

## Quality Assurance for Biomolecular Simulations

Stuart E. Murdock,<sup>†,‡</sup> Kaihsu Tai,<sup>\*,§</sup> Muan Hong Ng,<sup>‡</sup> Steven Johnston,<sup>‡</sup> Bing Wu,<sup>§,||</sup> Hans Fangohr,<sup>‡</sup> Charles A. Laughton,<sup>⊥</sup> Jonathan W. Essex,<sup>\*,†</sup> and Mark S. P. Sansom<sup>§</sup>

*School of Chemistry and School of Engineering Sciences, University of Southampton, United Kingdom, Department of Biochemistry and Oxford e-Science Centre, University of Oxford, United Kingdom, and School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, United Kingdom*

Received May 15, 2006

**Abstract:** Contemporary structural biology has an increased emphasis on high-throughput methods. Biomolecular simulations can add value to structural biology via the provision of dynamic information. However, at present there are no agreed measures for the quality of biomolecular simulation data. In this Letter, we suggest suitable measures for the quality assurance of molecular dynamics simulations of biomolecules. These measures are designed to be simple, fast, and general. Reporting of these measures in simulation papers should become an expected practice, analogous to the reporting of comparable quality measures in protein crystallography. We wish to solicit views and suggestions from the simulation community on methods to obtain reliability measures from molecular-dynamics trajectories. In a database which provides access to previously obtained simulations—for example BioSimGrid (<http://www.biosimgrid.org/>)—the user needs to be confident that the simulation trajectory is suitable for further investigation. This can be provided by the simulation quality measures which a user would examine prior to more extensive analyses.

\* Corresponding author phone: +44 23 80592794; e-mail: j.w.essex@soton.ac.uk (J.W.E.) and e-mail: kaihsu.tai@bioch.ox.ac.uk (K.T.).

<sup>†</sup> School of Chemistry, University of Southampton.

<sup>‡</sup> School of Engineering Sciences, University of Southampton.

<sup>§</sup> Department of Biochemistry, University of Oxford.

<sup>||</sup> Oxford e-Science Centre, University of Oxford.

<sup>⊥</sup> University of Nottingham.

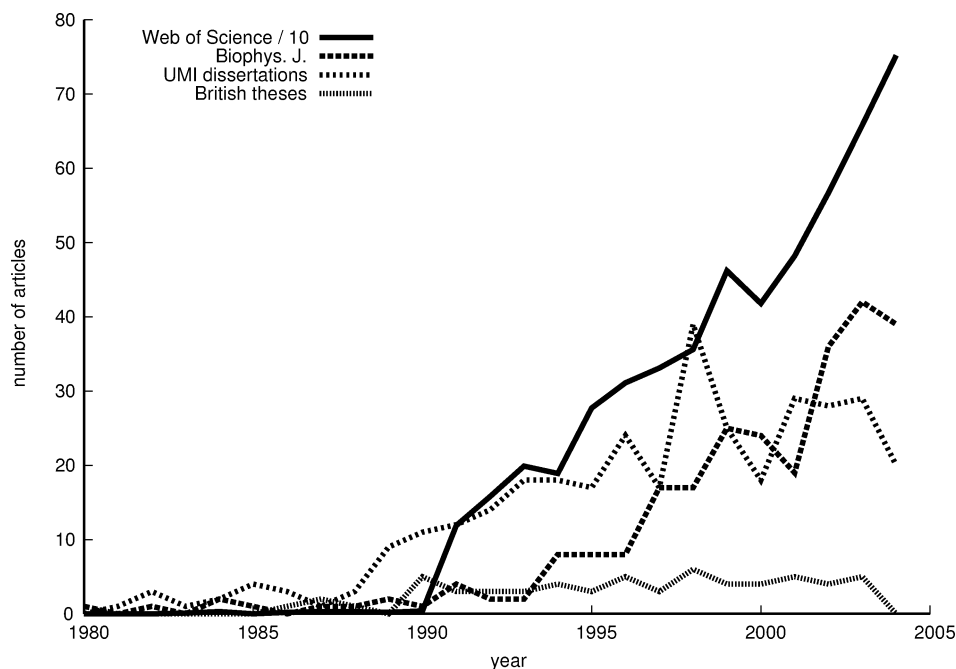
<sup>\*</sup> Current address: Schrödinger Inc., 101 SW Main Street, Suite 1300, Portland, OR 97204.

## 1. OVERVIEW

For the past quarter century, biomolecular simulations have been adding value to structural biology via the provision of dynamic information.<sup>1</sup> As genomics move from sequencing to structural and dynamical considerations, and high-throughput technologies advance from crystallography to molecular-dynamics (MD) simulation, this process is occurring with vigor. As the bibliometric data in Figure 1 show, MD simulation of biopolymers is now becoming a routine technique. To help this maturation process, standardized practice should be established in the simulation community, similar to that in crystallography.<sup>2,3</sup> It is already regular practice to print quality measures in a formulaic table in published articles reporting crystallographic results—indeed, it is surprising if such a table is missing, and the referees would readily reject the manuscript.

We are hereby initiating a discussion on the appropriate measures of quality and convergence<sup>4</sup> for MD simulation trajectories of biopolymers. The process of calculating these measures is designed to be automated for large numbers of trajectories; hence the set of analyses used for this description should be general, with minimal interaction of a human curator. The scientist can then use these measures, along with sensible comparisons with known experimental data (which we recognize as essential), to decide whether a specific trajectory is suitable for further investigation. Our purpose is to solicit feedback from the simulation community with regard to the analyses we have chosen and to obtain further suggestions. We invite the community to express their views on our choices of measures.

We are motivated to do this by our work in building BioSimGrid,<sup>5</sup> a distributed environment for archiving and analyzing biopolymer simulations. Other similar databases are emerging<sup>6</sup> (personal communications with Valerie Daggett and Modesto Orozco, <http://mmb.pcb.ub.es/MODEL/>); these also require some quality-assurance measures. The BioSimGrid project was implemented to satisfy the growing demand for the storage of large amounts of simulation data which is currently being produced within a number of laboratories. This environment enables the storage and analysis of large biopolymer MD trajectories, making previously logistically difficult comparative analyses and data curation easier. For example, analyzing large numbers of trajectories distributed across many laboratories has now become as seamless through BioSimGrid, as if the data were produced in the same laboratory. Traditionally, timeseries of millions of degrees-of-freedom are collected in a biomolecular MD simulation, but only a small fraction of these are presented in the resulting paper, due to the lifetime of the project. In this scenario, other laboratories may be



**Figure 1.** Bibliometrics shows that molecular-dynamics simulation has the potential to become a routine scientific technique for investigating biopolymer dynamics: search results for protein molecular dynamics simulations in Web of Science, *Biophysical Journal*, UMI ProQuest Digital Dissertations, and Index to Theses in Great Britain. Web of Science search for protein molecular dynamics simulations “TS = ((molecular SAME dynamics) AND simulation\* AND protein\*)” at <http://wok.mimas.ac.uk/>. Title/abstract search in the *Biophysical Journal* for keywords ‘molecular dynamics’ and ‘simulation’ at <http://www.biophysj.org/>. Abstract search “‘molecular dynamics’ and ‘protein\*’” in the UMI ProQuest Digital Dissertations (mainly North American) at <http://www.lib.umi.com/dissertations/>. Search “‘molecular dynamics’ and ‘protein\*’” in the Index to Theses in Great Britain (false positives in the 1980s removed) at <http://www.theses.com/>. Data up-to-date mid-2004.

interested in a trajectory of a biomolecule for which MD has been done, but without a clearinghouse, access is difficult to obtain. BioSimGrid aims to make the results of large-scale computer simulations of biomolecules more accessible to the biological community. As a comprehensive simulation data-management system with many analysis tools, BioSimGrid provides scientists access to trajectories stored throughout many laboratories, once public availability has been granted by the owner of the data. An early exemplar application is comparisons among enzymes of similar active sites.<sup>7</sup>

Some may hold the opinion that, based on the motivation of the scientist carrying out the work, MD simulations may be split into two categories: equilibrium and nonequilibrium. The nonequilibrium simulations are those where the scientist wished to explore unfolding pathways or large conformational changes not manifest in known crystallographic structures. The investigations of biomolecular dynamics in an equilibrium state fall into the other category. It follows that the quality-assurance measures and analyses methods for these different classes have to be quite different. We take a more agnostic approach at the original motivation that brought the trajectory into existence. There are many types of MD simulations being performed under varying degrees of nonequilibrium conditions; any boundaries or distinctions imposed a priori may turn out inappropriate. Another drawback of such a priori description is that it requires speculation about the scientist’s state-of-mind: the decision reached thus may not always be accurate.

The following quality-assurance measures we introduce are not restricted to proteins, but any polymers, and may be readily applied to nucleic acids, sugars (polysaccharides),

and even nonbiological polymers where the monomers can be clearly identified. To keep our convergence measures general, so they may be automated, we include all atoms in each biopolymer within the trajectory, and we perform the analysis by first performing quaternion least-squares fitting<sup>8</sup> for the atoms with respect to the initial configuration of the stored trajectory. These measures could appropriately be split into three different classes: quality, convergence, and structural;<sup>9,10</sup> they are described in detail in the sections below.

## 2. PROVENANCE METADATA, WHOSE REPORTING SHOULD BE OBLIGATORY

Two particular sets of metadata that describe the provenance of the trajectory should be reported. The provenance, or ontogeny, metadata tell how and whence the trajectory came about. The first concerns the preprocessing before the initial structure for the MD simulation can be obtained; the second, the particular setup for the simulation. As is typical for reporting of experimental procedures in scientific literature, these should be reported to the extent that an unrelated laboratory will be able to reproduce the simulation.

The first set, presentable in free-styled text, includes the following: the Protein Data Bank identification code for the crystallographic structure on which the simulation is based; the procedure for reconstructing residues and side chains which were not observed in the crystallographic structure; determination of protonation states; ligand insertion; solvation (including the retention of crystallographic water molecules and addition of ions) or insertion into medium (for example, the procedure of solvation in water and that for insertion

**Table 1:** Formulaic Table Providing Some of the Provenance Metadata for a Molecular-Dynamics Simulation<sup>11</sup>

global trajectory identifier	BioSimGrid_GB-OXF_9
trajectory name	outer-membrane phospholipase A
trajectory type	membrane-bound protein
method	molecular dynamics
time-step	2 fs
sampling frequency	10 ps
total number of frames	566
computational platform	commodity Intel-based personal computer
software package	GROMACS 2.0
ensemble	<i>NpT</i> (isothermal–isobaric)
thermostat	Berendsen, 298 K
thermostat relax time	0.1 ps
barostat	Berendsen (anisotropic), 100 kPa
barostat relax time	1.0 ps
boundary condition	periodic
unit cell	cuboid
force field	GROMOS87
solvent	water
solvent force field	SPC
electrostatics treatment	cutoff, 1.8 nm
source Protein Data Bank identifier	1QD5
enzyme classification	3.1.1.32

into a lipid bilayer); and the equilibration protocol (including the preequilibration of solvents and ions).

The second set, mostly presentable in a formulaic table (Table 1), includes the following: MD software package (name and version); computer hardware and operating system used (general timing information if appropriate); force fields for both solute and solvent (name and version, including any special modifications); boundary condition and shape of unit cell; electrostatics treatment; ensemble; barostat (if used: type and target pressure); thermostat (if used: type and target temperature); constraints; time-step; snapshot sampling frequency; duration of simulated trajectory; and special restraints and interactive MD protocol.

Calls for a standard in the output of MD packages are particularly poignant here: all the items enumerated in the second set need to be reported in the entirety to facilitate comparison; this is not routinely done in the literature. The most convenient and sensible spot to capture these metadata is at the point of generation; that is, at the output of the MD package. However, at the moment, the defaults of each MD package (which may be different) are often not written out explicitly and have to be speculated downstream. Capturing metadata in detail also helps in avoiding inappropriate setup parameters from being used—for example, with the help of a validating program that alerts the scientist about inappropriate combinations of electrostatics treatment and solvent force field (among others).

### 3. THERMODYNAMIC MEASURES OF QUALITY, WHOSE REPORTING SHOULD BE OBLIGATORY

When MD simulations are performed, various quantities are calculated and written to output files. Some of these are well-understood and may be used to observe what is happening in a simulation as an indication of quality. For instance, we might expect the temperature in a simulation

with a thermostat applied to be fluctuating about a constant value, within a small range, over time; if this is not the case, then the quality of the simulation is suspicious. We have decided that these quality-indicating measures which we consider should be the following: temperature, pressure, potential energy, kinetic energy, number density, volume, cell dimensions, and specific heat capacity. There are some redundancies in reporting; this is to avoid the need for the users to recalculate often-used values, but the values also need to be verified to ensure consistency. “Quality” here means “overall thermodynamic stability” and does not necessarily guarantee the simulation’s accuracy in reproducing physical phenomena or the “usefulness” of the trajectory: Even an ill-parametrized force field can yield to stable but unphysical trajectories. Further, anomalies may occur due to local unbalanced distributions of kinetic energy; a careful human curator, rather than the reported values here, will have to be relied upon to catch these.

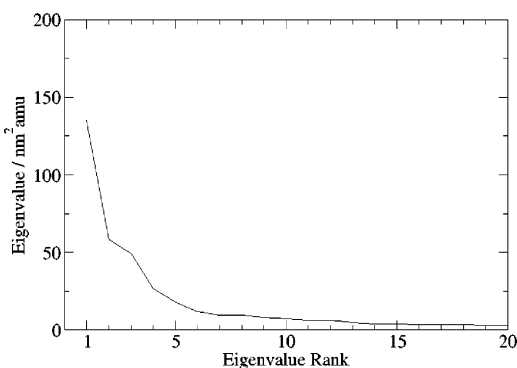
As their behavior is well-understood and they depend on the simulation conditions, these measures may be called quality measures for the trajectory. These should be obtained from the simulation output files, though we note that current MD packages usually do not write these out for every time-step by default; this arguably should be rectified to avoid data loss. These quantities should be plotted as a time-series. The mean and standard deviation should be reported, so automatic filters can be easily applied downstream to select only the trajectories over a quality threshold.

