of 4 × 4 × 6 unit cells. The size of the simulation box is therefore $a = 80.088$ Å, $b = 79.596$ Å, $c = 80.298$ Å. In all simulations, 3×10^7 trial insertions are used (eq 20). The result obtained by the Ewald summation ($\alpha = 0.3$ Å^{-1}, $k = 20$, $r_{\mathrm{cut}} = 12$ Å) is also included.

changes one can easily calculate the new values for $\mathbf{R}(x(\mathbf{k}))$ and $\mathbf{I}(x(\mathbf{k}))$ by subtracting the contributions of the old configuration and adding the contributions for the new configuration. This has to be done only for atoms that have a different position in the old and new configuration.

Recently, Wolf and co-workers[15] have proposed a pairwise alternative for the Ewald summation. In this method, the Coulombic interactions are damped using a complementary error function that is truncated and shifted at distance r_{cut} resulting in the following pair potential ($r_{ij} \leq r_{\mathrm{cut}}$):

$$\varphi_{\mathrm{Wolf}}(r_{ij}) = q_i q_j \left[\frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\mathrm{erfc}(\alpha r_{\mathrm{cut}})}{r_{\mathrm{cut}}}\right] \tag{30}$$

Fennell and Gezelter[16] have developed a slightly modified damping function ($r_{ij} \leq r_{\mathrm{cut}}$)

$$\varphi_{\mathrm{FG}}(r_{ij}) = q_i q_j \left[\frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\mathrm{erfc}(\alpha r_{\mathrm{cut}})}{r_{\mathrm{cut}}} + (r_{ij} - r_{\mathrm{cut}}) \left[\frac{\mathrm{erfc}(\alpha r_{\mathrm{cut}})}{r_{\mathrm{cut}}^2} + \frac{2\alpha \exp[-\alpha^2 r_{\mathrm{cut}}^2]}{\sqrt{\pi} r_{\mathrm{cut}}}\right]\right] \tag{31}$$

Compared to the Ewald summation, the long-range Fourier part (eq 26) is omitted, which is only strictly correct in the limits $\alpha \rightarrow 0$ and $r_{\mathrm{cut}} \rightarrow \infty$. To make a comparison between the results obtained from the Ewald summation and the Wolf method, one should not only consider the computationally efficiency of the methods but also how the computed averages obtained from the methods depend on the control parameters (for the Wolf method, r_{cut}, α and for the Ewald summation r_{cut}, α and the number of k-vectors in each reciprocal direction (k)). As the cutoff radius is limited by the size of the simulation box (e.g., when the nearest image convention is used, r_{cut} is limited to half the boxsize[14]), a larger value of r_{cut} may imply that a zeolite consisting of more unit cells is required.

Although there have been several studies on the comparison between the Wolf method and the Ewald summation,[16,42–44] these studies mainly focus on reproducing the total energy of the simulation box at finite temperature and not on the energy change when a polar molecule is inserted into a simulation box. We have tested the Ewald summation and the Wolf method for calculating the heat of adsorption of CO_2 at zero loading in all silica MFI-type zeolite using the recently proposed force field of ref 45 (calculated using eq 22). As our goal is to investigate the differences between the Wolf method and the Ewald summation, we have eliminated all Lennard-Jones interactions, and all atoms in the system are modeled as hard spheres with a diameter of 3 Å. The partial charges of the atoms are taken as follows: $q_{\mathrm{O(zeo)}} = -1.025e$, $q_{\mathrm{Si(zeo)}} = +2.05e$, $q_{\mathrm{C(CO_2)}} = +0.6512e$, $q_{\mathrm{O(CO_2)}} = -0.3256e$. Grid interpolation techniques[46,8] to compute pair potentials have not been used. As the self-interactions of the Wolf method for CO_2 and the zeolite are constant and independent of the reference state, only pair interactions between CO_2 and the zeolite (eqs 30 and 31) need to be considered for the Wolf method.

In Figure 5, we have plotted the energy difference $\langle U_1\rangle_1 - \langle U_0\rangle_0 - \langle U_g\rangle$ calculated using eq 22 for both the Wolf method and the modification by Fennell and Gezelter (eq 31) as a function of α for various values of the cutoff radius. In the same figure, the result obtained from the Ewald summation is also included. Clearly, the result strongly depends on the precise value of α and r_{cut}, especially for small values of r_{cut}. The value of $\alpha = 0.2$ Å^{-1} as suggested by some authors[16,43] leads to incorrect results for this system. At this point it is important to note that for practical usage of the Wolf method, the computed results should not depend too strongly on the precise values of r_{cut} and α. From Figure 5 it becomes clear that that there is only a small range of α which leads to results consistent with the Ewald summation and that a cutoff radius of at least 25 Å is required to avoid a large sensitivity for the precise values of α and r_{cut}. This holds for both the original Wolf method and the modification by Fennell and Gezelter. Although for some values of α and

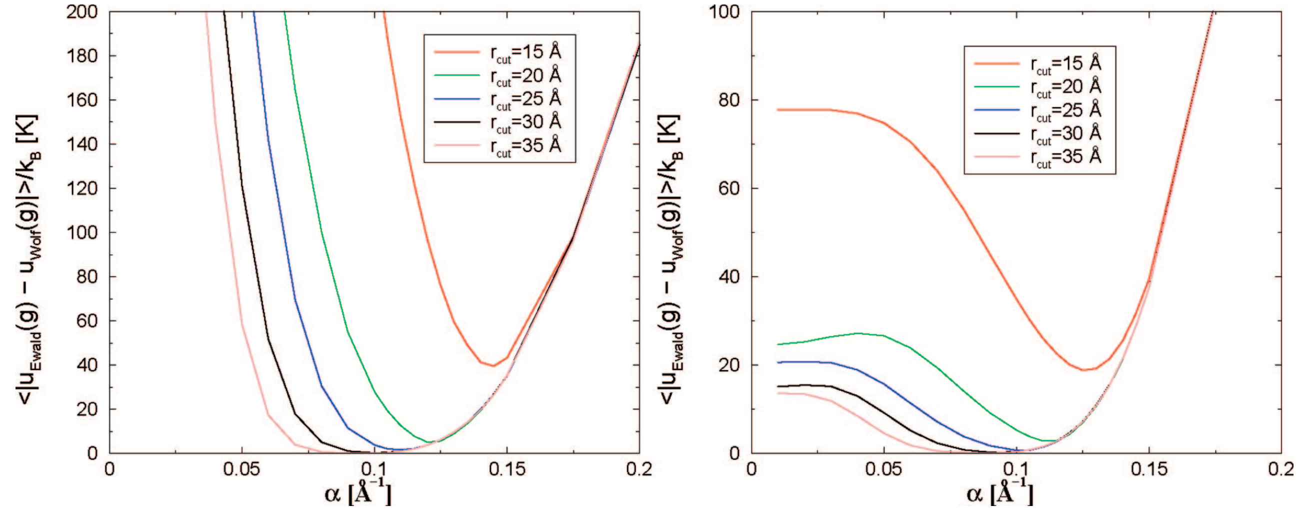


Figure 6. Absolute energy difference between the Ewald summation and the Wolf method for individual configurations of CO_2 (denoted as *g*) in MFI-type zeolite, averaged over all possible configurations of CO_2 that do not overlap with the zeolite. Left: Original Wolf method (eq 28). Right: The modification by Fennell and Gezelter (eq 29). The zeolite consists of 4 × 4 × 6 unit cells. The size of the simulation box is therefore $a = 80.088$ Å, $b = 79.596$ Å, $c = 80.298$ Å. In all simulations, 3×10^7 trial insertions are used (eq 20).

r_{cut} the average heat of adsorption computed using the Wolf method is identical to the one computed using the Ewald summation, this does not guarantee that the energy of each individual configuration of CO_2 computed using both methods is identical. In Figure 6, we plotted the average absolute energy difference between the adsorption energy computed using the Ewald summation ($u_{Ewald}(g)$) and the adsorption energy computed using the Wolf method ($u_{Wolf}(g)$), calculated for the same configurations (g) of CO_2 and averaged over all possible configurations of CO_2 that do not overlap with the zeolite. For the optimal value of α for $r_{cut} = 20$ Å, the average absolute energy difference between the Wolf method and the Ewald summation for a single configuration of CO_2 is on average 4%, while the difference in average energy is zero for the optimal value of α. This means that although the Wolf method may predict correct averages for the optimal α and r_{cut}, one should also test whether the energy of individual configurations is computed correctly (i.e., identical to the result obtained by the Ewald summation). If this is not the case, one can expect artifacts in the simulation results. From Figures 5 and 6 it becomes clear that the Wolf method only converges to the correct energy for each configuration of CO_2 for (1) a cutoff radius of at least 25 Å and (2) a narrow range of α.

The results of the Ewald summation are not very sensitive to the precise values of α and r_{cut}, see Table 1. It is well-known that if only nearest images are used to compute the real-space part of the energy (eq 24), the total energy of the system as a function of k at constant α should always converge to the same value for large k, provided that α is large enough. The difference between the averages in Table 1 is less than 1% which is approximately equal the error in any of the reported values in Table 1. We verified that the energy of each single configuration of CO_2 is within 0.1% for any of the values of α, r_{cut}, and k reported in Table 1. It is important to note that the finite-size effect is absent for our system when charges are handled using the Ewald summation, and therefore one can use a small simulation box (2 × 2 × 3 unit cells). For the Wolf method however, this is not possible as a cutoff radius of at least 25 Å is needed. According to the nearest image convention the size of the simulation box needs to be at least twice the value of r_{cut}. Using such a large cutoff radius is inconvenient, as the usual sorbate−zeolite and sorbate−sorbate Lennard-Jones type interactions have a cutoff radius of typically 12 Å.[11,12,17,47] We compared the CPU time required for the Wolf method for $\alpha = 0.1$ Å^{-1} and $r_{cut} = 25$ Å (4 × 4 × 6 unit cells) with the Ewald summation ($\alpha = 0.3$ Å^{-1}, $k = 15$, $r_{cut} = 12$ Å, 2 × 2 × 3 unit cells) for the same number of trial insertions of CO_2. Although in this simulation using the Ewald method 88% of the CPU time is spent in calculating the Fourier energy (eq 24), the simulation using the Ewald method is still 7 times faster due to the fact that a smaller cutoff radius can be used.

Table 1. Average Energy Difference $\langle U_1\rangle_1 - \langle U_0\rangle_0 - \langle U_g\rangle$ for CO_2 in All Silica MFI-Type Zeolite Computed Using the Ewald Summation[a]

no. of unit cells	α [Å^{-1}]	k	r_{cut} [Å]	$(\langle U_1\rangle_1 - \langle U_0\rangle_0 - \langle U_g\rangle)/k_B$ [K]
4×4×6	0.4	30	20	−110.4
2×2×3	0.3	15	12	−108.9
2×2×3	0.3	20	12	−109.8
2×2×3	0.3	30	12	−109.6
2×2×3	0.4	10	12	−110.6
2×2×3	0.4	20	12	−110.1
2×2×3	0.3	10	18	−109.2
2×2×3	0.3	20	18	−109.2
2×2×3	0.3	30	18	−109.2
2×2×3	0.5	20	18	−109.6

[a] In all simulations, 3×10^7 trial insertions are used (eq 22).

While other studies found quite good agreement between the Wolf method and the Ewald summation for condensed phases,[16,42–44] our findings (Figures 5 and 6) show that convergence to the Ewald result is more difficult for zeolites. The reason is that the Wolf method critically depends on the screening of electrostatics around a central ion. Due to the screening the effective range of the potential will be

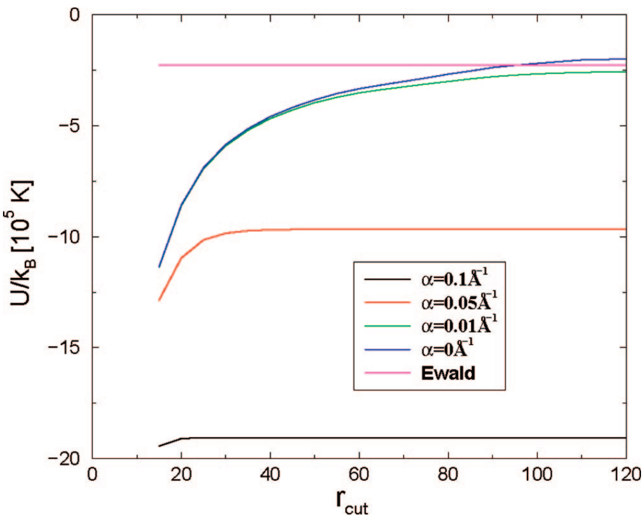


Figure 7. Total energy of a single configuration of 100 randomly placed Na^+ ions and 100 randomly placed Cl^- ions in a simulation box of 250 Å × 250 Å × 250 Å, computed using the Ewald summation, the Wolf method, and a direct r^{-1} summation ($\alpha = 0$ Å^{-1}). Hard-core overlaps are avoided in this single configuration.

reduced, and therefore a summation of pair interactions seems sufficient to describe effective interactions between ions. As a large part of the zeolite consists of empty pores (at low loading), the screening of Coulombic interactions is less efficient and a large cutoff radius is required.

It is instructive to test the convergence of the Wolf method for a worse-case scenario, i.e. a system that is highly disordered with large concentration gradients, even though the Wolf method was not designed for such a system. For molten NaCl at high density, it has been shown that the total energy computed by the Wolf method is in reasonable agreement with the Ewald summation.[44] Consider a system of 100 Na^+ ions and 100 Cl^- ions in a simulation box of 250 Å × 250 Å × 250 Å, the density of this system is extremely low. The ions are placed at random positions in the simulation box in such a way that hard-core overlaps (here: ion–ion distance smaller than 3 Å) are avoided. For a single configuration of the ions, we investigated the total energy computed using the Wolf method for various values of α[15]

$$E_{\mathrm{Wolf}} = -\left(\frac{\alpha}{\sqrt{\pi}} + \frac{\mathrm{erfc}(\alpha r_{\mathrm{cut}})}{2r_{\mathrm{cut}}}\right)\sum_{i=1}^{N} q_i^2 + \sum_{i<j} \varphi_{\mathrm{Wolf}}(r_{ij}) \qquad (32)$$

The results are shown in Figure 7. Clearly, for $\alpha = 0.1$ Å^{-1} and $\alpha = 0.05$ Å^{-1} the total energy of the Wolf method converges to an incorrect total energy. For $\alpha = 0.01$ Å^{-1}, the energy of the Wolf method converges to the result obtained by the Ewald summation. However, the convergence is just as slow as a direct r^{-1} summation of pair interactions ($\alpha = 0$ Å^{-1} in eq 28). This clearly shows that the Wolf method critically depends on the screening of electrostatic interactions and that for a given system one should always check whether such a screening is applicable.

It is well-known that a direct r^{-1} summation only converges for very large values of the cutoff radius,[14] for example more than 40 Å for neutral molecules. Damping the Coulombic potential reduces this range, although still a cutoff of at least 25 Å seems to be the minimum requirement, but at the cost of uncertainty in the damping constant α. Different values for α are recommended in the literature implying this damping is system specific and possibly dependent on the molecule type, framework topology, temperature, etc. The Ewald summation is able to compute the electrostatic energy exactly (to within any specified numerical precision) for a system of charges in an infinitely periodic structure at a cost less than quadratic with the number of particles ($N^{3/2}$ for Ewald). The energy and forces computed with the Ewald summation are well-defined and unique. They do not depend on a judicious choice of damping parameters and cutoff values, and Ewald type methods are at the same level of accuracy significantly cheaper.

6. Conclusions

We have introduced a new method to compute the heat of adsorption for adsorbates in framework structures with nonframework cations that requires only a single simulation of the host. The main difference with the conventional Widom's test particle method is that Widom's test particle method only provides the Henry coefficient at a certain temperature, and therefore, additional simulations at different temperatures and a numerical differentiation will be necessary to compute the heat of adsorption. The often used method based on energy differences is far more sensitive to the precise displacements of the nonframework cations than to the interactions with the adsorbate and therefore this method is unsuited for structures with nonframework cations. In addition, we have shown that for studying adsorption in zeolites, the Ewald summation is superior to the recently proposed Wolf method (and other direct summation methods), both in accuracy and speed. The Ewald method can be implemented very efficiently for Monte Carlo methods as only the change for moving atoms needs to be computed.

Acknowledgment. T.J.H.V. acknowledges The Netherlands Organization for Scientific Research (NWO-CW) for financial support through a VIDI grant. This work is supported by the Spanish "Ministerio de Educación y Ciencia" (CTQ2007-63229/BQU) and by the resources, technical expertise, and assistance provided by BSC-CNS, and by the National Science Foundation (CTS-0507013).

Appendix: Chemical Potential

The chemical potential of a chain molecule can be computed using the Rosenbluth scheme[36]

$$\beta\mu = -\ln\langle W\rangle \qquad (33)$$

in which W is the Rosenbluth weight. Consider the situation that we use f trial positions for the first bead and k trial positions for the other $N-1$ beads. The Rosenbluth weight W is then defined as

$$W = \frac{\prod_{i=1}^{N} \sum_{j} \exp[-\beta u_{ij}]}{fk^{N-1}} \qquad (34)$$

in which u_{ij} is the energy of the jth trial position of the ith chain segment. In this scheme, trial directions are selected with a probability

$$P_{\text{sel}} = \frac{\exp[-\beta u_{ij}]}{\sum_j \exp[-\beta u_{ij}]} \tag{35}$$

It is sometimes convenient to choose a different selection of trial segments than in the Rosenbluth scheme,

$$P^*_{\text{sel}} = \frac{\exp[-\beta u^*_{ij}]}{\sum_j \exp[-\beta u^*_{ij}]} \tag{36}$$

in which $u^*_{ij} \neq u_{ij}$. This is for example useful for charged systems, in which only the real-space part of the Coulombic potential is used to select trial segments. Another application is the use of a second cutoff radius to select trial segments.[27,37] Below, we will show that in the case of a different selection rule for trial segments, the chemical potential can be computed using

$$\beta\mu = -\ln\langle W\rangle = -\ln\left\langle \frac{\prod_{i=1}^{N}\sum_j \exp[-\beta u^*_{ij}]}{fk^{N-1}} \times \exp[-\beta\delta]\right\rangle = -\ln\langle W^* \times \exp[-\beta\delta]\rangle \tag{37}$$

in which δ is the energy difference $\delta = \sum_i [u_{ij(i)} - u^*_{ij(i)}]$ in which $j(i)$ is the selected trial direction for segment i. Note that if $u^*_{ij} = u_{ij}$, then we recover eq , and if $u^*_{ij} = 0$, then we have $W^* = 1$ and we recover the original Widom's test particle method.[14] To show that eq 37 is correct, consider the extension of a chain with a single trial segment. The probability of generating a given set of k trial directions with orientations $\mathbf{\Gamma}_1$ though $\mathbf{\Gamma}_k$ is

$$P_{\text{gen}}(\mathbf{\Gamma}_1)\cdots P_{\text{gen}}(\mathbf{\Gamma}_k)d\mathbf{\Gamma}_1\cdots d\mathbf{\Gamma}_k \tag{38}$$

The probability to select any of these trial segment follows from eq 36:

$$P^*_{\text{sel}} = \frac{\exp[-\beta u^*_{ij}]}{\sum_j \exp[-\beta u^*_{ij}]} = \frac{\exp[-\beta u^*_{ij}]}{w^*_i(\mathbf{\Gamma}_1\cdots\mathbf{\Gamma}_k)} \tag{39}$$

We have to consider the average of $w^*_i \exp[-\beta(u_{ij} - u^*_{ij})]$ over all possible sets of trial segments (j) and all possible choices of the trial segment (j'):

$$\begin{aligned}\left\langle \frac{w^*_i \exp[-\beta(u_{ij} - u^*_{ij})]}{k}\right\rangle &= \int \prod_{j=1}^{k} d\mathbf{\Gamma}_j P_{\text{gen}}(\mathbf{\Gamma}_j) \sum_{j'=1}^{k} P^*_{\text{sel}} \times \frac{w^*_i \exp[-\beta(u_{ij} - u^*_{ij})]}{k} \\ &= \int \prod_{j=1}^{k} d\mathbf{\Gamma}_j P_{\text{gen}}(\mathbf{\Gamma}_j) \sum_{j'=1}^{k} \frac{\exp[-\beta u^*_{ij}]}{w^*_i} \times \frac{w^*_i \exp[-\beta(u_{ij} - u^*_{ij})]}{k} \\ &= \int \prod_{j=1}^{k} d\mathbf{\Gamma}_j P_{\text{gen}}(\mathbf{\Gamma}_j) \sum_{j'=1}^{k} \frac{\exp[-\beta u_{ij}]}{k}\end{aligned} \tag{40}$$

As the labeling of trial segment is arbitrary, this leads to

$$\left\langle \frac{w^*_i \exp[-\beta(u_{ij} - u^*_{ij})]}{k}\right\rangle = \int d\mathbf{\Gamma} P_{\text{gen}}(\mathbf{\Gamma}) \exp[-\beta u_i] \tag{41}$$

By definition, we have[14]

$$\left\langle \frac{w_i}{k}\right\rangle = \int d\mathbf{\Gamma} P_{\text{gen}}(\mathbf{\Gamma}) \exp[-\beta u_i] \tag{42}$$

in which

$$w_i = \sum_{j=1}^{k} \exp[-\beta u_{ij}] \tag{43}$$

This proves that

$$\langle W\rangle = \langle W^* \exp[-\beta\delta]\rangle \tag{44}$$

In a similar way, one can show that

$$\langle F \times W\rangle = \langle F \times W^* \exp[-\beta\delta]\rangle \tag{45}$$

in which F can be an arbitrary function that only depends on the position of the generated chain.

References

(1) Baerlocher, Ch.; McCusker, L. B.; Olson, D. H. *Atlas of Zeolite Framework Types*, 6th ed.; Elsevier: Amsterdam, 2007.

(2) Bezus, A. G.; Kiselev, A. V.; Lopatkin, A. A.; Du, P. Q. *J. Chem. Soc., Faraday Trans. II* **1978**, *74*, 367.

(3) Snurr, R. Q.; Bell, A. T.; Theodorou, D. R. *J. Phys. Chem.* **1993**, *97*, 13742.

(4) Smit, B.; Siepmann, J. I. *Science* **1994**, *264*, 1118.

(5) Smit, B.; Maesen, Th. L. M. *Nature* **1995**, *374*, 42.

(6) Maginn, E. J.; Bell, A. T.; Theodorou, D. N. *J. Phys. Chem.* **1995**, *99*, 2057.

(7) Macedonia, M. D.; Maginn, E. J. *Mol. Phys.* **1999**, *96*, 1375.

(8) Vlugt, T. J. H.; Krishna, R.; Smit, B. *J. Phys. Chem. B* **1999**, *103*, 1102.

(9) Fuchs, A. H.; Cheetham, A. K. *J. Phys. Chem. B* **2001**, *105*, 7375.

(10) Pascual, P.; Ungerer, P.; Tavitian, B.; Pernot, P.; Boutin, A. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3684.

(11) Dubbeldam, D.; Calero, S.; Vlugt, T. J. H.; Krishna, R.; Maesen, Th. L. M.; Smit, B. *J. Phys. Chem. B* **2004**, *108*, 12301.

(12) Dubbeldam, D.; Calero, S.; Vlugt, T. J. H.; Krishna, R.; Maesen, Th. L. M.; Beerdsen, E.; Smit, B. *Phys. Rev. Lett.* **2004**, *93*, 088302.

(13) Liu, B.; Smit, B.; Calero, S. *J. Phys. Chem. B* **2006**, *110*, 20166.

(14) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: from Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, CA, 2002.

(15) Wolf, D.; Keblinski, P.; Phillpot, S. R.; Eggebrecht, J. *J. Chem. Phys.* **1999**, *110*, 8254.

(16) Fennell, C. J.; Gezelter, J. D. *J. Chem. Phys.* **2006**, *124*, 234104.

(17) Calero, S.; Dubbeldam, D.; Krishna, R.; Smit, B.; Vlugt, T. J. H.; Denayer, J. F. M.; Martens, J. A.; Maesen, Th. L. M.

J. Am. Chem. Soc. **2004**, *126*, 11377.

(18) García-Pérez, E.; Dubbeldam, D.; Maesen, Th. L. M.; Calero, S. *J. Phys. Chem. B* **2006**, *110*, 23968.

(19) García-Pérez, E.; Dubbeldam, D.; Liu, B.; Smit, B.; Calero, S. *Angew. Chem.* **2006**, *45*, 276.

(20) Liu, B.; García-Pérez, E.; Dubbeldam, D.; Smit, B.; Calero, S. *J. Phys. Chem. C* **2007**, *111*, 10419.

(21) Wood, G. B.; Panagiotopoulos, A. Z.; Rowlinson, J. S. *Mol. Phys.* **1988**, *63*, 49.

(22) Dunne, J. A.; Mariwala, R.; Rao, M.; Sircar, S.; Gorte, R. J.; Myers, A. L. *Langmuir* **1996**, *12*, 5888.

(23) Smit, B.; Siepmann, J. I. *J. Phys. Chem.* **1994**, *98*, 8442.

(24) Karavias, F.; Myers, A. L. *Langmuir* **1991**, *7*, 3118.

(25) Widom, B. *J. Chem. Phys.* **1963**, *39*, 2802.

(26) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.

(27) Vlugt, T. J. H. *Mol. Simul.* **1999**, *23*, 63.

(28) Siepmann, J. I.; Frenkel, D. *Mol. Phys.* **1992**, *75*, 59.

(29) Frenkel, D.; Mooij, G. C. A. M.; Smit, B. *J. Phys.: Condens. Matter* **1992**, *4*, 3053.

(30) de Pablo, J. J.; Laso, M.; Suter, U.W. *J. Chem. Phys.* **1992**, *96*, 6157.

(31) Siepmann, J. I. In *Computer simulation of biomolecular systems: theoretical and experimental applications;* van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Escom Science Publisher: Leiden, 1993; p 249.

(32) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes: The art of scientific computing;* Cambridge University Press: Cambridge, 1986.

(33) Consta, S.; Wilding, N. B.; Frenkel, D.; Alexandrowicz, Z. *J. Chem. Phys.* **1999**, *110*, 3220.

(34) Consta, S.; Vlugt, T. J. H.; Wichers Hoeth, J.; Smit, B.; Frenkel, D. *Mol. Phys.* **1999**, *97*, 1243.

(35) Combe, N.; Vlugt, T. J. H.; ten Wolde, P. R.; Frenkel, D. *Mol. Phys.* **2003**, *101*, 1675.

(36) Rosenbluth, M. N.; Rosenbluth, A. W. *J. Chem. Phys.* **1955**, *23*, 356.

(37) Vlugt, T. J. H.; Martin, M. G.; Smit, B.; Siepmann, J. I.; Krishna, R. *Mol. Phys.* **1998**, *94*, 727.

(38) Pluth, J. J.; Smith, J.V. *J. Am. Chem. Soc.* **1980**, *102*, 4704.

(39) Patra, M.; Karttunen, M.; Hyvönen, M. T.; Falck, E.; Lindqvist, P.; Vattulainen, I. *Biophys. J.* **2003**, *84*, 3636.

(40) Patra, M.; Karttunen, M.; Hyvönen, M. T.; Falck, E.; Vattulainen, I. *J. Phys. Chem. B* **2004**, *108*, 4485.

(41) Ewald, P. P. *Ann. Phys.* **1921**, *64*, 253.

(42) Demontis, P.; Spanu, S.; Suffritti, G. B. *J. Chem. Phys.* **2001**, *114*, 7980.

(43) Zahn, D.; Schilling, B.; Kast, S. M. *J. Phys. Chem. B* **2002**, *106*, 10725.

(44) Avendano, C.; Gil-Villegas, A. *Mol. Phys.* **2006**, *104*, 1475.

(45) García-Pérez, E.; Parra, J. B.; Ania, C. O.; García-Sánchez, A.; van Baten, J. M.; Krishna, R.; Dubbeldam, D.; Calero, S. *Adsorption* **2007**, *13*, 469.

(46) June, R. L.; Bell, A. T.; Theodorou, D. N. *J. Phys. Chem.* **1992**, *96*, 1051.

(47) Martin, M. G.; Thompson, A. P.; Nenoff, T. N. *J. Chem. Phys.* **2001**, *114*, 7174.

CT700342K

J. Chem. Theory Comput. **2008,** *4,* 1119–1128

JCTC Journal of Chemical Theory and Computation

Ensuring Mixing Efficiency of Replica-Exchange Molecular Dynamics Simulations

Mark J. Abraham and Jill E. Gready*

Computational Proteomics Group, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra, ACT, 2601, Australia

Received January 16, 2008

Abstract: We address the question of constructing a protocol for replica-exchange molecular dynamics (REMD) simulations that make efficient use of the replica space, assess whether published applications are achieving such “mixing” efficiency, and provide a how-to guide to assist users to plan efficient REMD simulations. To address our first question, we introduce and discuss three metrics for assessing the number of replica-exchange attempts required to justify the use of a replica scheme and define a “transit number” as the lower bound for the length of an efficient simulation. Our literature survey of applications of REMD simulations of peptides in explicit solvent indicated that authors are not routinely reporting sufficient details of their simulation protocols to allow readers to make independent assessments of the impact of the method on their results, particularly whether mixing efficiency has been achieved. Necessary details include the expected or observed replica-exchange probability, together with the total number of exchange attempts, the exchange period, and estimates of the autocorrelation time of the potential energy. Our analysis of cases where the necessary information was reported suggests that in many of these simulations there are insufficient exchanges attempted or an insufficiently long period between them to provide confidence that the simulation length justifies the size of the replica scheme. We suggest guidelines for designing REMD simulation protocols to ensure mixing efficiency. Two key recommendations are that the exchange period should in general be larger than 1 ps and the number of exchange attempts should be chosen to significantly exceed the transit number for the replica scheme.

1. Introduction

Interest from chemists in replica-exchange techniques for molecular dynamics simulations of biomolecules has grown rapidly since the seminal paper in 1999,[1] although the method was in use in physics before then.[2] One attractive feature of the method is its generality. Implementations of the algorithm have been developed that exchange temperatures,[1] modified nonbonded potentials,[3] the solvent contribution to the potential energy surface,[4] graininess of solute structure,[5–7] other Hamiltonian features,[8–10] λ values in free-energy calculations,[11] and bias potentials as used in meta-dynamics simulations.[12,13] Even pseudoexchange with a previously generated ensemble of structures[5,14] has been explored.

Techniques have been developed to optimize the distribution of the temperatures of the replicas in the conventional algorithm[15] to reduce bottlenecks in temperature exchange, and to produce a set of replica temperatures that span a given temperature range to yield a given average exchange probability.[16] Other work to improve exchange efficiency through the use of hybrid implicit−explicit solvent Hamiltonians during the exchange attempts[17] has been reported. There has been much discussion of the relative efficiency of canonical sampling of biomolecular systems using replica-exchange and fixed-temperature equilibrium sampling.[18–23] However, there has been no discussion of the dependence of the efficiency of the chosen replica scheme on the number of exchanges attempted over a simulation, either for conformational searching or for canonical sampling.

---

* Corresponding author e-mail: Jill.Gready@anu.edu.au.

10.1021/ct800016r CCC: $40.75

There has been some discussion of the choice of period between exchange attempts (the *exchange period*). Zhang et al.[19] justify their choice of period (1 ps) by observing that, for their system, an alternative choice of 0.02 ps led to significant reduction in the traversal of temperature space by replicas, and that an alternative choice of 5 ps led to results qualitatively similar to those for 1 ps. They attributed the reduction in traversal to the reduced ability of replicas to stabilize in their new ensemble before being considered for exchange back to their former ensemble and recommended the use of an exchange period larger than 0.02 ps.

An alternative way of expressing the thought underlying this recommendation is that the probabilities for successive exchange events were time-correlated because the underlying potential energies and the configurations generating them were also time-correlated. Successive integration steps in molecular dynamics (MD) simulations of biomolecular systems occur in time steps of around 1−4 fs. These lead only to small changes in position and velocity coordinates and, because of the nature of the mathematical functions in the force fields, correspondingly small changes in potential energy. It follows that the correlation of an observable such as the potential energy with former values of itself (the *autocorrelation*) will be appreciable and will decay with increasing lengths of time between observations.

When measuring the value of an observable f such as potential energy in an MD simulation, we would like to make a set of Q independent (i.e., uncorrelated) measurements f_i. From these, we can form the estimator of the mean, $\overline{f} = 1/Q \sum_{i=1}^{Q} f_i$. The variance of that estimator, $\overline{f}$ (the square of the standard error), is then given by $\sigma^2(\overline{f}) = \sigma^2(f)/Q$ for $\sigma^2(f)$, the variance of f. It can be shown[24,25] that the variance of $\overline{f}$ for Q correlated samples collected at intervals Δt is given by

$$\sigma^2(\overline{f}) = \frac{2\tau_{\mathrm{int},f}\sigma^2(f)}{Q\Delta t} \tag{1}$$

for $\tau_{\mathrm{int},f}$, the *integrated autocorrelation time*. $\tau_{\mathrm{int},f}$ may be computed with

$$\tau_{\mathrm{int},f} = \Delta t\left(\frac{1}{2} + \sum_{t=1}^{\infty}\frac{C_f(t)}{C_f(0)}\right) \tag{2}$$

where $C_f(t)$ is the *autocorrelation function* of f. This latter function describes the extent to which successive measurements of f are correlated in time and is defined by

$$C_f(t) = \langle f(\tau)\, f(t+\tau)\rangle \tag{3}$$

averaging over all time origins τ. It is clear that the variance of the estimate of $\overline{f}$ of the correlated sample is a factor of $2\tau_{\mathrm{int},f}/\Delta t$ larger than it would be if the samples were independent (when $C_f(t)$ is a δ function). An alternative interpretation of eq 1 is that there are only $Q\Delta t/2\ \tau_{\mathrm{int},f}$ samples of f that are effectively statistically independent. It follows that the length of a molecular dynamics simulation for measuring f should be such that $Q\Delta t \gg \tau_{\mathrm{int},f}$ independent samples can be gathered. Conformational searching with replica-exchange molecular dynamics (REMD) seems most likely to be effective when successive exchange attempts make measurements of the potential energy U that are independent. Accordingly, the chosen exchange period should be at least as large as $\tau_{\mathrm{int},U}$.

There is further support for the above recommendation of Zhang et al.[19] from the findings of Periole and Mark,[22] who performed a comparative study of exchange periods of 0.1, 0.5, 2, and 5 ps in 10 ns REMD simulations over 20 replicas of a β-heptapeptide in explicit solvent. In this study, they observed a greater number of successful exchanges for 0.1 ps, which they attributed to an excessive correlation between successive exchange attempts. Further, they observed a higher frequency of "back exchange" following successful exchanges for both 0.1 and 0.5 ps, again suggesting correlation. These authors concluded that attempting replica exchanges with periods that are this short is not helpful in achieving the overall goal of a normal REMD simulation. They also tried a simulation where velocities were not scaled after successful exchanges and found that the correlation phenomena were not strongly affected by the use of velocity scaling. Further, they also presented results which indicated that the relative efficiency of various replica-exchange schemes for reaching the folding equilibrium across the generalized ensemble was greatest when the exchange period was large enough to prevent such correlation, but no larger.

Recent work by Sindhikara et al.[26] makes the claim that sampling efficiency in REMD simulations increases with a decreasing exchange-attempt period, even below the minimum period identified in refs 19 and 22. This claim is based on their observations of the variation of the dihedral angle root-mean-square deviation (rmsd) distribution of some model peptides with the exchange-attempt period, in each case compared with a reference REMD simulation. They use implicit-solvent simulations and chose as their comparison metric a 36 × 36 2D histogram of the φ and ψ dihedral angles. These were normalized, and the rmsd between the test and reference histograms was formed and then averaged for each residue. They did not discuss alternative metrics. It is possible for averaging over time and residue to mask significant variation in the distribution of dihedral angles. However, it is not possible to assess whether the observed variation in the values of the metric is significant, although this could be done by demonstrating how much variation occurs between replicates of the reference REMD simulation. It would be of interest to see if the above claims can be substantiated for explicit-solvent simulations or for other metrics.