### 4. CONVERGENCE AND FLUCTUATION MEASURES, WHOSE REPORTING SHOULD BE OBLIGATORY

Once the quality of the simulation has been determined by examining the thermodynamic measures, it is instructive to inspect some measures that reflect the convergence and the fluctuations. As not all simulations are meant to be converging, these are not necessarily measures of the quality, but unexpected behaviors can lead to interesting investigations or possible problems. These useful measures are the root-mean-square deviation and fluctuation and the radius of gyration.

The root-mean-square deviation (RMSD) time-series throughout a trajectory can be used to understand by how much the conformation of a biopolymer changes with respect to time. The RMSD provides a measure of conformational stability or drift. For a converging simulation, we would expect the RMSD to increase and then start to plateau. (However, RMSD plateauing does not necessarily indicate convergence.) As we wish the same analysis to be performed automatically, we want a set of basic, specific rules that can be adhered to. First, all atoms in the biopolymer molecule are used in the least-squares fitting procedure to remove the translational and rotational degrees of freedom. The RMSD is calculated with respect to the initial configuration in the trajectory, and all atoms are used to calculate the RMSD. In addition, it may be useful to report RMSD for only the backbone or  $\alpha$ -carbon atoms of proteins or to exclude some loops with large fluctuations in RMSD calculations.

The root-mean-square fluctuation (RMSF), like the RMSD, is calculated for all atoms in each biopolymer. Before this calculation, all atoms included in the trajectory are used to remove the translational and rotational degrees of freedom.



**Figure 2.** The 20 eigenvalues with the largest magnitude for a 9.6 ns simulation of a prion protein are plotted to obtain a scree plot to determine the most prominent modes of motion.

The RMSF gives a measure of the fluctuation of atoms around the average position, and any large fluctuations should be understood in light of the crystallographic B factors.<sup>9,12</sup>

The radius of gyration gives a measure of how the mass of a group of atoms is distributed around their center of mass. For converged trajectories, the radius of gyration time-series of a biopolymer should also reach a plateau.

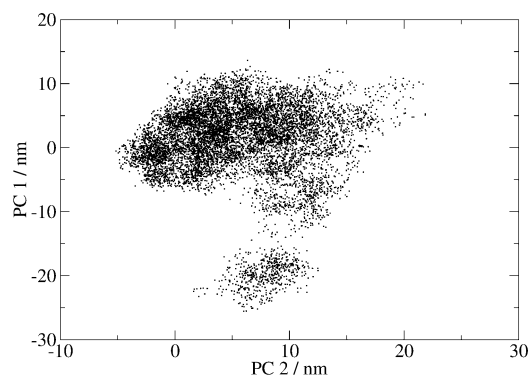
Further convergence measures such as the cosine content<sup>10,14</sup> and overlap measures from block (“windowed”) analyses<sup>15,16</sup> can be obtained from principal component analysis. For example, comparison of the first and last parts of a trajectory might help to detect inadequacies in equilibration or convergence that calls for a longer simulation time. If desirable, the obligation to report these will emerge from the simulation community.

#### 5. STRUCTURAL MEASURES, WHOSE REPORTING IS NOT OBLIGATORY BUT INFORMATIVE

A plot of eigenvalues from principal component analysis, ranked in decreasing magnitude, should usually show a “scree” shape; that is, an initial sharp drop in magnitude followed by a leveling of small eigenvalues. This indicates that only a few modes of motions have large length-scales, as the magnitudes of the eigenvalues tell the contribution to the total motion from the corresponding principal component. From this plot the appropriateness of the “essential dynamics” analysis,<sup>13</sup> where only the motion modes with the largest length-scales receive attention, may be determined. As the motion of all the constituent atoms is to be determined the mass-weighted covariance matrix<sup>14</sup> is used for the analysis. A plot where such a “scree” shape can be seen is shown in Figure 2.

For each of the few principal components with large eigenvalues (as shown in the scree plot), it is recommended to plot a graph of the projection of the trajectory onto the component as a time-series; here the projection in a particular principal component shows the degree of sampling within that component. This plot reveals the global-conformational “states”, or clusters of the same projections, visited. As the trajectory becomes longer, previously unvisited states can become visited (exploration); also, previously visited states may be visited again (revisiting).

Projections of the trajectory onto a pair of principal components with large eigenvalues provide a good idea of the phase-space sampled. With an energy value plotted on



**Figure 3.** Projection of a trajectory for a 9.6 ns simulation of a prion protein onto the two principal components with the largest eigenvalues showing the sampling in the first (PC 1) and second (PC 2) components.

the third axis, this is a way to visualize the energy landscape.<sup>17</sup>

An example of the comparison of the projections onto the two principal components with the largest eigenvalues is shown in Figure 3. In this example, two distinct regions in the phase-space have been sampled. It is possible to read from this graph whether there are revisiting events in some areas in the phase-space. This may be done by clustering or block analysis (discussed above); however, this process is highly dependent on the clustering algorithm, so we do not consider it to be obligatory. Again, if these are desirable, the obligation to report these can emerge from the simulation community.

To obtain more specific information on the structure of the biopolymers, it is advisable to perform analyses designed for the type of biopolymer in question. For example, secondary-structure determination over the trajectory can be done for proteins,<sup>18</sup> or the Curves analysis set for nucleic acids.<sup>19</sup> These may provide a more detailed picture on the quality of the simulation. In addition, quality indices used by the crystallography, nuclear magnetic resonance, and bioinformatics communities may be considered to complement the indices here.<sup>20</sup>

#### 6. CONCLUSION

In this Letter, we have suggested a scheme where a general understanding of a biosimulation trajectory may be obtained, with a set of strictly defined analyses. This set gives the researcher a reasonable overview to decide whether to investigate further. We have tried to design the analyses to be as simple and general as possible, while maintaining enough complexity to understand and distinguish the simulations.

We have implemented these measures in the BioSimGrid toolkit. This toolkit has been developed to enable users to perform predefined analyses easily within the analysis environment. When the archiving of a trajectory is complete, these analyses will automatically be performed on the recently deposited data via the standard tools included in the toolkit.

We consider the set of analyses suggested here to be adequate and hope that the simulation community will adopt it as a basis for development and discussion. Suggestions may be sent to the Editors of this journal or to the authors.

## ACKNOWLEDGMENT.

We thank all our colleagues for discussion, especially Oliver Beckstein, Richard Boardman, Leo Caves, Simon J. Cox, David Moss, Adrian Mulholland, and Paul Jeffreys. We also thank the anonymous referee for helpful comments. This project is funded by the Biotechnology and Biological Sciences Research Council, United Kingdom.

## REFERENCES

- (1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (2) Brown, I. D.; McMahon, B. CIF: the computer language of crystallography. *Acta Crystallogr.* **2002**, *B58*, 317–324.
- (3) Friend, G. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **1990**, *8*, 52–56.
- (4) van Gunsteren, W. F.; Mark, A. E. Validation of molecular dynamics simulation. *J. Chem. Phys.* **1998**, *108*, 6109–6116.
- (5) Tai, K.; Murdock, S.; Wu, B.; Ng, M. H.; Johnston, S.; Fangohr, H.; Cox, S. J.; Jeffreys, P.; Essex, J. W.; Sansom, M. S. P. BioSimGrid: towards a worldwide repository for biomolecular simulations. *Org. Biomol. Chem.* **2004**, *2*, 3219–3221.
- (6) Wozniak, J. M.; Brenner, P.; Thain, D.; Striegel, A.; Izaguirre, J. A. Generosity and Gluttony in GEMS: Grid Enabled Molecular Simulations. In Proceedings of 14th IEEE International Symposium on High-Performance Distributed Computing 2005, pp 191–200. DOI: 10.1109/HPDC.2005.1520959.
- (7) Tai, K.; Baaden, M.; Murdock, S.; Wu, B.; Ng, M. H.; Johnston, S.; Boardman, R.; Fangohr, H.; Cox, K.; Essex, J. W.; Sansom, M. S. P. Active-site dynamics of hydrolases: comparison of simulations using the BioSimGrid database. *J. Mol. Graphics Modell.* **2006**. DOI: 10.1016/j.jmgl.2006.08.010.
- (8) Mackay, A. L. Quaternion transformation of molecular orientation. *Acta Crystallogr.* **1984**, *A40*, 165–166.
- (9) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; Vol. LXXI of Wiley Series on Advances in Chemical Physics; Wiley-Interscience: New York, 1988.
- (10) Hess, B. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E* **2000**, *62*, 8438–8448.
- (11) Baaden, M.; Meier, C.; Sansom, M. S. P. A Molecular Dynamics Investigation of Mono and Dimeric States of the Outer Membrane Enzyme OMPLA. *J. Mol. Biol.* **2003**, *331*, 177–189.
- (12) McCammon, J. A.; Harvey, S. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, U.K., 1987.
- (13) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412–425.
- (14) Hess, B. Convergence of sampling in protein simulations. *Phys. Rev. E* **2002**, *65*, 031910.
- (15) Faraldo-Gomez, J. D.; Forrest, L. R.; Baaden, M.; Bond, P. J.; Domene, C.; Patargias, G.; Cuthbertson, J.; Sansom, M. S. P. Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations. *Proteins* **2004**, *57*, 783–791.
- (16) Flyvbjerg, H.; Petersen, H. G. Error estimates on averages of correlated data. *J. Chem. Phys.* **1989**, *91*, 461–466.
- (17) Becker, O. M. Principal coordinate maps of molecular potential energy surfaces. *J. Comput. Chem.* **1998**, *19*, 1255–1267.
- (18) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.
- (19) Stofer, E.; Lavery, R. Measuring the geometry of DNA grooves. *Biopolymers* **1994**, *34*, 337–346.
- (20) Kleywegt, G. J. Validation of protein crystal structures. *Acta Crystallogr. D* **2000**, *56*, 249–265.

CT6001708

## Feynman–Kleinert Linearized Path Integral (FK-LPI) Algorithms for Quantum Molecular Dynamics, with Application to Water and He(4)

Jens Aage Poulsen,<sup>\*,†</sup> Gunnar Nyman,<sup>†</sup> and Peter J. Rossky<sup>‡</sup>

*Physical Chemistry, Göteborg University, S-412-96, Göteborg, Sweden, and Institute for Theoretical Chemistry, Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712*

Received May 11, 2006

**Abstract:** The Feynman–Kleinert Linearized Path Integral (FK-LPI) representation of quantum correlation functions is extended in applications and algorithms. Diffusion including quantum effects for a flexible simple point charge model of liquid water is explored, including new tests of internal consistency. An ab initio quantum correction factor (QCF) is also obtained to correct the far-infrared spectrum of water. After correction, a spectrum based on a classical simulation is in good agreement with the experiment. The FK-LPI QCF is shown to be superior to the so-called harmonic QCF. New computational algorithms are introduced so that the quantum Boltzmann Wigner phase-space density, the central object in the implementation, can be obtained for arbitrary potentials. One scheme requires only that the standard classical force routine be replaced when turning from one molecular problem to another. The new algorithms are applied to the calculation of the Van Hove spectrum of liquid He(4) at 27 K. The spectrum moments are in very good agreement with the experiment. These observations indicate that the FK-LPI approach can be broadly effective for molecular problems involving the dynamics of light nuclei.

### 1. Introduction

Recently, a variety of computational schemes such as centroid molecular dynamics (CMD),<sup>1,2</sup> ring polymer molecular dynamics (RPMD),<sup>3,4</sup> forward–backward semiclassical dynamics,<sup>5</sup> the classical Wigner (CW) model,<sup>6–10</sup> mode coupling theory,<sup>11</sup> and analytic continuation methods<sup>12</sup> have been applied for modeling real many-body quantum dynamical processes in the condensed phase. Common to all of these methods is the focus on the time correlation function (CF) formalism: The process of interest is studied by evaluating its corresponding CF. For instance, the dynamic structure factor in neutron or X-ray scattering experiments is described by a Van Hove CF, and the rate of diffusion of a particle is obtained from its velocity CF. The accuracy of the different methods remains difficult to evaluate, but it appears that they

are all able to capture the main qualitative quantum effects that are relevant in a condensed phase, where quantum coherences are quenched because of the strong interatomic couplings.

The CW model, the subject of this paper, is perhaps conceptually the simplest of the methods considered above and has a rigorous derivation.<sup>9</sup> It has been successfully applied for obtaining vibrational relaxation rates of oxygen in liquid oxygen, determined by the golden rule force–force CF,<sup>10</sup> the diffusion coefficient for liquid para-hydrogen at 17 and 25 K, via the Kubo velocity CF,<sup>8</sup> neutron scattering via the Van Hove CF of liquid He(4) at 27 K,<sup>6</sup> and the diffusion coefficient plus a quantum correction factor for the infrared spectrum of water.<sup>7</sup> In short, the model splits the calculation of the quantum CF into two separate tasks, the generation of initial conditions and the propagation of dynamics. The sampling of the appropriate quantum initial conditions (here assumed to be given by a Boltzmann operator) is done through Wigner’s phase-space distribution,

\* Corresponding author e-mail: jens72@chem.gu.se.

<sup>†</sup> Göteborg University.

<sup>‡</sup> University of Texas at Austin.

while the dynamics are approximated as classical (the single, but great, simplification); see eq 1 below. To derive the CW method, one expresses the CF as a Feynman path integral (PI) and linearizes the action difference between the forward–backward time paths.<sup>9,13</sup> This *linearized path integral* (LPI) approximation, which is common to several recently proposed dynamics methods,<sup>9,13–16</sup> can be justified by a strong coherence loss for systems in which the dynamics of many degrees of freedom are relatively well-coupled to the degree of freedom being probed; see ref 9 for further details.

To make the computation of the Wigner-transformed Boltzmann operator tractable, one may recast the latter in a semianalytical form through the variational effective frequency theory as introduced by Giachetti and Tognetti<sup>17</sup> and independently by Feynman and Kleinert<sup>18</sup> (hereafter FK). A key ingredient in the FK theory is the concept of a Feynman path centroid and an effective potential, the centroid potential. Feynman’s centroid also plays a central role in CMD.<sup>1,2</sup> The special LPI implementation with its FK-based Wigner transform is denoted FK-LPI.<sup>9</sup> In the following, the acronyms LPI and CW will be used interchangeably, as they are formally equivalent.

In this paper, we apply the FK-LPI method to a realistic model of liquid water based on a simple point charge (SPC) model. We have already reported some results of a similar water study,<sup>7</sup> but here, the focus is a bit different: not only is the new water sample twice as large as the old one, but we also now focus explicitly on the consistency of the various FK-LPI CFs and compare the performance of a FK-LPI-based quantum correction factor (QCF) with the so-called harmonic QCF, when both are applied to correct the far infrared spectrum of water obtained by classical molecular dynamics. The consistency tests among alternative routes to a CF are not, in general, accessible in other schemes.

In the FK-LPI implementation, it is the computation of the centroid potential which is the computationally most demanding part, as discussed below. In some special cases, for example, for potentials that can be expressed by Gaussian functions or in a Fourier series, one may still derive analytic formulas for the centroid potential, but for general potentials, this is not possible. In this paper, we present computational schemes for determining the FK centroid potential—and thereby the Wigner-transformed Boltzmann operator—for systems described by *arbitrary* potentials. The implementations are sufficiently general to permit a user to make only small changes to a FK-LPI code when migrating from one physical problem to another. Indeed, one is only required to substitute one potential gradient routine for another, when shifting from, for example, SPC water to a He(4) liquid. We report such “black box” implementations of the FK-LPI theory to the determination of the spectrum of density fluctuations in liquid He(4) at 27 K, the first realistic problem to which the FK-LPI theory was applied.<sup>6</sup>

This paper is structured as follows. In section 2, we describe the FK-LPI theory and its general implementation. Section 3 contains a description of the specific FK-LPI water and He(4) implementations. In section 4, we report and compare FK-LPI results for the water quantum liquid structure, nuclear velocity, and infrared spectrum to both the

predictions of classical mechanics and experimental results. Also, we present the results of the “black box” implementations of FK-LPI when applied to the He(4) problem. We conclude in section 5 with a summary and further comment on the outlook for future applications of the method.

## 2. Theory and Implementation.

In this section, we present the basic equations of the FK-LPI method and review the iterative FK equations which must be solved in order to set up the Boltzmann Wigner transform. The “black box” versions of the FK equations are introduced, and finally, a brief introduction to the He(4) dynamic structure factor is given.

**2.1. The Classical Wigner or LPI Approach.** The LPI approach to the computation of quantum CFs is perhaps the simplest one to express. The precise expression that is derived from the linearization approximation seems to have appeared first in the work of Hernandez and Voth<sup>14</sup> and was developed independently as the linearized semiclassical initial value representation by Sun and Miller.<sup>19</sup> The LPI approximation to the CF  $\langle \hat{A}(0)\hat{B}(t) \rangle$  is simply

$$\langle \hat{A}(0)\hat{B}(t) \rangle \approx \frac{1}{(2\pi\hbar)^{3N}} \int \int \frac{dq dp}{Z} (\exp(-\beta\hat{H})\hat{A})_W[q, p] (\hat{B})_W[q_t, p_t] \quad (1)$$

which follows *directly* from the exact PI expression when carrying out the linearization procedure. Equation 1 may be interpreted/implemented as follows: Phase-space points  $(q, p)$  are sampled from the Wigner transform of  $\exp(-\beta\hat{H})\hat{A}$ , the transform being defined for an arbitrary operator  $\hat{C}$  by

$$(\hat{C})_W[x, p] \equiv \int_{-\infty}^{+\infty} d\eta \exp(-ip\eta/\hbar) \left\langle x + \frac{1}{2}\eta \left| \hat{C} \left| x - \frac{1}{2}\eta \right. \right. \right\rangle \quad (2)$$

$(q, p)$  are evolved classically according to the Hamiltonian  $H(q, p)$  to  $(q_t, p_t)$ , which serve as the phase-space arguments of  $(\hat{B})_W[q_t, p_t]$ .  $3N$  is the dimensionality of the problem.

**2.2. Feynman–Kleinert Wigner Transform.** Recently, we suggested a practical route to the Wigner transform of the  $\exp(-\beta\hat{H})\hat{A}$  operator, which is required by LPI.<sup>9</sup> This approach was based on combining the novel effective frequency variational theory of Feynman and Kleinert (FK)<sup>18</sup> with the quasi-density operator formalism of Jang and Voth.<sup>1,20</sup> The approach exploits the classical centroid phase-space variables  $(x_c, p_c)$ , where

$$x_c = \frac{1}{\beta\hbar} \int_0^{\beta\hbar} dt x(t) \quad (3)$$

and  $p_c$  is the corresponding centroid momentum. Presented in formal terms, in one dimension, one may approximate the Boltzmann operator by<sup>9</sup>

$$\exp(-\beta\hat{H}) \approx \int \int dx_c dp_c \rho_{\text{FK}}(x_c, p_c) \hat{\delta}_{\text{FK}}(x_c, p_c) \quad (4)$$

where  $\rho_{\text{FK}}(x_c, p_c)$  is the FK approximation to the centroid phase-space density:

$$\rho_{\text{FK}}(x_c, p_c) = \frac{1}{2\pi\hbar} \exp\left(-\beta \frac{p_c^2}{2M}\right) \exp[-\beta W_1(x_c)] \quad (5)$$

and  $W_1(x_c)$  is the corresponding FK approximation to the centroid potential. The operator  $\hat{\delta}_{\text{FK}}(x_c, p_c)$  is the so-called effective frequency quasi-density operator (QDO):

$$\hat{\delta}_{\text{FK}}(x_c, p_c) = \int \int dx dx' \sqrt{\frac{M\Omega(x_c)}{\pi\hbar\alpha}} \left| x' \right\rangle \left\langle x \right| \exp \left\{ i \frac{p_c}{\hbar} (x' - x) - \frac{M\Omega(x_c)}{\hbar\alpha} \left( \frac{x+x'}{2} - x_c \right)^2 - \frac{M\Omega(x_c)\alpha}{4\hbar} (x' - x)^2 \right\} \quad (6)$$

where  $\alpha$  is a function of the effective frequency,  $\Omega(x_c)$ , through the relation

$$\alpha = \coth \left[ \frac{\Omega(x_c)\hbar\beta}{2} \right] - \frac{2}{\Omega(x_c)\hbar\beta} \quad (7)$$

and the quantity  $\alpha$  is related to the smearing width  $a^2(x_c)$ <sup>18</sup> through

$$a^2(x_c) = \hbar\alpha/2M\Omega(x_c) \quad (8)$$

This width measures the thermal quantum ‘‘fuzziness’’ around the classical-like position  $x_c$ . To close the equations, the effective frequency is given by the mass-weighted classical Hessian averaged over the length  $a^2(x_c)$ :

$$\Omega^2(x_c) = \frac{1}{M} \int dy \frac{1}{\sqrt{2\pi a^2(x_c)}} V''(x_c + y) \exp \left[ -\frac{1}{2} y^2/a^2(x_c) \right] \quad (9)$$

This FK prescription gives the best local harmonic description of the potential surface, on the basis of a system free energy criterion.<sup>18</sup> To obtain  $\Omega^2(x_c)$  and  $a^2(x_c)$ , eqs 7–9 must be solved iteratively, as discussed in the next section. Wigner-transforming eq 4 then amounts to transforming  $\hat{\delta}_{\text{FK}}$ , eq 6, which can be done analytically:

$$[\hat{\delta}_{\text{FK}}(x_c, p_c)]_W[q, p] = \frac{2}{\alpha} \exp \left[ -\frac{M\Omega(x_c)}{\hbar\alpha} (q - x_c)^2 - \frac{1}{M\Omega(x_c)\alpha\hbar} (p - p_c)^2 \right] \quad (10)$$

Further, if  $\hat{A}$  is a relatively simple phase-space operator, the transform of  $\exp(-\beta\hat{H})\hat{A}$  can also be obtained. For details, we refer to ref 9.

**2.3. Feynman-Kleinert Iterative Equations.** It is in principle easy to generalize the effective frequency theory to many dimensions; see, for example, ref 24. In the multidimensional formulation, the effective frequency matrix  $\underline{\Omega}^2(\bar{z}_c)$  and the smearing width matrix  $\underline{\mathbf{A}}(\bar{z}_c)$  [the  $3N \times 3N$  dimensional generalizations of  $\Omega^2(x_c)$  and  $a^2(x_c)$ , respectively], defined for a position  $\bar{z}_c$ , are determined through two self-consistent equations:

$$\underline{\Omega}^2(\bar{z}_c) = \int d\bar{z} \frac{1}{\sqrt{\|2\pi\underline{\mathbf{A}}(\bar{z}_c)\|}} \underline{\mathbf{M}}^{-1/2} \underline{\mathbf{H}}_V(\bar{z}) \underline{\mathbf{M}}^{-1/2} \exp \left[ -\frac{1}{2} (\bar{z} - \bar{z}_c)^T \underline{\mathbf{A}}(\bar{z}_c)^{-1} (\bar{z} - \bar{z}_c) \right] \quad (11)$$

and

$$\underline{U}^\dagger(\bar{z}_c) \underline{\mathbf{M}}^{1/2} \underline{\mathbf{A}}(\bar{z}_c) \underline{\mathbf{M}}^{1/2} \underline{\mathbf{U}}(\bar{z}_c) = \underline{\Lambda} \quad (12)$$

where

$$\underline{\Lambda}_{ij} = \delta_{ij} \left\{ \frac{k_B T}{\Omega_i^2(\bar{z}_c)} \left[ \frac{\hbar\Omega_i(\bar{z}_c)}{2k_B T} \coth \left( \frac{\hbar\Omega_i(\bar{z}_c)}{2k_B T} \right) - 1 \right] \right\} \quad (13)$$

In these equations,  $\underline{\mathbf{H}}_V(\bar{z})$  is the  $3N \times 3N$  classical Hessian matrix of the potential,  $\underline{\mathbf{M}}^{1/2}$  is the diagonal matrix of the square root of masses, and  $\Omega_i(\bar{z}_c)$ ,  $i = 1$  and  $3N$ , is the square root of eigenvalues of  $\underline{\Omega}^2(\bar{z}_c)$ .  $\underline{\mathbf{U}}(\bar{z}_c)$  is the matrix which diagonalizes  $\underline{\Omega}^2(\bar{z}_c)$  and hence defines the effective frequency normal modes through

$$\bar{\eta} = \underline{U}^\dagger(\bar{z}_c) \underline{\mathbf{M}}^{1/2} \bar{z} \quad (14)$$

where  $\bar{z}$  holds the usual Cartesian coordinates.

In practice, eqs 11–12 are solved as follows: Let  $\bar{z}_c$  be given. Assume that the correct effective frequency matrix  $\underline{\Omega}^2(\bar{z}_c)$  has been found in a geometry  $\bar{z}'_c$  close to  $\bar{z}_c$ , for example, from a previous step in a Metropolis Monte Carlo (MC) walk on the centroid potential. Then, the steps are (i) diagonalize the effective frequency matrix to obtain an approximate  $\underline{\mathbf{U}}(\bar{z}_c)$  matrix and vector of frequencies  $\Omega_i(\bar{z}_c)$ . (ii) Use eq 13 followed by eq 12 to obtain  $\underline{\Lambda}(\bar{z}_c)$ . (iii) Equation 11 gives a new effective frequency matrix  $\underline{\Omega}^2(\bar{z}_c)$ . One then returns to step i until self-consistency for  $\underline{\Omega}^2(\bar{z}_c)$  and  $\underline{\Lambda}(\bar{z}_c)$  is reached. In practice, only a few iterations are necessary. Once the effective frequencies and normal modes are obtained, the centroid potential is calculated according to<sup>18</sup>

$$W_1(\bar{z}_c) = k_B T \sum_{i=1}^{3N} \ln \left( \frac{\sinh \left[ \frac{\hbar\Omega_i(\bar{z}_c)}{2k_B T} \right]}{\frac{\hbar\Omega_i(\bar{z}_c)}{2k_B T}} \right) + V_{A(\bar{z}_c)} - \frac{1}{2} \sum_{i=1}^{3N} \Lambda_{ii} \Omega_i^2(\bar{z}_c) \quad (15)$$

with the smeared potential given by the Gaussian average

$$V_{A(\bar{z}_c)} = \int d\bar{z} \frac{1}{\sqrt{\|2\pi\underline{\mathbf{A}}(\bar{z}_c)\|}} V(\bar{z}) \exp \left[ -\frac{1}{2} (\bar{z} - \bar{z}_c)^T \underline{\mathbf{A}}(\bar{z}_c)^{-1} (\bar{z} - \bar{z}_c) \right] \quad (16)$$

Also, one may set up the Boltzmann Wigner transform, compare eq 10:

$$[\hat{\delta}_{\text{FK}}(\bar{z}_c, \bar{p}_c)]_W[\bar{q}, \bar{p}] = \prod_{i=1}^{3N} \frac{2}{\alpha_i} \exp \left[ -\frac{\Omega_i(\bar{z}_c)}{\alpha_i \hbar} (\eta_i - \eta_{i,c})^2 - \frac{1}{\Omega_i(\bar{z}_c) \alpha_i \hbar} (\nu_i - \nu_{i,c})^2 \right] \quad (17)$$

with the mass-weighted position and momentum normal modes,  $\eta_i$  and  $\nu_i$ , being the  $i$ th entries of

$$\bar{\eta} = \underline{U}^\dagger(\bar{z}_c) \underline{\mathbf{M}}^{1/2} \bar{q}, \quad \bar{\nu} = \underline{U}^\dagger(\bar{z}_c) \underline{\mathbf{M}}^{-1/2} \bar{p} \quad (18)$$

respectively. Two corresponding relations define  $\eta_{c,i}$  and  $\nu_{c,i}$  from the Cartesian centroid coordinates and momenta:



$$\vec{\eta}_c = \underline{U}^\dagger(\vec{z}_c) \underline{\mathbf{M}}^{1/2} \vec{z}_c, \quad \vec{v}_c = \underline{U}^\dagger(\vec{z}_c) \underline{\mathbf{M}}^{-1/2} \vec{p}_c \quad (19)$$

**2.4. General Implementations of the FK Equations.** The major difficulty encountered in applying the FK-LPI method is the evaluation of eq 11, the smeared Hessian, and to a smaller extent, eq 16, the smeared potential. In this paper, we consider liquids described by nonpolarizable potentials, where efficient analytical expressions for these equations can be derived, see below. However, the derivation of the analytic equations require a fair amount of algebra, and alternative ways of solving the iterative FK equations for general potentials are desirable. Below, we consider two new implementations of the iterative equations which—to various degrees—eliminate the algebraic problem and extend the types of potentials accessible.