Sindhikara et al. also demonstrated that the distribution of potential energies remains Boltzmann over the range of exchange-attempt periods they studied. They then argued that, because the average probability of an exchange acceptance[27,28] $\overline{P}_{\mathrm{acc}}$ may be expressed as

$$\overline{P}_{\mathrm{acc}}(\beta_1, \beta_2) = \int_1 \int_2 P(U_1)\, P(U_2) \min(1, \mathrm{e}^{\Delta\beta\Delta U})\, \mathrm{d}U_2\, \mathrm{d}U_2 \tag{4}$$

and the probability distributions of potential energies $P(U)$ are independent of the exchange-attempt period, this implies that the value of this integral is also necessarily independent of the exchange-attempt period. This is not true for exchange-attempt periods shorter than the autocorrelation time of the potential energy. Following a successful exchange, $P(U_1)$ and $P(U_2)$ are correlated even though they are both samples from correct Boltzmann distributions. That correlation will increase with decreasing exchange-attempt period. Accordingly, there must be a dependence of $\overline{P}_{\mathrm{acc}}$ on the exchange-attempt period, which agrees with the findings of Periole and Mark above.[22]

In this work, we introduce the distinct but related concepts of *thermodynamic efficiency* and *mixing efficiency* of a REMD simulation. Such a simulation attains thermodynamic efficiency when it converges the observables of the ensembles of interest, and mixing efficiency when each replica has experienced the full range of replica conditions; that is, the replica-exchange system is “well mixed”. We note that it is quite possible to achieve either thermodynamic or mixing efficiency without achieving the other. In a case where the free energy surface is flat with respect to the range of conformations accessible to replicas, the extra sampling ability of the higher-range replicas is wasted, and thermodynamic efficiency will be attained if the simulation is long enough. Even if mixing efficiency was attained in such a case by simulating long enough with replicas spaced closely enough, the simulation would overall still be wasteful of resources. In a case where the free energy surface is rugged with respect to the range of conformations accessible to replicas, it can be difficult to attain thermodynamic efficiency, even if, as above, mixing efficiency is attained.

Although the existence of thermodynamic efficiency of REMD is just as difficult to assess as ensemble convergence in normal MD simulations, recent work by Lyman and Zuckerman has shown some progress.[29] They proposed a method for assessing structural convergence in a knowledge- and parameter-free way and suggested that it could be applied to trajectories from replica-exchange simulations as they traverse temperature space. We agree with their expectation that, in an efficiently mixed replica-exchange simulation, the “decorrelation time” they define would be uniform across the replicas. They also expected that this decorrelation time will be shorter for replica-exchange simulations than for standard ones. In this work, we will consider the impact of the size of the replica-exchange scheme on the decorrelation time.

Thermodynamic and mixing efficiency both contribute to determining the total cost of a REMD simulation. An effective REMD simulation needs a replica scheme whose range of conditions is sufficient for a convergence of observables and needs to be long enough to achieve this convergence. A REMD simulation that is both effective and minimizes the total cost requires that the range of conditions in the replica scheme be sufficient but not excessive,[23] that it be no longer than required for observable convergence, and that it be long enough for good mixing to occur. Our objective in this work is to provide guidance on assessing the third of these requirements, that is, ensuring mixing efficiency. Accordingly, we will introduce and compare three metrics for assessing mixing efficiency, survey some REMD simulations reported in the literature, make some observations about replica-exchange periods and potential energy autocorrelation times, and provide some practical guidelines for future REMD simulations.

2. Methods

2.1. Measuring Mixing Efficiency of REMD Simulations. As a basis for proposing three metrics for measuring the efficiency of replica-exchange schemes, we start by briefly reviewing replica-exchange methods. These methods are examples of generalized ensemble schemes, in which multiple independent simulations are carried out using different ensembles. It is desired that each independent simulation samples states according to some weight function that corresponds to a thermodynamic ensemble of choice. Periodically, exchanges of configurations between pairs of ensembles are attempted. For the generalized ensemble to converge to its equilibrium state, the condition of detailed balance must be observed for such exchanges. This can be satisfied by the use of a generalized Metropolis criterion for exchanges where, for Boltzmann-weighted states, the probability of exchange between replicas i and $i + 1$ from ensembles with temperatures and potential energies T_i, T_{i+1}, U_i, and U_{i+1} respectively, is

$$P(\text{exchange}) = \min\left\{1, \exp\left[\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_{i+1}}\right)(U_i - U_{i+1})\right]\right\} \tag{5}$$

where k_B is Boltzmann’s constant. The temperatures are usually chosen such as to keep the average exchange probability constant over replica space. Accordingly, and in the absence of known bottlenecks in the free energy surface,[15] the temperatures are distributed exponentially over a chemically relevant range.

2.1.1. Metric One. We start with the straightforward Metric One. It is normally true that one of the extremal ensembles of a replica-exchange scheme is the target ensemble from which the simulation data are to be collected (the *minimum*), and the other extreme is expected to have the greatest capacity to enable the generalized ensemble to sample the target ensemble in an efficient manner (the *maximum*; e.g., Observation 3 in ref 20). We define a *transit* to be the process of a replica visiting the maximum and subsequently the minimum. A given replica scheme will only result in an efficient computation when there are sufficient exchange events to ensure significant probability that replicas have undergone transit(s). Given an estimate of the average probability of exchange occurring across replica space, it is possible to calculate the number of exchange attempts that are required for a given probability of at least one transit occurring. We define the *transit number* of the replica scheme to be the number of replica-exchange attempts required for a 95% probability that at least one replica has visited the maximum before visiting the minimum.

This transit number may be interpreted as the minimum number of exchange attempts that is necessary in order to begin to justify the use of that particular replica scheme. Any simulation that attempts fewer exchanges will not be making effective use of the replica space, and thus computing power. We note, however, that the transit number merely establishes a lower bound and implies nothing for the convergence of the ensemble, or the autocorrelation time of any observables. In practice, in order to explore the conformational space efficiently and thoroughly, enough exchange attempts will be required such that multiple replicas will have the opportunity to undergo many transits. That number of exchange attempts will be dependent on the nature of the generalized free energy surface. Depending on the nature of that surface and the number of replicas, it may not be necessary to achieve transits in order to achieve sufficient sampling of the target ensemble. However it *is* necessary to achieve transits in order to sample efficiently, because otherwise the high-temperature replicas are not useful, and the computational cost could be reduced by eliminating those replicas. One might think it "fairer" to define transits as those between a top and bottom subset of replicas, particularly in cases where data are collected from more than the lowest-temperature replica, or where the higher temperatures are known to be excessive. Because all replicas that reach an extremum have done so after entering the subset for that extremum, such a replica system will have an effective transit number identical with that of a smaller system with single-replica subsets at the extrema.

It is straightforward to model the time evolution of a replica scheme with M replicas with average exchange probability p. The "odd" pairs of replicas are the pairs (1, 2), (3, 4),..., $(2n+1,\ 2n+2)$ for $0 \le n \le {}^1/_2(M-2)$, and the "even" pairs are (2, 3), (4, 5),..., $(2n,\ 2n+1)$ for $0 \le n \le {}^1/_2(M-1)$, where the integers indicate the index of the ensemble within the ordered set. In line with the implementation of the REMD method in the molecular simulation package GROMACS,[30–32] we limit our study to a scheme where only the odd pairs may exchange at one attempt, and only the even pairs at the next, alternating indefinitely. There are alternatives here, such as permitting exchange attempts between all neighboring replicas, or all possible replicas. If an exchange attempt was made between replicas i and $i+1$ and a further attempt was made at that time between $i+1$ and $i+2$, then the probabilities for all of these exchanges would depend on the energy of all three replicas. Exchange between i and $i+2$ would be much more unlikely than between neighboring pairs, however. Considering more than neighboring pairs thus increases the complexity of the computation at each exchange attempt for very few extra successful exchanges. In a practical REMD simulation, this increase in complexity would increase the length of time spent by the parallel computing system testing for exchanges (from M sends and receives of single messages, to $2M$ or M^2), with a corresponding decrease in efficiency of the whole calculation.

So, when developing a model of a REMD simulation, we can assign each replica to an ensemble, and note that after each exchange attempt, on average, each replica will be in the same ensemble with probability $1-p$ and at its exchange partner's ensemble with probability p. After further exchanges, the distribution of all of the replicas across the generalized ensemble is known probabilistically. In order to study the transit number as a function of M and p, it is not necessary to track the evolution of the probability distribution of each of the M replicas. It suffices to "mark" all of the probability density at the maximum after each exchange attempt and to track only the amount of marked and unmarked density present in each ensemble. Likewise, on reaching the minimum, any marking of probability density is removed, so that it is able to return to the maximum unmarked and transit again (however, it is unlikely that such density will affect the transit number because the simulation will likely be too short to permit this). It should be clear that the origin of any probability density is not of concern for measuring mixing efficiency. Neither is it of concern whether the pathway from the maximum to the minimum was direct or meandering, fast or slow; only whether it has been to the maximum before reaching the minimum matters. A running total of the amount of density from which marking was removed can be kept, and the transit number can then be calculated.

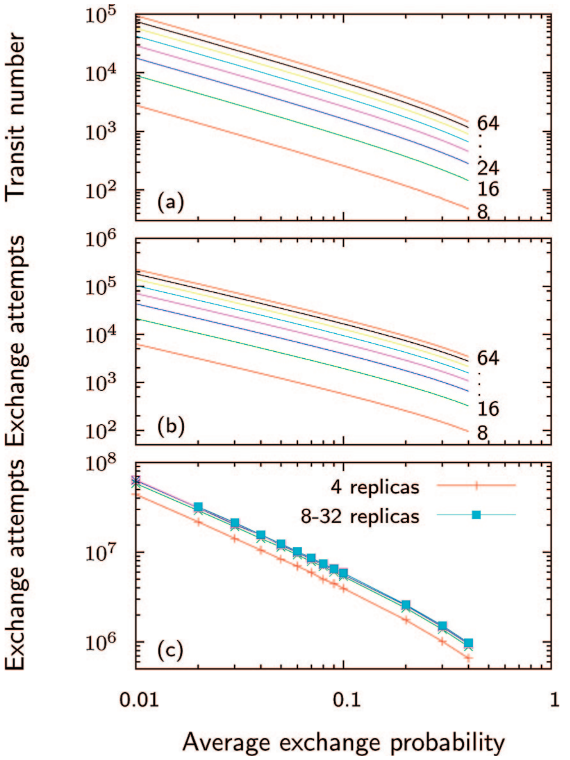


Figure 1. Illustrations of metrics (a) one, (b) two, and (c) three. These show either transit numbers or the number of exchange attempts required for convergence of the indicated simulated replica-scheme sizes and average exchange probabilities. For c, the mean number of exchange attempts required to achieve convergence of the observed replica distribution is shown, and for each datum, at least 1000 iterations were performed, so that in all cases the standard error of the sample mean was less than 1% of the sample mean. Thus, the error bars on these data are insignificant on the scale of the ordinate axis and are not shown. All axes are logarithmic.

A C++ program was written to implement the above algorithm. An initial density of 1 was assigned to each ensemble, and the transit number was observed after at least 0.95 density had been unmarked. The transit number was found for a range of average exchange probabilities and replica-scheme sizes. The results are plotted in Figure 1a.

The gradients of the lines on the log−log plot of Figure 1a clearly indicate the inverse proportionality between average exchange probability and transit number over a common range of probabilities and numbers of replicas. The spacing between the lines indicates that the transit number increases sublinearly with the number of replicas also (data not shown). Together, these trends illustrate that the minimum length of a simulation with an efficient replica-exchange scheme increases with the number of replicas and in roughly inverse proportion with the exchange probability.

It would be possible to calculate the number of exchange attempts necessary for a given probability of a number of transits occurring. Because multiple transits by different replicas can overlap in time, this will not be a simple function of the transit number. It is not clear how many transits indicate an efficient replica-exchange scheme.

We note also that the above is only a satisfactory model of an efficient REMD simulation when the latter achieves independence of replica-exchange events by simulating for a sufficiently long period between exchange attempts.

2.1.2. Metric Two. A second metric to test for efficient conduct of a REMD simulation can be derived from the expectation that, in the limit of a large number of exchange attempts, the probability that a replica in ensemble $i \in \{1,..., M\}$ began the simulation in any ensemble $j \in \{1,..., M\}$ approaches $1/M$. This convergence can again be studied in a simulation. We define the *replica probability matrix* $p(k)$ as the matrix whose $p(k)_{ij}$ elements are the probabilities that, after k exchange attempts, the replica in ensemble i began the simulation in ensemble j. $p(0)_{ij}$ is a diagonal matrix with unity on the leading diagonal, and in the limit of large k, all elements approach $1/M$.

A second C++ program was written to characterize this behavior. The criterion for assessing convergence after k exchange attempts over M replicas was the rmsd of $p(k)$ from its limiting value. This value is initially $\sqrt{(M-1)}/M$ and approaches zero as the replica system mixes. The convergence criterion was defined as

$$\sqrt{\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\frac{1}{M} - p(k)_{ij}\right)^2} < 10^{-3} \tag{6}$$

The number of exchange attempts necessary to achieve this convergence for a range of probabilities and replica-scheme sizes is shown in Figure 1b. Note again that the number of exchange attempts required to converge the replica probability matrix does not ensure that all replicas will have been exposed to all ensembles equally; it merely sets a lower bound below which reasonably equal exposure cannot have occurred.

The same kind of proportionality as for Metric One is evident, with the number of steps for convergence inversely proportional to average exchange probability.

The arbitrary nature of the convergence criteria chosen in constructing the plots in Figure 1a and b means that the actual number of exchange attempts on the ordinate axes is not significant. The value of these plots is in making clear how the effectiveness of the replica-exchange scheme depends on the number of replicas and the average exchange probability.

2.1.3. Metric Three. We carried out one further simulation of a replica-exchange process. This study followed the progress of the replicas and recorded the number of exchange attempts needed to converge the observed frequency of each replica in each ensemble to $1/M$. The criterion for assessing convergence after k exchange attempts over M replicas was the approach of the rmsd to zero, namely

$$\sqrt{\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\frac{1}{M} - f(k)_{ij}\right)^2} < 10^{-3} \tag{7}$$

where $f(k)_{ij}$ is the observed relative frequency with which the replica that began the simulation in ensemble i is in ensemble j after k exchange attempts. The number of exchange attempts required for convergence over a number of iterations was noted, and the mean and standard error of this sample were calculated for a range of exchange probabilities and sizes of replica schemes, as shown in Figure 1c. This plot again shows that the number of exchange attempts required increases inversely with the average exchange probability. However, in this model of replica exchange, there is very little dependence on the number of replicas; the reasons for this are not clear.

2.2. MD Protocol for Measuring Potential Energy Autocorrelation Time. The 357-atom, 23-residue N-terminal fragment of the N-terminal H4 histone tail studied by Lins and Rothlisberger[33] was solvated in a cubic box of normal TIP3P[34] water of side length 5.9 nm with nine chloride ions to neutralize the total charge. The all-atom CHARMM22 forcefield[35] in GROMACS 3.3.1[30−32] was used to model the peptide, with smooth particle-mesh Ewald (PME)[36] treatment of long-range electrostatics in an NpT ensemble. Berendsen coupling[37] to independent thermal baths was used for solute and nonsolute groups, both with a time constant of 0.5 ps, and the Berendsen algorithm was also used for pressure coupling with a time constant of 1 ps. SETTLE[38] and LINCS[39] were used to constrain all bond lengths in the solvent and solute, respectively. The system was equilibrated for 10 ps in NVT with position restraints and then for 50 ps in NpT, beginning from a canonical α helix. The length of the NpT production simulation was 20 ns. In the PME algorithm, the real-space summation was cut off at 1.0 nm, and the Ewald parameter was chosen to bound the error in the real-space summation such that $\mathrm{erfc}(\beta r)/r < 10^{-5}$ for r greater than the above cutoff. In the reciprocal-space approximation, the β-spline interpolation order was 6, and a Fourier grid spacing of 0.10 nm was used. van der Waals interactions were modeled with a 6,12 Lennard-Jones potential truncated at a maximum interaction length of 1.2 nm, with a long-range dispersion correction applied.

3. Analysis of Efficiency of REMD Simulations Reported in the Literature

We performed a thorough survey of published temperature REMD simulations on protein or peptide systems using explicit solvent models. Although all authors reported the number of replicas, temperature range, and simulation length, there was wide variation in the reporting of the expected and observed ranges of the frequency of successful exchange attempts. Several authors made no mention at all of these latter values, which makes it extremely difficult to assess whether the use of the particular REMD protocol was justified. Some authors neglected to record either the number of exchange attempts or the period between them. For those who did report these quantities, their values fell into ranges of 400−500 000 and 0.01−5 ps, respectively. A significant number of authors did not report the frequency with which they collected data from the simulation for their analyses; however, this is not necessarily pertinent to the validity of their conclusions.

In the following sections, we discuss several aspects of this survey. We propose and test a method for estimating the average exchange probability where the authors did not report it. We then use this method to increase the size of our data set for a comparison of published simulations, using Metric One to indicate the extent to which these simulations achieved mixing efficiency.

3.1. Estimating Average Exchange Probabilities. Although we wanted to study the number of exchange attempts used in reported REMD simulations as a function of the number of replicas and the average exchange probability, we were hampered by the shortage of published data with a complete description. Consequently, we then aimed to extend the size of this data set available in the literature, by testing a method of estimating the average exchange probability to use in cases where that value was not given. It is possible to estimate the frequency of exchanges in the following way,[40] by using properties of the molecular system and the reported temperature range. From the equipartition theorem, we can approximate

$$U_i = U_{i+1} N_{\mathrm{df}} \frac{c}{2} k_{\mathrm{B}} (T_i - T_{i+1}) \qquad (8)$$

where N_{df} is the number of degrees of freedom for the system, and c is 2 for protein−water systems.[40] If replica temperatures are distributed approximately exponentially (as is commonly the case in REMD simulations), so that successive temperatures have a constant ratio $1 + \varepsilon$, that is, $T_{i+1} = (1 + \varepsilon)T_i$, then from the definition of the replica-exchange probability in eq 5 we have

$$P(\text{exchange}) = \min\left(1, \exp\left[-\frac{\varepsilon^2 N_{df}}{1+\varepsilon}\right]\right) \qquad (9)$$

In order to test the above estimation process, a data set of 33 simulations from 21 papers[18,41−60] from 2001 was selected, representing cases where the authors reported either expected or observed exchange probabilities and enough descriptive data to generate an *a priori* estimate of the exchange probabilities. The value of ε, where not reported, was inferred from the reported number of replicas and temperature range, under the supposition that the temperatures were distributed exponentially. The number of degrees of freedom of a condensed-phase system is significantly lower than the familiar $3N - 6$ used in single-molecule analyses. Many degrees are correlated, and some are eliminated methodologically, such as by the use of bond constraint algorithms. Not all authors reported their use of bond constraints in either the solute or solvent parts of their systems. The assumption of constraints on all bonds implies the approximation of $N_{\mathrm{df}} \approx 2N_{\mathrm{atoms}}$, which leads to poor correlation between reported and estimated exchange probabilities for the same simulation. Approximating $N_{\mathrm{df}} \approx N_{\mathrm{atoms}}$ led to a better correlation, as shown in Figure 2. This demonstrates some qualitative predictive value for this crude approximation. Hence, we used it to generate estimates of the average exchange probability in cases where the authors had not specified the information.

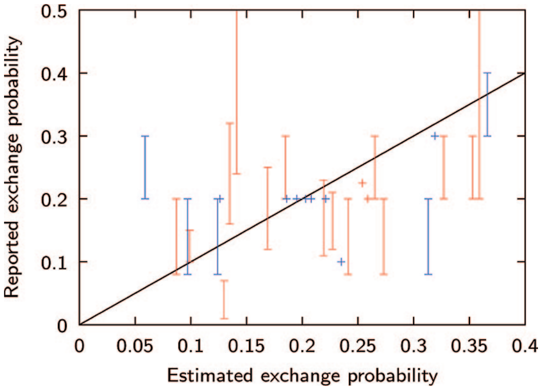


Figure 2. Comparison of estimated and reported average exchange probabilities. Reported replica-exchange probabilities for simulations described in refs 18, 41–60 are compared *a priori* with our estimates based on eq 9. Cases where the source reported only the replica-exchange probabilities expected by the authors in constructing the replica-exchange scheme are shown in blue, and those where the observed probabilities were reported are shown in red. Cases where the source reported a range of probabilities are indicated with a vertical bar, but those where the source only reported an average probability are plotted with a single point. The black line indicates perfect correlation. A version of this figure with the sources of the data points labeled is given in the Supporting Information.

3.2. Comparison of Mixing Efficiency of Simulations from the Literature. A second, intersecting data set of 41 simulations from 23 papers[22,41−54,56−59,61−64] was constructed, for which the authors had reported the exchange attempt frequency, the exchange period, and either the expected or observed exchange probability, or a sufficient atomic description to estimate that probability using eq 9. For these simulations, the reported number of exchange attempts and the exchange period were plotted in Figure 3 against the average expected, observed, or estimated exchange probability, and the transit number lines from Figure 1a are shown for reference.

It is desirable for a simulation to have an exchange attempt period that is long enough to ensure uncorrelated exchanges, and for that simulation to attempt sufficient exchanges for exploration of the replica space to take place, that is, to mix efficiently. By our construction in section 2.1.1, the transit number is a lower bound below which a simulation cannot be considered efficient from a computational point of view. It is clear from Figure 3b that all of the simulations used sufficient exchange attempts to exceed the transit number for their exchange probability and number of replicas. However, with one exception, all simulations were within 2 orders of magnitude above the transit number. We contend

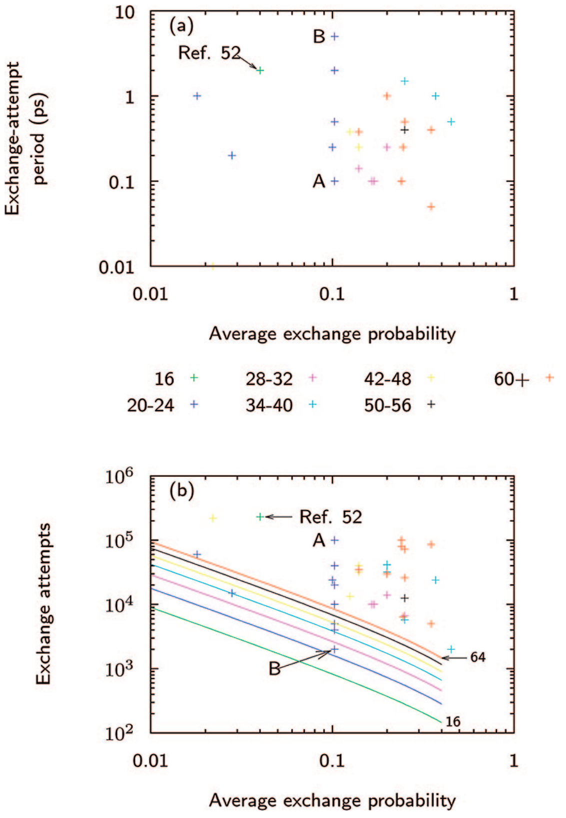


Figure 3. Comparison of replica scheme parameters from the literature. The (a) exchange period and (b) number of exchange attempts, reported in refs 22, 41–54, 56–59, 61–64, are plotted as a function of the average exchange probability (reported or estimated) and number of replicas (indicated by different symbols). Also superposed in b are some of the transit number lines from Figure 1a, indicatively labeled by replica number. The exchange attempt data are grouped according to the number of replicas in the simulation, in subranges of width 8. The corresponding transit number line has the same color for comparison. All axes are logarithmic. A version of b with the sources of the data points labeled is given in the Supporting Information.

that many transits are necessary to achieve mixing efficiency, and that this requires a simulation to exceed the transit number by a considerable degree. Accordingly, we recommend that a simulation exceed the transit number by at least 2 orders of magnitude.

Figure 3a illustrates that some simulations exceeded the transit number, but they did so with a questionably short exchange attempt period, as mentioned earlier and discussed further below. Of interest was the group of simulations with an average exchange probability of 0.1 using 24 replicas, which were all from the comparative study of Periole and Mark discussed earlier.[22] The data with high numbers of exchange attempts had correspondingly low exchange periods (labeled A in Figure 3), and vice-versa (labeled B in Figure 3). Thus, the appearance in these simulations of either a suitably long exchange attempt period in Figure 3a or high number of exchange attempts in Figure 3b is offset by a deficit in the other quantity, and none approaches the quality of the exceptional simulation.[52]

That exceptional simulation[52] mentioned above had an average exchange probability of 0.04, a number of exchange attempts of 230 000, and an exchange period of 2 ps. We believe that the other 40 simulations studied here used a number of exchange attempts that was too few to make effective use of their entire replica space, or they did so with a questionably low exchange-attempt period. Consequently, these simulations did not mix efficiently, even if they achieved thermodynamic efficiency by sampling the ensembles in a way suitable for the observables needed for their conclusions.

Some authors provided plots of the random walks in temperature space taken by one or several of their replicas as an indication that the replica-exchange method was working properly. Although helpful, we suggest these should be accompanied by a report of (say) the number of transits for the generalized ensemble. This could be sobering, particularly in cases where "random walk plots" reveal only one or two transits per replica.

The scaling behavior we have demonstrated for our metrics shows that the simulation length required for mixing efficiency increases with the number of replicas. This has implications for the decorrelation time of Lyman and Zuckerman[29] referred to earlier. The simulation length required for convergence of the individual replica trajectory decorrelation times will increase as the number of replicas increases. The value to which the decorrelation times converge can only decrease, however. Thus, we anticipate the existence of a negative correlation between our *a priori* transit number and their *a posteriori* decorrelation time. It would be interesting to test for this correlation by undertaking some REMD simulations on suitable model peptides. Various authors have observed the tendency of REMD simulations on peptides to predict an unphysical flattening of melting curves;[44,52] a method to address this issue partially exists.[65] Inadequate mixing of replicas could contribute to this effect. We propose that a suitable adaptation of this "decorrelation time" method evaluated at constant temperature, despite the discontinuities, might prove to be a useful indicator of per-replica-temperature thermodynamic convergence.

Finally, we need to reiterate that the data underlying Figure 1a−c were all generated on the assumption that each exchange attempt was statistically independent from all other attempts. This will only be true in a real REMD simulation if the period of simulation time between exchange attempts is greater than the autocorrelation time of the potential energy of the system. This is almost certainly not true for some of the shorter exchange periods for the peptide systems in our sample literature data set, illustrated in Figure 3a. If it were true, then sufficient configurational sampling would likely be occurring at the minimum, and in such cases, application of the replica-exchange method would not be useful. As no data are available to calculate such autocorrelation times, we cannot comment further on the data in Figure 3, except to observe that, for a system where the exchange probabilities are correlated between successive exchanges, the number of exchange attempts required to exceed the transit number will be higher than that for the uncorrelated case. Thus, the general trend, where the number of exchange attempts reported by these authors is possibly too few to justify the number of replicas (and, thus, the temperature space spanned),

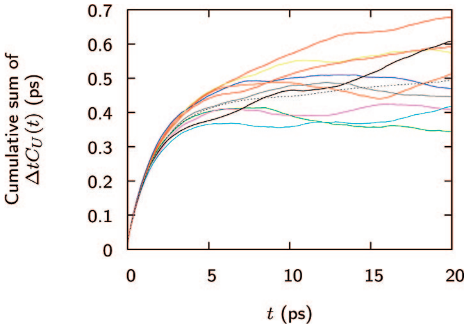


Figure 4. Behavior of potential energy autocorrelation functions. The product of the sampling interval Δt and the cumulative sum of the potential energy autocorrelation time $C_U(t)$ is plotted against time t for 0−20 ps. Plots for 1 ns autocorrelation functions based on 10 disjointed 2 ns fragments are shown in color, and the sum for the 10 ns autocorrelation function based on the whole 20 ns simulation is shown as a dotted line.

is likely to be even worse than it appears in Figure 3. We explore the behavior of the autocorrelation of potential energy further in the next section.

3.3. Estimation of Potential Energy Autocorrelation Time. To investigate further the interpretation of Periole and Mark[22] of the possible impact of correlation on the mixing efficiency of simulations, as discussed in the Introduction and above, we performed a MD simulation without replica exchange to estimate the autocorrelation time of the potential energy. The system studied was a 23-residue, 357-atom N-terminal fragment of the N-terminal H4 histone tail used in the comparative analysis of electrostatic treatments by Lins and Rothlisberger.[33] Over the 20 ns simulation, the initial helix unfolded progressively from the N-terminal end, retaining a helix turn at the C-terminal end. This is in qualitative agreement with the findings of Lins and Rothlisberger (see Figure 2C of ref 33) for this peptide under PME electrostatics.

The potential energy U of this system was recorded at $\Delta t = 0.01$ ps intervals, and the autocorrelation function $C_U(t)$ was found over the time range $\{0... P\Delta t\}$ from a simulation fragment of length $Q\Delta t$ (with $P \leq {}^{Q}/_{2}$) using

$$C_U(j\Delta t) = \frac{1}{P} \sum_{i=0}^{Q-1-P} U(i\Delta t)\, U[(i+j)\Delta t] \qquad (10)$$

Although other algorithms for computing an autocorrelation function exist,[40] this one has the virtue that the same number of simulation observations (P) contribute to each point of the function, and so the expectation of statistical error is consistent along its length. The 20 ns simulation was broken into 10 disjointed 2 ns fragments, and a 1 ns autocorrelation was derived for each (thus $Q = 2 \times 10^5$ and $P = 10^5$ for each). At large times, the exponential decay of $C_U(t)$ is swamped by the constant noise from the simulation. This is clearly seen in Figure 4, which plots the product of Δt and the cumulative sum of $C_U(t)$ against t.

Determining an integrated autocorrelation time can be difficult because the variance of the truncated-series estimator of $\tau_{\text{int},f}$ from eq 2 diverges as $t \rightarrow \infty$ because of the aforementioned noisy behavior of $C_f(t)$ at large t. Thus, the series needs to be truncated before this noise dominates. Averaging over the 10 2 ns autocorrelation functions produced above, and applying eq 2 with time truncated at 10 ps, we find $\tau_{\text{int},u} = 0.47$ ps. Our choice of 10 ps is justified by the curves in Figure 4, which suggests that noise is dominant at or about this time. Further, we note that the value of $C_U(t)$ first becomes negative (indicating dominance of noise) at 9.96 ps and 11.96 ps, respectively, for the 10 ns autocorrelation and the average over 10 2 ns autocorrelation functions. This noise could be reduced by extending the length of the autocorrelation; this is confirmed by the plot based on a 10 ns autocorrelation, which is dotted in Figure 4. However, the present results are sufficient for the purpose of indicating the size of the integrated autocorrelation time of the potential energy.

Thus, we conclude that successive measurements of U for such peptide-in-water systems are not independent unless at least around 0.5 ps has passed. This agrees with the findings of Zhang et al.[19] and Periole and Mark[22] that an exchange period under 1 ps is too short for the replica-exchange attempts to be regarded as independent. As may be seen in Figure 3a, there are several published studies performed on simulations that have strong similarities to the one studied here, for which the replica-exchange periods are below 1 ps. At the very least, these replica schemes did not achieve the mixing efficiency indicated by the number of exchanges they attempted.

4. How-To Guide for REMD with Suitable Mixing Efficiency

On the basis of our analyses and observations, we have developed a general strategy guide for designing REMD simulations so as to ensure a demonstrable level of mixing efficiency. This is described below.

1. Choose the simulation system and parameters to arrive at a full description, including the thermodynamic ensemble, solvent model, ionic concentration, and periodic box size.

2. Having judged the nature of the expected free energy surface, choose a range of temperatures over which to distribute replicas that is sufficient to permit replicas to cross the relevant activation barriers, and not greatly excessive.[23]

3. Given a number of available processors, *either* distribute a given number of replicas over the temperature range either exponentially or according to some algorithm to deal with generalized free energy surface bottlenecks,[15] and either estimate or test for the average exchange probability, *or*, given a desired average exchange probability, find a distribution of replicas that achieves this, either empirically or using an algorithm[16] as implemented at http://folding.bmc.uu.se/remd/index.php.

4. Calculate or estimate the potential energy autocorrelation time of the system at the temperature(s) of interest, and choose an exchange period at least this long. As a general rule for all-atom simulations of peptides in explicit solvents, we recommend no smaller than 1 ps.

5. Given the average exchange probability and the number of replicas, look up the transit number in Figure 1a, and choose a number of exchange attempts considerably larger

than this. Together with the exchange period, this determines the minimum simulation length for mixing efficiency.

If the above scheme leads to a simulation cost that is too great for the available resources, then reducing the size of the replica range is the simplest way to reduce both the number of replicas and the length of the simulation. The impact of such a reduction on the thermodynamic efficiency of the method will normally be difficult to judge *a priori*.

5. Conclusion

In summary, from the literature survey we conducted for this work, it is clear that authors are not routinely reporting information necessary for readers to assess the quality of the data underlying the analysis. The expected or observed replica-exchange probability, together with the total number of exchange attempts, the exchange period, and the estimates of the autocorrelation time of the potential energy should be given. We have provided a step-by-step guide to help in constructing protocols for REMD simulations that can mix efficiently. Two key recommendations are that the exchange period should not be smaller than about 1 ps and the number of exchange attempts should be chosen to significantly exceed the transit number for the replica-exchange scheme.

Replica-exchange methods offer great promise for improving the ability of computational chemists to sample thermodynamic ensembles of protein and peptide systems with efficiency. After a slow start in the literature, their recent implementation in common MD simulation programs has made them easily accessible, and they have become a popular method to improve sampling efficiency. However, the lesson of our work is that it is necessary to use these methods in a thoughtful and measured fashion in order to ensure that the desired increase in sampling is actually realized with mixing efficiency, and to properly report the simulation details.

Acknowledgment. We acknowledge support from the Australian National University Institute for Advanced Studies (IAS) block grant, and the APAC (Australian Partnership for Advanced Computing) for computer-time grants. We thank the reviewers for their constructive comments.

Supporting Information Available: Versions of Figures 2 and 3 labeled by reference number. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(2) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.

(3) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. *J. Chem. Theory Comput.* **2006**, *2*, 217–228.

(4) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.

(5) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. *Phys. Rev. Lett.* **2006**, *96*, 028105.

(6) Christen, M.; van Gunsteren, W. F. *J. Chem. Phys.* **2006**, *124*, 154106.

(7) Liu, P.; Voth, G. A. *J. Chem. Phys.* **2007**, *126*, 045106.

(8) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058–9067.

(9) Jang, S. M.; Shin, S.; Pak, Y. *Phys. Rev. Lett.* **2003**, *91*, 8305.

(10) Cheng, X. L.; Cui, G. L.; Hornak, V.; Sinnnerling, C. *J. Phys. Chem. B.* **2005**, *109*, 8220–8230.

(11) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B.* **2003**, *107*, 13703–13710.

(12) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.

(13) Piana, S.; Laio, A. *J. Phys. Chem. B.* **2007**, *111*, 4553–4559.

(14) Li, H. Z.; Li, G. H.; Berg, B. A.; Yang, W. *J. Chem. Phys.* **2006**, *125*, 144902.

(15) Trebst, S.; Troyer, M.; Hansmann, U. H. *J. Chem. Phys.* **2006**, *124*, 174903.

(16) Patriksson, A.; van der Spoel, D. *Phys. Chem. Chem. Phys* **2008**, *10*, 2073–2077.

(17) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 420–433.

(18) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins* **2002**, *46*, 225–234.

(19) Zhang, W.; Wu, C.; Duan, Y. *J. Chem. Phys.* **2005**, *123*, 154105.

(20) Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1200–1202.

(21) Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1693.

(22) Periole, X.; Mark, A. E. *J. Chem. Phys.* **2007**, *126*, 014903.

(23) Nymeyer, H. *J. Chem. Theory Comput.* **2008**, *4*, 626–636.

(24) Wolff, U. *Comput. Phys. Commun.* **2004**, *156*, 143–153.

(25) Madras, N.; Sokal, A. D. *J. Stat. Phys.* **1988**, *50*, 109–186.

(26) Sindhikara, D.; Meng, Y.; Roitberg, A. E. *J. Chem. Phys.* **2008**, *128*, 024103.

(27) Kofke, D. A. *J. Chem. Phys.* **2002**, *117*, 6911–6914.

(28) Rathore, N.; Chopra, M.; de Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 024111.

(29) Lyman, E.; Zuckerman, D. M. *J. Phys. Chem. B* **2007**, *111*, 12876–12882.

(30) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(31) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.

(32) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(33) Lins, R. D.; Rothlisberger, U. *J. Chem. Theory Comput.* **2006**, *2*, 246–250.

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

(35) 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.; 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.

(36) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(37) Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(38) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.

(39) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(40) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. Gromacs User Manual, version 3.3; University of Groningen: Groningen, The Netherlands, 2005.

(41) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345–354.

(42) Zhou, R. H.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931–14936.

(43) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782–2787.

(44) Garcia, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13898–13903.

(45) Ghosh, T.; Garde, S.; Garcia, A. E. *Biophys. J.* **2003**, *85*, 3187–3193.

(46) Gnanakaran, S.; Garcia, A. E. *Biophys. J.* **2003**, *84*, 1548–1562.

(47) Zhou, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.

(48) Garcia, A. E. *Polymer* **2004**, *45*, 669–676.

(49) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C. K.; Li, M. S. *Proteins* **2005**, *61*, 795–808.

(50) Nguyen, P. H.; Mu, Y. G.; Stock, G. *Proteins* **2005**, *60*, 485–494.

(51) Nishino, M.; Sugita, Y.; Yoda, T.; Okamoto, Y. *FEBS Lett.* **2005**, *579*, 5425–5429.

(52) Seibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. *J. Mol. Biol.* **2005**, *354*, 173–183.

(53) Yoshida, K.; Yamaguchi, T.; Okamoto, Y. *Chem. Phys. Lett.* **2005**, *412*, 280–284.