**2.4.1. Hessian Sampling.** A general way to evaluate eqs 11 and 16 would be to sample the classical Hessian and potential numerically. Thus, first, the multidimensional Gaussian sampling function, which is identical in these equations, is rephrased in normal mode coordinates, eq 14, which makes the Gaussian factorize. Then, one samples normal mode displacements from this Gaussian via Box–Muller sampling,<sup>25</sup> and these are converted back to Cartesian coordinates for evaluation of the Hessian or potential. As we shall see in a moment, this procedure works extremely well if one has an analytic Hessian at hand, which is usually possible if the potential is nonpolarizable. The method then needs routines directly providing both the classical Hessian and potential.

**2.4.2. Gradient Sampling.** This procedure only requires a routine that provides the potential and its gradient. These are available in any molecular dynamics code. To derive this method, we start by integrating eq 11 by parts, which leads immediately to

$$\underline{\underline{\Omega}}^2(\vec{z}_c) = \int d\vec{z} \frac{1}{\sqrt{||2\pi\mathbf{A}(\vec{z}_c)||}} \underline{\mathbf{M}}^{-1/2} \underline{\Delta}(\vec{z}, \vec{z}_c) \underline{\mathbf{M}}^{-1/2} \exp\left[-\frac{1}{2}(\vec{z} - \vec{z}_c)^T \mathbf{A}(\vec{z}_c)^{-1} (\vec{z} - \vec{z}_c)\right] \quad (20)$$

where

$$\underline{\Delta}_{ij}(\vec{z}, \vec{z}_c) = \left\{ \sum_{k=1}^{3N} (\underline{\mathbf{A}}^{-1})_{kj} (z_k - z_{c,k}) \right\} \times \frac{\partial}{\partial z_i} V(\vec{z}) \quad (21)$$

Hence, the scheme for calculating the smeared Hessian requires sampling vectors  $\vec{z}$  around  $\vec{z}_c$  and averaging to obtain the  $(3N)^2$  quantities

$$\Gamma_{ij} = \left\langle (z_i - z_{c,i}) \frac{\partial}{\partial z_j} V(\vec{z}) \right\rangle \quad (22)$$

The  $(i,j)$ th element of the smeared Hessian is then given by

$$\underline{\underline{\Omega}}_{ij}^2(\vec{z}_c) = m_i^{-1/2} m_j^{-1/2} \sum_{k=1}^{3N} \{ \underline{\mathbf{A}}^{-1} \}_{kj} \Gamma_{ki} \quad (23)$$

In practice, because of statistical errors in any finite sampling sequence,  $\underline{\underline{\Omega}}_{ij}^2(\vec{z}_c)$  will not be symmetric, and this can be enforced by a “symmetrization” procedure:

$$\underline{\underline{\Omega}}_{ij}^2(\vec{z}_c) \rightarrow [\underline{\underline{\Omega}}_{ij}^2(\vec{z}_c) + \underline{\underline{\Omega}}_{ji}^2(\vec{z}_c)]/2 \quad (24)$$

**2.5. Dynamic Structure Factor.** The spectrum of density fluctuations—or equivalently the dynamic structure factor—may be defined for the reciprocal lattice vector  $\vec{\mathbf{Q}}$  through<sup>26</sup>

$$S(\vec{\mathbf{Q}}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dt \exp(-i\omega t) S(\vec{\mathbf{Q}}, t) \quad (25)$$

where the Van Hove CF is given by<sup>26</sup>

$$S(\vec{\mathbf{Q}}, t) = \frac{1}{N} \sum_{j,k=1}^N \langle \exp[-i\vec{\mathbf{Q}} \cdot \vec{r}_j(0)] \exp[i\vec{\mathbf{Q}} \cdot \vec{r}_k(t)] \rangle \quad (26)$$

The dynamic structure factor is experimentally accessible through neutron or X-ray scattering experiments. From the experiment,<sup>27</sup> one may extract the first three moments (time derivatives) of  $S(\vec{\mathbf{Q}}, \omega)$  ( $S(\vec{\mathbf{Q}}, t)$ ). These moments thus serve as a link between the experiment and simulation.

In any liquid model that employs periodic boundary conditions, the reciprocal lattice vector must fulfill the Laue equation:  $Q \times L = 2\pi n$ , where  $Q$  is a component of  $\vec{\mathbf{Q}}$ ,  $L$  is the box length in that direction, and  $n$  is an integer.

### 3. Water and He(4) FK-LPI Molecular Dynamics Implementation.

**3.1. Water Potential and Its Implementation in the Iterative FK Equations.** The liquid water model is constructed by adopting the SPC model of Berendsen et al.<sup>28</sup> In the SPC model, the water interactions are described by Lennard-Jones (LJ) interactions between oxygen atoms ( $\epsilon = 78.22$  K;  $\sigma = 3.165$  Å) and an electrostatic Coulomb potential between atoms on different molecules, the latter defined by charges  $0.41e$  and  $-0.82e$ , on hydrogen and oxygen, respectively. A flexible water model is introduced by adopting the harmonic part of the local mode intramolecular water potential of Reimers and Watts,<sup>29</sup> see eq 30 below. The water monomer geometry is put equal to the equilibrium geometry of the original rigid SPC model. In the following, we will refer to the model as SPCf (f for flexible). We will make no effort to reparametrize the SPCf potential parameters in the current work, in order, for example, to develop a model which agrees better with the experiment after quantization.

To handle the long-range electrostatic interactions, the Ewald summation method is adopted with conducting tin foil boundary conditions. For an exceptionally clear and pedagogical account of the Ewald summation technique, see the user’s guide and manual of the MOSCITO simulation package.<sup>30</sup> In the real space part, we employ the standard approximation of including only the central box in the summation. A total of 337 reciprocal space vectors corresponding to  $\vec{\mathbf{k}} = 2\pi(n_x, n_y, n_z)/L$ ,  $n_x^2 + n_y^2 + n_z^2 \leq 27$  are utilized in the Ewald reciprocal space sum, and the Gaussian charge distribution parameter  $\alpha$  is set to  $0.43135$  Å<sup>-1</sup> in accordance with the general suggestion<sup>30</sup> ( $5.37/L$ ), with  $L = 12.4255$  Å being the length of the simulation box. We consider 64 water molecules equivalent to a density of  $0.99$  g/cm<sup>3</sup>. The minimum image convention together with a spherical cutoff at half the box length is adopted for the short-

**Table 1.** Intramolecular Potential Parameters in SPCf Simulation

$i$	$D/\text{KJ mol}^{-1}$	$\alpha_i$
1	549.0586	2.13498 Å <sup>-1</sup>
2	414.5975	0.70337

ranged part of the potential. A time step of 1 fs is used for dynamics, and all velocity CFs are calculated up to 2.5 ps.

To derive an analytic expression for the smeared Hessian, we first decompose the total SPCf potential as

$$V_{\text{H}_2\text{O}} = V_{\text{LJ}}^{\text{OO}} + V_{\text{Ewald}} + V_{\text{intra}} \quad (27)$$

with

$$V_{\text{LJ}}^{\text{OO}} = 4\epsilon \left\{ \left( \frac{\sigma}{r_{\text{OO}}} \right)^{12} - \left( \frac{\sigma}{r_{\text{OO}}} \right)^6 \right\} \quad (28)$$

and

$$V_{\text{Ewald}} = V_{\text{realspace}} + V_{\text{rec.space}} - V_{\text{self}} - V_{\text{molc.selfint.}} \quad (29)$$

$$V_{\text{intra}} = V_{\text{OH1}} + V_{\text{OH2}} + V_{\Theta} = D_1 \alpha_1^2 (r_{\text{OH1}} - r_{\text{eq}})^2 + D_1 \alpha_1^2 (r_{\text{OH2}} - r_{\text{eq}})^2 + D_2 \alpha_2^2 (\Theta_{\text{HOH}} - \Theta_{\text{eq}})^2 \quad (30)$$

The Ewald real space part, reciprocal space part, and charge self-interaction and molecular self-interaction parts are given by<sup>30</sup>

$$V_{\text{realspace}} = \sum_i \sum_{j>i} q_i q_j \frac{\text{erfc}(\alpha r_{ij})}{r_{ij}} \quad (31)$$

and

$$V_{\text{rec.space}} = \frac{1}{2} \sum_{\mathbf{k} \neq 0} \sum_{ij} 4\pi \frac{q_i q_j}{V \|\mathbf{k}\|^2} \exp(-\|\mathbf{k}\|^2/4\alpha^2) \exp[-i\mathbf{k}(\vec{r}_i - \vec{r}_j)] \quad (32)$$

$$V_{\text{self}} = \frac{\alpha}{\sqrt{\pi}} \sum_i q_i^2 \quad (33)$$

$$V_{\text{molc.selfint.}} = \sum_{\text{pairs}(i,j)} \frac{q_i q_j}{r_{ij}} \text{erf}(\alpha r_{ij}) \quad (34)$$

respectively. The parameter values for the intramolecular potential,  $D_1$ ,  $D_2$ ,  $\alpha_1$ , and  $\alpha_2$ , are listed in Table 1; these are obtained by requiring eq 30 to match the harmonic part of the local mode intramolecular potential of Reimers and Watts.<sup>29</sup> The values of  $r_{\text{eq}}$  and  $\Theta_{\text{eq}}$  are however kept equal to their original SPC values:<sup>28</sup> an angle of 109.5° and a bond length of 1.0 Å. To proceed, we express the OH local mode intramolecular energy and oxygen LJ interactions in a “sum over Gaussian” form, that is

$$V_{\text{OH}_j} = \sum_{i=1,3} \beta_i \exp\left(-\frac{1}{2} r_{\text{OH}_j}^2 / \zeta_i\right) + \zeta, \quad j = 1, 2 \quad (35)$$

and

$$V_{\text{LJ}}^{\text{OO}} = \sum_{i=1,3} \gamma_i \exp\left(-\frac{1}{2} r_{\text{OO}}^2 / \alpha_i\right) \quad (36)$$

Also, the Ewald real space part and Ewald molecular self-interaction term are put in a “sum over Gaussian” form by writing

$$\frac{\text{erfc}(\alpha r_{ij})}{r_{ij}} = \sum_{l=1,3} \delta_l \exp\left(-\frac{1}{2} r_{ij}^2 / \eta_l\right) \quad (37)$$

and

$$\frac{\text{erf}(\alpha r_{ij})}{r_{ij}} = \sum_{l=1,2} \tau_l \exp\left(-\frac{1}{2} r_{ij}^2 / \zeta_l\right) \quad (38)$$

respectively. The Gaussian parameters are collected in Table 2. The Gaussian representation of the classical electrostatic energy is tested by comparing its predicted energy to the energy of the original representation, eqs 31–34, for two randomly sampled configurations of the liquid. For this purpose, we use the MOSCITO 4 simulation package<sup>30</sup> with the same Ewald setup. An acceptable energy error of 0.1% was found for both geometries. Because Gaussian averages of multidimensional Gaussian and complex exponentials (the Ewald reciprocal space term) can be performed algebraically, see, for example, ref 24, it follows that  $V_{\text{H}_2\text{O}, \underline{\Delta}(\vec{z}_c)}$  and  $\underline{\Omega}^2(\vec{z}_c)$  can be evaluated analytically, except for the intramolecular bending mode term, the latter being a function of all three water monomer atoms. To obtain the smeared Hessian and smeared potential of  $V_{\Theta}$ , we adopt the following two approximations: First, the smeared Hessian of  $V_{\Theta}$  is approximated by its classical Hessian, evaluated at the centroid. Second, we notice that the smeared potential, eq 16, may be expanded using the smearing width matrix as a smallness parameter:

$$V_{\Theta, \underline{\Delta}(\vec{z}_c)} = V_{\Theta}(\vec{z}_c) + \frac{1}{2} \sum_{ij} \underline{A}(\vec{z}_c)_{ij} \frac{\partial^2}{\partial z_{c,i} \partial z_{c,j}} V_{\Theta}(\vec{z}_c) + \dots \quad (39)$$

Obviously, if we truncate after the first nontrivial term, being linear in the smearing width matrix, we obtain an expression that requires only the analytic *classical* Hessian of the bending mode potential. Both approximations are exact for harmonic systems and for high temperatures. The Gaussian potentials in eqs 35–38 were also adopted in the classical propagation part of the LPI procedure.

We use two cycles, i–iii, of the FK iterative equations (section 2.3); for each cycle, the  $576 \times 576$  effective frequency matrix is diagonalized. After self-consistency has been reached, the centroid potential and Wigner-transformed quasi-density operator are obtained, see eqs 15, 17–19 or ref 9 for more details. The temperature  $T$  is set to 296 K.