(54) Baumketner, A.; Bernstein, S. L.; Wyttenbach, T.; Lazo, N. D.; Teplow, D. B.; Bowers, M. T.; Shea, J. E. *Protein Sci.* **2006**, *15*, 1239–1247.

(55) Gnanakaran, S.; Nussinov, R.; Garcia, A. E. *J. Am. Chem. Soc.* **2006**, *128*, 2158–2159.

(56) Juraszek, J.; Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15859–15864.

(57) Mu, Y. G.; Nordenskiold, L.; Tam, J. P. *Biophys. J.* **2006**, *90*, 3983–3992.

(58) Sugita, Y.; Miyashita, N.; Yoda, T.; Ikeguchi, M.; Toyoshima, C. *Biochemistry* **2006**, *45*, 11752–11761.

(59) Wickstrom, L.; Okur, A.; Song, K.; Hornak, V.; Raleigh, D. P.; Simmerling, C. L. *J. Mol. Biol.* **2006**, *360*, 1094–1107.

(60) Zhang, J.; Qin, M.; Wang, W. *Proteins* **2006**, *62*, 672–685.

(61) Gustiananda, M.; Liggins, J. R. R.; Cummins, P. L. L.; Gready, J. E. E. *Biophys. J.* **2004**, *86*, 2467–2483.

(62) Kawashima, Y.; Sugita, Y.; Yoda, T.; Okamoto, Y. *Chem. Phys. Lett.* **2005**, *414*, 449–455.

(63) Nymeyer, H.; Woolf, T. B.; Garcia, A. E. *Proteins* **2005**, *59*, 783–790.

(64) Baumketner, A.; Bernstein, S. L.; Wyttenbach, T.; Bitan, G.; Teplow, D. B.; Bowers, M. T.; Shea, J. E. *Protein Sci.* **2006**, *15*, 420–428.

(65) van der Spoel, D.; Seibert, M. M. *Phys. Rev. Lett.* **2006**, *96*, 0238102.

CT800016R

J. Chem. Theory Comput. **2008,** *4,* 1129–1137 **1129**

JCTC Journal of Chemical Theory and Computation

Quantum Chemical Modeling of the Dehalogenation Reaction of Haloalcohol Dehalogenase

Kathrin H. Hopmann and Fahmi Himo*

Department of Theoretical Chemistry, School of Biotechnology, Royal Institute of Technology, AlbaNova University Center, SE-106 91 Stockholm, Sweden

Received February 12, 2008

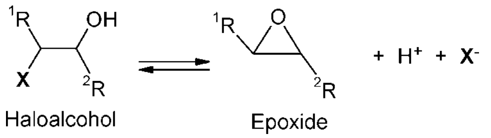
Abstract: The dehalogenation reaction of haloalcohol dehalogenase HheC from *Agrobacterium radiobacter* AD1 was investigated theoretically using hybrid density functional theory methods. HheC catalyzes the enantioselective conversion of halohydrins into their corresponding epoxides. The reaction is proposed to be mediated by a catalytic Ser132−Tyr145−Arg149 triad, and a distinct halide binding site is suggested to facilitate halide displacement by stabilizing the free ion. We investigated the HheC-mediated dehalogenation of (*R*)-2-chloro-1-phenylethanol using three quantum chemical models of various sizes. The calculated barriers and reaction energies give support to the suggested reaction mechanism. The dehalogenation occurs in a single concerted step, in which Tyr145 abstracts a proton from the halohydrin substrate and the substrate oxyanion displaces the chloride ion, forming the epoxide. Characterization of the involved stationary points is provided. Furthermore, by using three different models of the halide binding site, we are able to assess the adopted modeling methodology.

I. Introduction

Haloalcohol dehalogenases (also known as halohydrin dehalogenases, halohydrin epoxidases, or hydrogen-halide lyases) are microbial enzymes that catalyze the reversible dehalogenation of vicinal haloalcohols, thereby forming the corresponding epoxides (Scheme 1). Haloalcohol dehalogenase activity has been identified in multiple microbial sources, including *Corynebacterium* sp.,[1] *Arthrobacter* sp.,[2,3] *Agrobacterium* radiobacter,[3] and *Mycobacterium* sp.[3] During recent years, haloalcohol dehalogenases have acquired increasing interest. It has been realized that they are of importance for the degradation of various halogenated environmental pollutants.[4] In addition, they have received recognition as promising biocatalysts that can be used for the kinetic resolution of epoxides and halohydrins, and for the preparation of various substituted alcohols.[5] Haloalcohol dehalogenases have broad substrate specificities, and in addition to the dehalogenation reaction, they are also able to catalyze the irreversible ring-opening of epoxides with the nonhalide nucleophiles NO_2^-, CN^-, and N_3^-.[5–11]

Sequence alignments suggested that the haloalcohol dehalogenases are structurally and mechanistically similar to the family of short-chain dehydrogenases/reductases (SDRs).[3] SDRs exhibit a conserved catalytic Ser−Tyr−Lys/(Arg) triad that also can be identified in haloalcohol dehalogenases.[3] However, the reactions catalyzed by the two families are very different. SDRs catalyze NAD(P)H-dependent oxidation and reduction reactions, while the haloalcohol dehalogenases catalyze a cofactor-independent dehalogenation reaction.

Scheme 1. Haloalcohol Dehalogenase Reaction



Determination of the crystal structure of the haloalcohol dehalogenase HheC from *Agrobacterium radiobacter* AD1 confirmed structural similarities to the SDRs.[12] However, instead of the NAD(P)H binding site found in SDRs, the structure of HheC revealed a distinct halide binding site.[12] The putative catalytic triad, Ser132−Ty145−Arg149, is positioned above the halide binding site (Figure 1). This supports the proposed mechanism shown in Scheme 2.[3,12] Tyr145 is proposed to be the catalytic base that abstracts a proton from the halohydrin substrate. The substrate oxyanion

---

* Corresponding author e-mail: himo@theochem.kth.se.

10.1021/ct8000443 CCC: \$40.75 © 2008 American Chemical Society
Published on Web 06/13/2008

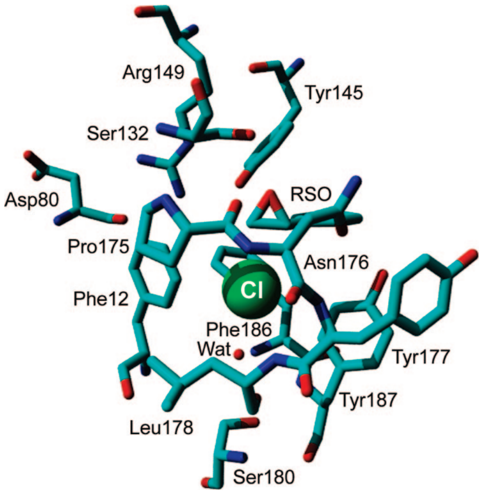


Figure 1. Active site of HheC (from the X-ray crystal structure of HheC in complex with (*R*)-styrene oxide and chloride, PDB: 1PWZ).[12] The chloride ion is shown in ball representation.

then displaces the chloride ion through an intramolecular S_N2 reaction. Arg149 is suggested to activate Tyr145 by abstracting a proton. Ser132 seems to be of importance for binding of the substrate as well as for stabilization of the emerging oxyanion.[3,12] Also, the conserved residue Asp80 has been shown to be of importance for the activity of HheC,[12] as its backbone carbonyl interacts with Arg149 and might be important for its positioning. Additionally, the Asp80 carboxyl side chain is proposed to be involved in a proton transfer pathway, which removes excess protons from the active site.[12] The displaced chloride ion is stabilized through interactions in the halide binding site. The crystal structures of HheC in complex with either bromide or chloride reveal several distinct interactions that seem to stabilize the free halide.[12] These include hydrogen-bonding interactions with the backbone amides of Tyr177 and Leu178, as well as with an ordered water molecule. Other residues suggested to be important for the formation of the halide binding site are Phe12, Pro175, Asn176, Phe186, and Tyr187 (Figure 1).[12]

It is today a quite common procedure to use relatively small models of enzyme active sites and apply quantum chemical methods to study their reaction mechanisms.[13] In this approach, the rest of the enzyme is usually treated using a homogeneous polarizable medium with some assumed dielectric constant. Density functional theory (DFT), and in particular the B3LYP functional,[14] has become the method of choice in this kind of study. It has been shown in many examples that the calculated energies often are sufficient to verify or rule out a suggested reaction mechanism, and in general, the experience has been that a well-chosen quantum model is able reproduce the chemistry taking place at an enzyme active site to such a high degree that it can provide detailed insight into mechanisms and selectivities.[15]

In the present study, we use this approach to examine the reversible dehalogenation reaction of the haloalcohol dehalogenase HheC. The transformation of (*R*)-2-chloro-1-phenylethanol to yield the epoxide (*R*)-styrene oxide (RSO) is studied with HheC active site models of different sizes. In similar studies, we have previously investigated the reaction mechanisms of two epoxide-hydrolyzing enzymes, namely, limonene epoxide hydrolase[16] and soluble epoxide hydrolase.[17] These investigations have established the adequacy of this approach to study mechanisms of this kind. However, the current HheC case presents an additional challenge in that a free halide ion is released at the active site. It is not *a priori* evident that a relatively small quantum chemical active site model, in conjunction with implicit solvation models, will be sufficient to describe the energetics of this system correctly. Previous calculations on other enzymatic dehalogenation reactions have generally employed molecular dynamics, quantum mechanics/molecular mechanics (QM/MM), or semiempirical methods.[18] We chose instead to study the reaction of HheC with the more accurate B3LYP method. This, however, puts some limitations on the size of the models that can be employed. Here, we have constructed and compared three different models of the HheC active site. The aim of the study is to shed more light on the dehalogenation reaction mechanism of HheC and also to assess the applicability of the adopted methodology.

II. Computational Details

All calculations were performed using the hybrid density functional theory method B3LYP,[14] as implemented in Gaussian 03.[19] Geometries were optimized with the 6-31G(d,p) basis set. On the basis of these geometries, more accurate energies were calculated by performing single-point calculations using the 6-311+G(2d,2p) basis set. Basis set superposition errors were not explicitly considered in this study.

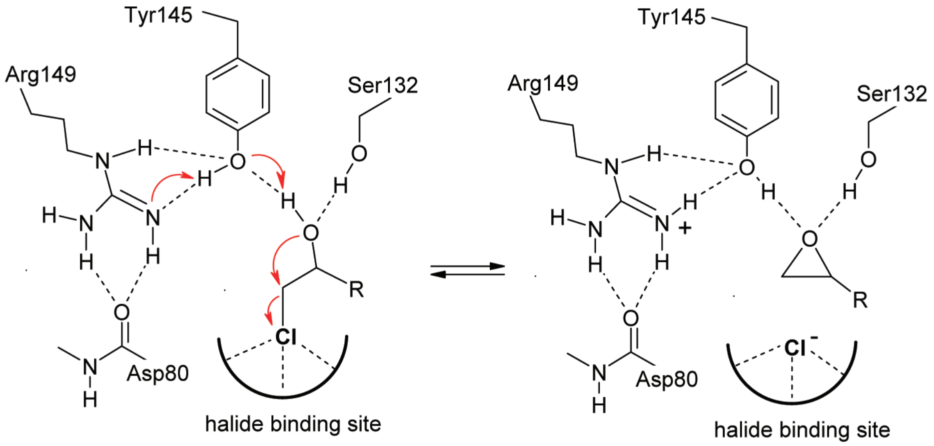
For Model A (see below), zero-point vibrational (ZPV) energies were calculated by performing frequency calculations on the optimized structures at the same level as optimizations. The size of Models B and C (see below) did not permit the calculation of frequencies. Instead, the ZPV effects were transferred from Model A. The solvation corrections were calculated at the same level of theory as optimizations by performing single-point calculations using the conductor-like polarizable continuum model (CPCM).[20] The energies presented here are the large basis set energies corrected for ZPV effects.

To preserve the overall spatial conformation of the models, the truncation points of residues included in the active site models were kept fixed to their X-ray positions during the geometry optimizations. This procedure gives rise to a number of small imaginary frequencies, typically around eight of them, all below $70i\ \mathrm{cm}^{-1}$. These frequencies do not contribute significantly to the zero-point energies and can thus be tolerated. However, they render the calculated harmonic entropy contributions unreliable. Thus, the reported energies here correspond to the enthalpies only.

III. Results and Discussion

III.a. Model A. We first devised a model of the HheC active site consisting of 83 atoms (called Model A). It was built on the basis of the crystal structure of HheC in complex with (*R*)-styrene oxide and chloride (PDB: 1PWZ; 2.50Å resolution).[12] The model contains parts of the side chains of the suggested Arg149−Tyr145−Ser132 catalytic triad and the backbone carbonyl of Asp80, which interacts with

Scheme 2. Proposed Dehalogenation Mechanism of HheC[3,12]



Arg149. A very small model of the halide binding site was used, consisting of the backbone of Leu178, parts of Phe186, and the crystallographically observed water molecule (Figure 2). As a substrate, we used (*R*)-2-chloro-1-phenyl-ethanol (RCPE). The dehalogenation of RCPE results in the formation of RSO and free chloride, corresponding to the complex observed in the 1PWZ X-ray crystal structure of HheC.[12]

As suggested from the proposed mechanism (Scheme 2), Arg149 and Tyr145 were modeled in their neutral forms. However, calculations showed that a more stable structure would be obtained if the two residues instead were modeled as protonated Arg149 and deprotonated Tyr145 ($Arg^{+}-Tyr^{-}$). For Model A, the charge-separated reactant was calculated to be 3.0 kcal/mol (using $\epsilon = 4$) lower in energy than the reactant with a neutral Arg−Tyr pair and was thus taken as the starting point for our investigation of the dehalogenation reaction. It can be noted in this context that the optimum pH of HheC for the dehalogenation reaction is between pH 8 and 9.[12] This is in line with a scenario in which a protonated Arg149 residue stabilizes an anionic Tyr145. In contrast, the reverse epoxide-opening reaction of HheC has a pH optimum of around 4−5,[12] indicating that both residues would be protonated in this case, allowing Tyr145 to function as a catalytic acid instead.

In the optimized reactant structure of Model A with RCPE (Figure 2B), the hydroxyl group of the substrate interacts with Tyr145 and Ser132. The chloro substituent is hydrogen-bonded to the amide backbone of Leu178 and to the water molecule.

Starting from this structure, we have located the transition state for the dechlorination reaction (shown in Figure 2C). The reaction turns out to occur in one single concerted step in which the alcohol proton is transferred to the phenolate of Tyr145, the alcohol oxygen performs a nucleophilic attack at the vicinal carbon, and the chlorine is displaced. Important optimized bond distances are displayed in Figure 2C. The O−C distance is 1.80 Å, and the C−Cl distance is 2.51Å. The alcohol proton is 1.04 Å from Tyr145 and 1.42 Å from the substrate oxygen. The hydrogen bond between Ser132 and the substrate oxygen is only slightly shorter than in the reactant structure. The hydrogen bonds from water and the Leu178 amide to the chloride are shortened significantly, from 3.44 and 4.25 Å, respectively, in the reactant to 3.20 and 3.63 Å at the transition state (hydrogen-bond length refers here to the distance from donor to acceptor atom).

In the product state with (*R*)-styrene oxide, the proton from the substrate has been fully transferred to Tyr145 (Figure 2D). The chloride ion remains hydrogen-bonded to water and the amide nitrogen. The lengths of the two hydrogen bonds have shortened further to 3.13 and 3.25 Å, respectively.

Without the inclusion of solvation effects, the calculated barrier for this reaction is 23.0 kcal/mol and the reaction energy is +17.5 kcal/mol (Table 1). From Figure 2C and D, it is seen that, despite the two hydrogen bonds, the chlorine ion appears quite exposed in the transition state and product structure of Model A. This suggests that the inclusion of solvation will have pronounced effects on the calculated energetics.

To account for the parts that are not explicitly included in the model, we have, as a first approximation, assumed that the surrounding is a homogeneous polarizable medium. This approximation has been used frequently in combination with quantum chemical active site models.[13,15] The solvation energies were computed using the CPCM method[20] on the optimized geometries. The choice of the dielectric constant is somewhat arbitrary, but the value $\epsilon = 4$ is frequently used in similar studies. In the present study, we chose to compare five different dielectric constants ($\epsilon = 2$, 4, 8, 16, and 80) to investigate the influence of the choice on the energetics. The calculated energies are presented in Table 1. As expected, solvation effects lower the energies of both the transition state and the product structures compared to the reactant. The barrier decreases from 23.0 kcal/mol without solvation to 15.0 and 12.2 kcal/mol using $\epsilon = 4$ and $\epsilon = 80$, respectively. The effect on the reaction energy is even more pronounced, resulting in a decrease from +17.5 kcal/mol without solvation to +3.4 and −1.2 kcal/mol using $\epsilon = 4$ and $\epsilon = 80$, respectively.

The experimentally determined rate constant for the HheC-catalyzed dehalogenation of RCPE is 48.5 s^{-1}.[3] Using classical transition state theory, this rate can be converted to a barrier of ca. 15 kcal/mol. It has been shown that halide release is rate-limiting in HheC.[21] It can thus be concluded that the barrier for the chemical step should be equal to or less than 15 kcal/mol. In addition, it has been shown experimentally that the equilibrium of HheC

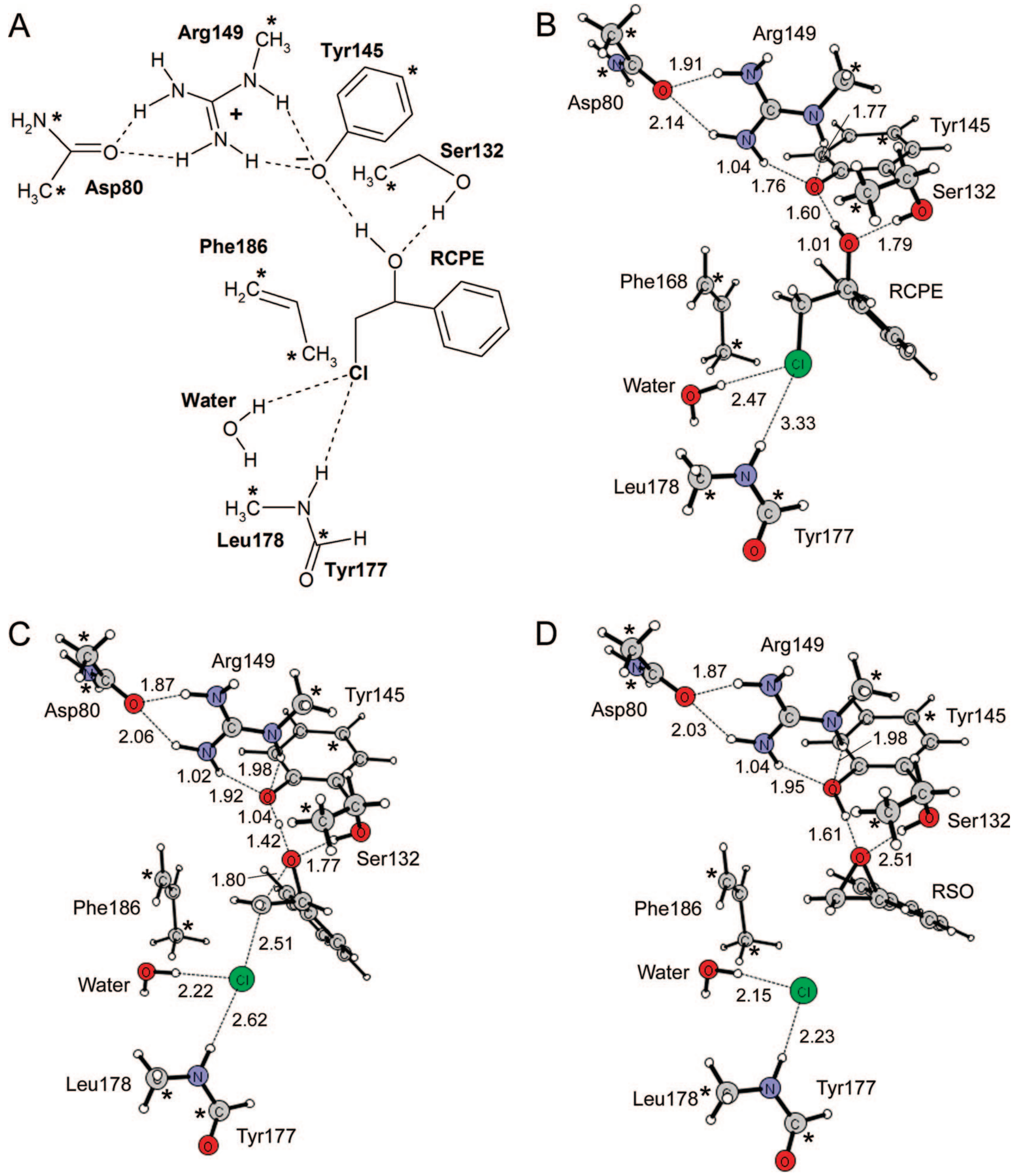


Figure 2. Optimized geometries for the formation of (*R*)-styrene oxide (RSO) from (*R*)-2-chloro-1-phenyl-ethanol (RCPE) in HheC active site Model A. (A) Schematic representation of the model. (B) Reactant geometry. (C) Transition state. (D) Product geometry. Distances are in ångstroms. Asterisks indicate atoms that are kept fixed to their crystallographically observed positions in calculations.

Table 1. Summary of the Calculated Energetics (kcal/mol) for the HheC-Catalyzed Transformation of RCPE to RSO

	Model A (83 atoms)		Model B (112 atoms)		Model C (161 atoms)	
	barrier	reaction energy	barrier	reaction energy	barrier	reaction energy
no solvation	23.0	+17.5	17.9	+14.1	18.2	+5.5
$\epsilon = 2$	17.8	+8.2	15.8	+8.3	17.7	+4.9
$\epsilon = 4$	15.0	+3.4	14.6	+5.2	17.4	+4.5
$\epsilon = 8$	13.5	+1.0	14.0	+3.6	17.1	+4.3
$\epsilon = 16$	12.8	−0.3	13.7	+2.8	17.0	+4.2
$\epsilon = 80$	12.2	−1.2	13.5	+2.2	17.0	+4.2

with styrene oxides and chloride is toward epoxide formation.[7] When (*R*)-*p*-nitrostyrene oxide and NaCl were used, the equilibrium constant K_{eq} = [NaCl] × [epoxide]/[chloroalcohol] was measured to be 40 mM.[7] From this, it can be concluded that the reaction should be exothermic. A similar conclusion can be reached considering the fact that HheC was successfully cocrystallized with RSO and chloride,[12] indicating that the enzyme-bound product state is energetically more favorable than the reactant state with the haloalcohol bound.

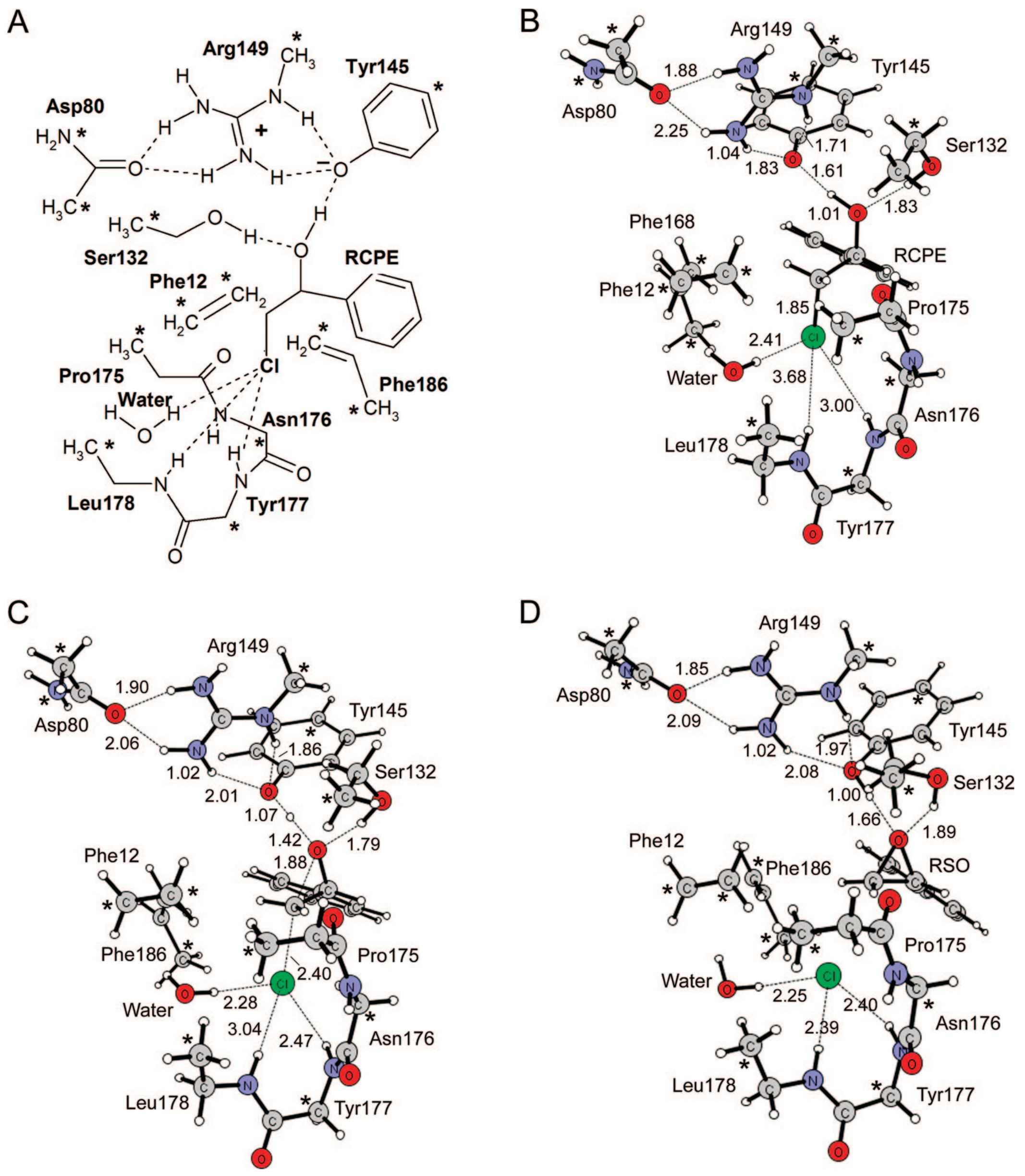


Figure 3. Optimized geometries for Model B. (A) Schematic representation of the model. (B) Reactant geometry. (C) Transition state. (D) Product geometry.

Our analysis shows that Model A fails to reproduce the experimental values without the inclusion of solvation effects, due to its insufficient description of the halide binding site. The gas-phase barrier is about 8 kcal/mol above the experimental value, while the reaction energy indicates that formation of the epoxide is highly endothermic, in contrast with experimental findings. However, with the inclusion of dielectric effects, especially employing large dielectric constants ($\epsilon > 8$), the reaction was found to proceed with a plausible barrier and reaction energy, in agreement with experimental results. Thus, despite its relative smallness, Model A provides support to the concerted mechanism shown in Scheme 2.

The poor description of the halide binding site makes the solvation effects in Model A very large, amounting to more than 14 kcal/mol for the reaction energy using $\epsilon = 4$. To test the validity and reliability of the above results, we therefore devised two larger models (referred to as Model B and Model C, respectively), in which the halide binding site was successively increased.

III.b. Model B. Model B consists of 112 atoms and contains, in addition to Model A, parts of Pro175, Asn176, Tyr177, Leu178, and Phe12, which gives a better description of the halide binding site (see Figure 3).

In the reactant structure, the chloride of RCPE forms hydrogen bonds as before to the water molecule and the backbone amide of Leu178. In addition, a hydrogen bond between Cl and the backbone of Tyr177 is observed (Figure 3B). The alcohol oxygen forms hydrogen bonds to Ser132 and Tyr145, as in Model A. At the transition state for epoxide formation (Figure 3C), the critical O−C and C−Cl distances are 1.88 and 2.40 Å, respectively. The alcohol proton is 1.07 Å from Tyr145 and 1.42 Å from the substrate oxygen. The three hydrogen bonds to the chloride have decreased significantly in length compared to the reactant, since the interactions with the partially displaced chloride ion are

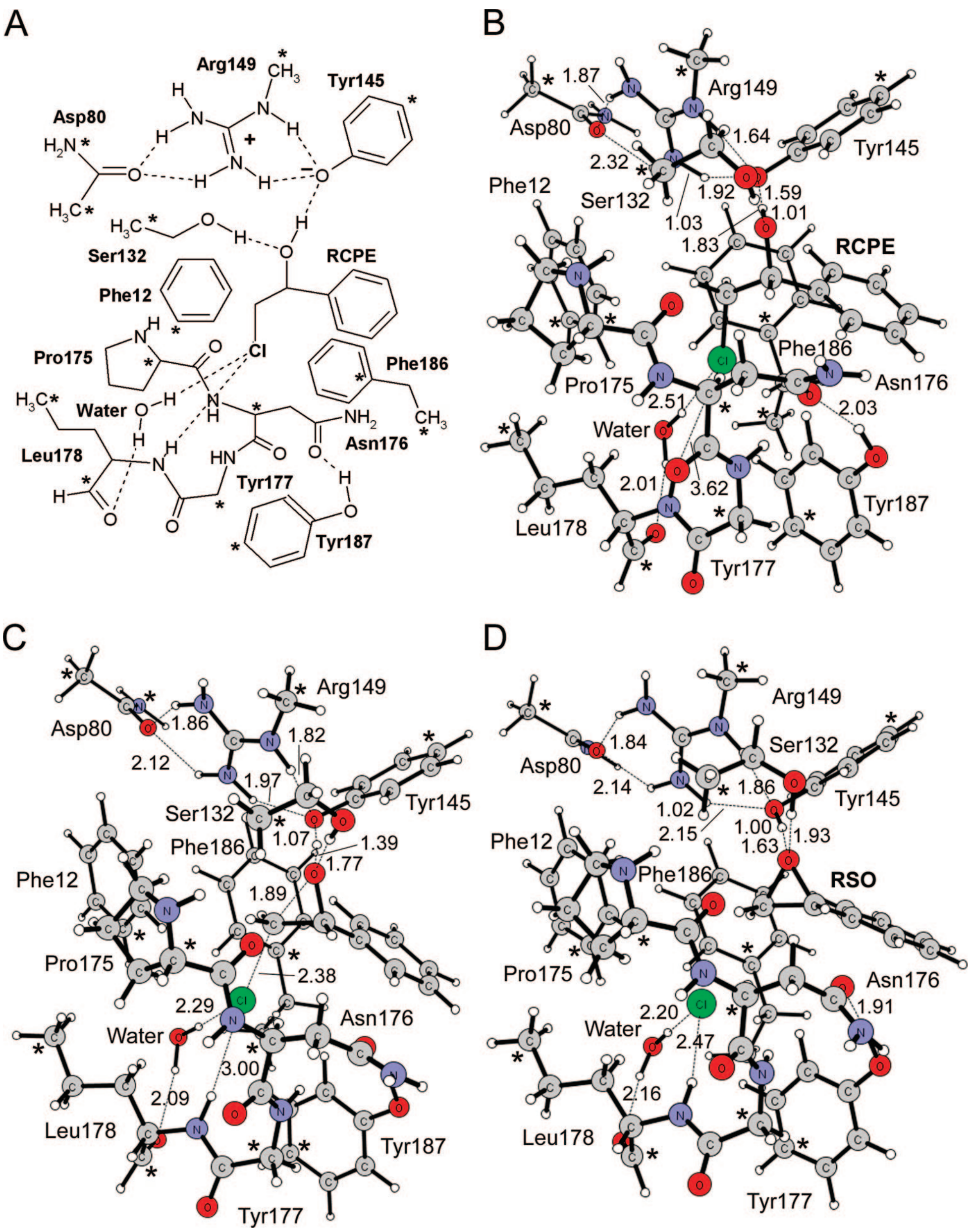


Figure 4. Optimized geometries for the largest model, Model C. (A) Schematic representation of the model. (B) Reactant geometry. (C) Transition state. (D) Product geometry.

stronger than with the neutral side chain. These interactions are further enhanced in the product state (Figure 3D), where the chloride ion has been fully liberated.

The calculated barrier for Model B is 17.9 kcal/mol, and the reaction energy is +14.1 kcal/mol (see Table 1). Both of these values are lower than the ones calculated for Model A (23.0 and 17.5 kcal/mol, respectively). Upon inclusion of solvation effects, both the barrier and the reaction energy decrease further. For example, the barrier decreases to 14.6 and 13.5 kcal/mol using $\epsilon = 4$ and $\epsilon = 80$, respectively, and the reaction energy decreases to +5.2 and +2.2 kcal/mol using $\epsilon = 4$ and $\epsilon = 80$, respectively. Although still very significant, the effects are smaller than for Model A due to the increased size of the halide binding site.

Overall, after consideration of the solvation effects, Model B conveys a very similar picture of the reaction mechanism of HheC to that of Model A.

III.c. Model C. The solvation effect on the reaction energy in Model B is still rather large (8.9 kcal/mol using $\epsilon = 4$, see Table 1), indicating that also this model lacks important parts of the halide binding site. Therefore, we devised an even larger model (referred to as Model C) which consists of 161 atoms and contains in principle the full side chains of the main residues forming the halide binding site, i.e. Pro175, Asn176, Tyr187, Phe12, and Phe186. Also, most of the Leu178 side chain is included (see Figure 4).

In the reactant geometry, the chloride forms hydrogen bonds with the water molecule and with the backbone amide of Leu178 (Figure 4B), but no interaction with the backbone amide of Tyr177 is observed, in contrast to Model B. The optimized transition state for chloride displacement exhibits very similar distances to those of Model B (Figure 4C). The carbon−chloride bond length is 2.38 Å, and the forming carbon−oxygen bond length is 1.89 Å. In the product state, the liberated chloride ion forms hydrogen bonds to the water molecule and to the backbone of Leu178, but not to the Tyr177 amide (Figure 4D). A possible reason is that the Tyr187 side chain introduced in Model C sterically hinders the chloride from approaching the Tyr177 backbone amide.

The barrier computed for Model C is 18.2 kcal/mol and the reaction energy is +5.5 (Table 1). Since the halide binding site is essentially saturated, the computed solvation effects are very small. For example, in going from no solvation to solvation with $\epsilon = 80$, the barrier drops only by 1.2 kcal/mol and the reaction energy by 1.3 kcal/mol.

Similar to the previous models, the results of Model C demonstrate the energetic feasibility of the proposed reaction mechanism. However, the barrier is slightly overestimated and the reaction energy is positive, in disagreement with experimental results. Several sources for this disagreement can be envisaged. One reason could be that the subsequent regeneration of the anion at the Tyr145 residue through the release of a proton to the bulk is an exothermic process that thus reduces the overall reaction energy. Also, release of the chloride ion could be energetically favorable, which would lower the reaction energy. Moreover, the calculated energies correspond to enthalpies. It is possible that the addition of entropic effects would lower the product state somewhat compared to the reactant state. However, another reason could be due to the chosen model. In Model C, the halide binding site is quite large, and in order to keep it intact in the model, it is necessary to lock a number of centers (indicated by asterisks in Figure 4) such that no artificial hydrogen-bond interactions or large movements take place. This procedure is, many times, very successful, but in the case of Model C, it might have introduced too much strain in the system and, therefore, resulted in a slightly endothermic reaction instead of a slightly exothermic one. Releasing constraints completely, however, is not an option in this model, because it will cause a major disruption to the halide binding site. The various groups will move freely and form artificial hydrogen bonds that are not present in the enzyme. To release the constraints will thus add more uncertainty to the results rather than make them more accurate.

IV. Conclusions

In the present paper, three quantum chemical active site models of different sizes have been employed to investigate the dehalogenation reaction of the haloalcohol dehalogenase HheC enzyme. The computed energies for the three models are summarized in Table 1. The conclusions can be divided into two parts, one related to the specific reaction under study, that is, the HheC reaction mechanism, and one related to the methodology used.

The three models give a consistent picture of the reaction mechanism taking place. The dehalogenation is shown to occur in a single concerted step in which Tyr145 abstracts a proton from the halohydrin substrate and the substrate oxyanion displaces the chloride ion through an intramolecular S_N2 reaction. Arg149 is suggested to activate Tyr145 by abstracting a proton, and Ser132 seems to be of importance for both binding of the substrate and stabilization of the oxyanion at the transition state. The activation barriers and reaction energies calculated for all three models (especially when solvation is considered) show that this reaction mechanism is energetically plausible.

Due to the large solvation energy of free halide ions, dehalogenation reactions are particularly challenging to treat using quantum chemical models in conjunction with implicit solvent. Unless the model is carefully chosen and the surrounding is properly accounted for, small models could yield unrealistic energies. It is here constructive to compare the energies obtained using the three models in order to gain deeper insight into the strengths and shortcomings of this approach.

Model A has a very small halide binding site, consisting of essentially two hydrogen bonds. As expected, the energies obtained with this small model are highly sensitive to the added solvation (see Table 1). Without solvation, this model yields a reaction barrier and energy that are highly overestimated compared to experiments. However, when the implicit solvation is added, this model reproduces the experiments quite nicely. This demonstrates that the implicit solvation model indeed is able to provide a rather good estimation of the solvation effect that the lacking protein surroundings would provide.

Model B provides an improved description of the halide binding site, which is immediately evident from the obtained energies without the solvation added. Both the barrier and the reaction energy are lower than for Model A (without solvation). Also, the solvation effects are smaller than for Model A, but they are still quite large.

For both Model A and Model B, we notice that the solvation effects on the energies saturate very quickly as a function of the dielectric constant (compare, e.g., the values for $\epsilon = 4$ and $\epsilon = 80$ in Table 1). This result can easily be rationalized by considering the simple Onsager model of a dipole in a spherical cavity, for which the solvation energy is proportional to $(\epsilon - 1)/(2\epsilon + 1)$. This simple fact is important to realize when choosing which dielectric constant to use in the modeling of enzymatic reactions (see below).

In Model C, the halide binding site is essentially saturated. This results in energies that are practically the same with and without solvation correction. Hence, the inaccuracy of the approximation of using a homogeneous polarizable continuum model decreases as the model size increases. The further away from the active site the truncation is made, the better this approximation performs, because most of the polarization effects on the reactive parts are already explicitly included in the quantum calculations. However, Model C fails to reproduce the fact that the enzyme−product complex (epoxide + chloride ion) should be more stable than the enzyme−substrate complex (chloroalcohol). As discussed above, this could be due to the proton release step, but most likely it is a result of the locking scheme used.

By comparing the results of Models A and B to those of Model C, one can conclude that a relatively small model of the enzyme active site, in combination with implicit solvent models, can be very useful in studying the enzyme reaction mechanism also for dehalogenation reactions. In fact, larger models could suffer from limitations that smaller models do not have. For example, it is difficult to ensure that different stationary points are in the same local minima with respect to the parts that are not directly involved in the reaction. In the case of Model C, it was essential to keep a number of atoms fixed to their crystallographic positions in order to avoid that artificial movements of the various groups at the active site take place. This procedure led to a somewhat rigid halide binding site, which resulted in a raised reaction barrier compared to the smaller models. The alternative approach is of course to use QM/MM methods where fixing truncation atoms is not necessary and the flexibility of the surroundings is properly accounted for.[22] However, in QM/MM, it becomes even more difficult to deal with the multiple minima problem.

A final remark concerns the choice of dielectric constant when computing solvent corrections to quantum chemical active site models. Typically, an $\epsilon = 4$ value is used.[13,15] This value is considered to correspond to a mixture of the protein itself (ϵ of 2−3) and water ($\epsilon = 80$).[23] The exact choice is thus somewhat arbitrary.

By comparing the results of Models A and B to those of Model C (Table 1), it can be seen that with $\epsilon = 4$ the various models give results that do not differ much. This shows that the choice of $\epsilon = 4$ works reasonably well. Also, as pointed out above, the solvation energy saturates quickly as a function of ϵ. In going from the cluster model (no solvation) to $\epsilon = 4$, most of the solvation is accounted for. The difference between $\epsilon = 4$ and $\epsilon = 80$ is not very large. For Model A, for example, the energy difference between no solvation and solvation with $\epsilon = 4$ is 14.1 kcal/mol for the reaction energy, while it is only 4.6 kcal/mol between $\epsilon = 4$ and $\epsilon = 80$. For Model B, the corresponding values are 8.9 and 3.0 kcal/mol, respectively. It is very important to point out here that the continuum solvation is an approximation, and the conclusion drawn about reaction mechanisms should not be very sensitive to the exact choice of the dielectric constant or else the approximation cannot be trusted.

Acknowledgment. We gratefully acknowledge financial help from The Swedish Research Council, The Wenner-Gren Foundations, The Carl Trygger Foundation, and The Magn Bergvall Foundation.

References

(1) Nakamura, T.; Nagasawa, T.; Yu, F.; Watanabe, I.; Yamada, H. *Appl. Environ. Microbiol.* **1994**, *60*, 1297–1301.

(2) van den Wijngaard, A. J.; Reuvekamp, P. T. W.; Janssen, D. B. *J. Bacteriol.* **1991**, *173*, 124–129.

(3) van Hylckama Vlieg, J. E. T.; Tang, L.; Lutje Spelberg, J. H.; Smilda, T.; Poelarends, G. J.; Bosma, T.; van Merode, A. E.; Fraaije, M. W.; Janssen, D. B. *J. Bacteriol.* **2001**, *183*, 5058–66.

(4) Fetzner, S.; Lingens, F. *Microbiol. Rev.* **1994**, *58*, 641–685.

(5) Janssen, D. B.; Majerić-Elenkov, M.; Hasnoui, G.; Hauer, B.; Lutje Spelberg, J. H. *Biochem. Soc. Trans.* **2006**, *34*, 291–295.

(6) Majerić-Elenkov, M.; Hauer, B.; Janssen, D. B. *Adv. Synth. Catal.* **2006**, *348*, 579–585.

(7) Lutje Spelberg, J. H.; van Hylckama Vlieg, J. E. T.; Tang, L.; Janssen, D. B.; Kellogg, R. M. *Org. Lett.* **2001**, *3*, 41–43.

(8) Lutje Spelberg, J. H.; Tang, L.; van Gelder, M.; Kellogg, R. M.; Janssen, D. B. *Tetrahedron Asym.* **2002**, *13*, 1083–1089.

(9) Majerić-Elenkov, M.; Tang, L.; Hauer, B.; Janssen, D. B. *Org. Lett.* **2006**, *8*, 4227–29.

(10) Hasnaoui, G.; Lutje Spelberg, J. H.; de Vries, E.; Tang, L.; Hauer, B.; Janssen, D. B. *Tetrahedron Asym.* **2005**, *16*, 1685–1692.

(11) Nakamura, T.; Nagasawa, T.; Yu, F.; Watanabe, I.; Yamada, H. *Biochem. Biophys. Res. Commun.* **1991**, *180*, 124–130.

(12) de Jong, R. M.; Tiesinga, J. J.; Rozeboom, H. J.; Kalk, K. H.; Tang, L.; Janssen, D. B.; Dijkstra, B. W. *EMBO J.* **2003**, *22*, 4933–44.

(13) See for example the following reviews: (a) Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Phys. Chem. B* **2001**, *105*, 9375. (b) Himo, F.; Siegbahn, P E. M. *Chem. Rev.* **2003**, *103*, 2421. (c) Siegbahn, P. E. M. *Q. Rev. Biophys.* **2003**, *36*, 91. (d) Noodleman, L.; Lovell, T.; Han, W.-G.; Li, J.; Himo, F. *Chem. Rev.* **2004**, *104*, 459. (e) Himo, F. *Biophys. Biochim. Acta* **2005**, *1707*, 24. (f) Himo, F. *Theor. Chem. Acc.* **2006**, *116*, 232.

(14) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–89. (b) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098–3100. (c) Becke, A. D. *J. Chem. Phys.* **1992**, *96*, 2155–2160. (d) Becke, A. D. *J. Chem. Phys.* **1992**, *97*, 9173–77. (e) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–52.

(15) (a) Velichkova, P.; Himo, F. *J. Phys. Chem. B* **2005**, *109*, 8216. (b) Himo, F.; Guo, J.-D.; Rinaldo-Matthis, A.; Nordlund, P. *J. Phys. Chem. B* **2005**, *109*, 20004. (c) Velichkova, P.; Himo, F. *J. Phys. Chem. B* **2006**, *110*, 16. (d) Chen, S.-L.; Fang, W.-H.; Himo, F. *J. Phys. Chem. B* **2007**, *111*, 1253. (e) Sevastik, R.; Himo, F. *Bioorg. Chem.* **2007**, *35*, 444. (f) Hopmann, K. H.; Guo, J.-D.; Himo, F. *Inorg. Chem.* **2007**, *46*, 4850.

(16) Hopmann, K. H.; Hallberg, B. M.; Himo, F. *J. Am. Chem. Soc.* **2005**, *127*, 14339.

(17) (a) Hopmann, K. H.; Himo, F. *Chem.—Eur. J.* **2006**, *12*, 6898. (b) Hopmann, K. H.; Himo, F. *J. Phys. Chem. B.* **2006**, *110*, 21299.

(18) (a) Hur, S.; Kahn, K.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2215–9. (b) Lau, E. Y.; Kahn, K.; Bash, P. A.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9937–42. (c) Lightstone, F. C.; Zheng, Y. J.; Maulitz, A. H.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 8417–20. (d) Lightstone, F. C.; Zheng, Y.-J.; Bruice, T. C. *J. Am. Chem. Soc.* **1998**, *120*, 5611–5621.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; 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.03; Gaussian, Inc.: Pittsburgh, PA, 2003.

(20) (a) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799–805. (b) Andzelm, J.; Kölmel, C.; Klamt, A. *J. Chem. Phys.* **1995**, *103*, 9312–9320. (c) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001. (d) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669–681.

(21) Tang, L.; Lutje Spelberg, J. H.; Fraaije, M. W.; Janssen, D. B. *Biochemistry* **2003**, *42*, 5378–5386.

(22) For a recent comprehensive review on QM/MM methods for biological systems, see:(a) Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173.

(23) Siegbahn, P. E. M.; Blomberg, M. R. A. *Chem. Rev.* **2000**, *100*, 421–437.

CT8000443

JCTC Journal of Chemical Theory and Computation

A Potential Energy Function for Heterogeneous Proton-Wires. Ground and Photoactive States of the Proton-Wire in the Green Fluorescent Protein

Oriol Vendrell,[†] Ricard Gelabert,[*,‡] Miquel Moreno,[‡] and José M. Lluch[‡,§]

Departament de Química and Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain, and

Theoretische Chemie, Physikalisch-Chemisches Institut, Universität Heidelberg, Im Neuenheimer Feld 229, 69120 Heidelberg, Germany

Received March 5, 2008

Abstract: In this paper an EVB-based method to describe the energetics of operation of arbitrary-length heterogeneous proton-wires is described. The method keeps the number of fittable parameters low by exploiting the idea of “protonation states”. The method is applied to describe the 3-proton proton-wire described in Green Fluorescent Protein (GFP), and two sets of parameters have been obtained, one for the electronic ground state and another for the photoactive excited electronic state, of a chemical model including the groups supporting the proton-wire and based on CASPT2//CASSCF quality reference energies. The fitted EVB functions are analyzed in static terms. In this way, it is seen that only a minimum exists in S_0 while two exist in S_1: one for the photoproduct and one for the reactant in the excited state, even though consideration of the Franck−Condon excitation energy predicts an effective barrier under 1 kcal mol^{-1}. Topological analysis of the transition state structure reveals a concerted but asynchronous motion of the protons, where the chromophore’s proton lags behind, and the final proton of the wire that goes from Ser205 to Glu222 leads the process. Inclusion of nuclear dynamic efects causes this small effective barrier to vanish and predicts an essentially barrierless process in the excited state.

Introduction

Proton transfer is perhaps the single most important and ubiquitous process in chemistry and biochemistry. In a proton transfer process, a hydrogen atom bound to an electronegative atom and thus, with a local positive charge density, transfers and binds to a neighboring electronegative atom with electron lone pairs available. The structural properties of some molecules or residues which are able to act both as proton donors and proton acceptors make it possible that, under certain circumstances, large hydrogen bond networks can be created. This is the case in bulk protic solvents like water, where an extensive hydrogen-bonded network exists that connects neighboring solvent molecules. When such extensive hydrogen-bonded networks come into existence, proton transfer over large distances occurs as a series of single proton transfer processes, either concerted or stepwise, in which several protons move a single step each in the network. This is part of the well-known Grotthuss mechanism of proton transfer in bulk water[1] and explains, for instance, why the transport of protons is generally much faster in liquid water than that of other cations or anions of similar charge and size to the hydronium cation, the smallest hydrated form of proton stable in water.

Theoretical methods have become invaluable tools to understand how enzyme reactions work.[2–4] In recent years proton transfer, and especially proton transfer along a

---

[*] Corresponding author e-mail: ricard.gelabert@uab.cat.

[†] Universität Heidelberg.

[‡] Departament de Química, Universitat Autònoma de Barcelona.

[§] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona.

10.1021/ct800075w CCC: $40.75

sequential arrangement of proton donors and acceptors, the so-called "proton wires," "proton relays", or "proton channels," has been receiving increasing attention from researchers in the field.[5–20] This singular kind of arrangements can be found for instance inside transmembrane proteins. In this case, an excess proton is "pumped" from one side of the membrane to the opposite, being then responsible for the generation of electrochemical potentials accross the membranes. Examples of this kind of proton transfer in biological systems are the F_0 component of ATP synthase;[21] the M2 proton channel of the A and B influenza viruses, which forms a proton channel necessary for the process of the uncoating of the virus;[22,23] the vacuolar H^+-ATPase (V-ATPase) and vacuolar H^+-pyrophosphatase (V-PPase), two proton-pump proteins crucial for the process of pH-homeostasis in endomembrane systems;[24] and the redox-driven proton pump in cytochrome *c* oxidase (C*c*O), a cornerstone of the energy transformation processes in biological systems.[25]

Transport of protons along a proton wire can also be initiated through photoactivation.[14,15,26–28] In this case, the trigger of the global proton traslocation is an excited-state proton transfer (ESPT): one of the transferable protons in the wire starts off bound to a photoactive molecule (chromophore). Absorption of a photon by the chromophore results in this case in increased acidity of the proton in the photoactive electronic excited state. The bond between chromophore and proton is much weakened, and proton transfer can take place. In some cases, the acidity of the proton in the photoactive state is so large that proton transfer in the excited state turns out to be barrierless, producing as a result ultrafast processes.[14,15,26–33] Given a proper arrangement of other proton acceptors and donors in the vicinity of the chromophore, proton transfer over large distances can occur also in an electronic excited state. Several cases of photoactivated proton-transfer processes involving a proton wire have been documented: the photocycle of retinal and bacteriorhodopsin[34] and the photocycle of the Green Fluorescent Protein (GFP)[35] and the asFP595 fluorescent protein.[36]

Because of the intrinsic interest of systems where multiple proton transfer occurs and the sometimes difficult interpretation of related experiments, detailed computational studies of these systems are necessary to lay the foundations of a sound understanding of their intricacies. After photoexcitation a series of ultrafast events may be triggered. They can be understood if a precise picture of the *real time* dynamics after photoexcitation (and therefore on one or several electronic excited-state PES) is available. Indeed, direct comparison of simulations to experimental measurements, e.g. pump–probe femtosecond spectra, is usually achieved via the computation of real-time correlation functions.[37] The inspection of the dynamics of the system after photoexcitation can also provide insightful mechanistical details just by following the evolution of classical trajectories or quantum wavepackets.

The problem of simulating such ultrafast processes can be conceptually divided in two major parts: 1) the computation of the underlying PES and 2) the solution of the dynamics of the nuclei on it. If classical dynamics, or more generally classical-trajectory based methods,[1,38] are to be used in the simulation, *ab initio* MD (AIMD)[39] methods may be employed with the features that the PES is only calculated there where it is needed and that they are full-dimensional in nature. These methods offer a convenient and combined solution to points 1) and 2). For ground electronic-state processes it is also possible to use polarizable force fields[40] or the empirical valence bond (EVB) method of Warshel[2,41] that accounts for bond forming and breaking at a much more reduced computational cost than AIMD methods. More recently another method has been proposed, which is suitable for condensed phase simulations and that combines Hartree–Fock molecular orbital (MO) and valence-bond (VB) theories to describe the PES of reactive systems (hence, MOVB method).[42–44]

It remains still the issue of whether classical-trajectory based methods can provide an accurate picture of the real-time dynamics following a photoexcitation event when light nuclei, e.g. protons, are involved.

Quantum dynamics is the correct theory to describe the motion of nuclei. However, converged quantum dynamics studies of large systems are nowadays not possible, and models of reduced dimensionality must be formulated when treating more than 6 or 7 atoms. The necessity to reduce the dimensionality of the problem arises basically from two factors. First, wavepacket methodologies based on grids require a global knowledge of the PES which must be computed on a grid of the order of 10^N points, where N is the dimensionality of the nuclear problem. Second, one must keep the problem of a size that is solvable by modern wavepacket methodologies such as the multiconfiguration time-dependent Hartree (MCTDH) method.[45] In this context, quantum dynamical, reduced-dimensionality investigations of processes involving the motion of light nuclei, such as proton-wires, appear as complementary and even necessary with respect to classical dynamics approaches. Reduced dimensionality studies of proton-transfer processes have been successfully carried on in the past on parametrized PESs and have the potential to unravel many important details of the systems under consideration.[46,47]

In this paper we present a convenient way to obtain a *parametrized* PES based on an affordable number of *ab initio* single-point calculations which is tailored to the simulation of proton-wires of the kind mostly found in biological systems, in which the proton-donor and proton-acceptor species in the wire can be rather heterogeneous. This covers our previous point 1) above. We formulate it bearing in mind that our goal is to perform quantum dynamics on the obtained PES, that is, our point 2). The PES we propose is part of a planar model of the proton-wire in which the proton positions and the donor–acceptor distances are the dynamical variables. The proposed PES and model are motivated by the fact that nowadays accurate quantum dynamical simulations are already feasible for 10 to 15 coupled degrees of freedom, which include many proton-wires within the range, while the computational cost of computing the full PES by DFT or *ab initio* methods for such systems is still far beyond reach. In a proton-wire with m transferable protons $2m$ bonds can be formed and broken. A PES able to describe such a

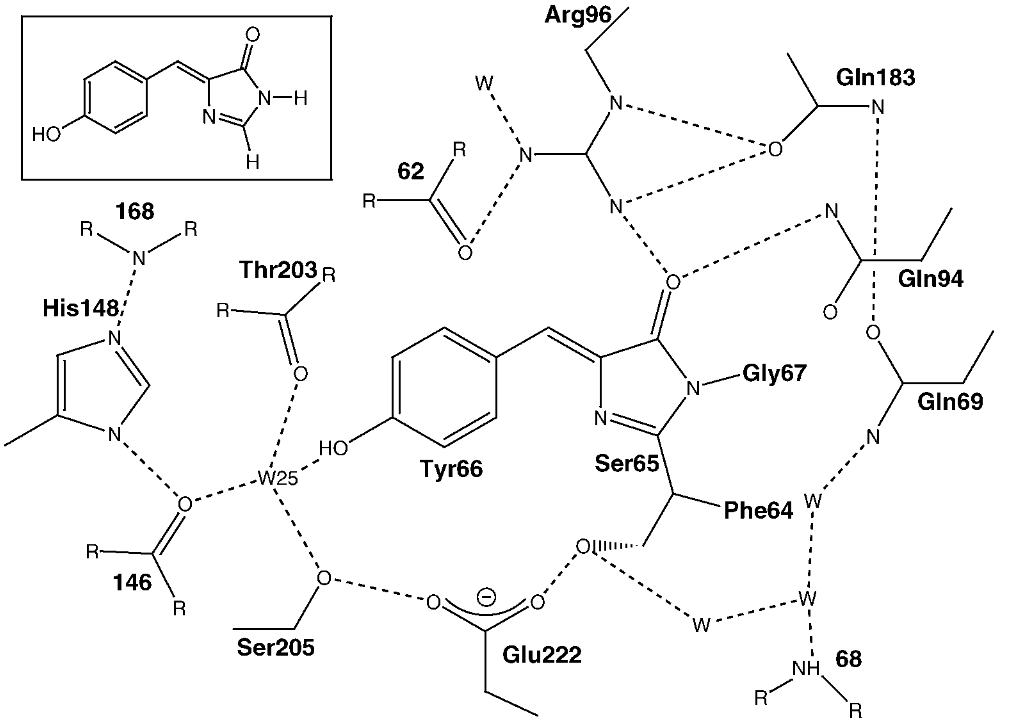


Figure 1. Schematic representation of the prevailing interactions surrounding the GFP chromophore. In the inset the isolated chromophore, *p*-hydroxybenzylideneimidazolinone, is shown.

system must then be able to describe bond breaking and forming correctly. To do so we base our approach on the EVB method[2,41] and present a convenient scheme to determine its parameters for the proton-wire at hand. It should be noted that a covalent bond or a hydrogen bond between a proton and a heavy atom in the wire will have different properties depending on the configuration of the rest of the wire. This could lead to an enormous number of parameters to describe all the different possibilities. The PES model is formulated such that the number of parameters involved keeps relatively low by exploiting the idea of protonation states. The number of *ab initio* reference points employed to fit the model PES is also kept as reduced as possible.

As an example of this computational approach the method will be applied to a system that has received much attention recently, the Green Fluorescent Protein (GFP) and its associated proton-wire.[35] GFP is a protein that occurs naturally in the North Pacific jellyfish *Aequorea Victoria* and that has found many uses as a biological fluorescent marker.[35] GFP absorbs in the blue region of the visible spectrum and has green fluorescence with a high quantum yield (0.72−0.85).[48–50] The structure of the protein has been described as a rigid cylinder made up of 11 β-sheets that shelter an α-helix supporting the chromophore roughly in the center of the cavity. The chromophore has been characterized as *p*-hydroxybenzylideneimidazolinone,[51,52] yet synthetic model compounds of the chromophore show no fluorescence,[53] indicating a strong influence of the microenvironment of the chromophore within the protein on the fluorescence. A detailed crystallographic description of GFP has been available for awhile[54] and has established that the chromophore is hydrogen-bonded to an internally caged water molecule (Wat25), which is bound to a serine residue (Ser205), which is finally bound to a glutamate residue (Glu222). While other interactions are present that help to stabilize the chromophore and keep it in place, there is nowadays consensus that the above-mentioned residues are the most relevant. See Figure 1 for a graphical description of these interactions. Wild-type GFP (wt-GFP) absorbs mainly at 397 nm, this having been identified with the absorption of the chromophore. The most widely accepted interpretation of the photophysical behavior of GFP is that photoexcitation of the chromophore triggers the operation of the proton-wire formed by the chromophore's phenolic group, Wat25, Ser205, and Glu222, including a total of three protons. The final product of this proton transfer, a protonated Glu222 residue, has been detected spectroscopically.[55,56] The green fluorescence emitted at 510 nm has been identified with the emission from the anionic (phenolate) form of the chromophore.[57] As described, the mechanism to induce fluorescence in GFP involves a heterogeneous proton-wire of three transferrable protons, with a total of four different chemical species supporting all the donor and acceptor atoms.

Our group has contributed to the theoretical knowledge on the dynamics of GFP previously. Static studies revealed that *in vacuo* the proton-transfer from the chromophore to the water molecule could occur involving a crossing of electronic states.[58] Later on, an extensive static study of the potential energy landscape controlling the triple proton-transfer in a model of the chromophore and intervening residues was undertaken.[18] This latter work is used in this paper to fit the potential energy surface functional form. The EVB approach described in the present work and in particular the PES obtained for the GFP proton-wire have been used recently to carry out a 6-dimensional quantum dynamical study of the proton-wire in GFP, in which the protein environment has not been included.[59] However, we note that, without loss of generality, the procedure described in this paper can be used to fit energy values coming from QM/MM methods.

The rest of this paper is organized as follows: section 2 discusses the theory, and in particular subsection 2.2 presents the methodological development of an EVB functional form for heterogeneous proton-wires; section 3 describes the application to the proton-wire in GFP; and finally section 4 concludes.

2. Theory

2.1. General EVB Potential. Each possible protonation state of the proton-wire is described by an EVB state. A protonation state is formally defined as a particular configuration of the hydrogen-bond system where each proton can formally be bonded to either the heavy atom at its left or at its right side. A proton-wire having m transferable protons is thus described, by $n = 2^m$ EVB states.

In the EVB theory the EVB states $|\phi_i\rangle$ appear only as a conceptual device, but they are neither directly calculated nor used. It is assumed that these states conform an orthonormal basis, $\langle\phi_i|\phi_j\rangle = \delta_{ij}$. The matrix elements of the EVB states $H_{ij} \equiv \langle\phi_i|\hat{H}|\phi_j\rangle$ are the actually relevant quantities, which are assumed to be functions of the positions of the nuclei, $H_{ij} \equiv H_{ij}(\mathbf{q})$. These matrix elements are typically described by parametrized functions similar to the usual molecular force fields. In practice, one constructs the EVB representation of the electronic Hamiltonian operator

$$\mathbf{H}(\mathbf{q}) = \begin{pmatrix} H_{11}(\mathbf{q}) & \cdots & H_{1n}(\mathbf{q}) \\ \vdots & & \vdots \\ H_{n1}(\mathbf{q}) & \cdots & H_{nn}(\mathbf{q}) \end{pmatrix} \quad (1)$$

and then diagonalizes it to get its eigenvalues and eigenvectors (From now on in the discussion, $\mathbf{q}$ may be sometimes dropped for the sake of clarity.). The only purpose of the EVB approach is to reproduce the PES of a particular electronic state (usually the ground electronic state, but not necessarily, as in this work) of a molecular system, with the extra property over traditional force fields that it is able to reproduce covalent bond breakage and formation. The PES of interest, representing the ground electronic state of the system or an excited electronic state, will be given invariably by the lowest eigenvalue of the $\mathbf{H}$ matrix, since the H_{ij} elements are parametrized for this purpose. It is important to stress at this point that the EVB form of the potential described above has nothing to do with the treatment of several electronic states of the molecular Hamiltonian in a diabatic or adiabatic representation. The approach is conceived to reproduce *only* a single PES as the lowest eigenvalue of the $\mathbf{H}$ matrix. We emphasize here that the use of an EVB-type PES to represent the potential of an excited electronic state is not new. It has been already used by some of us to describe an excited-state proton-transfer process in the condensed phase through a two-state EVB model.[33] More recent examples can also be found.[60] If the coupling between different electronic states were to come into play the present formulation should have to be modified.

The EVB PES is given by

$$V(\mathbf{q}) = \langle 0|\hat{H}|0\rangle \quad (2)$$

$$V(\mathbf{q}) = \sum_{i=1}^{n}\sum_{j=1}^{n} H_{ij} c_i^{(0)} c_j^{(0)} \quad (3)$$

which is equivalent to writing the eigenvalues equation for the EVB matrix

$$\mathbf{H}\,\mathbf{c}^{(0)} = E^{(0)}\,\mathbf{c}^{(0)} \quad (4)$$

where $\mathbf{c}^{(0)}$ is the eigenvector corresponding to the lowest eigenvalue, $E^{(0)} \equiv E^{(0)}(\mathbf{q}) \equiv V(\mathbf{q})$ and $|0\rangle \equiv \mathbf{c}^{(0)}$.

2.2. EVB Matrix-Elements for an Heterogeneous Proton-Wire. *2.2.1. Diagonal Elements.* The diagonal element H_{ii} describes the potential of the EVB state i which, as mentioned above, corresponds to a particular protonation state of the wire. In a wire of m protons there are m covalent bonds per protonation state that will be described by some sort of function. In addition we include functions that take into account nonbonded van der Waals interactions. The diagonal term describing the ith protonation state is given in general by

$$H_{ii} = \sum [\text{bond. interac.}] + \sum [\text{nonbond. interac.}] + \Delta_i + \Delta_{tot} \quad (5)$$

The different terms appearing in the diagonal of the EVB matrix are discussed in what follows.

Bonding Interactions. A possible strategy is to consider each bond in each protonation state as different, which means described by a different parametrized function. In a wire with m transferable protons the number of protonation (EVB) states is given by $n = 2^m$. For each protonation state one must specify m bonding interactions, one for each proton, leading to a total of $m \times 2^m$ functions to describe bonding interactions. The number of bonding interactions to be parametrized grows exponentially under such scheme. The exponential growth may be cut down to linear by assuming that each heavy atom in the wire is able to form two types of bonds with the transferring protons: For a given protonation state a nonterminal residue can have 0, 1, or 2 protons attached. In the first case that residue holds a hole of the wire and a formal negative charge. In the last case it holds the extra proton and a formal positive charge. Only when a residue has 1 or 2 protons attached can a covalent bond be conceived, being the bond, in principle, of a different nature in each case. Thus, for every different nonterminal residue two new types of covalent bonds are taken into account. Using a similar argument the terminal residues add a new type of bond each. Adopting this model for the bonding interactions, it is easy to see that a proton-wire with m transferable protons, where all the residues are different, has a total number of $2m$ bond types. The other limiting case, a system where all the residues are of the same kind (e.g., water), could be described by *only* 2 different bond types.[61]

Morse potentials are usually employed for the bonding interactions when constructing EVB potentials

$$M(d) = D[1 - e^{-\beta(d-d^0)}]^2 \quad (6)$$

where d is the distance from the proton to the heavy atom.

Besides, bond-angle and torsional deformations are also counted among the bonding interactions. Usually, the bond-angle deformations are described using a harmonic-like potential term

$$A(\theta) = \frac{1}{2}k_\theta(\theta - \theta^0)^2 \tag{7}$$

where θ would be the angle formed by three consecutive atoms in the wire. As for torsional contributions to bonding interactions, they are usually parametrized with terms of the form

$$D(\varphi) = k_\varphi[1 + \cos(n\varphi - \varphi^0)] \tag{8}$$

where n represents the periodicity of the torsional potential.

Nonbonding Interactions. For the nonbonding interactions, Coulomb and/or van der Waals interactions can be considered. In the case of the Coulomb interactions, a charge is assigned to each atom in every protonation state, and the charges are adjusted along with the rest of parameters

$$E = \sum \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_iq_j}{r_{ij}} \tag{9}$$

where the summation runs over all nonbonded atom pairs.

van der Waals interactions are important to maintain the structure of the wire and to represent the nonbonding interactions of the protons and heavy atoms, namely the hydrogen bonds. van der Waals interactions between protons and heavy atoms to which they are not bonded force the potential energy to increase when a proton tries to enter the vicinity of this heavy atom to which it is not bonded. Interactions between heavy atoms keep them from flying apart and control the donor−acceptor distance which is important to have well-described hydrogen bonds. The model of nonbonding interactions is as follows: a proton has a nonbonding interaction with a neighbor heavy atom if this proton has no covalent bond with it. The type of the nonbonding interaction is independent of the configuration of the rest of protons in the wire and depends only on the type of the residue. In a chain with m protons this gives rise to $2m$ different van der Waals interactions, two for each proton, one with each different neighbor. The interaction of a pair of adjacent heavy atoms gives rise to a particular van der Waals type, independent of the protonation state of the system. In a chain of m protons this results in m different van der Waals interactions between the heavy atoms. The 12−6 van der Waals function has been adopted to describe the nonbonding interactions

$$W(d) = 4\varepsilon[(\frac{\sigma}{d})^{12} - (\frac{\sigma}{d})^6] \tag{10}$$

Shifts. Each EVB state is added a shift term Δ_i. This term is meant to give flexibility to the EVB states so that the relative energies between them can adjust without placing strain on the parameters of the bonding and nonbonding interactions. Moreover, adding or subtracting the same constant k to all Δ_i values shifts the whole potential by k. It is important that all Δ_i's can vary independently since this degree of freedom allows the potential as a whole to approach the reference energies without changing its shape. In the event that the *shape* of the potential would be good but its absolute values be shifted by some amount k from the reference energies, a simple shift in all Δ_i's by an amount k would suffice to improve the overall fit. Instead of allowing all Δ_i's to change independently, one of them is always kept fixed at 0, so that the $(n-1)$ Δ_i values turn into relative shifts with respect to the fixed one. The degree of freedom that has been lost in parameter space is recovered by introducing a total shift, Δ_{tot}. By introducing a single parameter that controls the total shift the fitting procedure becomes more stable.

2.2.2. Coupling Terms. The EVB coupling terms are the nondiagonal elements of the EVB matrix. If the EVB matrix is taken to be real and symmetric, the number of different coupling terms is $n(n-1)/2$, n being the number of EVB states. Using again the relation between n and the number of transferrable protons m, the number of coupling terms is given by $2^{2m-1}-2^{m-1}$, again an exponential growth with the number of transferrable protons. In order to avoid the inclusion of so many coupling terms, and hence of many parameters into the functional form of the PES being adjusted, only the coupling terms between protonation states interconvertible by transfer of *a single proton* are considered. Such assumption in the functional form does not represent a problem in the description of concerted proton motions: If data points are included belonging to such regions of configurational space, the functional form as a whole will adjust to represent them correctly as the full set of parameters is optimized. Within this model the number of coupling terms is cut down to $m2^{m-1}$, which is however still exponential but with a much smaller growth rate. This scheme also facilitates the consideration of wires where all the residues are equal, since a single type of coupling needs to be considered: that corresponding to the transfer of a proton from a residue to another (which still appears in $m \times 2^{m-1}$ positions at each side of the diagonal in the EVB matrix).

The coupling terms can be treated in different ways. One possibility is to use Gaussian functions. This approach has the advantage that far from the strong interaction region the coupling vanishes smoothly and the equilibrium regions are cleanly described by one single EVB state. Thus, the EVB diagonal terms can be general force fields that were previously parametrized and are not being refitted. The couplings take care only of the interaction region. This is not our case, since we fit all the parameters simultaneously with respect to a set of reference *ab initio* energies. In such a case constant coupling terms for each electronic state are justified, which add a single parameter each to the functional form and facilitate the fitting. Using constant couplings with general force fields may destroy the properties of the force field in the noninteracting regions, but again this is not the case here, since we are only interested in having a global potential, flexible enough and yet tailored to reproduce a proton-wire potential after determination of its parameters.

3. Application to the GFP Proton-Wire

3.1. Specification of the EVB Matrix Elements. The proton-wire in GFP includes three transferrable protons. Hence, there are $2^3 = 8$ possible protonation states. This leads to a real symmetric 8-states EVB matrix:

Table 1. Equivalence between EVB States—Diagonal Elements in the EVB Matrix, Eq *11*—and Protonation States, Here Given as a 3-Digit Binary Number[a]

	protonation state	
EVB state	this work	refs 18 and 59
1	000	CWS
2	100	WWS
3	010	CSS
4	001	CWG
5	110	WSS
6	101	WWG
7	011	CSG
8	111	WSG

[a] See the main text for details. In the third column, the denomination of each protonation state is also given as in our related work in GFP in refs 18 and 59: each protonation state of the wire is referenced by a three-letter code where the first, second, and third letter identify the heavy atom supporting the first, second, and third proton of the wire, respectively, following this key: *C*: chromophore, *W*: Wat25, *S*: Ser205, and *G*: Glu222.

$$\mathbf{H}(\mathbf{q}) = \begin{pmatrix} H_{11} & H_{12} & H_{13} & H_{14} & 0 & 0 & 0 & 0 \\ H_{12} & H_{22} & 0 & 0 & H_{25} & H_{26} & 0 & 0 \\ H_{13} & 0 & H_{33} & 0 & H_{35} & 0 & H_{37} & 0 \\ H_{14} & 0 & 0 & H_{44} & 0 & H_{46} & H_{47} & 0 \\ 0 & H_{25} & H_{35} & 0 & H_{55} & 0 & 0 & H_{58} \\ 0 & H_{26} & 0 & H_{46} & 0 & H_{66} & 0 & H_{68} \\ 0 & 0 & H_{37} & H_{47} & 0 & 0 & H_{77} & H_{78} \\ 0 & 0 & 0 & 0 & H_{58} & H_{68} & H_{78} & H_{88} \end{pmatrix} \tag{11}$$

Table 1 indexes each diagonal term in the EVB matrix in eq 10 with the corresponding protonation state of the GFP proton-wire. In order to label in a systematic way the protonation states in GFP, we propose a naming scheme that is concise and can easily be extended to represent any conceivable linear proton wire. The labeling scheme plays with the idea of protonation states, a protonation state being a structure in which each proton is within bonding distance to either its donor or acceptor atom. The protonation state is represented by a binary number with 3 digits, each of them assigned, respectively, to the Cro-Wat25, Wat25-Ser205, and Ser205-Glu222 hydrogen bond. In case that one of the three protons is in bonding distance to its donor atom it is represented by “0”. If the proton is within bonding distance to the acceptor atom it is represented by “1”. Since there are three transferable protons, the protonation states can be referred to as a binary number of three digits with a total of $2^3 = 8$ protonation states. The idea of protonation states we are using does not assume anything about their stability; it is simply a convenient way of referring to the different structures relevant to the proton-wire. Using such convention, reactants are represented by (000), products are represented by (111), the situation in which only the proton shared by the chromophore and Wat25 has transferred would be given by (100), and so on. This naming scheme can straightforwardly be generalized to a proton-wire of *m* protons, by using an *m*-digit binary number to represent the 2^m protonation states.

Even though all elements in the matrix in eq 11 could be functions of all coordinates **q**, a different choice has been adopted here. As explained in the previous section, the nondiagonal terms in eq 11 are just parameters (that is, not functions of the coordinates **q**). The diagonal terms are functions of the coordinates and are given by

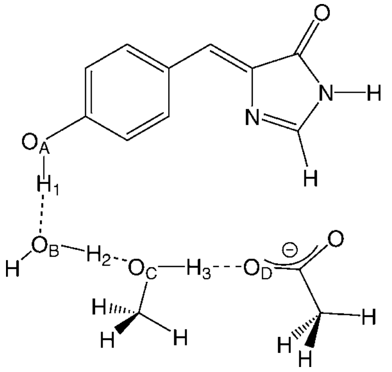


Figure 2. Chemical model used in ref 18 and in this work to compute the potential energies of selected structures used in the fit.

$$H_{ii} = B_i + W_i + W_{\text{chain}} + \Delta_i + \Delta_{tot} \tag{12}$$

where B_i is the bonding contribution of the protonation state *i*, W_i is the nonbonding contribution between protons and oxygens forming hydrogen bonds in the *i*th protonation state, and W_{chain} are the nonbonding contributions between neighboring oxygen atoms, which are the same in each protonation state. Δ_i is the state-specific energy-shift parameter and Δ_{tot} is the total shift parameter.