The integration over centroid positions, compare eq 4, is done by Metropolis MC,<sup>31</sup> utilizing the centroid density as a weight function. The trial moves involve simultaneous displacements of *all* atoms and are decomposed by the following sequential operations: (i) a uniform random translation of each atom, (ii) a uniform random rotation of

**Table 2.** Gaussian Fit Parameters Used in SPCf Simulation<sup>a</sup>

<i>i</i>	1	2	3
$\beta/\text{Hartree}$	7.1122	-8.73264	1.12917
$\gamma/\text{Hartree}$	1.77205	2.13115	89.459
$\delta/\text{bohr}^{-1}$	0.19716251	0.3483857	0.088230
$\tau/\text{bohr}^{-1}$	0.10209604	0.15546671	
$\zeta/\text{bohr}^2$	2.14823	2.81753	3.1732
$\alpha/\text{bohr}^2$	2.1315	13.4881	1.11393
$\eta/\text{bohr}^2$	3.36579	1.3012714	7.443253
$\varsigma/\text{bohr}^2$	13.79489	101.44204	

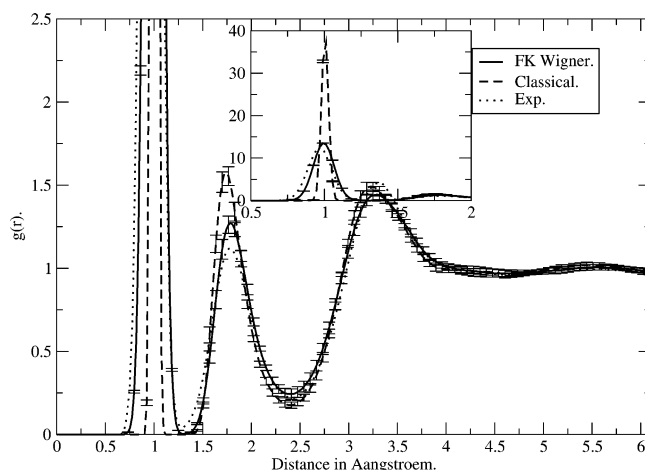
<sup>a</sup> The value of  $\zeta$  in eq 35 is  $\zeta = 0.892\,724$  Hartree.

each molecule, (iii) a uniform random translation of each molecule, and finally (iv), again, a uniform random translation of each atom. Thus, the transition probability for trial configurational moves is symmetric in the configurational state, and we may utilize the usual Metropolis criterion for acceptance; see page 121 of ref 31. The step parameters are 0.0104 Å, 0.0334 rad, and 0.00175 Å for center-of-mass translation, rotation, and atomic translation, respectively.

To estimate the uncertainties in the calculated quantities, both classical and quantum calculations are repeated using 10 (7) different equilibrated initial water configurations for the classical (quantum) simulations. For the quantum simulations, 90 000 MC centroid steps are utilized to equilibrate the liquid, starting from a classical equilibrium geometry. For the production runs, for each geometry, the number of centroid MC steps is about 30 000–45 000 for the quantum calculations and 45 000–90 000 for the classical simulations. For each 20th configuration, the corresponding Wigner-transformed QDO is constructed and 10 sets of position and momenta are sampled for subsequent classical dynamics. For the quantum calculations, 45 000 MC samplings took approximately 6 days using a computer equipped with a single AMD 2.2 GHz Opteron processor.

**3.2. He(4) Model.** The calculation of the spectrum of density fluctuations for He(4) is done at 27 K and a density of 0.25 g/cm<sup>3</sup>. The considered  $\mathbf{Q}$  vector is 1.64 Å<sup>-1</sup>. The liquid model/calculation is exactly as that in ref 6. That is, the He(4)–He(4) interactions are determined by the HFD-B2 potential of Aziz and co-workers,<sup>32</sup> which is rewritten as a sum over Gaussian functions to derive analytical expressions for eqs 11 and 16. The new calculation differs however from the original calculation in one significant way: instead of 64 He(4) atoms, we now consider 57 atoms, which makes the Laue equation exactly fulfilled. The new 57 atom box length is 11.494 Å, and hence,  $\mathbf{Q} = 1.64$  Å<sup>-1</sup>. The Laue condition was only approximately satisfied in the old work.<sup>6</sup>

We perform three different calculations, which use (i) the original analytical Gaussian-fit code,<sup>6</sup> (ii) the Hessian sampling strategy, and (iii) the gradient procedure, eqs 20–23. Both numerical methods, ii–iii, still utilize the Gaussian representation of the He(4) potential, so that any discrepancy between results of i and ii–iii may be ascribed to shortcomings of the new numerical implementations. In method ii, 160 and 1000 Box–Mueller samplings are adopted to evaluate eqs 11 and 16, respectively, while in method iii,



**Figure 1.** Radial distribution function  $g_{\text{OH}}(r)$  for SPCf water model, evaluated via alternative methods, as indicated in the legend. Experimental data from ref 33.

2500 and 1000 Box–Mueller samplings are adopted to evaluate eqs 22 and 16, respectively. We utilize  $3 \times 300\,000$  centroid MC steps for methods i and ii, while 135 000 centroid MC steps were performed for the more time-consuming method iii.

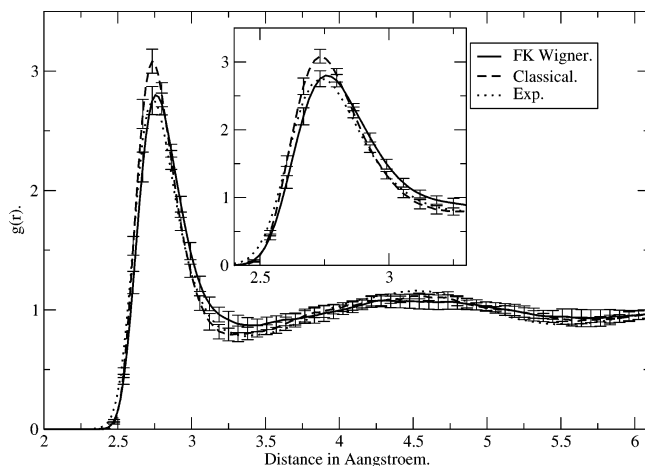
These calculations are performed on a single computer equipped with a single 2 GHz Xeon processor. The analytic and the Hessian sampling codes use less than 1 week for 300 000 MC steps including classical dynamics, while the gradient sampling implementation requires about 30 days to perform 135 000 centroid MC steps (including dynamics).

## 4. Results

In this section, we present the structural and dynamic results of the water and He(4) simulations.

**4.1. Equilibrium Properties of the Water Model.** As in ref 7, one may consider the distribution of the FK effective frequencies and compare it with the density of classical instantaneous normal-mode frequencies. The result for the present model is, practically speaking, identical to the distribution given earlier for the 32-molecule study.<sup>7</sup> This means that the FK Wigner transform is well-defined for all centroid positions. As discussed in ref 7, for the vast majority of cases with one or more imaginary FK frequencies, the equations provide no ambiguity, and for the rare cases that do, a natural assignment (of zero momentum) is available for that mode.

The classical, FK Wigner-based, and experimental<sup>33</sup> radial distribution functions (RDFs) for O–H and O–O atom pairs,  $g_{\text{OH}}(r)$  and  $g_{\text{OO}}(r)$  are shown in Figures 1 and 2, respectively. While these results can be obtained essentially exactly via path integral Monte Carlo simulation (see, for example, ref 34), here, these calculations test the accuracy of the effective frequency local harmonic approximation. As is evident, the general trend is that which is expected: the quantum water RDFs are less structured and the peaks are shifted to longer distances as compared to those of the classical RDFs.<sup>21</sup> Beginning with  $g_{\text{OH}}(r)$ , we see that the Wigner-based  $g_{\text{OH}}(r)$  agrees well with the experimental



**Figure 2.** Radial distribution function  $g_{OO}(r)$  for SPCf water model, evaluated via alternative methods, as indicated in the legend. Experimental data from ref 33.

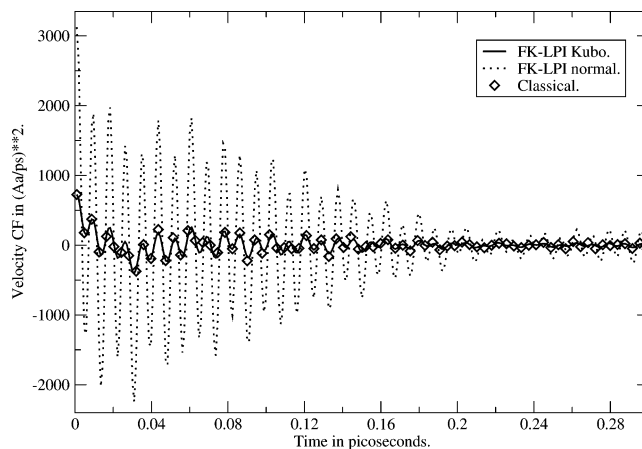
$g_{OH}(r)$ . Not only are the experimental and Wigner intramolecular distributions (at about 1 Å) in good agreement with each other, so are the peak heights of the next two peaks, corresponding to the two hydrogen atoms belonging to a water molecule in the first solvation shell. More specifically, classically, the first intermolecular peak is higher than the second, whereas both the quantum and experimental RDFs predict the reverse order for amplitudes. This is because the closer, hydrogen-bonded, H at about 2 Å is broadened significantly by zero-point effects. From Figure 2, we observe that  $g_{OO}(r)$  matches reasonably with the height of the experimental first peak, while the classical RDF exaggerates the structure of the liquid.

**4.2. Dynamical Results for Water Model.** *4.2.1. Diffusion.* We start by considering the diffusion coefficient  $D$  of liquid water, which can be extracted from the oxygen or hydrogen velocity CF (VCF) through the transport coefficient expression

$$D = \frac{1}{3} \int_0^{\infty} dt \langle \vec{v}(t) \vec{v}(0) \rangle \quad (40)$$

In Figure 3, we display the hydrogen atom VCF obtained directly from the FK-LPI approximation and also by classical MD. We further show the so-called Kubo-transformed VCF as calculated by FK-LPI.<sup>8</sup> This special VCF resembles the classical VCF in many respects: it is real-valued, and it shares the same initial value as its classical counterpart. Still, the integrated Kubo-transformed VCF yields the correct quantum, not the classical diffusion coefficient, see refs 8 and 12. From Figure 3, we observe the larger initial amplitude of the FK-LPI normal VCF as compared to the formally equal values for the classical and Kubo FK-LPI VCF. The intramolecular motion is also much more visible in the FK-LPI normal VCF, a fact which may be ascribed to zero-point motion.

In Table 3, the values for  $D$  obtained from classical MD, FK-LPI, and the FK-LPI Kubo VCF are presented. All values are extracted using the oxygen VCF. We show also the value of  $D$  reported by Lobaugh and Voth,<sup>35</sup> using CMD with two



**Figure 3.** Classical, normal FK-LPI, and FK-LPI Kubo hydrogen velocity correlation functions for SPCf water model.

flexible SPC potentials (SPC/F and SPC/F2), and by Miller and Manolopoulos,<sup>4</sup> employing RPMD with the rigid SPC/E potential. First, we observe the required consistent agreement between the two FK-LPI diffusion coefficients, which are also identical, within statistical accuracy, to the values obtained from integrating the hydrogen atom Kubo and normal VCFs. A straightforward comparison between the FK-LPI and CMD/RPMD values are difficult for several reasons. First, the  $D$ s depend—not surprisingly—on the chosen water potential. This is clearly illustrated by considering the results of the SPC/F and SPC/F2 potentials, which differ only by the form of their intramolecular force field. As seen from Table 3, there is a large difference between the classical  $D$ s obtained from these potentials. Second, as shown by Yeh and Hummer,<sup>36</sup>  $D$  increases linearly with the negative inverse of the box length, and using the viscosity of the liquid, one should first extrapolate to infinite system size in order to establish a correct model value for  $D$ . In this connection, it is interesting to note that Miller and Manolopoulos<sup>4</sup> found that the *ratio* of quantum to classical diffusion coefficients is almost constant as a function of the system size. Perhaps, this ratio is also the most interesting quantity when comparing quantum and classical models, because the models here do not produce an accurate agreement with the experiment. In any case, it appears that the quantum effect predicted by FK-LPI is somewhat larger than that for the other methods.