In a previous work in which a lengthy molecular-dynamics simulation of GFP was performed, we showed that the proton-wire was well formed over time and that it remained approximately planar.[18] Thus, it seems reasonable to drop from the bonding interactions those that imply departure from planarity or torsional potential energy terms. Furthermore, as will be explained soon, the reference data against which the EVB functional form will be fitted will come from a reduced model of the proton-wire, used before in our previous study[18] and represented here in Figure 2. In this model, the effect of the rigid stucture of the protein surrounding the proton-wire is not accounted for, and, thus, terms of bond-angle deformation of the heavy atoms (i.e., depending on $O_AO_BO_C$ or $O_BO_CO_D$—see Figure 2) cannot be meaningfully included in the bonding interactions, because the rigid structure of the cavity where the proton-wire lies is missing in the reduced model. For this study, we choose as bonding contributions those that come directly from the stretching of the bonds between the protons and the donor and acceptor atoms. Within this approach, the seven atoms forming the wire being restricted to lie in a plane require $2 \times 7 = 14$ degrees of freedom, from which only 11 correspond to internal motion (2 for the center of mass on the plane and one for in-plane rotation). If the bending motion of the heavy atoms is not allowed, the number of degrees of freedom remaining is $11 - 2 = 9$.

The bonding contributions are then given by

$$\begin{array}{lcllll}
B_1 &=& M_A(d_{A1}) &+& M_{B-}(d_{B2}) &+& M_{C-}(d_{C3}) \\
B_2 &=& M_{B+}(d_{1B}) &+& M_{B+}(d_{B2}) &+& M_{C-}(d_{C3}) \\
B_3 &=& M_A(d_{A1}) &+& M_{C+}(d_{2C}) &+& M_{C+}(d_{C3}) \\
B_4 &=& M_A(d_{A1}) &+& M_{B-}(d_{B2}) &+& M_D(d_{3D}) \\
B_5 &=& M_{B-}(d_{1B}) &+& M_{C+}(d_{2C}) &+& M_{C+}(d_{C3}) \\
B_6 &=& M_{B+}(d_{1B}) &+& M_{B+}(d_{B2}) &+& M_D(d_{3D}) \\
B_7 &=& M_A(d_{A1}) &+& M_{C-}(d_{2C}) &+& M_D(d_{3D}) \\
B_8 &=& M_{B-}(d_{1B}) &+& M_{C-}(d_{2C}) &+& M_D(d_{3D})
\end{array} \qquad (13)$$

with

$$M_{X*}(d_{Xn}) = D_{X*}[1 - e^{-\beta_{X*}(d_{Xn} - d^0_{X*})}]^2 \qquad (14)$$

where d_{Xn} (and d_{nX}) is the distance between oxygen X and proton n (with A being the phenolic oxygen in Cro, B the oxygen in Wat25, C the oxygen in Ser205, and D the acceptor oxygen in Glu222; proton 1 is shared between Cro-Wat25, proton 2 between Wat25-Ser205, and proton 3 between Ser205-Glu222: see Figure 2), D_{X*}, β_{X*}, and d^0_{X*} are parameters, and $*$ may be $+$, $-$, or blank: $+$ and $-$ refer to the Morse potentials of the nonterminal residues supporting 2 and 1 protons, respectively, and a blank refers to the Morse potential for terminal residues supporting a proton.

Let us consider the nonbonding contributions. In our case, the consideration of electrostatic terms (i.e., charges) was found to be redundant with the bonding interactions and not necessary to describe the proton-wire, since each atom moves in a relatively small region of configurational space. It has to be recalled that the final goal is to achieve a good description of the PES by fitting the EVB potential to some reference data, and the inclusion of charges led in this case to an overparametrization of the model. As a consequence, the only terms actually considered in the nonbonding contributions are van der Waals terms. Within these considerations, the nonbonding contributions of each O—H hydrogen bond are given by

$$\begin{array}{lcllll}
W_1 &=& W_{1B}(d_{1B}) &+& W_{2C}(d_{2C}) &+& W_{3D}(d_{3D}) \\
W_2 &=& W_{A1}(d_{A1}) &+& W_{2C}(d_{2C}) &+& W_{3D}(d_{3D}) \\
W_3 &=& W_{1B}(d_{1B}) &+& W_{B2}(d_{B2}) &+& W_{3D}(d_{3D}) \\
W_4 &=& W_{1B}(d_{1B}) &+& W_{2C}(d_{2C}) &+& W_{C3}(d_{C3}) \\
W_5 &=& W_{A1}(d_{A1}) &+& W_{B2}(d_{B2}) &+& W_{3D}(d_{3D}) \\
W_6 &=& W_{A1}(d_{A1}) &+& W_{2C}(d_{2C}) &+& W_{C3}(d_{C3}) \\
W_7 &=& W_{1B}(d_{1B}) &+& W_{B2}(d_{B2}) &+& W_{C3}(d_{C3}) \\
W_8 &=& W_{A1}(d_{A1}) &+& W_{B2}(d_{B2}) &+& W_{C3}(d_{C3})
\end{array} \qquad (15)$$

The O—O nonbonding contributions are equal for all protonation states and are given by

$$W_{\text{chain}} = W_{AB}(d_{AB}) + W_{BC}(d_{BC}) + W_{CD}(d_{CD}) \qquad (16)$$

where d_{XY} is the distance between oxygen atoms X and Y. The 6−12 van der Waals terms have been adopted for the nonbonding interactions

$$W_{XY}(d_{XY}) = 4\varepsilon_{XY}[(\frac{\sigma_{XY}}{d_{XY}})^{12} - (\frac{\sigma_{XY}}{d_{XY}})^{6}] \qquad (17)$$

Globally the potential energy surface is obtained as the lowest eigenvalue of the matrix in eq 11 and depends on a total of 9 independent variables: the distances between each donor and acceptor oxygen atom (d_{AB}, d_{BC}, d_{CD}), the distances between donor atom and the transferring proton (d_{A1}, d_{B2}, d_{C3}), and the distances between acceptor atom and the transferring proton (d_{1B}, d_{2C}, d_{3D}):

$$V = V(d_{AB}, d_{BC}, d_{CD}, d_{A1}, d_{B2}, d_{C3}, d_{1B}, d_{2C}, d_{3D}) \qquad (18)$$

This model as it stands is valid for a system where all proton-wire atoms are contained in the same plane and where bending of the heavy atoms is not allowed.

3.2. Fitting of the EVB Functional Form. To save computational effort each of the three hydrogen bonds in GFP has been kept linear (that is, the proton moves in a straight line from donor to acceptor atom). Because in the model each hydrogen bond is now restricted to remain linear, distances between donor and acceptor, proton and acceptor, and proton and donor of each hydrogen bond are related:

$$\begin{array}{l}
d_{AB} = d_{A1} + d_{1B} \\
d_{BC} = d_{B2} + d_{2C} \\
d_{CD} = d_{C3} + d_{3D}
\end{array} \qquad (19)$$

This reduces effectively the number of degrees of freedom of the system to 6.

The reference data points used to adjust the parameters of the PES have been partially published in a related work.[18] A brief summary of the level of calculation and strategy followed is given here: The potential energy was evaluated for both the S_0 and S_1 states at the CASPT2 level with an active space of six electrons in six π molecular orbitals on 252 different geometries of the chain.[18] The system was modeled to reduce the number of atoms involved and is depicted in Figure 2.

The chemical modelization included forced planarity of the system and substitution of the Ser205 residue by a methanol and of the Glu222 residue by an acetate. The restriction of planarity was justified by the results of a long-time molecular dynamics simulation of the ground state.[18] The geometries were generated along each possible path involving a single proton transfer event (giving a total of 12 paths: see Figure 3) and for three different oxygen—oxygen distances in each case. Seven equidistant positions of the transferring proton were scanned for each path and each donor—acceptor distance. As for the rest of the atoms in the model, their relative positions were held fixed at their values in the minimum in the ground electronic state, S_0. When oxygen atoms were displaced along the wire (change of a d_{XY} coordinate) the potential energy was computed for a structure where the distance between the oxygen atoms and their supporting residues was kept fixed (that is, the residues were translated alongside with the oxygen atom with fixed relative coordinates). Using this procedure, the potential energies of the ground electronic state and the photoactive $\pi\pi^*$ state have been obtained. To this large number of single-point calculations an extra set of 86 additional points has been calculated to improve the description of some areas of configurational space that were not relevant for a static study (such as that carried out in ref 18) but that have a certain effect in the asymptotic regions of the potential energy function that is fitted in this work, which become relevant in dynamical studies, where high-energy regions can support

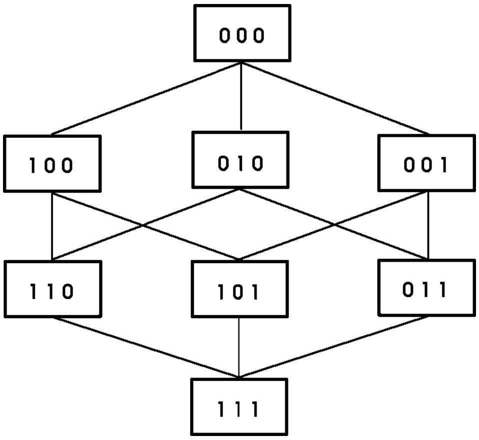


Figure 3. Schematic representation of the different paths connecting the eight possible protonation states of the proton-wire.

a certain amount of density over time.[59] Besides, we have added structures where all protons are exactly in the middle of their respective hydrogen bonds, and structures where two given protons are exactly in the middle of their respective hydrogen bonds, while the third is bound to either the donor or the acceptor atom. These structures are important to describe the three possible concerted double proton transfers and the also possible concerted triple proton transfer. Details on the structures and energetics of this extra set are included in the Supporting Information. Overall, the total number of reference data point against which the potential energy functions have been fitted amounts to 338.

Next, the error function

$$E(\mathbf{p}) = \sqrt{\frac{\sum_i^{N_p} w_i (V(\mathbf{q}_i;\mathbf{p}) - V_{ref}(\mathbf{q}_i))^2}{N_p}} \tag{20}$$

which depends solely on the set of parameters **p**, is minimized using a simplex minimization algorithm.[62] The EVB functional form of each surface depends on a total of 56 adjustable parameters. N_p is the number of configurations for which a reference energy is known (338 in our case). w_i is a weight function that is introduced in order to favor the fit in regions of low potential energy with respect to high energy less relevant regions. The weight is defined as follows

$$w_i \equiv w(v_i) = \begin{cases} 1 & \text{if } v_i \le v_c \\ e^{-(v_i - v_c)/\alpha} & \text{if } v_i > v_c \end{cases} \tag{21}$$

where v_i is the *ab initio* potential energy value at point q_i, v_c is a cutoff potential that controls the domain of action of the weight, and α controls the rate at which the weight goes to 0 as the energy increases. At this point we highlight that of all the energy shifts Δ_i, one is taken as fixed in our model per each electronic state. For the PES of S_0, Δ_1 has been held fixed at 0, because this corresponds to the 000 (CWS) protonation state of S_0, known to be the most stable. Likewise, in S_1 it is Δ_8 that has been held fixed at 0, since it corresponds to the 111 (WSG) species, which has been described as the most stable protonation state in S_1.[18]

The fit was carried out for S_0 and S_1. In both cases, v_c was made equal to the lowest energy of the *ab initio* reference data for the corresponding electronic state *plus* 15 kcal mol^{-1}, and $\alpha = 10$ kcal mol^{-1}. The optimization procedure was carried out until E in eq 20 was minimized. The parameters obtained are collected in Tables 2 for S_0 and 3 for S_1.

Table 2. Parameters of the Potential Energy Surface for the Ground State S_0

parameter	value	units	parameter	value	units
D_A	141.38	kcal/mol	σ_{C3}	1.36	Å
β_A	2.024	Å $^{-1}$	ε_{AB}	8.48	kcal/mol
d_A^0	0.98	Å	σ_{AB}	2.34	Å
D_{B-}	81.30	kcal/mol	ε_{BC}	4.32	kcal/mol
β_{B-}	2.51	Å $^{-1}$	σ_{BC}	2.40	Å
d_{B-}^0	0.98	Å	ε_{CD}	0.90	kcal/mol
D_{B+}	114.33	kcal/mol	σ_{CD}	2.67	Å
β_{B+}	1.88	Å $^{-1}$			
d_{B+}^0	1.010	Å	H_{12}	17.69	kcal/mol
D_{C-}	188.22	kcal/mol	H_{13}	33.62	kcal/mol
β_{C-}	1.52	Å $^{-1}$	H_{14}	58.03	kcal/mol
d_{C-}^0	0.95	Å	H_{25}	79.41	kcal/mol
D_{C+}	155.32	kcal/mol	H_{26}	20.90	kcal/mol
β_{C+}	1.27	Å $^{-1}$	H_{35}	90.92	kcal/mol
d_{C+}^0	0.34	Å	H_{37}	52.75	kcal/mol
D_D	414.44	kcal/mol	H_{46}	45.18	kcal/mol
β_D	0.79	Å $^{-1}$	H_{47}	48.02	kcal/mol
d_D^0	0.81	Å	H_{58}	99.85	kcal/mol
ε_{1B}	0.0032	kcal/mol	H_{68}	18.94	kcal/mol
σ_{1B}	1.58	Å	H_{78}	29.33	kcal/mol
ε_{2C}	4.41	kcal/mol	Δ_1	0 (fixed)	kcal/mol
σ_{2C}	1.01	Å	Δ_2	78.48	kcal/mol
ε_{3D}	1.22	kcal/mol	Δ_3	10.54	kcal/mol
σ_{3D}	1.09	Å	Δ_4	62.03	kcal/mol
ε_{A1}	−0.0016	kcal/mol	Δ_5	87.63	kcal/mol
σ_{A1}	1.18	Å	Δ_6	41.65	kcal/mol
ε_{B2}	0.0828	kcal/mol	Δ_7	40.84	kcal/mol
σ_{B2}	1.21	Å	Δ_8	50.02	kcal/mol
ε_{C3}	0.0859	kcal/mol	Δ_{tot}	44.18	kcal/mol

Table 3. Parameters of the Potential Energy Surface for the Photoactive Excited State S_1

parameter	value	units	parameter	value	units
D_A	232.20	kcal/mol	σ_{C3}	1.09	Å
β_A	1.31	Å $^{-1}$	ε_{AB}	1.35	kcal/mol
d_A^0	0.87	Å	σ_{AB}	2.58	Å
D_{B-}	65.32	kcal/mol	ε_{BC}	1.93	kcal/mol
β_{B-}	2.54	Å $^{-1}$	σ_{BC}	2.48	Å
d_{B-}^0	0.99	Å	ε_{CD}	1.43	kcal/mol
D_{B+}	287.23	kcal/mol	σ_{CD}	2.54	Å
β_{B+}	1.38	Å $^{-1}$			
d_{B+}^0	−0.28	Å	H_{12}	142.19	kcal/mol
D_{C-}	113.46	kcal/mol	H_{13}	18.36	kcal/mol
β_{C-}	2.32	Å $^{-1}$	H_{14}	3.56	kcal/mol
d_{C-}^0	0.98	Å	H_{25}	193.63	kcal/mol
D_{C+}	53.00	kcal/mol	H_{26}	−123.23	kcal/mol
β_{C+}	1.35	Å $^{-1}$	H_{35}	70.47	kcal/mol
d_{C+}^0	0.86	Å	H_{37}	92.51	kcal/mol
D_D	169.94	kcal/mol	H_{46}	168.46	kcal/mol
β_D	1.53	Å $^{-1}$	H_{47}	−37.96	kcal/mol
d_D^0	0.88	Å	H_{58}	104.49	kcal/mol
ε_{1B}	0.20	kcal/mol	H_{68}	2.50	kcal/mol
σ_{1B}	1.38	Å	H_{78}	37.68	kcal/mol
ε_{2C}	0.94	kcal/mol	Δ_1	24.87	kcal/mol
σ_{2C}	−1.03	Å	Δ_2	−39.74	kcal/mol
ε_{3D}	0.26	kcal/mol	Δ_3	140.00	kcal/mol
σ_{3D}	1.22	Å	Δ_4	95.67	kcal/mol
ε_{A1}	5.75	kcal/mol	Δ_5	−10.03	kcal/mol
σ_{A1}	0.96	Å	Δ_6	−213.73	kcal/mol
ε_{B2}	0.39	kcal/mol	Δ_7	59.10	kcal/mol
σ_{B2}	−1.18	Å	Δ_8	0 (fixed)	kcal/mol
ε_{C3}	0.56	kcal/mol	Δ_{tot}	147.75	kcal/mol

Table 4. Energy-Dependent RMS of the Fitted Potential with Respect to the Reference *ab Initio* Points, for the Ground State S_0[a]

E_{lim}	RMS	N
10.0	0.946	30
20.0	1.390	96
30.0	1.575	175
40.0	1.737	238
50.0	1.933	292
60.0	2.483	319
70.0	2.839	336
80.0	2.836	338

[a] The RMS (second column, in kcal mol^{-1}) has been computed for all reference structures whose energy is within E_{lim} (first column, in kcal mol^{-1}) of that of the lowest-energy structure in S_0. N indicates how many structures meet the energy criterion.

Table 5. Energy-Dependent RMS of the Fitted Potential with Respect to the Reference *ab Initio* Points, for the Photoactive State S_1[a]

E_{lim}	RMS	N
80.0	1.325	10
90.0	1.397	75
100.0	1.503	201
110.0	1.841	261
120.0	2.522	305
130.0	3.458	332
140.0	3.672	338

[a] The RMS (second column, in kcal mol^{-1}) has been computed for all reference structures whose energy is within E_{lim} (first column, in kcal mol^{-1}) of that of the lowest-energy structure in S_0. N indicates how many structures meet the energy criterion.

Ever since the weights are introduced in the error function its value does not have the same direct meaning as an RMS, for instance. To assess the goodness of the fit, instead, a set of energy-dependent RMS of the fit with respect to the *ab initio* data has been computed and is shown in Tables 4 (S_0) and 5 (S_1). In this way, it is possible to appreciate that an acceptable RMS under 1.6 kcal mol^{-1} is achieved, for instance, for energies within 30 kcal mol^{-1} of the minimum in the ground state S_0 (comprising 175 structures). A comparable rms of 1.5 kcal mol^{-1} is achieved for up to 201 structures that are within 100 kcal mol^{-1} of the S_0 minimum, this time for the photoactive state S_1. Taking into account the inherent complexity of potential-fitting, we think that the fittings for S_0 and S_1 are satisfactory and thus that the potential energy functions obtained correctly and almost quantitatively describe the CASPT2 energetics of the low-lying areas of both the ground and photoactive states of the model system in Figure 2.

Figures 4 and 5 display the deviations from the fitted values with respect to the reference points. It can be seen that for both electronic states the fitted values agree within chemical accuracy (~1 kcal mol^{-1}) with reference data for structures within 10−20 kcal mol^{-1} of the lowest *ab initio* energy for that state. Agreement for the rest of the surface is good, as the RMS values indicate. The quality of the fitted energies decreases as absolute energy increases. This is due to the effect of the weights in the error function, which are introduced to allow a smoother fit for the relevant (i.e., low-energy) areas of each potential energy surface.

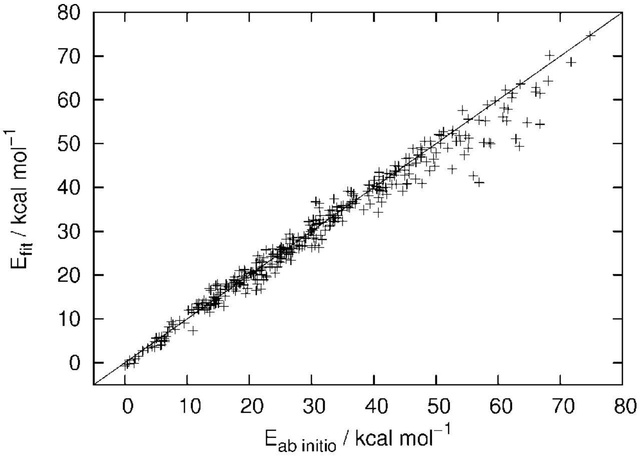


Figure 4. Deviation of the fitted values with respect to the *ab initio* reference points for the ground state S_0. Points on the diagonal are perfectly fitted to the *ab initio* reference data.

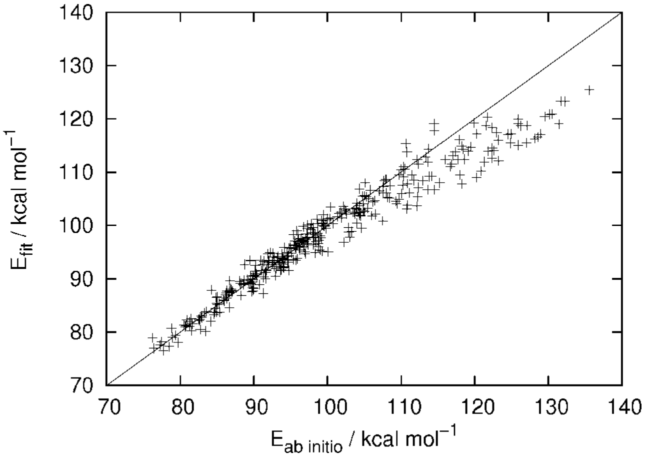


Figure 5. Deviation of the fitted values with respect to the *ab initio* reference points for the photoactive state S_1. Points on the diagonal are perfectly fitted to the *ab initio* reference data.

3.3. Analysis of the Fitted Potential Energy Surfaces. The availability of the 6-dimensional fitted surfaces for S_0 and S_1 makes an exhaustive exploration of the energetics of the wire feasible, such that the results would be, in quality terms, equivalent to an exploration of the chemical model (Figure 2) but avoiding the expensive optimizations involved. Besides, as will be discussed below, some of the energetic terms —specifically those including nuclear dynamic corrections—can only be computed from this model.

3.3.1. Static Analysis: Stationary Points. Even though a total of 8 protonation states can be thought of, it is not clear whether all of them exist as chemical entities or whether the same exist both in S_0 and S_1. An exhaustive search for the points with a null potential energy gradient has been undertaken. Figure 6 shows the energy of such points, and Figure 7 summarizes the geometry of the proton-wire for each of these stationary points.

S_0 has only one minimum where the reactants region is (000, or CWS structure). All attempts to locate other minima

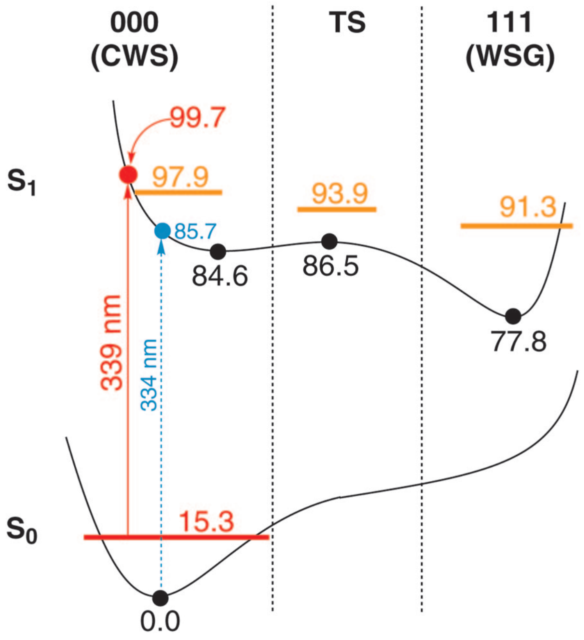


Figure 6. Schematic representation of the energetics of the operation of the proton-wire in GFP. The legend at the heading indicates the ‘chemical’ name of the different regions on the potential energy surface (PES). Black circles identify a stationary point, and figures in black denote the *Born*−*Oppenheimer potential energy* values of stationary points on the S_0 and S_1 PESs. The light blue arrow and the corresponding circle indicate the Franck−Condon transition without considering kinetic energy. The red arrow and corresponding circle indicate the Franck−Condon transition including kinetic energy. Horizontal segments identify the zero-point energy level of chemical species. Figures (and likewise, segments) in red represent exact values—within the model—for total energy obtained through dynamical simulations, while figures in orange represent estimated values for the total energy obtained assuming harmonic vibrations. All units, except where explicitly indicated, are kcal mol^{-1}.

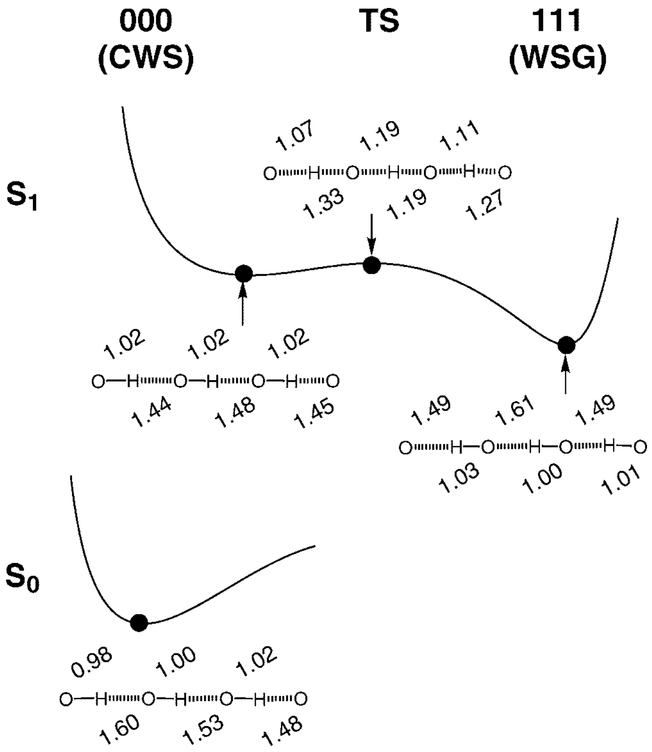


Figure 7. Schematic representation of the geometry of the proton-wire in the model in the stationary points found (solid circles) on the fitted potential energy surface.

corresponding to the remaining protonation states led invariably to this structure. In S_1 a total of two different minima have been found: the first corresponds to the reactants' structure (000, or CWS), +84.6 kcal mol^{-1} above the minimum in S_0, while the second corresponds to the final product of the wire (111, or WSG), +77.8 kcal mol^{-1} above the minimum in S_0 or 6.8 kcal mol^{-1} below the reactants' structure in S_1. Hence, the 000 → 111 process in S_1 is exoergic. The transition state structure connecting both minima has been found to be only +1.9 kcal mol^{-1} above the reactants' in S_1.

The structural changes brought up by photoexcitation are shown in Figure 7. Upon excitation, the three hydrogen bonds are shortened noticeably: Cro-Wat25 goes from 2.58 Å to 2.46 Å ($\Delta d_{AB} = -0.12$), Wat25-Ser205 from 2.53 Å to 2.50 Å ($\Delta d_{BC} = -0.03$), and Ser205-Glu222 from 2.50 Å to 2.47 Å ($\Delta d_{CD} = -0.03$). This goes in line with the elongation of the donor-proton distance, which is +0.04 Å for the chromophoric proton (Δd_{A1}) and +0.02 Å for the Wat25 proton (Δd_{B2}) and no appreciable change for the Ser205 proton (Δd_{C3}). This behavior is reasonable: accepting that photoexcitation is localized on the chromophore, the further away from Cro the proton is, the smaller the distortion of the donor−H bond is. In terms of proton motion, it seems that the chromophoric proton moves further when comparing the S_0 and S_1 000 (CWS) structures. The transition state structure for the 000 → 111 process in S_1 represents approximately a concerted process, although asynchronous. To measure the degree of asynchronicity, we define the position asymmetry coordinate δ_i as follows ($i = 1$ for the Cro-Wat25, 2 for Wat25-Ser205, and 3 for Ser205-Glu222 hydrogen bond, respectively):

$$\begin{aligned}\delta_1 &= d_{A1} - d_{1B}\\ \delta_2 &= d_{B2} - d_{2C}\\ \delta_3 &= d_{C3} - d_{3D}\end{aligned} \tag{22}$$

When $\delta_i = 0$ this indicates that the ith proton is exactly in the middle of the hydrogen bond, or, in other words, its transfer is 50% complete. The values determined for the transition state structure are $\delta_1 = -0.26$, $\delta_2 = 0.0$, and $\delta_3 = -0.16$. Even though all protons have initiated their motion to the acceptor oxygen, proton 2 (shared among Wat25-Ser205) is the one that is ahead, while proton 1 (shared among Cro-Wat25) is the one that lags behind. In the process of going from reactants in S_1 to the transition state structure a noticeable shortening of all hydrogen bonds is noticed: $\Delta d_{AB} = -0.06$, $\Delta d_{BC} = -0.12$, and $\Delta d_{CD} = -0.09$. Thus, summarizing, from a purely topological point of view (i.e., without considering dynamics of the nuclei), the changes brought up by the operation of the proton-wire in S_1 point in the direction of an overall concerted process but where the second proton—that shared among Wat25-Ser205—is ahead in the transfer, closely followed by the third proton—shared among Ser205-Glu222.

Lastly, when considering the chemical fate of the system in the excited state after photoexcitation, it is important to remember that the photoexcited system does not start off the reactants' minimum in S_1: following Franck−Condon's principle electronic reorganization occurs much faster than nuclear motion. A first-order estimate of the total energy of the system in the excited state right after photoexcitation consists of computing the energy of the structure of the ground-state reactant in S_1. In the current case this yields +85.7 kcal mol^{-1} above the ground-state minimum (see Figure 6). Accepting this value as an acceptable estimate of the energy of the system after photoexcitation, the effective potential energy barrier in S_1 for the 000 → 111 process is reduced to only +0.8 kcal mol^{-1}. The small magnitude of the effective potential energy barrier, combined with the fact that the process involves motion of light nuclei, makes it necessary to include dynamical considerations in the analysis.

3.3.2. Dynamical Corrections. A remarkable consequence of the quantum nature of matter is that bound atoms are never at rest. Thus, the energy values measured in the previous subsection can only be taken as estimations, because they consist only of the Born−Oppenheimer potential energy, that is, nuclei are held fixed in space. A formally correct estimate of the energy differences between different structures must include an estimation of the kinetic energy of the nuclei.

We have computed these dynamical corrections for all chemical entities (structures) present in Figure 6 as follows. In S_1 the only topologically stable (minimum potential energy) structures are reactants (000) and products (111). In these cases, the quantum kinetic energy of the nuclei comes from the vibrational motions of all bound normal modes, which within the harmonic oscillator approximation reads

$$\Delta E_{\mathrm{vibration}} = \frac{1}{2} h \sum_i \nu_i \tag{23}$$

where the summation runs over the six degrees of freedom. The vibrational frequencies ν_i are computed as

$$\nu_i = \frac{1}{2\pi}\sqrt{\lambda_i} \tag{24}$$

where λ_i are the eigenvalues of the mass-weighted Hessian matrix. The transition state structure for the interconversion 000 → 111 in S_1 cannot be treated in the same way because motion is topologically unbound along the reaction coordinate. In this case, the dynamic correction has been computed using eqs 23 and 24 but selecting only the five real frequencies: motion being unbound in one direction means that the Hessian matrix has one negative eigenvalue.

To complete the consideration of dynamical corrections, the question remains as to how to compute them for the Franck−Condon transition. The Franck−Condon principle states that nuclei do not change nuclear motion state in electronic transitions. To compute the dynamical correction to the Franck−Condon transition the nuclear motion wave function needs to be computed. A traditional approach in quantum dynamics simulations mimicks the Franck−Condon transition by promoting the ground vibrational state of the system (i.e., the most populated) in S_0 to the photoactive state. It is possible to determine the ground vibrational state by determining the time evolution in imaginary time (i.e., relaxation) with the S_0 Hamiltonian until invariance in the (imaginary) time-evolved state is achieved, in which case the ground vibrational state is obtained

$$t \rightarrow -i\infty \quad \chi(\mathbf{q}) \rightarrow \chi_0(\mathbf{q}) \tag{25}$$

where $\chi(\mathbf{q})$ is a reasonable guess (e.g., harmonic oscillator) to the ground-state wave function, and $\chi_0(\mathbf{q})$ the vibrational eigenstate. $\mathbf{q}$ is a vector that defines the nuclear configuration of the system. All that remains is to use an algorithm to solve the time-dependent Schrödinger equation for a multidimensional wave packet $\chi(\mathbf{q})$

$$i\hbar \frac{\partial}{\partial t}\chi(\mathbf{q}, t) = \hat{H}\chi(\mathbf{q}, t) \tag{26}$$

where $\hat{H}$ is the Hamiltonian (total energy) operator for S_0

$$\hat{H}^{S_0} = \hat{T} + \hat{V}^{S_0} \tag{27}$$

Here, $\hat{V}^{S_0}$ is the potential energy surface for the ground electronic state (which is the EVB functional form we have fitted in this work, with parameters shown in Table 2), and $\hat{T}$ is the kinetic energy operator for the system, using the same coordinate set. Recently one of us has published an accurate and successful quantum dynamical simulation of the transfer of an excess proton along a chain of water molecules in which a *linear* model for the complete proton-wire was adopted.[47] The linearity of the model restricted the motion of all atoms in the wire to take place along a line. This assumption effectively halved (at least) the number of degrees of freedom and thus is very appropriate for ulterior dynamical studies. This linear model has also been adopted here and in a related work where the PES presented here has been used.[59] Once the ground-state wave function has been obtained, the *total energy* including dynamical corrections to the Franck−Condon transition is obtained as

$$E_{\mathrm{FC}} = \langle \chi_0 | \hat{H}^{S_1} | \chi_0 \rangle \tag{28}$$

This procedure allows us to obtain a numerically exact (within the bounds of the model) value for the dynamical energy for the ground vibrational state in S_0 or zero-point energy (ZPE) as

$$E_{S_0\,000} = \langle \chi_0 | \hat{H}^{S_0} | \chi_0 \rangle \tag{29}$$

where the Hamiltonian corresponds now to the S_0 electronic state. In this way, a better estimate of the Franck−Condon transition is obtained. To obtain the nuclear motion vibrational ground state $|\chi_0\rangle$ through relaxation dynamics and then to compute E_{FC} (eq 28) and the ZPE for the 000 species in S_0 (eq 29), the Multiconfigurational Time-Dependent Hartree (MCTDH) wave packet propagation method has been used,[63−66] in particular, the Heidelberg MCTDH implementation.[67] The energy results including the dynamical corrections just described are represented also in Figure 6.

The most relevant change from the 'static' picture is the relative energetic ordering of the structures in S_1, because as expected, the small static barrier of +1.9 kcal mol^{-1}

measured from the 000 minimum in S_1 disappears once dynamical corrections are included: the S_1 000 structure is now 4.0 kcal mol^{-1} above the transition state. This means that the 000 species is not stable in S_1. Moreover, after dynamical corrections have been included in the Franck−Condon transition, the initial state in S_1 is now 1.8 kcal mol^{-1} above the 000 structure. The process in the photoactive state is predicted to occur without potential energy barrier.

Conclusions

An EVB-based approach to describe the energetics of operation of heterogeneous proton-wires of arbitrary length has been presented. The PES model described keeps the number of parameters to be fitted to reference data points low, by exploiting the idea of "protonation states," which makes this approach to construct PESes especially suited for long proton-wires.

The method has been used to construct a potential energy function for the ground and photoactive excited states of the three-hydrogen-bond proton-wire in the surroundings of the chromophore of the green fluorescent protein (GFP). This is achieved within a reduced dimensionality model of the EVB method described in this paper, in which the full proton-wire is restricted to remain in a plane and in which each hydrogen bond is restricted to be linear. As reference data, a set of 338 CASPT2//CASSCF energy values for S_0 and S_1 (with an active space of 6 electrons in 6 π molecular orbitals), covering all areas of configurational space relevant to the operation of the proton-wire, has been used. As a result, this paper contains two sets of 56 parameters, one for each electronic state, describing the operation of the proton-wire in GFP.

The fitted PESs have been analyzed in static terms (that is, in terms of stationary points of potential energy). S_0 is seen to contain only a potential energy minimum in the region of reactants. In S_1, besides a deep minimum in the region of photoproducts, a shallow minimum in the area of reactants is also found. Consideration of the Franck−Condon excitation energy (also in static terms) reduces the potential energy barrier to an effective value of only +0.8 kcal mol^{-1}. The transition state structure corresponds to a concerted, though asynchronous, motion of the three protons toward the respective acceptor atoms: the proton of the chromophore is the last to initiate its motion, while the proton of Ser205 at the other end of the proton wire leads the transfer.

Dynamical corrections computed either through the harmonic oscillator approximation or, where possible, through quantum dynamical wave packet relaxation dynamics reveal that the shallow minimum in S_1 in the region of reactants exists only at a purely topological (that is, potential energy) level and that the process is predicted to occur in an essentially barrierless manner in the excited state.

Acknowledgment. The authors are grateful for financial support from the "Ministerio de Educación y Ciencia" and the "Fondo Europeo de Desarrollo Regional" (project CTQ2005-07115/BQU) and from the "Generalitat de Catalunya" (project 2005SGR00400). O.V. acknowledges the European Commission for a Marie Curie individual fellowship.

Supporting Information Available: Table containing the energy of the 338 structures that have been used in the fit and a text file containing the geometrical parameters and energies used in the fit. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Marx, D.; Tuckerman, M.; Hutter, J.; Parrinello, M. *Nature* **1999**, *397*, 601–604.

(2) Warshel, A. In *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley and Sons: New York, U.S.A., 1991; pp 1−236.

(3) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523–2544.

(4) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.

(5) Åqvist, J.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 7–14.

(6) Pomes, R.; Roux, B. *Biophys. J.* **1996**, *71*, 19–39.

(7) Mei, H. S.; Tuckerman, M. E.; Sagnella, D. E.; Klein, M. L. *J. Phys. Chem. B* **1998**, *102*, 10446–10458.

(8) Decornez, H.; Drukker, K.; Hammes-Schiffer, S. *J. Phys. Chem. A* **1999**, *103*, 2891–2898.

(9) Brewer, M. L.; Schmitt, U. W.; Voth, G. A. *Biophys. J.* **2000**, *78*, 331A–331A.

(10) Brewer, M. L.; Schmitt, U. W.; Voth, G. A. *Biophys. J.* **2001**, *80*, 1691–1702.

(11) Pomes, R.; Roux, B. *Biophys. J.* **2002**, *82*, 2304–2316.

(12) Burykin, A.; Warshel, A. *Biophys. J.* **2003**, *85*, 3696–3706.

(13) Mann, D. J.; Halls, M. D. *Phys. Rev. Lett.* **2003**, *90*, 195503.

(14) Tanner, C.; Manca, C.; Leutwyler, S. *Science* **2003**, *302*, 1736–1739.

(15) Manca, C.; Tanner, C.; Coussan, S.; Bach, A.; Leutwyler, S. *J. Chem. Phys.* **2004**, *121*, 2578–2590.

(16) Braun-Sand, S.; Strajbl, M.; Warshel, A. *Biophys. J.* **2004**, *87*, 2221–2239.

(17) Agmon, N. *Biophys. J.* **2005**, *88*, 2452–2461.

(18) Vendrell, O.; Gelabert, R.; Moreno, M.; Lluch, J. M. *J. Am. Chem. Soc.* **2006**, *128*, 3564–3574.

(19) Kato, M.; Pisliakov, A. V.; Warshel, A. *Proteins* **2006**, *64*, 829–844.

(20) Riccardi, D.; König, P.; Guo, H.; Cui, Q. *Biochemistry* **2008**, *47*, 2369–2378.

(21) Boyer, P. D. *Annu. Rev. Biochem.* **1997**, *66*, 717–749.

(22) Pinto, L. H.; Lamb, R. A. *Photochem. Photobiol. Sci.* **2006**, *5*, 629–632.

(23) Pinto, L. H.; Lamb, R. A. *J. Biol. Chem.* **2006**, *281*, 8997–9000.

(24) Schumacher, K. *Curr. Opin. Plant Biol.* **2006**, *9*, 595–600.

(25) Busenlehner, L. S.; Salomonsson, L.; Brzezinski, P.; Armstrong, R. N. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15398–15403.

(26) Granucci, G.; Hynes, J. T.; Millié, P.; Tran-Thi, T.-H. *J. Am. Chem. Soc.* **2000**, *122*, 12243–12253.

(27) Domcke, W.; Sobolewski, A. L. *Science* **2003**, *302*, 1693–1694.

(28) Mohammed, O. F.; Pines, D.; Dreyer, J.; Pines, E.; Nibbering, E. T. J. *Science* **2005**, *310*, 83–86.

(29) Gelabert, R.; Moreno, M.; Lluch, J. M. *Chem. Phys. Chem.* **2004**, *5*, 1372–1378.

(30) Ortiz-Sánchez, J. M.; Gelabert, R.; Moreno, M.; Lluch, J. M. *J. Phys. Chem. A* **2006**, *110*, 4649–4656.

(31) Ortiz-Sánchez, J. M.; Gelabert, R.; Moreno, M.; Lluch, J. M. *Chem. Phys. Chem.* **2007**, *8*, 1199–1206.

(32) Ortiz-Sánchez, J. M.; Gelabert, R.; Moreno, M.; Lluch, J. M. *J. Chem. Phys.* **2007**, *127*, 084318.

(33) Vendrell, O.; Moreno, M.; Lluch, J. M.; Hammes-Schiffer, S. *J. Phys. Chem. B* **2004**, *108*, 6616–6623.

(34) Luecke, H.; Schobert, B.; Cartailler, J. P.; Richter, H. T.; Rosengarth, A.; Needleman, R.; Lanyi, J. K. *J. Mol. Biol.* **2000**, *300*, 1237–1255.

(35) Zimmer, M. *Chem. Rev.* **2002**, *102*, 759–781.

(36) Schäfer, L. V.; Groenhof, G.; Klingen, A. R.; Ullmann, G. M.; Boggio-Pasqua, M.; Robb, M. A.; Grubmüller, H. *Angew. Chem., Int. Ed.* **2007**, *119*, 536–542.

(37) Balint-Kurti, G. G.; Dixon, R. N.; Marston, C. C. *J. Chem. Soc., Faraday Trans.* **1990**, *86*, 1741–1749.

(38) Tuckerman, M. E.; Marx, D.; Klein, M. L.; Parrinello, M. *J. Chem. Phys.* **1996**, *104*, 5579–5588.

(39) Marx, D.; Hutter, J. Chapter 13. In *Ab Initio Molecular Dynamics: Theory and Implementation*; Grotendorst, J., Ed.; John von Neumann Institut für Computing: Jülich, Germany, 2000; pp 329−477.

(40) Ojamäe, L.; Shavitt, I.; Singer, S. J. *J. Chem. Phys.* **1998**, *109*, 5547–5564.

(41) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.

(42) Gao, J. L.; Garcia-Viloca, M.; Poulsen, T. D.; Mo, Y. R. *Adv. Phys. Org. Chem.* **2003**, *38*, 161–181.

(43) Mo, Y. R.; Gao, J. L. *J. Comput. Chem.* **2000**, *21*, 1458–1469.

(44) Mo, Y. R.; Gao, J. L. *J. Phys. Chem. A* **2000**, *104*, 3012–3020.

(45) Meyer, H.-D.; Worth, G. A. *Theor. Chim. Acta* **2003**, *109*, 251–267.

(46) Vendrell, O.; Moreno, M.; Lluch, J. M. *J. Chem. Phys.* **2002**, *117*, 7525–7533.

(47) Vendrell, O.; Meyer, H.-D. *J. Chem. Phys.* **2005**, *122*, 104505.

(48) Ward, W. W.; Cody, C. W.; Hart, R. C.; Cormier, M. J. *Photochem. Photobiol.* **1980**, *31*, 611–615.

(49) Morise, H.; Shimomura, O.; Johnson, F. H.; Winant, J. *Biochemistry* **1974**, *13*, 2656–2662.

(50) Kummer, A. D.; Kompa, C.; Lossau, H.; Pollinger-Dammer, F.; Michel-Beyerle, M. E.; Silva, C. M.; Bylina, E. J.; Coleman, W. J.; Yang, M. M.; Youvan, D. C. *Chem. Phys.* **1998**, *237*, 183–193.

(51) Shimomura, O. *FEBS Lett.* **1979**, *104*, 220–222.

(52) Cody, C. W.; Prasher, D. C.; Westler, W. M.; Prendergast, F. G.; Ward, W. W. *Biochemistry* **1993**, *32*, 1212–1218.

(53) Niwa, H.; Inouye, S.; Hirano, T.; Matsuno, T.; Kojima, S.; Kubota, M.; Ohashi, M.; Tsuji, F. I. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13617–13622.

(54) Yang, F.; Moss, L. G.; Phillips, G. N. *Nat. Biotechnol.* **1996**, *14*, 1246–1251.

(55) van Thor, J. J.; Zanetti, G.; Ronayne, K. L.; Towrie, M. *J. Phys. Chem. B* **2005**, *109*, 16099–16108.

(56) Stoner-Ma, D.; Jaye, A. A.; Matousek, P.; Towrie, M.; Meech, S. R.; Tonge, P. J. *J. Am. Chem. Soc.* **2005**, *127*, 2864–2865.

(57) Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormö, M.; Remington, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2306–2311.

(58) Vendrell, O.; Gelabert, R.; Moreno, M.; Lluch, J. M. *Chem. Phys. Lett.* **2004**, *396*, 202–207.

(59) Vendrell, O.; Gelabert, R.; Moreno, M.; Lluch, J. M. *J. Phys. Chem. B* **2008,** *112*, 5500–5511.

(60) Cembran, A.; Gao, J. *Theor. Chem. Acc.* **2007**, *118*, 211–218.

(61) Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. B* **1998**, *102*, 5547–5551.

(62) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. Minimization or Maximization of Functions. In *Numerical Recipes in Fortran*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 1992; pp 387−448.

(63) Meyer, H.-D.; Manthe, U.; Cederbaum, L. S. *Chem. Phys. Lett.* **1990**, *165*, 73–78.

(64) Manthe, U.; Meyer, H.-D.; Cederbaum, L. S. *J. Chem. Phys.* **1992**, *97*, 3199–3213.

(65) Beck, M. H.; Jäckle, A.; Worth, G. A.; Meyer, H.-D. *Phys. Rep.* **2000**, *324*, 1–105.

(66) Meyer, H.-D.; Worth, G. A. *Theor. Chem. Acc.* **2003**, *109*, 251–267.

(67) Worth, G. A.; Beck, M. H.; Jäckle, A.; Meyer, H.-D. The MCTDH Package, Version 8.2 (2000). H.-D. Meyer, Version 8.3 (2002). See http://www.pci.uni-heidelberg.de/tc/usr/mctdh (accessed Apr 18, 2008).

CT800075W

J. Chem. Theory Comput. **2008,** *4,* 1151–1161

JCTC Journal of Chemical Theory and Computation

The pDynamo Program for Molecular Simulations using Hybrid Quantum Chemical and Molecular Mechanical Potentials

Martin J. Field*

Laboratoire de Dynamique Moléculaire Institut de Biologie Structurale — Jean-Pierre Ebel (CEA/CNRS/UJF — UMR 9075), 41 Rue Jules Horowitz, F - 38027 Grenoble, Cedex 01, France

Received March 19, 2008

Abstract: The pDynamo program has been developed for the simulation of molecular systems using hybrid quantum chemical (QC) and molecular mechanical (MM) potentials. pDynamo is written in a mixture of the computer languages Python and C and is a successor to the previous version of Dynamo, now denoted fDynamo, that was written in Fortran 90 (*J. Comput. Chem.* **2000**, *21*, 1088). The current version of Dynamo has a similar range of functionality to the older one but extends it in some significant ways, including the addition of a density functional theory QC capability. This paper gives a general description of pDynamo and outlines some of the advantages and disadvantages that have been encountered in switching computer languages. Some technical aspects of the implementation of pDynamo's algorithms are also discussed and illustrated with the results of example calculations. pDynamo is available on the Web at the address http://www.pdynamo.org and is released under the CeCILL license which is equivalent to the GNU general public license but conforms to the principles of French law.

1. Introduction

Hybrid quantum chemical (QC) and molecular mechanical (MM) potentials, which were first introduced by Warshel and Levitt in 1976[1] in their study of the catalytic mechanism of the enzyme, lysozyme, have become standard tools for the investigation of reaction processes in condensed phase systems. As witness to their utility are the increasing number of modeling programs, both academic and commercial, that implement hybrid potentials and the large number of papers published that employ them (see refs 2 and 3 for recent reviews).

In spite, or because, of this success, hybrid potential techniques are the target of intense research which aims at improving their precision and speed as well as the range of problems to which they can be applied. Hybrid potentials also represent a well-studied and relatively well-understood example of a multiscale approach, in which methods that operate at different scales—in this case electrons and atoms—are coupled together and which are gaining in importance in many areas of computational science.

For many years, our laboratory has been using hybrid potentials to study the mechanisms of enzyme reactions. All our work in this area is done with a program, Dynamo, that we have written and which incorporates all the hybrid potential methods that we have developed and with which all our hybrid potential simulations are performed. Our policy has been to make Dynamo freely available, with source code, in the hope that it could prove useful to the wider scientific community.

The aim of this paper is to introduce a new version of our program, called pDynamo, that is written in Python[4] and C, and which succeeds the previous version, called fDynamo, that was written in Fortran 90.[5] The outline of this paper is as follows. Section 2 gives a general description of the pDynamo program, section 3 discusses, with examples, certain aspects of its implementation, and section 4 summarizes.

---

* Corresponding author phone: (33)-4-38-78-95-94; fax: (33)-4-38-78-54-94; e-mail: martin.field@ibs.fr.

10.1021/ct800092p CCC: $40.75

2. The pDynamo Program

A recent monograph[6] that introduces some of the concepts of molecular simulation employs pDynamo for its example programs and explains many of its features. To avoid duplicating what was said there, we shall restrict ourselves in this paper to a brief overview of the pDynamo program and shall present some of the more technical aspects of the algorithms it implements. Interested readers seeking more should refer to ref 6.

The current version of pDynamo may be downloaded from the pDynamo Web site[7] and is released under the CeCILL license.[8] This is similar in many respects to the GNU general public license[9] but has been drafted to conform to the principles of French law.

2.1. Capabilities of the Program. pDynamo reproduces and extends the functionality of its predecessor fDynamo. Both programs permit the following general classes of calculation or operation:

- Calculation of the potential energy of a system using QC, MM, and hybrid QC/MM potentials. fDynamo employed the all-atom OPLS force field[10] as its MM potential and the semiempirical AM1, MNDO, and PM3[11–13] methods as its QC potentials.
- Geometry optimization of a system to locate both minima and saddle points.
- Reaction path determinations using a variety of methods, including the nudged-elastic-band method.[14]
- Normal mode analysis.
- Classical molecular dynamics simulations for aperiodic (vacuum) systems or for periodic systems in a variety of thermodynamical ensembles.
- Monte Carlo calculations in the canonical and isobaric−isothermal ensembles but limited to simulations of rigid molecules using MM potentials.
- Miscellaneous tools that enable the imposition of constraints, the manipulation of coordinates, and the analysis of simulation data.

In addition, pDynamo introduces some new features, including

- An easier and more powerful interface thanks to the use of the Python computer language.[4]
- A density functional theory (DFT) QC method which can be used alone or as part of a QC/MM hybrid potential.
- An increased range of semiempirical methods, including PDDG[15] and RM1.[16]
- The option of using the AMBER[17] and CHARMM[18] force fields in addition to OPLS.
- The possibility of performing geometry optimizations or molecular dynamics and Monte Carlo simulation in systems with arbitrary crystal symmetry and without the restriction of the minimum image convention.
- The ability to read, write, or otherwise transform molecular data between a larger range of standard molecular formats, including MDL MOL format[19] and SMILES.[20]

Two features currently missing from pDynamo that were present in fDynamo are the abilities to calculate analytical second derivatives for MM potentials and to perform path-integral molecular dynamics simulations.

2.2. General Structure of the Program. pDynamo is structured as a collection of modules and packages written in Python,[4] which is a dynamic, high-level, scripting language that is becoming increasingly popular in scientific computing due to its power and ease of use. pDynamo provides no interface language of its own, and so users perform pDynamo calculations by writing Python programs. Although users only ever see Python, many of the most computationally expensive parts of pDynamo's algorithms are implemented directly in C. The link between the Python and C portions of the code is effected with the glue language Pyrex.[21]

Currently pDynamo has about 140 Python modules made up of approximately 65000 lines of computer code (excluding third-party libraries) of which 35% are Python, 15% Pyrex, and 50% C. The modules are divided into four groups. Three of these are Python packages which in order of precedence are:

- **pCore** is the most fundamental package in the pDynamo program and contains modules that cope with basic programming and mathematical tasks, such as file handling, linear algebra, and the optimization of functions. The modules in this package have been designed to be independent of molecular simulation and could, if necessary, be employed in other Python programs.
- **pDynamo** has modules that deal with the definition of a molecular system and the calculation of its potential energy using either QC, MM, or hybrid potential approaches.
- **pBabel** defines the modules that handle the input, output, and transformation of molecular data between various standard representations or file formats.[22]

The fourth group of modules, pMoleculeScripts, does not comprise a package but consists of a series of miscellaneous Python scripts which carry out various common tasks in molecular simulation, such as geometry optimization, normal-mode analysis, and molecular dynamics simulation. Modules in pMoleculeScripts employ items from all of the packages described above.

Python is not a pure object-oriented language but it permits object-oriented programming, and it is this style which pDynamo employs. All items in pDynamo are organized into classes, and operations are performed by setting the attributes or invoking the methods of class instances. Although elegant, this approach can sometimes be cumbersome, and so it is convenient for certain classes to define "helper" functions that carry out particular operations without having to explicitly instantiate them.

To get a flavor of what this means and also of the flexibility that Python provides consider the pDynamo program shown in Figure 1. This calculates the energy and various other properties of a water molecule using the AM1, MNDO, and PM3 semiempirical QC methods. The salient points of this program are as follows, noting that all blank lines and lines starting with a hash character, #, are ignored:

- **Line 1** is a Python documentation line that gives a brief description of what the program does.
- **Lines 3−5** import items from the various pDynamo packages that are to be used later in the program.

```
 1  """Example water calculation."""
 2
 3  from pBabel  import MOLFile_ToSystem
 4  from pCore   import logfile
 5  from pDynamo import QCModelMNDO
 6
 7  # . Define the energy models.
 8  energymodels = [ QCModelMNDO ( "am1"  ), \
                     QCModelMNDO ( "mndo" ), \
                     QCModelMNDO ( "pm3"  )  ]
 9
10  # . Loop over the energy models.
11  results = []
12  for model in energymodels:
13      molecule = MOLFile_ToSystem ( "water.mol" )
14      molecule.DefineQCModel ( model )
15      molecule.Summary ( )
16      energy  = molecule.Energy ( )
17      charges = molecule.AtomicCharges ( )
18      dipole  = molecule.DipoleMoment  ( )
19      results.append ( ( model.label, energy, charges, dipole.Norm2 ( ) ) )
20
21  # . Output the results.
22  table = logfile.GetTable ( columns = [ 10, 20, 20, 20, 20, 20 ] )
23  table.Start  ( )
24  table.Title  ( "Energy Model Results for Water" )
25  table.Heading ( "Model"  )
26  table.Heading ( "Energy" )
27  table.Heading ( "Charges", columnspan = 3 )
28  table.Heading ( "Dipole" )
29  for ( label, energy, charges, dipole ) in results:
30      table.Entry ( label )
31      table.Entry ( "%.1f" % ( energy, ) )
32      for charge in charges: table.Entry ( "%.3f" % ( charge, ) )
33      table.Entry ( "%.3f" % ( dipole, ) )
34  table.Stop ( )
```

Figure 1. A pDynamo Python program for calculating the energy of a water molecule using the AM1, MNDO, and PM3 semiempirical methods.

- **Line 8** defines a Python list containing the QC energy models which are to be employed in the calculation. Each item in the list is created by instantiating the class QCModelMNDO with, as argument, a string that indicates the appropriate MNDO Hamiltonian to use.

- **Lines 12−19** form a loop in which the different energy models are iterated over. Line 12 initiates the loop by extracting, at each iteration, a model from the list of energy models, whereas lines 13−19 form the loop's body. In contrast to other languages, Python employs indentation to indicate which lines are to be iterated over, a return to the original indentation level (in this case line 22) denoting an end to the loop's scope.

- **Line 13** reads a MOL file of a water molecule using the MOLFile_ToSystem helper function of the MOLFileReader class. This function instantiates the class, opens, parses, and closes the file, and then returns the file's data as an instance of the class System which, in this case, is assigned to the variable molecule. System is in many respects the most important pDynamo class as its instances represent molecular systems that are to be simulated.

- **Lines 14−15** invoke methods of molecule to define its QC energy model and to output a concise summary of its composition.

- **Lines 16−19** calculate the potential energy, the atomic charges (using a Löwdin analysis),[23] and the dipole moment vector for molecule. The results, along with the name of the energy model (model.label), are saved on line 19 in the list called results which was created before entry to the loop on line 11. Note that the entire dipole moment vector is not saved but only its magnitude which is determined by invoking the vector's Norm2 method.

- **Lines 22−34** output the results of the calculation to a table. Line 22 creates the table by invoking the appropriate method of the variable logfile which, roughly speaking, is pDynamo's equivalent of standard output. Lines 23−28 set up the table and its headers whereas lines 29−33 form a loop that unpacks the results for each energy model from results and then prints them.

2.3. Python/C versus Fortran. There were a number of complementary reasons why the choice was made to switch to Python from Fortran for the new version of Dynamo: (i) Python is a high-level, dynamic language, with features such as object-orientation and garbage collection, that greatly simplifies the writing of programs, and (ii) Python is free and open-source and has implementations for many different platforms and operating systems. In constrast, even after more than 15 years, there are still no fully reliable, open-source Fortran 90 compilers, which has complicated the use of fDynamo. (iii) Python has a very active development community and an extensive standard module library that performs a wide range of different tasks. This is an advantage when compared not only to Fortran but also to other scripting languages, such as Ruby.[24] (iv) Running a fDynamo program requires a preliminary compilation step, but this is unnecessary for pDynamo as running a Python program requires no (explicit) compilation step and pDynamo's C extension modules are compiled when the program is installed.

On the downside, Python does have some disadvantages compared to Fortran (or other compiled languages, such as C). One of these is that many errors that would be caught at compilation time in Fortran programs do not appear until run time in Python because it is a dynamically typed language. This behooves users to check thoroughly their scripts on short simulations before running longer ones.

A more serious problem is one of execution speed as Python programs are normally much slower than programs of compiled languages. This drawback can be alleviated by coding the computationally demanding parts of algorithms in a compiled language (C in the case of pDynamo), but the process of linking Python and C code is nontrivial. The use of Pyrex helps greatly in this procedure, but it is yet another language that has to be mastered.[21]

Given that one of the most important roles of pDynamo in our laboratory is as a testbed for trying out new algorithms, Python's flexibility has proved more important than the performance penalty that its use incurs. In general, though, for standard types of calculation we aim to have no more than a factor of 2 speed difference between pDynamo and an equivalent algorithm coded in a compiled language.

3. Specific Topics

In this section, we illustrate with examples some aspects of pDynamo's QC and hybrid QC/MM potentials that have not been presented elsewhere and, in particular, its DFT capability. As yet, we have completed no full enzyme reaction study with pDynamo, but such work is underway and will be reported in due course.

3.1. Density Functional Theory Implementation. Semiempirical QC methods have proved very useful in hybrid potentials, but, clearly, there are situations in which ab initio or DFT methods are necessary, either for validating the results of semiempirical calculations or for treating systems for which semiempirical methods are not very reliable, such as those that contain transition methods. Given the success and widespread use of DFT methods, we chose to incorporate one of these techniques into Dynamo. Initial work was done in fDynamo but was transferred to pDynamo as this version developed.

We considered two approaches for adding a DFT method to Dynamo. The first was to loosely couple Dynamo with an existing stand-alone DFT program and enable communication either through files or a minimal interface. The second was to embed the DFT code into Dynamo in the same way as the code of its existing QC methods. There are advantages with both approaches. The former is easier because changes to each of the coupled programs can be minimized, whereas the latter allows a greater freedom in the implementation of the DFT hybrid potential. In the end we adopted the second approach because, when we started this work, we were unable to find a reliable, open-source DFT program that was appropriate for our needs, and so we decided to write our own. Subsequently, we have also employed the first approach, but we shall leave a discussion of this until section 3.3.

For our initial DFT implementation, we selected a method based upon Gaussian basis functions, as opposed to plane-waves or other types of basis function, as we deem Gaussians more suitable for the types of hybrid potential that we use and the systems that we study. Currently, pDynamo's DFT module has the following features:

- Standard local and nonlocal (or generalized gradient approximation) functionals are supported but only those that do not include Hartree−Fock exchange.[25] This means, for example, that BLYP is present but not B3LYP.[26]
- A Coulomb-fitting (or RI) approximation is employed to calculate the electron−electron repulsion terms in the Kohn−Sham equations.[27,28] All other terms are evaluated directly with the orbital basis. The Coulomb-fitting basis, like the orbital basis, uses Gaussians.
- Mixed Cartesian and real spherical-harmonic Gaussian basis sets can be treated for both the orbital and Coulomb-fitting functions. At present, integrals and their first derivatives may be calculated if these involve functions of *g* angular momentum or less, although extension to functions of higher angular momentum would be straightforward. pDynamo is distributed with a number of standard orbital and Coulomb-fitting bases.
- All integrals are evaluated at the start of each energy calculation and stored in memory. This improves speed but limits the size of calculation that can be done on a machine with 2 Gb of memory to one with approximately 400 orbital and 1000 Coulomb-fitting functions, respectively. It is possible that this restriction will be removed in the future by going to a direct scheme, in which integrals are re-evaluated when they are needed, or to one that uses external files.
- The DFT correlation and exchange terms are evaluated using Euler-Maclaurin radial quadrature and Lebedev angular integration on atom-centered grids combined using Becke weights.[29,30] A variety of grids is provided which give integrations of different accuracy.

The DFT implementation is reasonably efficient and, most importantly, is done in such a way that whenever a semiempirical QC calculation is performed, a DFT calculation may be carried out instead. Thus, for example, in the example of Figure 1 it would have been possible to add an extra item to the energy models list, defined on line 8, that corresponded to a DFT QC energy model with a particular functional, integration grid and orbital and Coulomb-fitting basis sets.

3.2. The Hybrid Potential. pDynamo's hybrid potential mirrors, in most respects, that of fDynamo[5] and follows closely the one developed by Field et al.[31] In brief, the potential partitions the atoms in a system into two regions, one QC and the other MM. The total potential energy of the system, E, can then be written as a sum of three terms—one for the atoms in the QC region, one for those in the MM region, and a term which describes the interactions between the two—as follows:

$$E[\Psi|\rho] = E_{\mathrm{QC}}[\Psi|\rho] + E_{\mathrm{MM}} + E_{\mathrm{QC/MM}}[\Psi|\rho] \qquad (1)$$

The energy expression for the atoms in the QC region, E_{QC}, takes the form that is appropriate for the QC method being used and will be a function(al) of the system's wave function, Ψ, for a molecular orbital method, or its density, ρ, for a DFT method. Likewise, the energy of the MM region, E_{MM}, is evaluated in the way appropriate for the force field being used. This term does not depend on the QC electronic variables, Ψ or ρ, because all the force fields currently available in pDynamo are nonpolarizable.

The QC/MM interaction energy, $E_{\mathrm{QC/MM}}$, consists of the following two terms:

$$E_{\mathrm{QC/MM}}[\Psi|\rho] = E^{\mathrm{el}}_{\mathrm{QC/MM}}[\Psi|\rho] + E^{\mathrm{LJ}}_{\mathrm{QC/MM}} \qquad (2)$$

The first term, $E^{\mathrm{el}}_{\mathrm{QC/MM}}$, is for the electrostatic interactions between the atoms of the QC and MM regions and will be described in detail in the next section as it is here that some changes have been made. The second term, $E^{\mathrm{LJ}}_{\mathrm{QC/MM}}$, is the Lennard-Jones interaction energy between the atoms of the two regions and is calculated in the usual way for the MM method.

Once the energy expression of eq 1 has been defined, it can be used to set up the equations that must be solved to obtain the wave function or electron density of the atoms in the QC region in the standard way. This gives rise to the Roothaan-Hall and Kohn−Sham equations for molecular orbital and DFT methods, respectively.[6,25] In practice, pDynamo employs a single spin-restricted or spin-unrestricted determinant to represent the electronic wave function or density for the system and solves the Roothaan-Hall or Kohn−Sham equations using a self-consistent field (SCF) technique, iterated to convergence, every time the geometry of the system changes or, in other words, at each step of a geometry optimization or of a molecular dynamics or Monte Carlo calculation. It is important to emphasize that the optimized wave function or density will be polarized by the MM environment because the electrostatic QC/MM interactions are included in the SCF procedure.

In those cases where there are covalent bonds between QC and MM atoms, pDynamo employs a link-atom technique which is identical, except for the electrostatic interactions, to that of fDynamo.[5] In this scheme, all MM atoms that have covalent bonds to QC atoms are marked as "boundary" atoms. Each boundary atom has two components—the MM atom, with its associated parameters, which represents the boundary atom in the MM calculation, and a hydrogen link-atom which enters into the QC calculation. The link-atom does not have an independent position of its own because its coordinates are constructed as an analytic function of the coordinates of its parent MM and QC atoms each time an energy is calculated. This is done by placing the link-atom along the QC-MM bond at a suitable distance (~1 Å) from the QC atom. The gradients of the energy with respect to the link-atom coordinates, if these are calculated, are easily partitioned between the parent MM and QC atoms by a straightforward application of the chain rule. The identification and handling of boundary and link-atoms is handled automatically by pDynamo and requires no intervention by the user. Although link-atom methods have been criticized, they are robust and easy to implement and, in our experience, are not consistently less accurate than alternative approaches.

3.2.1. Electrostatic Interactions. The principle change in the hybrid potential between pDynamo and fDynamo lies in the treatment of the electrostatic QC/MM interactions. In fDynamo, the QC/MM electrostatic energy was calculated with the following formula

$$E^{\mathrm{el}}_{\mathrm{QC/MM}} = \sum_m Q_m \left\{ \sum_q Z_q f_{\mathrm{nucleus}}(\boldsymbol{r}_q, \boldsymbol{r}_m) - \int d\boldsymbol{r}\,\rho(r) f_{\mathrm{electron}}(\boldsymbol{r}, \boldsymbol{r}_m) \right\} \qquad (3)$$

where ρ is the electron density of the QC atoms expressed in terms of the QC density matrix and products of orbital basis functions, $\boldsymbol{r}$ is a position vector, Z is a nuclear charge, Q is a partial charge, and the subscripts q and m refer to QC and MM atoms, respectively. The functions f_{nucleus} and f_{electron} determine the distance-dependence of the electrostatic interaction. For a DFT method they are both equal to the inverse distance (i.e., $1/|\mathbf{r} - \mathbf{r}_m|$), whereas for the semiempirical methods they have more complicated forms.

This expression is, of course, the natural, exact equation for the interaction energy between a set of point charges and a charge distribution consisting of a mixture of point charges (nuclei) and continuous densities (electrons). However, when extending our hybrid potentials with a DFT method, the question arose as to whether this is necessarily the best approach. First, the method is very expensive if there are thousands of MM atoms as is typically the case for the systems we study. Second, there is a great disparity in the representations of the QC and MM atom charge distributions as it is very detailed for the QC atoms and much more primitive for the MM atoms. This being so, is it worth calculating the interaction to such a high precision? Finally, there is the 'charge-sloshing' effect. In a pure QC calculation, the charge density of an atom is restricted by interaction with the charge densities of the QC atoms around it. In a hybrid potential (at least at the QC/MM boundary), charge can leak from the QC region and accumulate unphysically in the MM region. This is not so much of a problem for semiempirical QC methods because their basis functions are very localized but becomes one for ab initio and DFT calculations with

Gaussian basis sets, especially as the size and diffusiveness of the basis increases.

This led us to consider the following simplified expression for the interaction energy which is appropriate for both semiempirical and DFT hybrid potentials:

$$E^{\mathrm{el}}_{\mathrm{QC/MM}} = \sum_m \sum_q Q_m \widetilde{Q}_q[\rho] f_{\mathrm{charge}}(\boldsymbol{r}_m, \boldsymbol{r}_q) \qquad (4)$$

In this equation f_{charge} is the function that determines the distance-dependence of the electrostatic interaction and $\tilde{Q}_q$ is an effective charge for the QC atom q derived from the atom's nuclear charge and a population analysis of the electron density, ρ. Note that the charges, $\tilde{Q}_q$, are redetermined at each iteration of an energy calculation's SCF procedure, so that they reflect the changes that are occurring in the electron density as it converges.

Equation 4 addresses the three concerns cited above as follows: (i) it is cheaper to calculate, because population analysis is normally much less expensive than integral calculation even if it has to be done at each step of an SCF calculation; (ii) the expression treats the charge distributions of the atoms in the QC and MM atoms in a much more symmetrical way; and (iii) "charge-sloshing" can no longer occur because the QC atom charges are localized on their nuclei. The adoption of eq 4 also has the important advantage that many of the efficient methods that exist for the treatment of long-range interactions in MM calculations, whether these be truncation or Ewald techniques, can be adapted straightforwardly to the hybrid potential case. It is our intention to do this once we have confirmed to ourselves the robustness of the point charge approach.

Clearly, the accuracy of eq 4 depends crucially upon the form of the function f_{charge} and what type of population analysis is used to determine the charges $\tilde{Q}_q$. We have examined a large number of different approaches, and our work indicates that eq 4 can give results of an accuracy comparable to that of eq 3. We have also looked at the best way of calculating the electrostatic interactions for link atoms with this representation. An article on this research is in preparation and should be submitted shortly.[32] It is worth noting that although the current version of pDynamo employs eq 4 to describe the QC/MM electrostatic interactions, it is not excluded that future releases will also incorporate methods based upon eq 3 if these are deemed necessary.

3.2.2. Example Calculations. To give an illustration of the capabilities of pDynamo's hybrid potentials, we describe in this section calculations on two different types of periodic system, the first a cubic box of 215 water molecules solvating a methane molecule and the second a series of three molecular, organic crystals with a mixture of space groups. In each case, the calculations were repeated with a pure MM potential and with both semiempirical and DFT QC/MM hybrid potentials. The OPLS-AA force field[10] with the TIP3P water model[33] was taken for the MM potential, the AM1 method[11] for the semiempirical potential and, for the DFT method, the BLYP functional[26] with the 3−21G orbital[34] and DeMon Coulomb-fitting[35] basis sets.

All nonbonding (electrostatic and Lennard-Jones) interactions between MM atoms and between QC and MM atoms were treated with the atom-based, force-switching (ABFS) method[36] with inner and outer cutoffs of 8 and 12 Å, respectively. This means that the f_{charge} function in eq 4 was the one appropriate for ABFS electrostatic interactions. Effective charges for the QC atoms in these interactions were derived using a Löwdin population analysis of the QC density matrix at each iteration of the SCF calculation. The derivatives of the energy expression of eq 4 with Löwdin charges are straightforward to determine using standard techniques, both with respect to density matrix elements (required for construction of the Fock and Kohn−Sham matrices) and atomic coordinates (needed for calculation of the atomic forces).[37]

As stated in section 2.1, fDynamo required that the minimum image convention held when treating periodic systems and, in addition, that the nonbonding interaction cutoff was sufficiently large that QC atoms in the central, periodic image did not interact with QC atoms in neighboring images. In pDynamo this is no longer the case, and so it is common (depending upon nonbonding cutoff and system size) to have QC atoms of different images interact. These situations are straightforwardly handled by modifying the energy expression of eq 4 to include interactions of the form $\tilde{Q}_q\tilde{Q}_{q'}\, f_{\mathrm{charge}}\,(\mathbf{r}_{q'}, \mathbf{r}_{q'})$ where the subscripts q and q' refer to QC atoms in different images. This modification also has the consequence that it is possible to perform QC/MM calculations in which all atoms in the central image are defined to be in the QC region and there are no MM atoms at all! In these cases, the only manifestation of the MM potential is in the determination of the Lennard-Jones interactions between the (QC) atoms of the central image and their (MM) copies in the surrounding images. We shall see examples of this type of calculation later.

As a first illustration, we simulated a system consisting of a methane molecule solvated by 215 waters in a cubic box of approximate dimension 18.8 Å. Calculations were done with a pure MM potential and with hybrid potentials in which the atoms of methane were treated quantum chemically. Molecular dynamics simulations were performed using the constant pressure and temperature leapfrog Verlet algorithm of Berendsen et al.[38] and consisted of a short equilibration run of 10 ps followed by a data collection phase of 100 ps in which coordinate data were saved every 50 fs. Simulations were carried out at 1 atm and 300 K with pressure and temperature coupling constants of 2000 ps atm and 0.1 ps, respectively. The integration time step was 1 fs meaning that the simulations with each potential needed approximately 1.1×10^5 energy and force calculations.

The radial distribution functions (RDFs) between the methane carbon and the water oxygens determined from these simulations are shown in Figure 2. The curves are very similar with the first peak being shortest and broadest for the MM potential and sharpest in the QC(AM1)/MM case. The QC(DFT)/MM curve is intermediate between the two. The number of waters in the first solvation shell of methane may be estimated by integrating the RDFs and choosing the values that correspond to the distances of their first troughs (~5.5 Å). This gives approximately 21 waters for each simulation.