*4.2.2. Spectrum of the Hydrogen Velocity CF and Dipole Moment Quantum Correction Factor.* The water IR spectrum can be obtained once the water total system dipole moment CF,  $\langle \vec{M}(t) \vec{M}(0) \rangle_{\text{QM}}$ , has been computed,<sup>38,39</sup> for example, via a FK-LPI calculation. However, even on a classical level, the calculation of  $\langle \vec{M}(t) \vec{M}(0) \rangle_{\text{CL}}$  converges rather slowly,<sup>37</sup> and it is therefore desirable to circumvent a direct FK-LPI attack on this CF. As shown by Marti and co-workers,<sup>37</sup> one may utilize the hydrogen atom VCF for mimicking the total system dipole moment derivative CF for water,  $\langle \dot{\vec{M}}(t) - \dot{\vec{M}}(0) \rangle_{\text{QM}}$ . In this approximation, the ratio—or QCF  $Q(\omega)$ —between the quantum and classical absorption spectra is identical to the ratio between the quantum FK-LPI and classical hydrogen VCF spectra:

**Table 3.** Water Diffusion Coefficients Obtained from Various Methods<sup>a</sup>

	RPMD(SPC/E) <sup>b</sup>	FK-LPI(SPCf), 64 molecules	CMD(SPC/F) <sup>c</sup>	CMD(SPC/F2) <sup>c</sup>
$D/qm$	0.43	$0.50 \pm 0.06/0.49 \pm 0.06^d$	0.42	0.38
$D/cl$	0.29	$0.23 \pm 0.06$	$0.3 \pm 0.02$	$0.22 \pm 0.02$
ratio	1.48	2.2/2.1	1.4	1.7

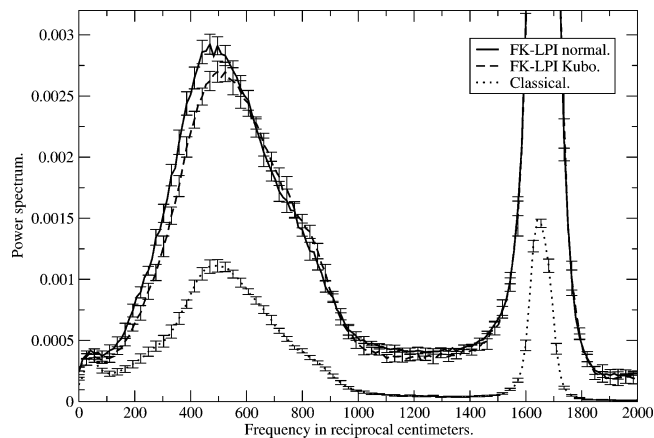
<sup>a</sup> Units Are  $\text{\AA}^2/\text{ps}$  <sup>b</sup> Data from ref 4 extrapolated to infinite system size. <sup>c</sup> Data from ref 35, using 125 molecules. <sup>d</sup> Two quantum values are the values obtained from FK-LPI Kubo and normal oxygen VCF, respectively.

$$\begin{aligned}
 Q(\omega) &= \frac{\int dt \exp(i\omega t) \langle \vec{M}(t) \vec{M}(0) \rangle_{\text{QM}}}{\int dt \exp(i\omega t) \langle \vec{M}(t) \vec{M}(0) \rangle_{\text{CL}}} = \\
 &= \frac{\int dt \exp(i\omega t) \langle \dot{M}(t) \dot{M}(0) \rangle_{\text{QM}}}{\int dt \exp(i\omega t) \langle \dot{M}(t) \dot{M}(0) \rangle_{\text{CL}}} \\
 &\approx \frac{\int dt \exp(i\omega t) \langle \vec{v}_H(t) \vec{v}_H(0) \rangle_{\text{FK-LPI}}}{\int dt \exp(i\omega t) \langle \vec{v}_H(t) \vec{v}_H(0) \rangle_{\text{CL}}} \equiv Q_{\text{FK-LPI}}(\omega)
 \end{aligned}
 \quad (41)$$

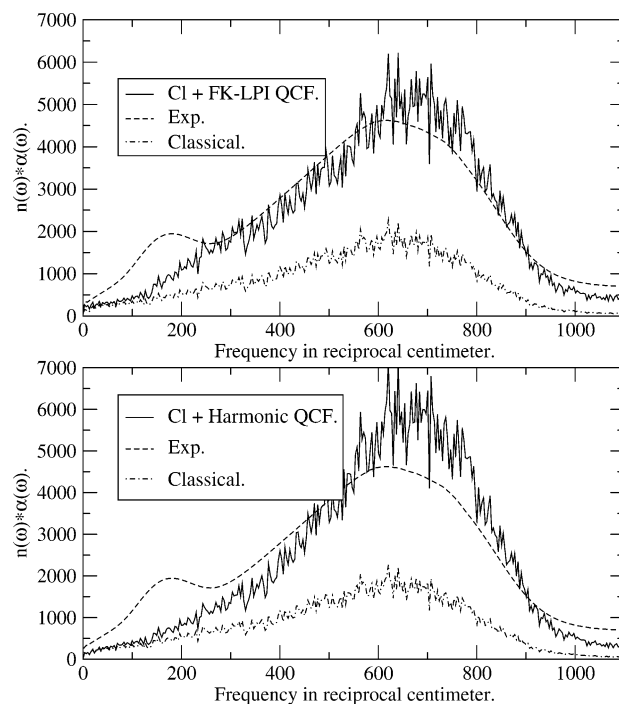
$Q_{\text{FK-LPI}}(\omega)$  is relatively inexpensive to obtain because it requires only a FK-LPI simulation of a single particle property. Knowing  $Q_{\text{FK-LPI}}(\omega)$ , thus, enables us to extract the quantum IR spectrum by performing only a classical simulation of the water total system dipole moment CF.

Referring to Figure 3, one sees that the Kubo and classical hydrogen VCF are rather close to each other. A comparison of the Fourier transform (spectrum) of the quantum hydrogen VCF as predicted by the FK-LPI VCF and the same function as derived from the FK-LPI Kubo-transformed VCF is informative. A standard Fourier relation may be used to map the FK-LPI Kubo VCF spectrum onto the normal VCF spectrum, see ref 40. Hence, the normal VCF spectrum can be obtained in two ways: directly or indirectly. For an exact theory, these two spectra would be identical. Here, the consistency is an important internal test of the theory, as there is no formal requirement that the two power spectra agree in a given approximation. In Figure 4 is shown a comparison between the classical hydrogen VCF spectrum, the normal FK-LPI hydrogen VCF spectrum, and the normal hydrogen VCF spectrum obtained from the FK-LPI Kubo hydrogen VCF. Clearly, good agreement between the two FK-LPI spectra is observed, especially at higher frequencies. The two FK-LPI spectra agree completely in the bending and OH stretching regions (the latter is not shown) of the spectrum, an observation which is consistent with the fact that the underlying FK-LPI theory is exact for harmonic potentials.

As discussed above, the ratio between the classical and FK-LPI hydrogen VCF spectrum may be used to define a FK-LPI QCF,  $Q_{\text{FK-LPI}}(\omega)$ . We apply  $Q_{\text{FK-LPI}}(\omega)$  as a multiplicative correction to the infrared spectrum obtained from classical molecular dynamics. We apply  $Q_{\text{FK-LPI}}(\omega)$ , obtained using the Kubo VCF for H, to the infrared spectrum reported by Jeon et al.,<sup>41</sup> who utilized a particular flexible SPC variant, often referred to as SPC/F.<sup>42</sup> The computed and experimental<sup>43</sup> IR spectra are shown in Figure 5, where also the spectrum obtained by applying (i) no correction (classical) and (ii) a correction based on the so-called harmonic QCF has been displayed; the last of which has very recently been



**Figure 4.** Spectra of hydrogen velocity correlation functions derived from classical, normal FK-LPI, and FK-LPI Kubo velocity correlation functions.



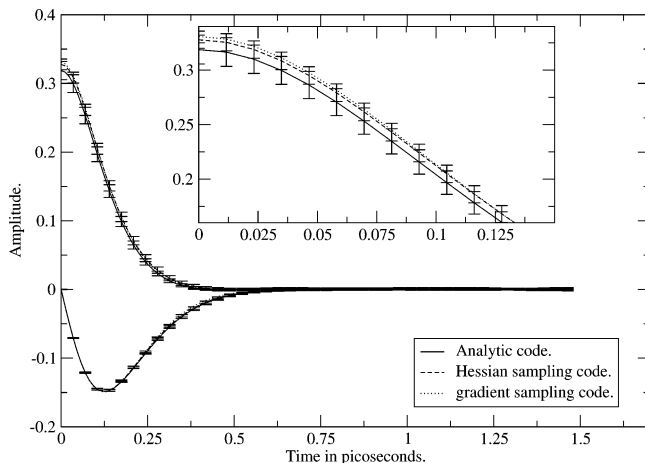
**Figure 5.** Comparison between water IR spectra derived by combining FK-LPI or “harmonic” quantum correction factors with the IR spectrum reported by Jeon et al.<sup>41</sup> obtained from classical molecular dynamics.

utilized to correct the IR spectrum of water.<sup>38</sup> The FK-LPI QCF is seen to perform better than the harmonic QCF, although the latter is clearly reasonable as a correction to the pure classical prediction. We emphasize that the FK-LPI QCF is computed ab initio and is determined only by the potential model chosen for water.

**Table 4.** Moments of the Van Hove Spectrum via Alternate Routes

	analytic <sup>a</sup>	Hessian numer. <sup>a</sup>	gradient <sup>a</sup>	Van Hove CF <sup>b</sup>	exptl. <sup>c</sup>	classical
$\langle\omega^1\rangle/\text{cm}^{-1}$	11.0	11.0	11.1	10.4	11.37	0
$\langle\omega^2\rangle$	1.9	1.9	1.9	1.7	1.8	1
$\langle\omega^3\rangle/\text{meV}^2$	134	134	131	107	123 ( $Q = 1.63 \text{ \AA}^{-1}$ )	

<sup>a</sup> Alternative sampling methods are “analytic”, analytical evaluation of averaged Hessian, eq 11; “Hessian, numer.”, numerical average of analytical Hessian; “gradient”, numerical average of position-weighted potential gradient, eq 22. <sup>b</sup> Data from ref 6. <sup>c</sup> Data from ref 27.



**Figure 6.** He(4) Van Hove correlation functions obtained from analytical Gaussian-fit, “Hessian sampling”, and “gradient sampling” codes. The upper and lower set of lines represent the real and imaginary parts of the CFs, respectively. The insert expands part of the upper set of lines. In the lower set, the lines fall on top of each other.

**4.3. Van Hove Spectrum of Liquid He(4).** We now consider the alternative numerical approaches to FK-LPI. In Figure 6 is shown the calculated Van Hove CF for all three implementations defined in section 3.2. All CFs were calculated up to 1.5 ps. Very good agreement is seen. We note that all of the methods agree within their variances, estimated from three individual 300 000-step calculations. As in ref 6, one may calculate the linear, quadratic, and cubic moments of the Van Hove CF as a test of accuracy. The procedure fits the real part of the CF to a sum over Gaussians, and then, the moments are found from eqs 30 and 12 in ref 6. The moments are listed in Table 4, where we also report the moments obtained by applying the same fitting procedure to the original Van Hove CF data in ref 6. From the table, one sees very good agreement between the experiment<sup>27</sup> and FK-LPI calculations. The classical values for the first and second moments (last column) reflect the size of the quantum effect. Also, we see that the new FK-LPI moments are in slightly better agreement with the experiment than the previously published values;<sup>6</sup> a fact that most likely is caused by the exact fulfillment of the Laue equation in the present work.

## 5. Conclusion and Outlook

In this paper, we have applied the FK-LPI approximation to CFs for models of water and He(4) in their liquid states. The water structural results are in reasonable agreement with experimental data, indicating that the FK-LPI local harmonic approximation is a good one. The water quantum diffusion

coefficients agree fairly well with CMD/RPMD simulations of similar water models, although the FK-LPI gives somewhat larger quantum effects. A consistency test is available for the FK-LPI case. It was found that the diffusion coefficients obtained from a direct calculation of the hydrogen or the oxygen atom velocity CFs agree very well with the values derived from a calculation of the Kubo-transformed CFs, which suggests that the values derived are accurate. The values must only be the same for an exact theory.

The He(4) Van Hove spectrum has been calculated using both a previously published implementation and two new and more general implementations of the FK-LPI theory. The results are, practically speaking, independent of the implementation and in very good agreement with the experiment. These explicitly demonstrate the computational feasibility of “black box” implementations of the underlying iterative Feynman–Kleinert equations, thereby enabling a general and simple coding of the theory.

The Hessian sampling approach has been demonstrated to be highly efficient, and for nonpolarizable potentials, where one has expressions for the classical Hessian, this implementation is efficient. The number of required samplings in eqs 11 and 16 would actually be *smaller* for water than for He(4), because the components of the smearing width matrix are smaller for water. The largest diagonal values for water are the hydrogen entries,  $\sim 0.01 \text{ \AA}^2$ , while those for He(4)  $\sim 0.03 \text{ \AA}^2$ . Hence, the sampling region in eqs 11 and 16 is simply smaller in the case of water.

The more general implementation is the gradient sampling technique, which only requires the classical potential and its gradient as input. Both are usual elements of any MD code. The results clearly show that the gradient implementation is slower than the Hessian sampling strategy, with eq 22 being the time-consuming part of the gradient calculation. However, it is immediately clear that this evaluation is highly suited for parallel computation. Hence, in terms of implementation, the gradient sampling approach also has this to recommend it.

**Acknowledgment.** J.A.P. and G.N. gratefully acknowledge support from the Swedish Research Council. P.J.R. gratefully acknowledges support from the National Science Foundation (CHE-0134775) and the Robert A. Welch Foundation. J.A.P. thanks Jonggu Jeon for sharing the SPC/F infrared spectrum with the authors.

## References

- (1) Jang, S.; Voth, G. A. *J. Chem. Phys.* **1999**, *111*, 2357–2370.
- (2) Hone, T. D.; Voth, G. A. *J. Chem. Phys.* **2004**, *64*, 6412–6422.

- (3) Craig, I. R.; Manolopoulos, D. E. *J. Chem. Phys.* **2005**, *121*, 3368–3373.
- (4) Miller, T. F., III; Manolopoulos, D. E. *J. Chem. Phys.* **2005**, *123*, 154504.
- (5) Nakayama, A.; Makri, N. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4230–4234.
- (6) Poulsen, J. A.; Nyman, G.; Rossky, P. J. *J. Phys. Chem. A* **2004**, *108*, 8743–8751.
- (7) Poulsen, J. A.; Nyman, G.; Rossky, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6709–6714.
- (8) Poulsen, J. A.; Nyman, G.; Rossky, P. J. *J. Phys. Chem. B* **2004**, *108*, 19799–19808.
- (9) Poulsen, J. A.; Nyman, G.; Rossky, P. J. *J. Chem. Phys.* **2003**, *119*, 12179–12193.
- (10) Shi, Q.; Geva, E. *J. Phys. Chem. A* **2003**, *107*, 9070–9078.
- (11) Rabani, E.; Reichman, D. R. *J. Chem. Phys.* **2004**, *120*, 1458–1465.
- (12) Rabani, E.; Reichman, D. R.; Krilov, G.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1129–1133.
- (13) Shi, Q.; Geva, E. *J. Chem. Phys.* **2003**, *118*, 8173–8184.
- (14) Hernandez, R.; Voth, G. A. *Chem. Phys.* **1998**, *223*, 243–255.
- (15) Wang, H.; Sun, X.; Miller, W. H. *J. Chem. Phys.* **1998**, *108*, 9726–9736.
- (16) Causo, M. S.; Ciccotti, G.; Montemayor, D.; Bonella, S.; Coker, D. F. *J. Phys. Chem. B* **2005**, *109*, 6855–6865.
- (17) Giachetti, R.; Tognetti, V. *Phys. Rev. Lett.* **1985**, *55*, 912–915.
- (18) Feynman, R. P.; Kleinert, H. *Phys. Rev. A* **1986**, *34*, 5080–5084.
- (19) Sun, X.; Miller, W. H. *J. Chem. Phys.* **1997**, *106*, 916–927.
- (20) We note that a Boltzmann Wigner transform based on effective frequency theory has previously been considered by Cuccoli, Tognetti, Verrucchi, and Vaia in *Phys. Rev. A* **1992**, *45*, 8418–8429. We thank Ruggero Vaia for bringing this to our attention. However, their paper does not use the quasi-density operator formalism.
- (21) Available quantum studies of water models predict an overall decrease in structure when quantum effects are included.<sup>22,34</sup> In an ab initio quantum MD study, Chen and co-workers<sup>23</sup> have observed generally comparable behavior to empirical models. However, they observed a more structured quantum OH radial distribution function, relative to classical, and rationalized this by an increase in polarity of the molecules, brought about by an increase in the anharmonic OH bond length when including quantum effects. We note that examination of such a quantum effect is not accessible in the present study, because the SPCf potential is harmonic in the OH stretch, and quantization of the nuclei will not change the bond lengths (in contrast to a Morse potential).
- (22) Stern, H. A.; Berne, B. J. *J. Chem. Phys.* **2001**, *115*, 7622–7628.
- (23) Chen, B.; Ivanov, I.; Klein, M. L.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *91*, 215503.
- (24) Cao, J.; Voth, G. A. *J. Chem. Phys.* **1994**, *101*, 6168–6183.
- (25) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. In *Numerical Recipes*, 2nd ed.; Cambridge University Press: Cambridge, U. K., 1992.
- (26) Lovesey, S. W. In *Theory of Neutron Scattering From Condensed Matter*; Oxford University Press: Oxford, U. K., 1986.
- (27) Verbeni, R.; Cunsolo, A.; Pratesi, G.; Monaco, G.; Rosica, F.; Masciovecchio, C.; Nardone, M.; Ruocco, G.; Sette, F.; Albergamo, F. *Phys. Rev. E* **2001**, *64*, 021203.
- (28) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Reidel: Dordrecht, The Netherlands, 1981.
- (29) Reimers, J. R.; Watts, R. O. *Mol. Phys.* **1984**, *52*, 357–381.
- (30) Paschek, D.; Geiger, A. *User's Guide and Manual MOSCITO 4*. <http://ganter.chemie.uni-dortmund.de/MOSCITO> (accessed Aug 10, 2006).
- (31) Allen, M. P.; Tildesley, D. J. In *Computer Simulations of Liquids*; Oxford University Press: Oxford, U. K., 1987.
- (32) Aziz, R. A.; Slaman, M. J.; Koide, A.; Allnatt, A. R.; Meath, W. J. *Mol. Phys.* **1992**, *77*, 321–337.
- (33) Soper, A. K. *Chem. Phys.* **2000**, *258*, 121–137.
- (34) Buono, G. S. D.; Rossky P. J.; Schnitker, J. *J. Chem. Phys.* **1991**, *95*, 3728–3737.
- (35) Lobaugh, J.; Voth, G. A. *J. Chem. Phys.* **1997**, *106*, 2400–2410.
- (36) Yeh, I.; Hummer, G. *J. Phys. Chem. B* **2004**, *108*, 15873–15879.
- (37) Marti, J.; Guardia, E.; Padro, J. A. *J. Chem. Phys.* **1994**, *101*, 10883–10891.
- (38) Iftimie, R.; Tuckerman, M. E. *J. Chem. Phys.* **2005**, *122*, 214508.
- (39) McQuarrie, D. A. In *Statistical Mechanics*; Harper and Row: New York, 1976.
- (40) Zwanzig, R. In *Nonequilibrium Statistical Mechanics*; Oxford University Press: New York, 2001.
- (41) Jeon, J.; Lefohn, A. E.; Voth, G. A. *J. Chem. Phys.* **2003**, *118*, 7504–7518.
- (42) Toukan, K.; Rahman, A. *Phys. Rev. B* **1985**, *31*, 2643–2648.
- (43) Bertie, J. E.; Lan, Z. *Appl. Spectrosc.* **1996**, *50*, 1047–1057.

## On the Competition between Predissociation and Direct Dissociation in Deuterium Chloride Ions ( $\text{DCI}^+$ )

Mikhail V. Korolkov<sup>†,‡</sup> and Karl-Michael Weitzel<sup>\*‡</sup>

*Stephanov Institute of Physics, National Academy of Science, Nezavisimosti ave. 70, 220602 Minsk, Republic of Belarus, and Philipps Universität Marburg, Fachbereich Chemie, Hans Meerwein Strasse, 35032 Marburg, Germany*

Received July 4, 2006

**Abstract:** The competition between direct dissociation and predissociation in  $\text{DCI}^+$  ions prepared in the  $A^2\Sigma^+$  state has been investigated numerically by solving the time dependent Schrödinger equation. This work concentrates on the rovibronic states ( $v_A; N_A$ ) with  $v_A = 8$  and  $N_A = 60\text{--}65$ , which are close to the top of a centrifugal barrier. We find that the relative yield of  $\text{D}^+$  (by direct dissociation) and  $\text{Cl}^+$  (by predissociation) strongly depends on the excitation frequency, intensity, and duration. Investigation of the time dependence of the product yields provides evidence for significant multichannel interactions.

### I. Introduction

The competition between different possible reaction channels lies at the heart of chemical research, with the ultimate goal to find selectivity for the formation of one specific product, e.g. in asymmetric catalysis<sup>1–3</sup> or to control the outcome of this competition.<sup>4–6</sup> One important piece of the mosaic is the understanding of competition at the molecular level, in particular the dynamical aspect. In the attempt to control chemical dynamics currently two types of approach are being pursued in the literature: the frequency domain approach<sup>7</sup> and the time domain approach.<sup>8</sup>

In the current work we investigate the competition of direct dissociation with predissociation in deuterium chloride ions ( $\text{DCI}^+$ ). The formation of  $\text{HCl}^+$  and  $\text{DCI}^+$  ions has been investigated extensively by resonance enhanced multiphoton ionization (REMPI)<sup>9–12</sup> but also single photon ionization in the VUV.<sup>13–16</sup> The internal state distribution of ions has been analyzed both by means of classical photoelectron spectroscopy<sup>17,18</sup> but also by photodissociation spectroscopy.<sup>19</sup> The spectroscopy of the ion is also well-known.<sup>19–21</sup> The energetically lowest dissociation limit in the  $\text{DCI}^+$  ion corresponds to the formation of  $\text{Cl}^+$  and D. The second dissociation limit corresponds to the formation of  $\text{D}^+$  and Cl. The energetics of these two dissociation channels and

the thresholds relevant for their competition are illustrated in Figure 1.

Here, the lowest fragmentation threshold  $T_1$  (formation of  $\text{Cl}^+$ ) is given by the energetic limit of the ion ground state ( $X^2\Pi_{3/2}$ ) and three repulsive potential curves ( $^4\Sigma^-$ ,  $^2\Sigma^-$ ,  $^4\Pi$ ). The dissociation limit of the first excited electronic state of  $\text{DCI}^+$  ( $A^2\Sigma^+$ ) converges to the formation of  $\text{D}^+$  (threshold  $T_3$ ). In the general situation of nonzero rotational angular momentum both bound potentials will exhibit a centrifugal barrier, denoted as  $T_2$  in the ion ground state and  $T_4$  in the A state. Note, that both  $T_2$  and  $T_4$  depend on the corresponding rotational quantum number  $N$ . Between the energies  $T_1$  and  $T_3$  formation of  $\text{Cl}^+$  is possible by direct dissociation of the ion ground state or by predissociation of the A state. The latter becomes possible by spin–orbit coupling to the three nominally repulsive electronic states,  $^4\Sigma^-$ ,  $^2\Sigma^-$ , and  $^4\Pi$ . Above  $T_3$  formation of  $\text{D}^+$  is accessible by direct dissociation from the A state, which will consequently compete with predissociation. Between  $T_3$  and  $T_4$  this direct dissociation requires tunneling, which is slow compared to predissociation. It is only in the region very close to the top of the barrier  $T_4$  that the direct dissociation of the A state can effectively compete with the predissociation. This competition is the subject of the current work.

### II. Computational Techniques

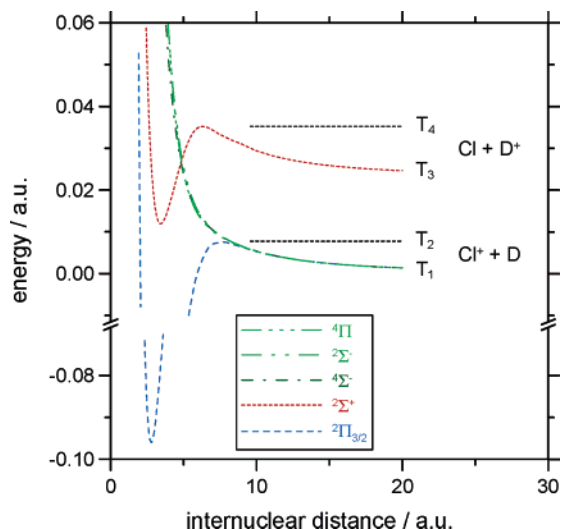
In the current work the dissociation dynamics of  $\text{DCI}^+$  ions have been investigated by numerical solution of time

\* Corresponding author e-mail: weitzel@chemie.uni-marburg.de.

<sup>†</sup> National Academy of Science.

<sup>‡</sup> Philipps Universität Marburg.





**Figure 1.** Schematic plot of the potential energy curves relevant for the fragmentation of DCI<sup>+</sup> together with 4 different thresholds. For further discussion see the text (1 au =  $a_0 = 0.5292 \text{ \AA}$ ).

dependent, coupled Schrödinger equations. Similar approaches have been described in previous papers.<sup>22–24</sup> A full account will be given below.

The input used for the wave packet calculations is based on ab initio data from Dalgarno et al.,<sup>25</sup> which were slightly adjusted in order to reproduce known spectroscopic parameters.<sup>14,19,26</sup> The current work takes into account those five different electronic states discussed above: (i) the ion ground state (X  $^2\Pi_{3/2}, j=1$ ), (ii) the first excited electronic state (A  $^2\Sigma^+, j=2$ ), and (iii) three repulsive states,  $^4\Sigma^-$  ( $j=3$ ),  $^2\Sigma^-$  ( $j=4$ ), and  $^4\Pi$  ( $j=5$ ). For these states coupled Schrödinger equations have been set up as given below<sup>27</sup>

$$i\hbar \frac{\partial \Psi_X(r,t)}{\partial t} = [\hat{T} + \tilde{V}_X] \Psi_X(r,t) + \eta_{X,A}(r) E_z(t) \Psi_A(r,t) + \eta_{X,4}(r) E_z(t) \Psi_4(r,t) \quad (1)$$

$$i\hbar \frac{\partial \Psi_A(r,t)}{\partial t} = [\hat{T} + \tilde{V}_A] \Psi_A(r,t) + \eta_{A,X}(r) E_z(t) \Psi_X(r,t) + \sum_{j=3}^5 H_{A,j}^{so}(r) \Psi_j(r,t) \quad (2)$$

$$i\hbar \frac{\partial \Psi_3(r,t)}{\partial t} = [\hat{T} + \tilde{V}_3] \Psi_3(r,t) + H_{A,3}^{so}(r) \Psi_A(r,t) \quad (3)$$

$$i\hbar \frac{\partial \Psi_4(r,t)}{\partial t} = [\hat{T} + \tilde{V}_4] \Psi_4(r,t) + \eta_{4,X}(r) E_z(t) \Psi_X(r,t) + H_{4,A}^{so}(r) \Psi_A(r,t) \quad (4)$$

$$i\hbar \frac{\partial \Psi_5(r,t)}{\partial t} = [\hat{T} + \tilde{V}_5] \Psi_5(r,t) + H_{5,A}^{so}(r) \Psi_A(r,t) \quad (5)$$

where the potential energy in the relevant electronic states is given by

$$\tilde{V}_x = V_x + \frac{J_X(J_X + 1) - \Omega_X^2}{2\mu \cdot r^2} \quad (6)$$

$$V_A = V_A + \frac{N_A(N_A + 1)}{2\mu \cdot r^2} + 0.5 \cdot \gamma \cdot N_A \quad (7)$$

$$\tilde{V}_j = V_j + \frac{J_j(J_j + 1) - \Omega_j^2}{2\mu \cdot r^2} \quad (8)$$

for  $j = 3$  ( $^4\Sigma^-$ ),  $4$  ( $^2\Sigma^-$ ), and  $5$  ( $^4\Pi$ ).

$V_X$ ,  $V_A$ , and  $V_j$  are the ab initio potential energies of the X-, A-, and the three repulsive states, respectively, to which the appropriate rotational energy is added. Here we use two different types of transitions: (i)  $J_X = N_A + 0.5$  ( $Q_1$  transition) and (ii)  $J_X = N_A - 0.5$  ( $R_1$  transition), further  $J_j = N_A + 0.5$ , and  $\gamma(\text{DCI}^+) = 0.25 \text{ cm}^{-1}$ . Note, that for the ground electronic X state and the repulsive  $^4\Sigma^-$ ,  $^2\Sigma^-$ , and  $^4\Pi$  states, Hund's case (a) representation was chosen, for the A state Hund's case (b) applies. The data shown in Figure 1 are for DCI<sup>+</sup> with  $N_A = 62$ . For the bound, ground state  $\Omega_X = 1.5$ , for the three repulsive states  $\Omega_j = 1.5$ . The reduced mass is  $\mu(\text{D}^{35}\text{Cl}^+) = 3542.23 \text{ au}$  (au = atomic units). In general the five states (and thus the Schrödinger equations) are coupled via spin-orbit interaction,  $H_{A,j}^{so}(r)$ , and through optical excitation, e.g.  $\eta_{X,A}(r) E_z(t)$ , where  $\eta_{X,A}(r)$  is the transition dipole moment connecting states X and A. The radial dependence of the latter is taken into account. Note, that the model is based on DCI<sup>+</sup> ions oriented along the  $x$  axis. The spin-orbit coupling matrix elements were taken from ref 23. The optical excitation will be discussed below.