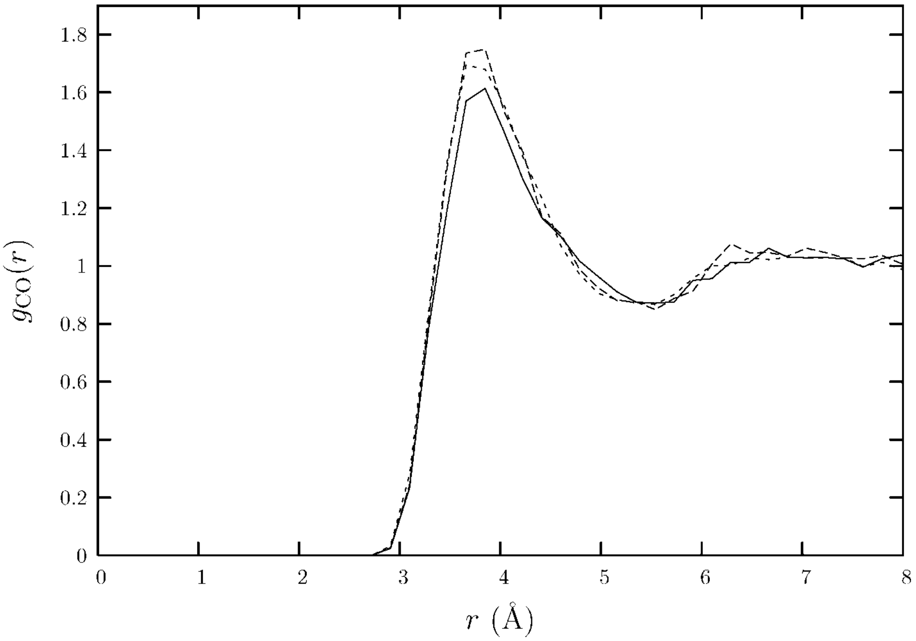


Figure 2. Radial distribution functions, $g_{\mathrm{CO}}(r)$, determined from simulations of a molecule of methane solvated in a cubic box of 215 water molecules with the following potentials: MM — solid line; QC(AM1)/MM — long dashed line; QC(DFT)/MM — short dashed line. In the last two calculations the methane molecule is in the QC region and the waters are in the MM region.

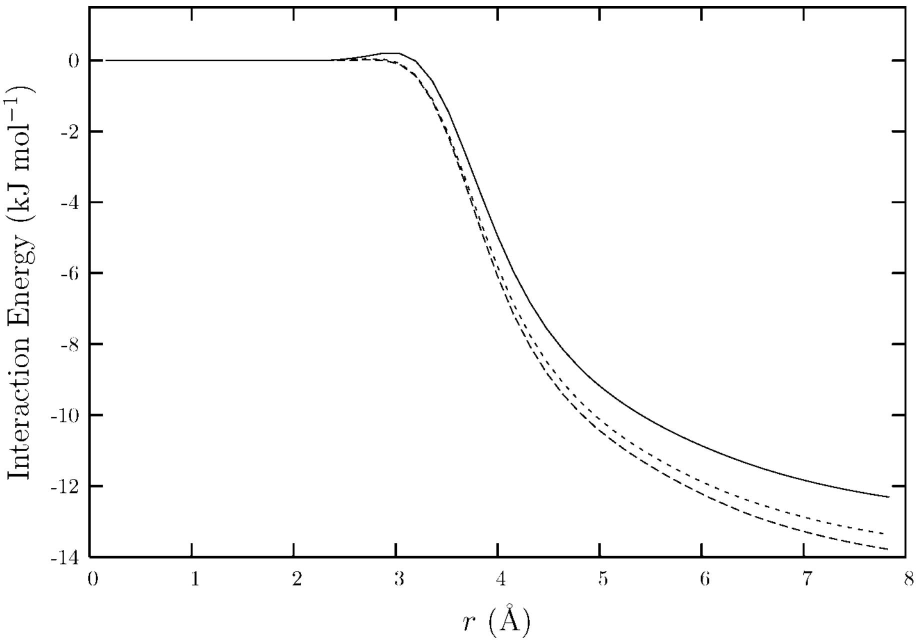


Figure 3. The average energy of interaction between methane and the surrounding water molecules as a function of cutoff distance determined from the simulations with the following potentials: MM — solid line; QC(AM1)/MM — long dashed line; QC(DFT)/MM — short dashed line.

Figure 3 shows the average energies of interaction between methane and the waters as a function of distance. The curves are again very similar although for all distances for which there is an interaction energy, the MM energies are the least stabilizing and the QC(AM1)/MM energies the most. At distances around 3 Å the MM energies are even slightly destabilizing, a feature which is absent from the other two calculations. Because there are 21 water molecules in the first solvation shell of methane in each simulation, methane's average energy of interaction with a water in that shell is approximately -0.5 kJ mol $^{-1}$.

As a final analysis, it is also possible to calculate methane's charge distribution from the hybrid potential simulations. The average Löwdin charges and their standard deviations on the carbon were -0.27 ± 0.02 and -0.37 ± 0.01 in the AM1 and DFT cases, respectively.

The second set of calculations involved a series of three molecular, organic crystals. These were obtained from the

Table 1. Results for the Crystal Geometry Optimization Calculations[a]

crystal	property	MM	AM1/MM	AM1	DFT/MM	DFT
HXACAN19	ΔE	−95.6	−175.8	−193.0	−152.0	−164.1
	RMSCD	0.27	0.18	0.06	0.18	0.13
	a	12.48 (0.39)	12.50 (0.37)	12.46 (0.41)	12.50 (0.37)	12.55 (0.32)
	b	9.17 (0.20)	9.14 (0.23)	9.73 (−0.36)	9.12 (0.25)	9.63 (−0.26)
	c	7.23 (−0.14)	7.25 (−0.17)	7.19 (−0.11)	7.26 (−0.17)	7.22 (−0.14)
	β	115.9 (−0.3)	115.9 (−0.2)	119.46 (−3.8)	115.7 (−0.1)	119.2 (−3.6)
LCDMPP10	ΔE	−172.6	−199.7	−302.2	−193.0	−268.9
	RMSCD	0.10	0.10	0.10	0.11	0.08
	a	7.62 (0.45)	7.62 (0.45)	7.66 (0.41)	7.66 (0.40)	7.61 (0.45)
	b	6.11 (−0.02)	6.10 (−0.01)	5.99 (0.09)	6.11 (−0.03)	6.04 (0.05)
	c	5.34 (−0.18)	5.34 (−0.19)	5.22 (−0.07)	5.33 (−0.18)	5.38 (−0.22)
	α	133.5 (−1.8)	133.6 (−1.9)	131.8 (−0.1)	133.7 (−2.0)	132.4 (−0.7)
	β	79.7 (2.7)	79.9 (2.5)	80.7 (1.7)	79.3 (3.1)	82.0 (0.4)
	γ	105.9 (0.7)	105.8 (0.8)	106.6 (0.0)	106.1 (0.4)	106.5 (0.1)
WABZOO	ΔE	−128.8	−134.9	−144.4	−105.7	−86.3
	RMSCD	0.37	0.20	0.23	0.25	0.34
	a	12.65 (−0.05)	12.64 (−0.05)	12.68 (−0.08)	12.59 (0.00)	12.59 (0.00)
	α	117.9 (0.1)	118.0 (−0.01)	117.9 (0.1)	117.9 (0.1)	117.9 (0.1)

[a] MM, AM1/MM, AM1, DFT/MM, and DFT refer to calculations with the following potentials: MM; QC(AM1)/MM with both QC and MM atoms in the central image; QC(AM1)/MM with no MM atoms in the central image; QC(DFT)/MM with both QC and MM atoms in the central image; and QC(DFT)/MM with no MM atoms in the central image. ΔE and RMSCD are the energy lowering (kJ mol^{-1}) and RMS coordinate difference (Å) between the experimental and geometry optimized structures. The RMSCDs were calculated after superimposing the optimized structure upon the experimental one. a, b, c, α, β, and γ refer to crystal unit cell distances (Å) and angles (°) for the optimized structures with the numbers in brackets giving the difference from the experimental values (experimental minus optimized).

Cambridge Structural Database[39] with codes HXACAN19, LCDMPP10, and WABZOO and correspond to the molecules paracetamol,[40] cyclo-L-alanyl-L-alanyl,[41] and (3aα, 4α,4α,7aα)-2,2-dimethyl-3a,4,4,7a-tetrahydro-1,3-benzodioxolane-4,7-diamine,[42] respectively. The numbers of atoms in the central image and the space groups are 20, 20, and 29 and *P*21*a*, *P*1, and *R*3, respectively.

Geometry optimizations of the crystals were done with a pure MM potential and with semiempirical and DFT QC/MM hybrid potentials in which both some and all of the atoms in the central image were included in the QC region. For those hybrid potential calculations in which there were both QC and MM atoms in the central image, the atoms in the QC region were the methyl group (HXACAN19), one of the two symmetry-equivalent methyls (LCDMPP10) and the five-membered ring along with its two methyl groups (WABZOO). These partitionings mean that there will be link atoms, one for each of HXACAN19 and LCDMPP10 and two for WABZOO. To avoid catastrophic electrostatic interaction between the partial charges of the link atom and its neighboring MM atom, we adopted the redistributed charge (RC) method of Lin and Truhlar[43] for these illustrative calculations. This involves an equal redistribution of the partial charge of the MM atom to the midpoints of the bonds between it and the MM atoms to which it is covalently bound. A full assessment of this and other approximations will be left to our future paper.[32]

Optimizations were carried out with a trust-radius, quasi-Newton algorithm with a starting second derivative matrix calculated with the MM potential at the experimental geometry. All degrees of freedom, molecular internal and crystal, were optimized with the restriction that the experimental space group was maintained. A summary of the results for the calculations is shown in Table 1. All optimizations converged and no significant difference was seen in the convergence behavior between the calculations with different potentials. Each of the potentials produced structures which remained close to the experimental one.

3.3. Coupling to External Programs. As an alternative to pDynamo's native DFT algorithms, it is also possible to employ third party QC programs. This can be advantageous when these programs are more efficient or provide complementary functionality, such as Hartree−Fock, post-Hartree−Fock, or excited-state methods. In this section, we illustrate what is possible by describing the coupling between pDynamo and the program ORCA.[44] This is a free, general purpose, quantum chemical package with a range of ab initio, DFT, and semiempirical QC algorithms that is available in binary form only.

Like many QC programs, ORCA requires a simple text input file in order to run and writes results to a text output file. Additional files are also produced in both text and binary formats that contain extra information about the calculation or quantities that could prove useful in future jobs (such as density matrices). This mode of functioning clearly suggests a model of coupling in which pDynamo writes an ORCA input file, launches the ORCA executable, waits for its termination, and then parses the output and other files for whatever information it requires.

In our current implementation, pDynamo controls all aspects of a simulation, and ORCA is only called when a QC energy and, optionally, its gradients are needed. All geometry optimization and other types of calculation, such as molecular dynamics integration and reaction-path finding, are done with pDynamo's own algorithms. The coupling between pDynamo and ORCA is transparent to users and means that ORCA can be invoked in exactly the same way as, and employed interchangeably with, pDynamo's other, in-built QC algorithms. Thus, in the example of Figure 1, to calculate the energy, atomic charges, and dipole moment of the water molecule with one of ORCA's QC methods, it suffices to instantiate the class corresponding to the ORCA

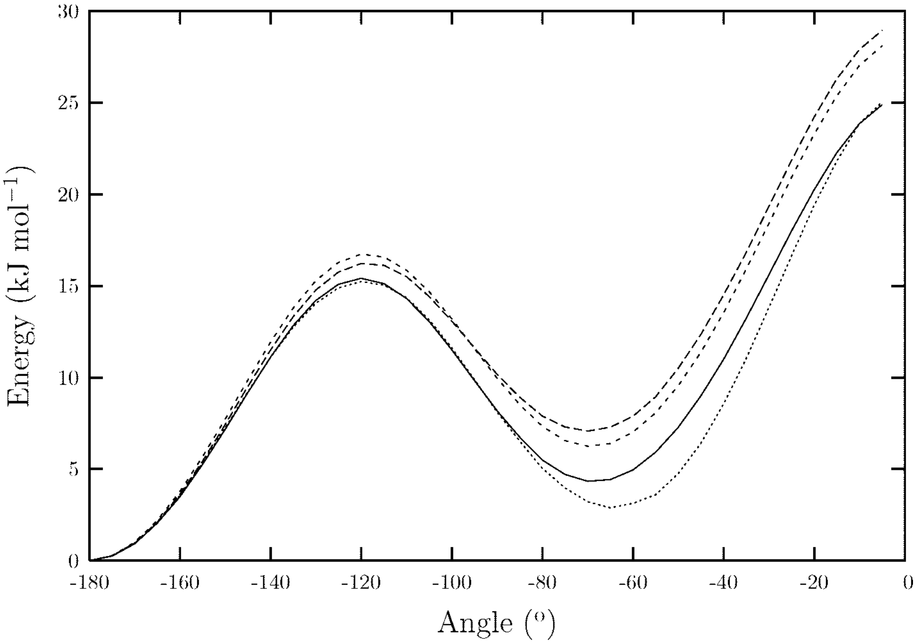


Figure 4. Energy profiles for rotation about the central C−C bond of butane calculated with the following potentials: MM — solid line; QC(AM1)/MM with half the molecule in the QC region — long dashed line; QC(MP2)/MM with half the molecule in the QC region — short dashed line; MP2 — dotted line. Only half the profile is shown as it is symmetrical.

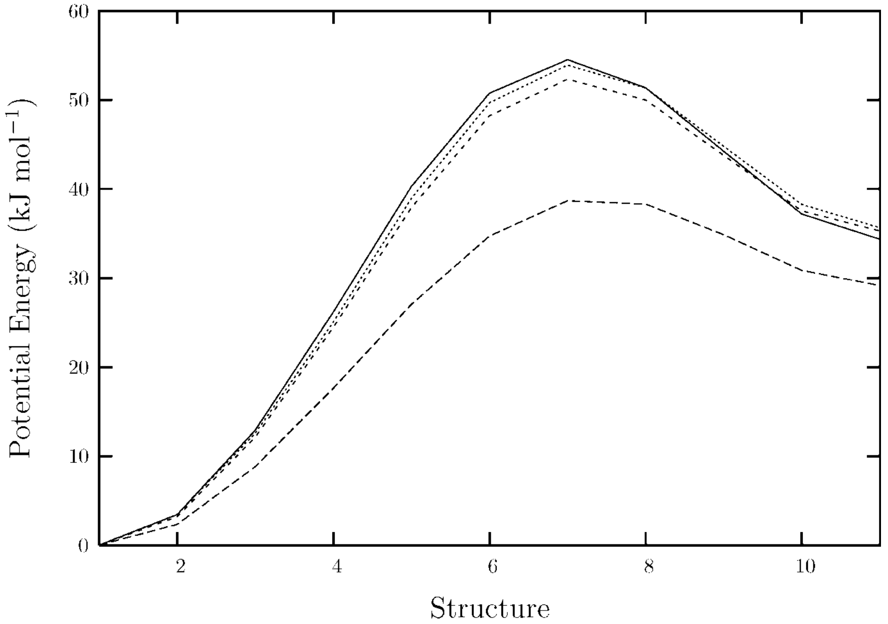


Figure 5. Energy profiles for the cyclohexane chair/twist-boat isomerization calculated with the following potentials: MM — solid line; QC(AM1)/MM — long dashed line; QC(HF)/MM — short dashed line; QC(MP2)/MM — dotted line.

QC model with the appropriate options and then add this to the energy models list.

The ORCA input file written by pDynamo when it invokes ORCA contains all the information necessary to perform an energy and gradient calculation. This includes a definition of the QC atoms and their coordinates along with various keywords that specify quantities such as the QC method, basis sets, and the system's charge and spin multiplicity. ORCA also permits QC calculations in the presence of non-QC point charges, and so, in a hybrid potential calculation, the input file contains the charges and the coordinates of the MM atoms as well.

Once the ORCA calculation is finished, pDynamo parses the appropriate files produced by ORCA during the run. These include the output file from which information, such as the atomic charges, is obtained and another file called the 'engrad' file. The latter contains the total energy of the system and the gradients with respect to the coordinates of the QC and, equally importantly, the non-QC (or MM) atoms. The fact that ORCA calculates these latter gradients is in marked contrast to most other QC programs and greatly simplifies its use as part of a hybrid potential.

The energy calculated by ORCA in the presence of point charges includes not only the interaction energy between the

QC atoms and the point charges but also the interaction energy between the point charges themselves. This means that the only electrostatic energy that pDynamo needs to calculate when using ORCA is that which would normally be excluded between MM atoms, i.e. that between MM atoms that are separated by one, two, and three bonds. This is rapid and straightforward to do and is subtracted from ORCA's total energy. All other energy terms, including the Lennard-Jones interactions involving QC and MM atoms, are handled by pDynamo in its normal fashion.

As a final point in the discussion of the technical aspects of the pDynamo/ORCA coupling, it is worth noting that hybrid potentials employing ORCA will be less flexible than those using pDynamo's in-built QC models because ORCA calculates the QC/MM electrostatic interactions following eq 3, not eq 4. This is time-consuming if there are large numbers of point charges; limits applications to 'vacuum' systems as periodic boundary conditions cannot be treated and may introduce artifactual boundary effects when simulating large systems due to the long-range nature of the electrostatic $1/r$ interaction.

We illustrate the coupling of pDynamo and ORCA with two sets of simple hybrid potential calculation, one that determines the energy profile for rotation about the central C−C bond of butane and the other for the chair/twist-boat isomerization in cyclohexane. In both sets of calculation we employed four potentials consisting of (i) a pure MM model using the OPLS-AA force field;[10] (ii) a QC(AM1)/MM hybrid potential using pDynamo's in-built AM1 method; (iii) a pDynamo/ORCA hybrid potential with a Hartree−Fock method and the 6−31G∗ basis set;[45] and (iv) a pDynamo/ORCA hybrid potential with an MP2 method and the 6−31G∗ basis set. To facilitate comparison between the different potentials all nonbonding interactions calculated by pDynamo were determined exactly (i.e., without truncation or the ABFS approximation), and the electrostatic interactions around link-atoms in ORCA were evaluated using the same RC approximation as described in section 3.2.2.

The energy profiles for rotation about the central C−C bond of butane were obtained by optimizing the butane structure with the appropriate potential while constraining the C−C−C−C torsion angle to have a specific value. Optimizations were done in 5^{o} increments in the range -180^{o} to 0^{o} (as the profiles are symmetric about 0^{o}). In the calculations with the three hybrid potentials, four different partitionings of the molecule were tried with increasing numbers of atoms in the QC region. These were a terminal methyl group, half the molecule, all but a terminal methyl group, and the complete molecule (i.e., equivalent to a pure QC calculation).

A selection of the profiles that were obtained is shown in Figure 4, whereas the full set of profiles are included as Supporting Information. From Figure 4 it can be seen that the pure MM and pure MP2 profiles are very similar although the energy of the MP2 s minimum (at -60^{o}) is slightly lower. Likewise the two hybrid potential profiles, one AM1 and one MP2, agree well although the energy barriers, and the energies of the second minimum are higher than for the MM and MP2 profiles.

As a second example, we calculated energy profiles for the isomerization of the chair to twist-boat forms of cyclohexane using the self-avoiding walk method of Elber et al.[46] In the cases where hybrid potentials were used, half the molecule was placed in the QC region (i.e., three consecutive CH_2 groups). Energy profiles were determined using 11 structures and are shown in Figure 5. It can be seen that the profiles obtained with the MM potential and with the two ab initio hybrid potentials (HF and MP2) agree very well, whereas the QC(AM1)/MM profile has a much lower barrier.

4. Summary

In this paper we have presented a new version of our Dynamo simulation program which is designed to study molecular systems with hybrid QC/MM potentials. The new version, pDynamo, is written in the languages Python and C and succeeds the previous version, fDynamo, which was written in Fortran 90. The switch to Python makes pDynamo much easier to use but has required that parts of the program be coded in C so that its computational efficiency approaches that of the Fortran version. pDynamo supports a similar range of molecular simulation algorithms as fDynamo but has some significant new capability, including the addition of a DFT QC method.

Acknowledgment. The author would like to thank Dr. Konrad Hinsen for introducing him to Python, Margherita Moreno and Dr. Guido Raos for help with the crystal calculations, and Dr. Troy Wymore and his colleagues at Pittsburgh Supercomputing Center for hosting the pDynamo web site. Acknowledgements also go to the Institut de Biologie Structurale − Jean-Pierre Ebel, the Commissariat à l'Energie Atomique, and the Centre National de la Recherche Scientifique for support of this work.

Supporting Information Available: Figures of all the rotational barriers calculated in section 3.3. This material is available free of charge via the Internet at http://pubs.acs.org.

References

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

(2) Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173.

(3) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.

(4) Python Programming Language. http://www.python.org (accessed April 29, 2008).

(5) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088.

(6) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 2007.

(7) Field, M. J. pDynamo Molecular Modeling Program. Hosted by Pittsburgh Supercomputing Center, Pittsburgh, U.S.A. http://www.pdynamo.org (accessed April 29, 2008).

(8) CeCILL, the French Free Software License. http://www.cecill.info (accessed April 29, 2008).

(9) The GNU General Public License. http://www.gnu.org/licenses/gpl.html. (accessed April 29, 2008).

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

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

(12) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

(13) (a) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209. *ibid* **1989**, *10*, 221.

(14) Fernandez-Galván, I.; Field, M. J. *J. Comput. Chem.* **2008**, *29*, 139.

(15) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601.

(16) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. *J. Comput. Chem.* **2006**, *27*, 1101.

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

(18) MacKerell, A. D.; Bashford, D.; Bellott, M., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.,., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J. E.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(19) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244.

(20) Weininger, D. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31.

(21) Pyrex — A Language for Writing Python Extension Modules. http://www.cosc.canterbury.ac.nz/greg.ewing/python/Pyrex (accessed April 29, 2008).

(22) The pBabel package is named after the well-known OpenBabel program but handles by no means as many different file formats, nor is it intended to. To find OpenBabel, see: Open Babel — The Open Source Chemistry ToolBox. http://openbabel.sourceforge.net (acccessed April 29, 2008).

(23) For MNDO semiempirical methods, such as AM1, Löwdin and Mulliken charge analyses are equivalent due to the fact that the overlap matrix is equal to the identity matrix.

(24) Ruby Programming Language. http://www.ruby-lang.org (accessed April 29, 2008).

(25) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: New York, NY, 2000.

(26) (a) BeckeA. D. *Phys. Rev. A* **1988**, *38*, 3098. (b) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648. (c) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(27) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. *J. Chem. Phys.* **1979**, *71*, 3396.

(28) Eichkorn, K.; Treutler, O.; Ohm, H.; Haser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283.

(29) Murray, C. W.; Handy, N. C.; Laming, G. J. *Mol. Phys.* **1993**, *78*, 997.

(30) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 2547.

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

(32) Lelimousin, M.; Field, M. J. Manuscript in preparation.

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

(34) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.

(35) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560.

(36) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667.

(37) See, for example:(a) Head-Gordon, M.; Pople, J. A. *J. Phys. Chem.* **1988**, *92*, 3063. (b) Field, M. J. *J. Phys. Chem.* **1991**, *95*, 5104.

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

(39) Allen, F. H. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380.

(40) Wilson, C. C. *Z. Kristallogr.* **2000**, *215*, 693.

(41) Benedetti, E.; Corradini, P.; Pedone, C. *Biopolymers* **1969**, *7*, 751.

(42) Mackay, M. F.; Bannell, M. G.; Richards, S. L. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **1993**, *49*, 556.

(43) Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991.

(44) Neese, F. ORCA Quantum Chemistry Program. http://www.thch.uni-bonn.de/tc/orca (accessed April 29, 2008).

(45) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

(46) Czermiński, R.; Elber, R. *Int. J. Quantum Chem.: Quantum Chem. Symp.* **1990**, *24*, 167.

CT800092P

JCTC Journal of Chemical Theory and Computation

Solvent-Driven Structural Changes in Anion−π Complexes

Dong Young Kim, N. Jiten Singh, Jung Woo Lee, and Kwang S. Kim*

Center for Superfunctional Materials, Department of Chemistry, Pohang University of Science and Technology, San 31, Hyojadong, Namgu, Pohang 790-784, Korea

Received April 14, 2008

Abstract: Among the π interactions, the anion−π interaction has been a novel type of interaction. In the cases of halide−π complexes, which are the most typical examples of the anion−π interaction, the theoretically predicted and experimentally observed structures in the gas phase are quite different from the most frequently observed crystal structures. We here investigate the structural changes in complexation of the F^-/Cl^- ion with triazine (TAz) as the number of water/acetonitrile molecules increases from 1 to 4. Both the covalent bonding type for F^-−TAz and the hydrogen-bonding type for Cl^-−TAz, which are the lowest-energy structures in the gas phase, change to the solvent-mediated anion−π-type or displaced anion−π-type complexes. This study explains why the (displaced) anion−π-type complexes with some flexible orientations are most common in many crystal structures.

Introduction

As intermolecular interactions involving π rings[1] play an important role in diverse chemical and biological systems, this understanding is vital for the design of nanomaterials and functional molecular systems.[2] Aromatic rings interact with other aromatic rings (π−π interactions),[3] cations (cation−π interactions),[4] hydrogen donors (H−π interactions),[5] lone pairs (lp−π interactions),[6] and anions (anion−π interactions),[7−11] depending on the nature of π-ring systems and molecular partner systems involved. Compared with π−π, cation−π, and H−π interactions, the lp−π and anion−π interactions have very recently been investigated, because aromatic rings were considered as electron sources which give repulsive interactions with anions. The anion−π interactions are highly contrasted to the cation−π interactions. Schneider et al.[7] initially published experimental evidence of π interactions between anions and aryl groups of several host−guest systems, and then they introduced the term "anion−π interaction". Gallivan and Dougherty and Danten et al. theoretically suggested the interaction of the negatively charged lone pair (lp) of water oxygen with the electron-deficient π ring of hexaflulorobenzene.[6a,b] Recently, anion−π interactions have been extensively investigated for the strategy toward a new type of anion recognition, host architecture, and supramolecular self-assembly.[8−11]

Halides can interact with π systems in a few different complexation forms: covalent-bonding type, hydrogen-bonding type,[12] and noncovalent-anion−π type. Mascal et al. and Berryman et al.[11] have shown that π-ring systems interacting with halides in the gas phase form either the H-bonding complexes or the covalent complexes (due to nucleophilic attack), whereas the complexes involving anion−π interactions are less stable in many cases. This has been successfully demonstrated by recent experimental studies by Schneider et al.[13] The covalent-type complexes generally referred to as "Mesenheimer" complexes[14] are found to be the key intermediates in the nucleophilic aromatic substitution (S_NAr) mechanism, and they have been characterized to be stable in the gas phase.[15] However, these are quite in contrast to the fact that the anions interacting with electron-deficient π-aromatic systems have relatively flexible orientations randomly distributed on the top of the ring in the Cambridge Structural Database.[8d,11b] These contrasting results between the crystal observation (noncovalent-bonding type) and the gas-phase structures (covalent-bonding/H-bonding type) demand further theoretical investigations of the anion−π systems in the presence of solvating molecules. Additionally, as the solid-phase anion−π complex structures are crystallized from the solvent medium, many of the crystal structures

---

* kim@postech.ac.kr.

10.1021/ct8001283 CCC: $40.75

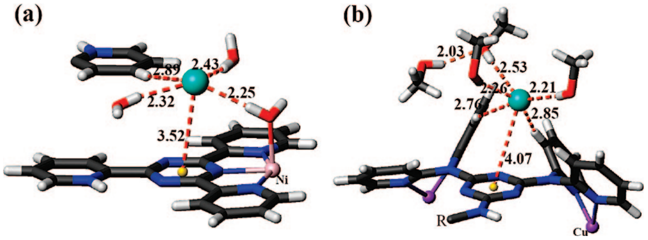


Figure 1. Fragments of the X-ray crystal structures of (a) diaqua-chloro-(2,4-bis(2-pyridyl)-6-(2-pyridinio)-1,3,5-triazine)-nickel (CCDC No. 254390) and (b) $[Cu_3Cl_5(opytrizediam)(MeOH)](Cl)-6.1CH_3OH-C_4H_8O$ (CCDC No. 188817) (refs 14 and 15). Distances are in Ångstroms.

are found to be cocrystallized with the solvent molecules; thereby, the anions interact with an electron-deficient π-aromatic system on one side and the solvent molecules and other coordinating moieties on the other side. Figure 1 displays the representative examples of S-triazine (TAz; $C_3N_3H_3$)-based compounds cocrystallized with the anions, coordinating metal ions, and solvent molecules.[16,17]

It is presumable that the crystal structure of the anion$-\pi$ complexes would reflect the way the anion interacts with the π system in the solvent medium or at the onset of the crystal packing. During the transition from the normal solvated state to the partially desolvated lattice (onset of crystal packing when solvent molecules start displacing out due to the evaporation of the solvent), this partially solvated status would allow both anions and π moieties to be more attractive. As the partially desolvated lattice slowly loses further solvent molecules, the interaction between the anion and π moiety would be stronger, which eventually forms the crystal packing structure. Yet, microscopic structures of the anion$-\pi$ complexes in the presence of solvating molecules are hardly predictable because the solvation of anions and π rings could give a drastic change in the complexion type as compared with the gas-phase structures. It is also conceivable that the electrostatic interaction, which dominates in the covalent type and H-bonding type, would become much less significant in the presence of polar solvent molecules due to the competition between the anion$-$solvent interaction and the anion$-\pi$ interaction. Thus, a theoretical understanding on structural changes of the anion$-\pi$ complexes in the presence of solvent molecules would help explain why covalent-bonding-type interactions are hardly observed, while the (displaced) anion$-\pi$ types are most commonly observed in many crystal structures of the anions interacting with electron-deficient π-aromatic systems. However, the effect including counter cations could also play a significant role, but this has not been considered in the present study since, in most cases, the counter cations tend to be solvated away from the anions in the condensed phase.

TAz is a remarkable unit for supramolecular assembly and is known to interact with anions.[18] In this regard, we have studied the model complexes of TAz and F^-/Cl^- solvated by up to four water/acetonitrile molecules [$TAz\cdots F^-/Cl^-\cdots(H_2O/CH_3CN)_{n=0\sim4}$] and show that, as the number of solvent molecules increases, the covalent- or H-bonding-type complex changes to the solvent-mediated anion$-\pi$ complex with the anion located above the π system.

Calculation Methods

We performed ab initio calculations using the Gaussian 03, Molpro, and Turbomole programs.[19] To search for low-lying energy structures, we have extensively investigated many possible configurations on the basis of our previous experiences to generate low-lying energy structures from various molecular clusters.[20] For hundreds of structural isomers, we have carried out geometry optimization using Moller$-$Plesset second-order perturbation theory (MP2) for smaller systems and the resolution of the identity approximation (RI) of MP2 (RI-MP2)[21] for larger systems with the 6-31+G* basis set. Then, a number of lower-energy structural isomers were optimized using the MP2 method with the aug-cc-pVDZ (abbreviated as aVDZ) basis set. The frequency analysis was made at the MP2/aVDZ level of theory to confirm the minimum-energy structures. For this low-energy conformer search, the cases for mono- to trisolvation are rather intuitive, while the cases for tetra-solvation require extensive investigation. The present study is indeed an extensive study to search for the low-lying energy structures for nanosolvation. Thus, we believe that the structures reported here would be the lowest-energy structures for each given cluster because practically all possible topological isomers have been investigated, and high-level ab initio calculations have been made along with the frequency analysis. Energies were obtained with the basis set superposition error correction. On the MP2/aVDZ geometries, we carried out the MP2/aug-cc-pVTZ (abbreviated as aVTZ) calculations, and the complete basis set (CBS) limit interaction energies at the MP2 level were obtained on the basis of the extrapolation method exploiting the fact that the electron correlation energy is proportional to N^{-3} for the aug-cc-pVNZ basis set.[22,23] Coupled cluster calculations with single and double excitations (CCSD) with aVDZ basis set on the MP2/aVDZ-optimized geometries were performed. Given the importance of a higher level of correlation energies, the CCSD/CBS limit interaction energies (ΔE_e^C) were evaluated from the MP2/CBS limit interaction energies by adding the difference between the MP2/aVDZ and CCSD/aVDZ interaction energies. Using the MP2/aVDZ zero-point energies (ZPE), the ZPE-corrected interaction energies (ΔE_0) at the CBS limit were evaluated. The amount of intermolecular charge transfer based on the natural bond orbital (NBO) analysis[24] was computed at the MP2/aVDZ level.

Results and Discussion

Figure 2 shows the low-lying energy structures of $TAz\cdots F^-/Cl^-\cdots(H_2O/CH_3CN)_{n=0\sim4}$ (a/b/c/d) complexes optimized at the MP2/aVDZ level of theory. TAz interacts with F^-/Cl^-, forming a few different kinds of isomers: (i) covalent interactions [C type] (where the F^- ion attacks a C atom of TAz, resulting in transformation from the sp^2 to sp^3 hybrid orbital), (ii) weakly covalent interactions [C^w type] (where the carbon atom of TAz to which the Cl^- ion is attached holds the sp^2 type), (iii) H-bonding interactions [H type]

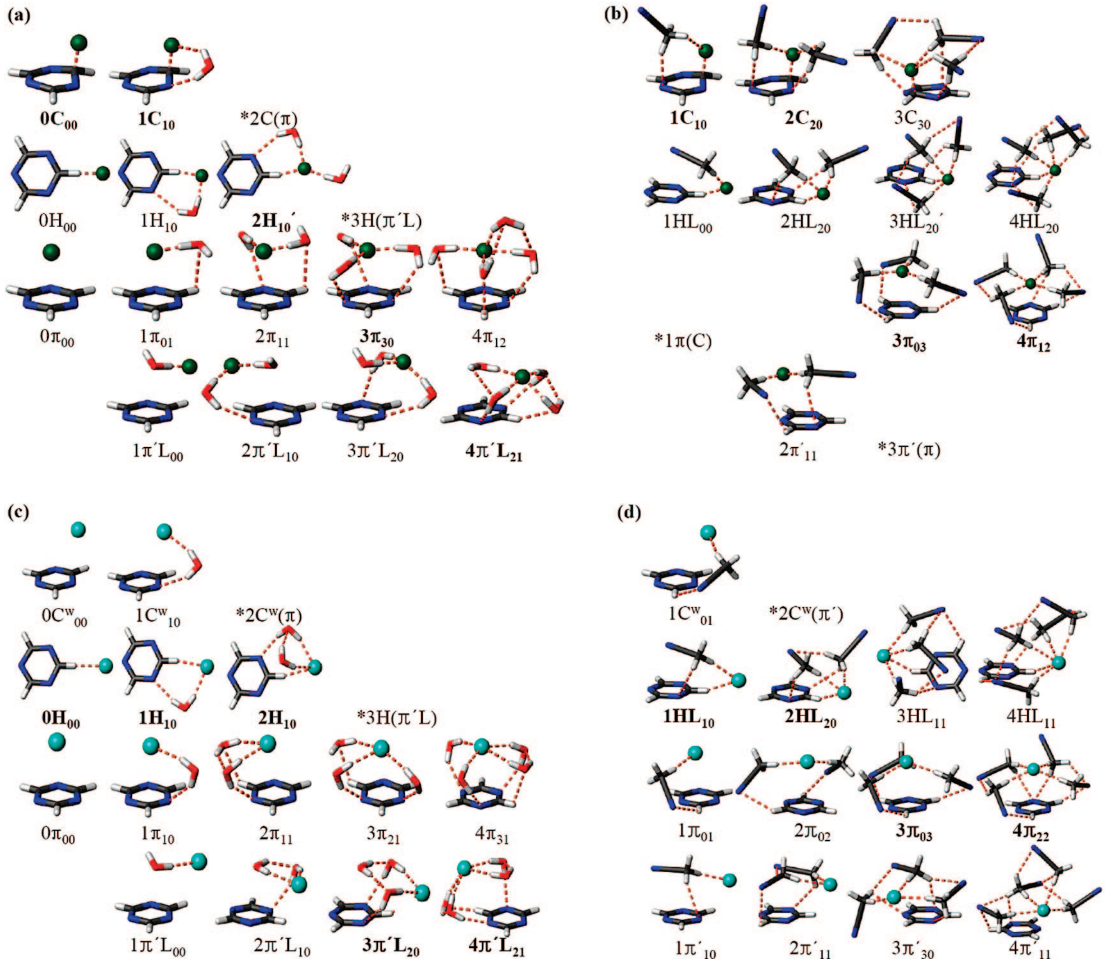


Figure 2. MP2/aVDZ low-lying energy structures of TAz···X$^-$···(Solv)$_{n=1\sim4}$, where X = F/Cl and Solv = H_2O/CH_3CN. (a) X/Solv = F/H_2O, (b) X/Solv = F/CH_3CN, (c) X/Solv = Cl/H_2O, and (d) X/Solv = Cl/CH_3CN. For each interaction type (C, covalent-bonding; C^w, weakly covalent-bonding; H, H-bonding; π, anion−π interaction; π', displaced anion−π interaction; HL, H-bonding and lp(solvent)−π combined interaction; π'L, displaced anion−π and lp(solvent)−π combined interaction), the prefix "n" and suffix subscripts "$n1/n2$" in notations "nY$_{n1n2}$" (Y, C$^{(w)}$/H/$\pi^{(\prime)}$/HL/π'L) are the number of solvent molecules and the number of H(solvent)−N(TAz)/lp(solvent)−H(TAz) H bonds (<3 Å), respectively. The lowest-energy conformer for each solvated structure is marked in bold. A structure marked "*" denotes that it changes to another lower-energy structure. Namely, in a, for $n = 2$, the C-type complex merges to the anion−π complex [$2\pi_{11}$], while for $n = 3$, the H-type complex merges to the π'L-type complex [$3\pi'$L$_{20}$]. In b, for $n = 1$, the π-type complex merges to the C-type complex, while for $n = 3$, the π'-type complex merges to the π-type complex [$3\pi_{03}$]. In c, for $n = 2$, the C^w-type complex merges to the anion−π complex [$2\pi_{11}$], while for $n = 3$, the H-type complex merges to the π'L-type complex [$3\pi'$L$_{20}$]. In d, for $n = 2$, the C^w-type complex merges to the π'-type complex [$2\pi'_{11}$].

(where F$^-$/Cl$^-$ interacts with a H atom of TAz by H-bonding), (iv) noncovalent anion−π interactions [π type] (where the F$^-$/Cl$^-$ ion is located above the centroid of TAz), and (v) solvent-mediated displaced anion−π interactions [π' type]. The F$^-$/Cl$^-$ ion and TAz are connected with water/acetonitrile molecules mainly with H(TAz)···N/O(solvents), N(TAz)···H(solvents), and H(solvents)$^{\delta+}$···F$^-$/Cl$^-$ H-bonding. In our notation of the structures, the solvent hydrogen bonds are not taken into account because these are always present. The O/N lone pair of a water/acetonitrile molecule hydrogen-bonded to a F$^-$/Cl$^-$ ion has a lp−π interaction [L type] with TAz.[6] These are represented as either a π'L type (displaced anion−π and lp(solvent)−π combined interaction) or an HL type (H-bonding and lp(solvent)−π combined interaction). In the figure, for each interaction type (Y: C/C^w/H/π/π'/π'L/HL), prefix "n" in notation "nY$_{n1n2}$" is the number of solvent molecules, and the suffix "$n1/n2$" is used to designate notation between different structures. "$n1/n2$" is correlated to the number of H(solvent)−N(TAz) or lp(solvent)−H(TAz) H bonds (<3.0 Å), respectively (but not the real number of H bonds). Although the O···O bond length indicating the radius of the first hydration shell (reflecting the direct interaction) in the liquid water is often given as 3.4−3.5 Å (or the H···O distance as 2.5 Å),[25] the H···N or lp···H distances within 3.0 Å are denoted as pseudo H bonds in Figure 2, because in clusters, even long H-bond distances show the direct interaction because of the absence of the second solvation shell. Since the distance to represent the on/off H bond cannot be clearly defined in clusters, the above H-bond distances are not used to count the correct number of H bonds but are used for the proper designation of each cluster to distinguish between different structures.

Table 1. CCSD/CBS Limit Interaction Energies (in kcal/mol)[a]

	$C^{(w)}$ type				H(L) type				π type				π'(L) type			
n	conf	Δd_c	ΔE_e^C	ΔE_0	conf	Δd_H	ΔE_e^C	ΔE_0	conf	Δd_{face}	ΔE_e^C	ΔE_0	conf	Δd_{cen}	ΔE_e^C	ΔE_0
$TAz\cdots F^-\cdots(H_2O)_{n=0\sim4}$																
0	**$0C_{00}$**	1.51	−33.99	−32.73	$0H_{00}$	1.53	−18.28	−18.70	$0\pi_{00}$	2.54	−11.27	−11.23				
1	**$1C_{10}$**	1.53	−17.04	−14.43	$1H_{10}$	1.66	−13.68	−11.48	$1\pi_{01}$	2.65	−10.00	−9.15	$1\pi'L_{00}$	3.50	−10.36	−9.46
2					**$2H_{10}'$**	1.75	−11.21	−10.18	$2\pi_{11}$	2.72	−9.35	−8.01	$2\pi'L_{10}$	3.59	−9.69	−8.27
3									**$3\pi_{30}$**	2.76	−8.71	−8.44	$3\pi'L_{20}$	3.67	−9.02	−8.30
4									$4\pi_{12}$	2.92	−7.65	−6.78	**$4\pi'L_{21}$**	3.78	−8.94	−7.28
$TAz\cdots F^-\cdots(CH_3CN)_{n=0\sim4}$																
1	**$1C_{10}$**	1.52	−19.39	−17.11	$1HL_{00}$	1.75	−11.91	−10.85								
2	**$2C_{20}$**	1.56	−11.25	−9.17	$2HL_{20}$	1.82	−9.92	−8.88					$2\pi'_{11}$	3.39	−8.53	−7.23
3	$3C_{30}$	1.60	−5.71		$3HL_{20}'$	1.88	−8.34		**$3\pi_{03}$**	2.79	−9.33					
4					$4HL_{20}$	2.00	−7.76		**$4\pi_{12}$**	2.79	−8.82					
$TAz\cdots Cl^-\cdots(H_2O)_{n=0\sim4}$																
0	$0C^w_{00}$	2.90	−7.04	−6.83	**$0H_{00}$**	2.30	−9.39	−9.25	$0\pi_{00}$	3.14	−6.63	−6.52				
1	$1C^w_{10}$	2.97	−8.23	−6.96	**$1H_{10}$**	2.31	−9.97	−8.75	$1\pi_{10}$	3.15	−7.48	−6.17	$1\pi'L_{00}$	4.24	−7.37	−6.73
2					**$2H_{10}$**	2.40	−9.84	−8.53	$2\pi_{11}$	3.21	−8.39	−7.19	$2\pi'L_{10}$	4.28	−8.79	−7.61
3									$3\pi_{21}$	3.22	−6.48	−6.18	**$3\pi'L_{20}$**	4.34	−7.57	−6.68
4									$4\pi_{31}$	3.25	−4.98	−5.66	**$4\pi'L_{21}$**	4.35	−6.57	−6.47
$TAz\cdots Cl^-\cdots(CH_3CN)_{n=0\sim4}$																
1	$1C^w_{01}$	3.04	−7.65	−6.92	**$1HL_{10}$**	2.40	−7.70	−7.07	$1\pi_{01}$	3.20	−6.94	−6.32	$1\pi'_{10}$	4.44	−6.97	−6.39
2					**$2HL_{20}$**	2.41	−7.78	−6.78	$2\pi_{02}$	3.24	−7.47	−6.29	$2\pi'_{11}$	4.36	−7.70	−6.54
3					$3HL_{11}$	2.49	−6.59		**$3\pi_{03}$**	3.31	−7.57		$3\pi'_{30}$	4.38	−6.86	
4					$4HL_{11}$	2.54	−6.80		**$4\pi_{22}$**	3.31	−7.50		$4\pi'_{11}$	4.21	−6.89	

[a] ΔE_0 is the ZPE-corrected ΔE_e^C (CCSD/CBS). The frequencies for ZPE were evaluated at the MP2/aVDZ level. The CCSD/CBS energies were obtained as [$E_{CCSD/CBS} = E_{MP2/CBS} + (E_{CCSD/aVDZ} - E_{MP2/aVDZ})$], where the MP2/CBS energies were estimated on the basis of the extrapolation method exploiting the fact that the electron correlation energy is proportional to N^{-3} for the aV*N*Z. The distance between F^- and C(TAz)/H(TAz)/Face(TAz)/Center(TAz) is $\Delta d_c/\Delta d_H/\Delta d_{face}/\Delta d_{cen}$. The lowest-energy conformers are denoted in bold characters.

We have listed the ΔE_e^C (CCSD/CBS energy) and ΔE_0 (ZPE-corrected CCSD/CBS energy) of the complexes (Figure 2) in Table 1.

We first consider the case of $TAz\cdots F^-\cdots(H_2O)_{n=0\sim4}$ complexes (Figure 2a). For $n = 0$ and 1, the C-type complexes [$0C_{00}$, $1C_{10}$] are the lowest-energy structures. As in the previous gas-phase ab initio studies by Mascal et al. and Berryman et al.,[11] we note that the C-type complex of $0C_{00}$ is more stable than the H-type-[$0H_{00}$]/π-type-[$0\pi_{00}$] complex by 22/14 kcal/mol in ΔE_0. For $n = 1$, $1C_{10}$ is lower than $1H_{10}/1\pi_{01}/1\pi'L_{00}$ by 3/5/5 kcal/mol. For $n = 2$, the H type [$2H_{10}'$] is the most stable, while the covalent-bonding type merges to the π type [$2\pi_{11}$]. For $n = 3$, the H type merges to the water-mediated displaced anion−π complex [$3\pi'L_{20}$], while $3\pi_{30}$ is the lowest-energy structure, which is nearly isoenergetic to $3\pi'L_{20}$. For $n = 4$, the displaced anion−π complex [$4\pi'L_{21}$] is nearly isoenergetic to or only 0.5 kcal/mol more stable than $4\pi_{12}$. This small energy difference between the displaced anion−π type and the anion−π type explains that these complexes would have relatively flexible orientations. The successive interaction energies [$\Delta\Delta E_0 = \Delta E_0(n) - \Delta E_0(n-1)$] of the most favorable complexes by the sequential addition of water molecules from $n = 1$ to $n = 4$ are 18.30, 4.25, 1.74, and 1.16 kcal/mol, showing a dramatic change in complex type (C type for $n = 1$, H type for $n = 2$, π type for $n = 3$, and π'L type for $n = 4$).

In the case of $TAz\cdots F^-\cdots(CH_3CN)_{n=0\sim4}$ complexes (Figure 2b), for $n = 0$ to $n = 2$, the C-type complexes [$0C_{00}$, $1C_{10}$, $2C_{20}$] are the lowest-energy structures. Although the anion−π complex for $n = 1$ merges to the covalent-bonding-type complex [$1C_{10}$], the anion−π complexes for $n = 3$ and 4 are the lowest-energy structures. The π-type complex of $3\pi_{03}/4\pi_{12}$ is 1 kcal/mol lower than the HL-type complex of $3HL_{20}'/4HL_{20}$.

In the case of $TAz\cdots Cl^-\cdots(H_2O)_{n=0\sim4}$ complexes (Figure 2c), for $n = 0$ to $n = 2$, the H types [$0H_{00}$, $1H_{10}$, $2H_{10}$] are the most stable complexes, while for $n = 3$ and 4, the π'L types are the most stable. In the gas phase, the H-type complex of $0H_{00}$ is 2/3 kcal/mol less stable than the C^w-type/π-type complex. For $n = 2$, the C^w-type complex merges to the π type, while for $n = 3$, the H-type complex merges to the solvent-mediated displaced anion−π complex [$3\pi'L_{20}$], which is the lowest-energy structure. The $\Delta\Delta E_0$ values of the most preferable structures by the sequential addition of water molecules ($n = 1-4$) are 0.5, 0.22, 1.85, and 0.21 kcal/mol, respectively. A large change from $\Delta E_0(2)$ to $\Delta E_0(3)$ reflects a conformational change from the H type to the π'L type.

In the case of $TAz\cdots Cl^-\cdots(CH_3CN)_{n=0\sim4}$ complexes (Figure 2d, Table 1), for $n = 0$, 1, and 2, the H-type complexes are the most stable isomers, while the lowest-energy structures for $n = 3$ and 4 are the anion−π-type complexes where the Cl^- ion is above the centroid of TAz, as in the case of the F^- complex.

In the absence of solvent molecules, the C type is the lowest-energy structure in the $TAz\cdots F^-$ complex, and the H type is the lowest-energy structure in the $TAz\cdots Cl^-$ complex. However, for both cases, in the presence of three water/acetonitrile molecules, the solvent-mediated anion−π-type complexes become the lowest-energy structures. These interesting trends are depicted in Figure 3.

For the $TAz\cdots F^-$ complexes, the covalent-bonding C-type interactions are characterized by the substantial

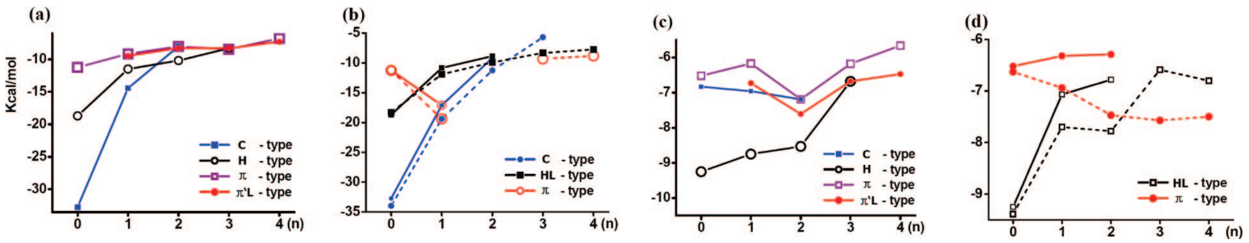


Figure 3. ZPE-corrected CCSD/CBS limit interaction energies [ΔE_0], (a) X/Solv = F/H_2O, (b) X/Solv = F/CH_3CN, (c) X/Solv = Cl/H_2O, and (d) X/Solv = Cl/CH_3CN for the low-lying energy structures of TAz···X^-···$(Solv)_{n=1\sim4}$. Dotted lines are for the ZPE-uncorrected values.

Table 2. Charge Transfer (in MP2/aVDZ NBO Charge in au) from the F^- Ion to the π System (Δq_π) and the Solvent Molecules (Water/Acetonitrile; Δq_s)

$F^-(solv)_n$		C type		H(L) type		π type		π'(L) type	
n	Δq_s	conf	$\Delta q_\pi/\Delta q_s$	conf	$\Delta q_\pi/\Delta q_s$	conf	$\Delta q_\pi/\Delta q_s$	conf	$\Delta q_\pi/\Delta q_s$
TAz···F^-···$(H_2O)_{n=0\sim4}$									
0		$0C_{00}$	0.51	$0H_{00}$	0.11	$0\pi_{00}$	0.02		
1	0.12	$1C_{10}$	0.46/0.04	$1H_{10}$	0.06/0.07	$1\pi_{01}$	0.02/0.11	$1\pi'L_{00}$	0.04/0.11
2	0.17			$2H_{>10}'$	0.05/0.12	$2\pi_{11}$	0.02/0.15	$2\pi'L_{10}$	0.02/0.15
3	0.18					$3\pi_{30}$	0.02/0.15	$3\pi'L_{20}$	0.02/0.17
4	0.18					$4\pi_{12}$	0.02/0.17	$4\pi'L_{21}$	0.02/0.16
TAz···F^-···$(CH_3CN)_{n=0\sim4}$									
1	0.10	$1C_{10}$	0.48/0.03	$1HL_{00}$	0.06/0.08				
2	0.13	$2C_{20}$	0.44/0.05	$2HL_{20}$	0.04/0.11			$2\pi'_{11}$	0.03/0.13
3	0.15	$3C_{30}$	0.39/0.07	$3HL_{20}'$	0.03/0.13	$3\pi_{03}$	0.01/0.15		
4	0.18			$4HL_{20}$	0.02/0.16	$4\pi_{12}$	0.01/0.17		
TAz···Cl^-···$(H_2O)_{n=0\sim4}$									
0		$0C^w{}_{00}$	0.07	$0H_{00}$	0.06	$0\pi_{00}$	0.02		
1	0.06	$1C^w{}_{10}$	0.05/0.05	$1H_{10}$	0.04/0.05	$1\pi_{10}$	0.01/0.05	$1\pi'L_{00}$	0.03/0.06
2	0.10			$2H_{10}$	0.04/0.08	$2\pi_{11}$	0.01/0.08	$2\pi'L_{10}$	0.02/0.08
3	0.10					$3\pi_{21}$	0.01/0.10	$3\pi'L_{20}$	0.03/0.10
4	0.11					$4\pi_{31}$	0.01/0.11	$4\pi'L_{21}$	0.01/0.11
TAz···Cl^-···$(CH_3CN)_{n=0\sim4}$									
1	0.05	$1C^w{}_{01}$	0.04/0.05	$1HL_{10}$	0.04/0.05	$1\pi_{01}$	0.02/0.05	$1\pi'_{10}$	0.03/0.05
2	0.08			$2HL_{20}$	0.02/0.09	$2\pi_{02}$	0.03/0.08	$2\pi'_{11}$	0.02/0.08
3	0.12			$3HL_{11}$	0.02/0.11	$3\pi_{03}$	0.02/0.11	$3\pi'_{30}$	0.02/0.11
4	0.14			$4HL_{11}$	0.02/0.13	$4\pi_{22}$	0.01/0.14	$4\pi'_{11}$	0.02/0.13

charge transfer from the F^- ion to the π system, whereas the charge transfer is small for the H type and insignificant for the π type and π'L type (Table 2). The anion−π interactions related to the π type/π'L type have substantial dispersion energies, while the H-bonding related to the H type is mostly electrostatic. In particular, the charge-transfer-driven C-type interactions are highly susceptible to polar solvents like water because these interactions compete with the charge-transfer-driven ionic H-bonding, similar to the special type of H-bonding species $(F-H-OH)^-$. In the presence of water molecules, for $n = 1$, the C-type interaction dominates (a charge transfer of 0.46 au from the F^- ion to the π system), so that the $(F-H-OH)^-$ ionic H-bonding interaction is not properly formed. However, for $n = 2$, the ionic H-bonding interaction dominates, so the carbon atom to which the F^- ion was attached when $n = 0/1$ is no longer of the sp^3 type, and it returns to the sp^2 type. As a result, in the presence of more than two water molecules, H/π/π'L-type interactions are more stable than the C-type interaction.

These charge-transfer-driven C-type interactions are less susceptible to the solvents of the relatively lower dielectric constant. Therefore, in contrast to the case in the aqua phase, in the presence of the acetonitrile molecules, the $(F-H-CH_2CN)^-$ ionic H-bonding interaction does not compete with the charge-transfer-driven C-type interaction. The charge transfer from F^- to TAz by forming the sp^3-type carbon is 0.48/0.44/0.39 au for $n = 1/2/3$. In the F^-···$(CH_3CN)_{n=1\sim3}$ complexes, the charge transfer from the F^- ion to the acetonitrile molecules is only 0.03/0.05/0.07 au. For $n = 3/4$, the complexation is characterized by the N(acetonitrile)···H(TAz) H-bonding and the acetonitrile-mediated π-type interaction. This is because the solvation by acetonitrile molecules drives the anion to be located above the π ring, allowing a large empty space for the stabilization of the excess electron around the anion.

For the TAz···Cl^- complexes, in the presence of solvent molecules, weakly covalent-bonding C^w-type interactions are characterized not only by small charge transfer (0.07/0.05 au for $n = 0/1$) from the Cl^- ion to the π system like the H-type interactions (0.06/0.04 au) but also by the sp^2-type carbon (different from the sp^3-type carbon in the F^- complexes). Also, for the TAz···Cl^- complexes, the covalent-bonding-type complexes are not the lowest-energy structures in the presence of water/acetonitrile molecules. For $n = 0/1/$

2, the complexation is characterized by the N(acetonitrile)$\cdots\pi$ lp$-\pi$ interaction and the H(solvent)$\cdots$N(TAz) and CH$^{\delta+}\cdots$Cl$^-$ hydrogen-bond interactions. The C^w-type complexes merge to the other type of complexes for $n = 2$. For $n = 3/4$, the anion is above the π ring as the Cl$^-$ ion interacts with water or acetonitrile molecules, like in the case of the F$^-$ complexes. In the presence of three and four water molecules, the complexation is characterized by both the H(water)$\cdots$N(TAz) and H(water)$^{\delta+}\cdots$Cl$^-$ H bonds and the O(water)$\cdots\pi$ lp$-\pi$ interaction. In the presence of three and four acetonitrile molecules, the complexation is characterized by the N(acetonitrile)$\cdots$H(TAz) H-bonding and the π-type interaction.

In crystal structures, each anion involved in the anion$-\pi$ interaction is surrounded by several other coordinating moieties. In addition, one or more solvent molecules per anion$-\pi$ complex are often present. Thus, the anion in the anion$-\pi$ complexes would have relatively flexible orientations toward the π-ring system, in agreement with a recent experiment.[26] Thus, the present results explain some of the real situations of the anion$-\pi$ interaction in crystals. Furthermore, the anion$-\pi$ complexes with a few water molecules or other solvent molecules in the gas phase can be experimentally observable, as in the case of halide$-$water clusters or benzene$-$water clusters.[27] Water molecules can act to pull the F$^-$/Cl$^-$ ion away from TAz, so that the TAz can be fully hydrated (with minimal anion$-\pi$ interaction) by a large number of water molecules in the aqua phase, consistent with the result of the diminished TAz$\cdots$Cl$^-$ interaction energy described on the basis of continuum solvent models.[11a]

Understanding the nature of anion$-\pi$ interaction in the presence microsolvents around the complex is important not only for the structural insight into these clusters, which would be observed in the gas-phase experiments, but also for the conceptual understanding of crystal growth in the solvent medium as the solvent molecules would eventually be displaced out during the crystal packing.

Concluding Remarks

It is known that there is a distinct difference between the theoretically predicted or experimentally observed most-stable form of halide$-\pi$ complexes in the gas phase and the most frequently observed crystal structures of halide$-\pi$ complexes. However, the anion binding patterns are expected to change significantly in the presence of polar solvent molecules. As a F$^-$/Cl$^-$ ion interacts with TAz, the binding pattern shows an intriguing structural change depending on the number of solvating molecules.

For the complexes with the F$^-$ ion, in the absence of solvent molecules, covalent bonding is clearly the most preferable because the F$^-$ ion attacks an sp^2-type C atom in TAz and transforms it into an sp^3-type C atom. This covalent-type complex is well-known as a "Jackson$-$Meisenheimer" or "Meisenheimer" complex. In the presence of a monohydrated F$^-\cdots$TAz system, the covalent-bonding type is still the lowest-energy structure. In the presence of two water molecules, the covalent-bonding type changes to the H-bonding type, which competes with the anion$-\pi$ type and the water-mediated displaced anion$-\pi$ type. For three and four water molecules, the displaced anion$-\pi$ interaction is most preferable. In the presence of acetonitrile molecules, for one and two acetonitrile molecules, the covalent-bonding-type complexes are the most preferable, while for the complexes with three and four acetonitrile molecules, the anion$-\pi$-type structures are the most stable. For the complexes with the Cl$^-$ ion, the H-bonding type is the most preferable up to two solvent molecules. However, in the presence of three solvent molecules, the H-bonding type changes to the solvent-mediated anion$-\pi$ type. For three and four solvent molecules, the anion$-\pi$ types are the lowest-energy structures. Since both the anion$-\pi$ type and the displaced anion$-\pi$ type are nearly isoenergetic, it explains why both types with relatively flexible orientations are most common in many crystal structures.

The intriguing role of hydration/solvation on the structure of anion$-\pi$ complexes will shed light on various π systems interacting with anions and help resolve the differences between the gas-phase results and the crystal data. Furthermore, the understanding of microscopic structures of anions interacting with π systems in the solution phase would practically aid the design and development of novel anion receptors.[28]

Acknowledgment. This work was supported by GR-L(KICOS) and BK21(KRF). Most calculations were carried out using supercomputers at KISTI.

Supporting Information Available: Structures and interaction energies of low lying energy complexes. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Brutschy, B. *Chem. Rev.* **2000**, *100*, 3891. (b) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. *Chem. Rev.* **2000**, *100*, 4145. (c) Hobza, P.; Selzle, H. L.; Schlag, E. W. *Chem. Rev.* **1994**, *94*, 1767. (d) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690. (e) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525. (f) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.

(2) (a) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210. (b) Hong, B. H.; Lee, J. Y.; Lee, C.-W.; Kim, J. C.; Bae, S. C.; Kim, K. S. *J. Am. Chem. Soc.* **2001**, *123*, 10748. (c) Hong, B. H.; Bae, S. C.; Lee, C.-W.; Jeong, S.; Kim, K. S. *Science* **2001**, *294*, 348. (d) Cockroft, S. L.; Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *J. Am. Chem. Soc.* **2005**, *127*, 8594. (e) Hong, B. H.; Small, J. P.; Purewal, M. S.; Mullokandov, A.; Sfeir, M. Y.; Wang, F.; Lee, J. Y.; Heinz, T. F.; Brus, L. E.; Kim, P.; Kim, K. S. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 14155. (f) Hoeben, F. J. M.; Jonkheijm, P.; Meijer, E. W.; Schenning, A. P. H. J. *Chem. Rev.* **2005**, *105*, 1491.

(3) (a) Lee, E. C.; Kim, D.; Jurečka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. *J. Phys. Chem. A* **2007**, *111*, 3446. (b) Kim, K. S. et al. *J. Am. Chem. Soc.* **2002**, *124*, 14268. (c) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 3500. (d) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887. (e) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142.

(4) (a) Dougherty, D. A.; Stauffer, D. *Science* **1990**, *250*, 1558. (b) Kumpf, R. A.; Dougherty, D. A. *Science* **1993**, *261*, 1708. (c) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303. (d) Cubero, E.; Luque, F. J.; Orozco, M. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 5976. (e) Kim, K. S.; Lee, J. Y.; Lee, S. J.; Ha, T.-K.; Kim, D. H. *J. Am. Chem. Soc.* **1994**, *116*, 7399. (f) Lee, J. Y.; Lee, S. J.; Choi, H. S.; Cho, S. J.; Kim, K. S.; Ha, T. K. *Chem. Phys. Lett.* **1995**, *232*, 67. (g) Kim, D.; Hu, S.; Tarakeshwar, P.; Kim, K. S.; Lisy, J. M. *J. Phys. Chem. A* **2003**, *107*, 1228.

(5) (a) Vaupel, S.; Brutschy, B.; Tarakeshwar, P.; Kim, K. S. *J. Am. Chem. Soc.* **2006**, *128*, 5416. (b) Cockroft, S. L.; Hunter, C. A. *Chem. Commun.* **2006**, 3806. (c) Lee, E. C.; Hong, B. H.; Lee, J. Y.; Kim, J. C.; Kim, D.; Kim, Y.; Tarakeshwar, P.; Kim, K. S. *J. Am. Chem. Soc.* **2005**, *127*, 4530. (d) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. *J. Am. Chem. Soc.* **2001**, *123*, 3323. (e) Tarakeshwar, P.; Lee, S. J.; Lee, J. Y.; Kim, K. S. *J. Chem. Phys.* **1998**, *108*, 7217. (f) Nakamura, K.; Houk, K. N. *Org. Lett.* **1999**, *1*, 2049.

(6) (a) Gallivan, J. P.; Dougherty, D. A. *Org. Lett.* **1999**, *1*, 103. (b) Danten, Y.; Tassaing, T.; Besnard, M. *J. Phys. Chem. A* **1999**, *103*, 3530. (c) Egli, M.; Sarkhel, S. *Acc. Chem. Res.* **2007**, *40*, 197. (d) Mooibroek, T. J.; Teat, S. J.; Massera, C.; Gamez, P.; Reedijk, J. *Cryst. Growth Des.* **2006**, *6*, 1569.

(7) (a) Schneider, H. J. *Angew. Chem.* **1991**, *103*, 1419. (b) Schneider, H. J.; Blatter, T.; Palm, B.; Pfingstag, U.; Ruediger, V.; Theis, I. *J. Am. Chem. Soc.* **1992**, *114*, 7704. (c) Schneider, H. J.; Werner, F.; Blatter, T. *J. Phys. Org. Chem.* **1993**, *6*, 590.

(8) (a) Alkorta, I.; Rozas, I.; Elguero, J. *J. Org. Chem.* **1997**, *62*, 4687. (b) Alkorta, I.; Rozas, I.; Elguero, J. *J. Am. Chem. Soc.* **2002**, *124*, 8593. (c) Mascal, M. *Angew. Chem., Int. Ed.* **2006**, *45*, 2890. (d) Quiñonero, D.; Garau, C.; Rotger, C.; Frontera, A.; Ballester, P.; Costa, A.; Deyà, P. M. *Angew. Chem., Int. Ed.* **2002**, *41*, 3389. (e) Quiñonero, D.; Garau, C.; Frontera, A.; Ballester, P.; Costa, A.; Deyà, P. M. *Chem. Phys. Lett.* **2002**, *359*, 486. (f) Garau, C.; Frontera, A.; Ballester, P.; Quiñonero, D.; Costa, A.; Deyà, P. M. *Eur. J. Org. Chem.* **2005**, 179. (g) Frontera, A.; Saczewski, F.; Gdaniec, M.; Dziemidowicz-Borys, E.; Kurland, A.; Deya, P. M.; Quiñonero, D.; Garau, C. *Chem—Eur. J.* **2005**, *11*, 6560. (h) Quiñonero, D.; Garau, C.; Frontera, A.; Ballester, P.; Costa, A.; Deyà, P. M. *J. Phys. Chem. A* **2005**, *109*, 4632. (i) Hermida-Ramón, J. M.; Estevéz, C. M. *Chem.—Eur. J.* **2007**, *13*, 4743. (j) Escudero, D.; Frontera, A.; Quiñonero, D.; Costa, A.; Ballester, P.; Deyà, P. M. *J. Chem. Theory Comput.* **2007**, *3*, 2098.

(9) (a) Kim, D.; Tarakeshwar, P.; Kim, K. S. *J. Phys. Chem. A* **2004**, *108*, 1250. (b) Kim, D.; Lee, E. C.; Kim, K. S.; Tarakeshwar, P. *J. Phys. Chem. A* **2007**, *111*, 7980.

(10) (a) Demeshko, S.; Dechert, S.; Meyer, F. *J. Am. Chem. Soc.* **2004**, *126*, 4508. (b) de Hoog, P.; Gamez, P.; Mutikainen, I.; Turpeinen, U.; Reedijk, J. *Angew. Chem., Int. Ed.* **2004**, *43*, 5815. (c) Rosokha, Y. S.; Lindeman, S. V.; Rosokha, S. V.; Kochi, J. K. *Angew. Chem., Int. Ed.* **2004**, *43*, 4650. (d) Schottel, B. L.; Chifotides, H. T.; Shatruk, M.; Chouai, A.; Pérez, L. M.; Bacsa, J.; Dunbar, K. R. *J. Am. Chem. Soc.* **2006**, *128*, 5895. (e) Mascal, M.; Yakovlev, I.; Nikitin, E. B.; Fettinger, J. C. *Angew. Chem., Int. Ed.* **2006**, *45*, 2890. (f) Tanaka, D.; Masaoka, S.; Horike, S.; Furukawa, S.; Mizuno, M.; Endo, K.; Kitagawa, S. *Angew. Chem., Int. Ed.* **2006**, *45*, 4628. (g) Gorteau, V.; Bollot, G.; Mareda, J.; Matile, S. *Org. Biomol. Chem.* **2007**, *5*, 3000. (h) Perez-Velasco, A.; Gorteau, V.; Matile, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 921.

(11) (a) Mascal, M.; Armstrong, A.; Bartberger, M. *J. Am. Chem. Soc.* **2002**, *124*, 6274. (b) Berryman, O. B.; Bryantsev, V. S.; Stay, D. P.; Johnson, D. W.; Hay, B. P. *J. Am. Chem. Soc.* **2007**, *129*, 48. (c) Gamez, P.; Mooibroek, T. J.; Teat, S. J.; Reedijk, J. *Acc. Chem. Res.* **2007**, *40*, 435.

(12) Bryantsev, V. S.; Hay, B. P. *J. Am. Chem. Soc.* **2005**, *127*, 8282.

(13) Schneider, H.; Vogelhuber, K. M.; Schinle, F.; Weber, M. *J. Am. Chem. Soc.* **2007**, *129*, 13022.

(14) (a) Terrier, F. *Chem. Rev.* **1982**, *82*, 78. (b) Artamkina, G. A.; Egorov, M. P.; Beletskaya, I. P. *Chem. Rev.* **1982**, *82*, 427. (c) Ingemann, S.; Nibbering, N. M. M.; Sullivan, S. A.; DePuy, C. H. *J. Am. Chem. Soc.* **1982**, *104*, 6520. (d) Sun, H.; DiMagno, S. G. *Angew. Chem., Int. Ed.* **2006**, *45*, 2720.

(15) (a) Chiavaino, B.; Crestoni, M. E.; Fornarini, S.; Lanucara, F.; Lemaire, J.; Maître, P. *Angew. Chem., Int. Ed.* **2007**, *46*, 1995. (b) Hiraoka, K.; Mizuse, S. *J. Chem. Phys.* **1987**, *86*, 4102.

(16) Zibaseresht, R.; Hartshorn, R. M. *Aust. J. Chem.* **2005**, *58*, 345. CCDC Deposition number: 254390.

(17) de Hoog, P.; Gamez, P.; Roubeau, O.; Lutz, M.; Driessen, W. D.; Spek, A. L.; Reedijk, J. *New J. Chem.* **2003**, *27*, 18.

(18) Mooibroek, T. J.; Gamez, P. *Inorg. Chim. Acta* **2007**, *360*, 381.

(19) (a) Frisch, M. J. et al. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004. (b) Werner, H.-J. *MOLPRO, a package of ab initio programs*, version 2006.1; University College Cardiff Consultants Limited: Cardiff, Wales, 2006. (c) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.

(20) (a) Lee, S.; Kim, J.; Lee, S. J.; Kim, K. S. *Phys. Rev. Lett.* **1997**, *79*, 2038. (b) Lee, H. M.; Suh, S. B.; Lee, J. Y.; Tarakeshwar, P.; Kim, K. S. *J. Chem. Phys.* **2000**, *112*, 9759. (c) Lee, H. M.; Kim, D.; Kim, K. S. *J. Chem. Phys.* **2002**, *116*, 5509. (d) Singh, N. J.; Park, M.; Min, S. K.; Suh, S. B.; Kim, K. S. *Angew. Chem., Int. Ed.* **2006**, *45*, 3795.

(21) Weigend, F.; Haser, M. *Theor. Chem. Acc.* **1997**, *97*, 331.

(22) Helgaker, T.; Ruden, T. A.; Jorgensen, P.; Olsen, J.; Klopper, W. *J. Phys. Org. Chem.* **2004**, *17*, 913.

(23) (a) Lee, H. M.; Kim, D.; Singh, N. J.; Kolaski, M.; Kim, K. S. *J. Chem. Phys.* **2007**, *127*, 164311. (b) Min, S. K.; Lee, E. C.; Lee, H. M.; Kim, D. Y.; Kim, D.; Kim, K. S. *J. Comput. Chem.* **2008**, *29*, 1208.

(24) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.

(25) (a) Soper, A. K. *Chem. Phys.* **2000**, *258*, 121. (b) Kim, K. S. *Chem. Phys. Lett.* **1989**, *159*, 261. (c) Sharp, K. A.; Madan, B.; Manas, E.; Vanderkooi, J. M. *J. Chem. Phys.* **2001**, *114*, 1791.

(26) Albrecht, M.; Wessel, C.; Groot, M.; de Rissanen, K.; Lüchow, A. *J. Am. Chem. Soc.* **2008**, *130*, 4600.

(27) (a) Robertson, W. H.; Johnson, M. A. *Science* **2002**, *298*, 69. (b) Robertson, W. H.; Johnson, M. A. *Annu. Rev. Phys. Chem.* **2003**, *54*, 173. (c) Loh, Z. M.; Wilson, R. L.; Wild, D. A.; Bieske, E. J.; Lisy, J. M.; Njegic, B.; Gordon, M. S. *J. Chem. Phys.* **2006**, *110*, 13736. (d) Reimann, B.; Buchhold, K.; Barth, H. D.; Brutschy, B.; Tarakeshwar, P.; Kim, K. S. *J. Chem. Phys.* **2002**, *117*, 8805. (e) Carbarcos, O. M.;

Weinheimer, C. J.; Lisy, J. M.; Xantheas, S. S. *J. Chem. Phys.* **1999**, *110*, 5. (f) Pribble, R. N.; Zwier, T. S. *Science* **1994**, *265*, 75.

(28) (a) Mizuno, T.; Wei, W. H.; Eller, L. R.; Sessler, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 1134. (b) Yoon, J.; Kim, S. K.; Singh, N. J.; Kim, K. S. *Chem. Soc. Rev.* **2006**, *35*, 355. (c) Chellappan, K.; Singh, N. J.; Hwang, I.-C.; Lee, J. W.; Kim, K. S. *Angew. Chem., Int. Ed.* **2005**, *44*, 2899; *Angew. Chem.* **2005**, *117*, 2959. (d) Ihm, H.; Yun, S.; Kim, H. G.; Kim, J. K.; Kim, K. S. *Org. Lett.* **2002**, *4*, 2897. (e) Gale, P. A. *Coord. Chem. Rev.* **2000**, *199*, 181.

CT8001283