In previous work we have shown that the predissociation lifetime can be reasonably approximated by a free decay model in which optical excitation is neglected. In the current work we are explicitly interested in the optical excitation spectrum inducing the dissociation.<sup>26,19,28,20</sup> Therefore we are looking at the decay of a wave function in a laser field. For this approach optical excitation starting from the ion ground state is taken into account by means of a realistic laser pulse. Spin-orbit coupling is taken into account from the beginning.

The laser field was represented by the following function

$$E(t) = \begin{cases} E_0 \cdot \sin^2\left[\frac{\pi \cdot t}{t_1}\right] \cdot \sin(\omega t) & \text{for } 0 \leq t \leq t_1 \\ 0 & \text{for } t > t_1 \end{cases} \quad (9)$$

where  $t_1$  is characterizing the width of the laser pulse at the base. Typically  $t_1 = 80 \text{ ps}$  was chosen in this work. This corresponds to a FWHM of the intensity profile of 29 ps.

Direct optical coupling is only taken into account between states A and X and  $^2\Sigma^-$  and X, respectively. Spin-orbit coupling is only taken into account between the A state and the three repulsive states but not between the A state and the X state. The latter would only be relevant, if we considered emission processes, which is not the case, and also only at significantly longer times.

The five nuclear wave functions are represented on an equidistant  $M$ -point spatial grid with  $r = (r_1, r_2, \dots, r_M)$ ,  $r_1 = 1.5a_0$ ,  $r_{i+1} - r_i = 0.02a_0$ , for  $4096 < M < 32768$ , which allows to use fast Fourier transform (FFT) methods for the transformation between coordinate and momentum space.<sup>29</sup> A combination of the split-operator method<sup>30</sup> and the integral

equation method<sup>31,32</sup> is used for the propagation of the wave functions (given in eqs 1–5) in time with time steps of  $\Delta t \leq 1$  atomic unit (1 au  $\approx 0.024$  fs). The total propagation time (identical to the laser pulse duration) is chosen as appropriate for the corresponding fragmentation dynamics. Typically this total time is about 1 order of magnitude larger than the respective lifetime. This ensures that the line width derived does not depend on the laser pulse duration. The absorbing boundary<sup>33</sup> is used to prevent the artificial reflection of all  $\Psi_j(r, t)$  wave functions at the edge of the grid, where necessary.

All calculations deal with the dynamics of the vibronic  $A^2\Sigma^+$  ( $v_A=8$ ) state, with rotational angular momentum ranging from  $N_A = 60$  to  $N_A = 65$ .

### III. Results

**III.1. Results for Target States ( $v_A=8; N_A$ ) in Weak Laser Fields.** We start this result section with the discussion of the ( $v_A; N_A$ ) = (8;60) state under conditions of excitation by a very weak laser field. This state is relatively short-lived.<sup>27</sup> The  $N_A = 60$  state is located roughly 2400  $\text{cm}^{-1}$  above the limit for direct dissociation ( $T_3$ ) but 450  $\text{cm}^{-1}$  below the top of the centrifugal barrier ( $T_4$ ).

In Figure 2 we show the normalized absorption spectrum (yield spectrum),  $Y_{\text{abs}} = (1 - \langle \Psi_X | \Psi_X \rangle)$ , calculated at the end of propagation with the target state  $N_A = 60$ . The spectrum also shows the relative contribution of the three predissociation channels  $k_j$ . This relative contribution is defined as

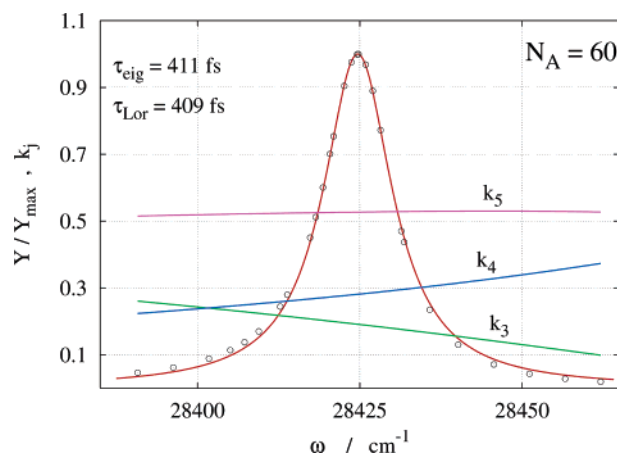
$$k_j = \frac{\text{Flux}(\Psi_j(r_o, t))}{\sum_{j=3}^5 \text{Flux}(\Psi_j(r_o, t))}$$

where the flux is calculated at an internuclear separation of  $r_o = 20$  au

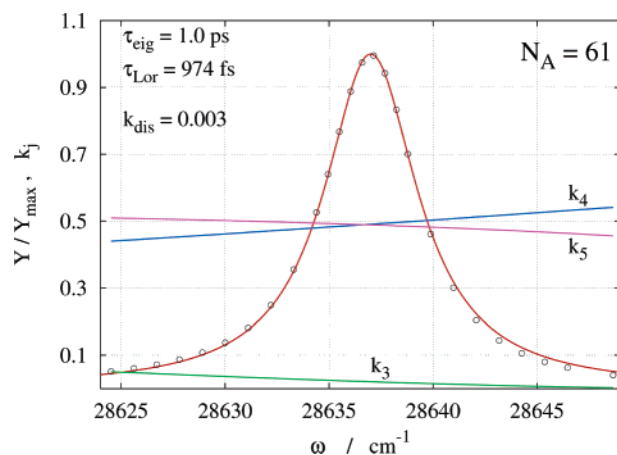
$$\text{Flux}(\Psi_j(r_o, t)) = \frac{\hbar}{2 \cdot i \cdot m} \left[ \Psi_j^*(r_o, t) \cdot \left( \frac{\partial \Psi_j(r, t)}{\partial r} \right)_{r=r_o} - \Psi_j(r_o, t) \cdot \left( \frac{\partial \Psi_j^*(r, t)}{\partial r} \right)_{r=r_o} \right]$$

Despite the fact that the (8;60) state is lying above  $T_3$ , only predissociation occurs. This implies that direct dissociation is much slower than predissociation. This in turn is due to tunneling being required for the direct process. Among the three predissociation channels state  ${}^4\Pi$  ( $j=5$ ) clearly dominates. This is connected to the crossing point between the bound A state and the repulsive states, where at the relevant internuclear distance of about 4.5 au the spin-orbit coupling matrix element is largest for this channel  $j = 5$ . Evidently the fraction  $k_5$  is basically independent of the excitation frequency,  $k_4$  slightly increases, and  $k_3$  slightly decreases with excitation frequency.

By fitting a Lorentzian function to the numerical data we derive a lifetime of 409 fs. For comparison we performed calculations for the free decay of eigen states. In the latter case the optical excitation step is not included in the

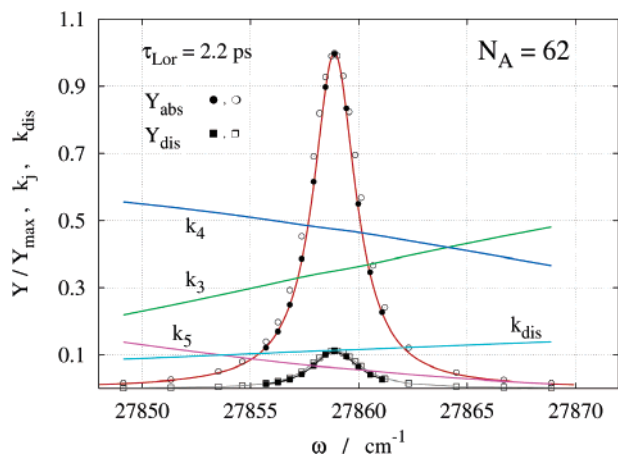


**Figure 2.** Normalized absorption spectrum of the DCI<sup>+</sup> into the  $v_A = 8$ ,  $N_A = 60$  state ( $Q_1$  transition): symbols: numerical data; line: Lorentzian function. Relative contribution of the three repulsive states to the overall predissociation process as a function of the excitation frequency (labeled  $k_3$ ,  $k_4$ , and  $k_5$ ).



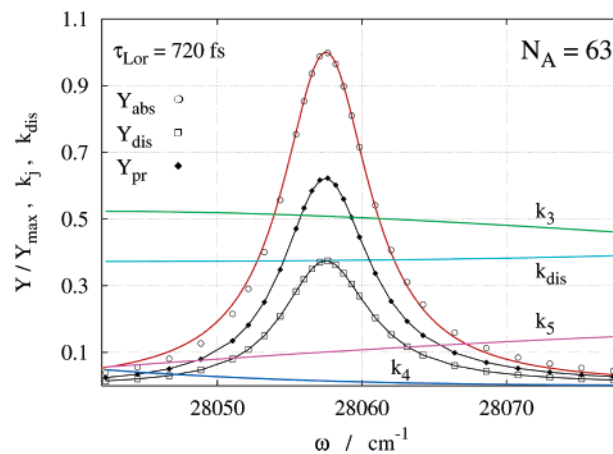
**Figure 3.** Normalized absorption spectrum of the DCI<sup>+</sup> into the  $v_A = 8$ ,  $N_A = 61$  state ( $R_1$  transition): symbols: numerical data; line: Lorentzian function. Relative contribution of the three repulsive states to the overall predissociation process as a function of the excitation frequency (labeled  $k_3$ ,  $k_4$ , and  $k_5$ ).

numerical calculations.<sup>23,27</sup> For this free decay a lifetime of 411 fs is obtained in very good agreement with the optically induced decay. We note, however, that the numerical data in Figure 2 show a small deviation from Lorentzian line shape. On the low-frequency wing of the spectrum, the numerical data are above; on the high-frequency wing of the spectrum the numerical data are below the best Lorentzian line fit. An experiment aimed at detecting this deviation from Lorentzian line shape would require a high signal/noise ratio on the order of 50. The absorption spectrum with the target state  $N_A = 61$  is shown in Figure 3. This rotational state,  $N_A = 61$ , is still about 330  $\text{cm}^{-1}$  below the top of the barrier ( $T_4$  for  $N_A = 61$ ). For this and all higher states predissociation and direct dissociation can effectively compete. This competition is the topic of the current manuscript. Here, for  $N_A = 61$  the contribution from direct dissociation is only 0.3%, despite the fact that the state  $N_A = 61$  is relatively long-

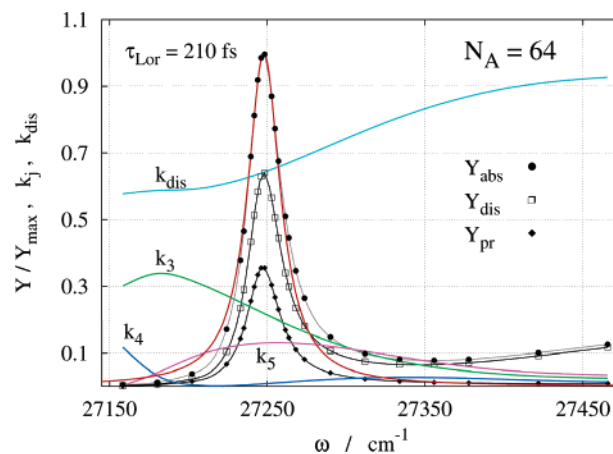


**Figure 4.** Normalized absorption spectrum of the DCI<sup>+</sup> into the  $v_A = 8$ ,  $N_A = 62$  state ( $Q_1$  transition): symbols: numerical data (● ■: 80 ps, ○ □: 40 ps); red line: Lorentzian function fitted to ●. The line through the ■ and □ data is the result of spline. Relative contribution of the three repulsive states to the overall predissociation process as a function of the excitation frequency (labeled  $k_3$ ,  $k_4$ , and  $k_5$ ).

lived. A Lorentzian fit to the numerical data leads to  $\tau = 974$  fs. The free decay of eigenstates leads to  $\tau = 1$  ps, again in good agreement. While channel 5 clearly dominated for  $N_A = 60$ , here predissociation is dominated by channel  $j = 4$  and  $j = 5$ , and channel  $j = 3$  contributes less than 5%. The frequency dependence of all  $k_j$  is weak. However, there is again a noticeable asymmetry in the frequency dependence of  $Y_{\text{abs}}$ , similar to the situation for  $N_A = 60$ . Figure 4 shows the absorption spectrum leading to the target state  $N_A = 62$ . This state still lies  $216 \text{ cm}^{-1}$  below  $T_4$  ( $N_A = 62$ ). During the course of the investigation it turned out, that this is the longest living state considered in this work. Thus it was deemed appropriate to check for possible contributions of the laser pulse duration on the calculated lifetime. Therefore numerical calculations have been performed for two different total laser pulse durations (propagation times), i.e., 40 and 80 ps ( $\tau(\text{FWHM}) \approx 14.6$  and 29 ps). Evidently, the absorption profile ( $Y_{\text{abs}}$ ) is slightly more narrow for 80 ps compared to 40 ps. Since the difference is small, we conclude that 80 ps are long enough to accurately represent the dynamics. Consequently a lifetime of  $\tau = 2.2$  ps has been derived from a Lorentz fit to the numerical absorption spectrum. The contribution of direct dissociation to the total flux is about 10%. A free decay of eigenstates has no meaning in this case and is therefore not available for comparison. In terms of channel competition the total flux in  $N_A = 62$  is dominated by channels  $j = 3$  and  $j = 4$ . Channel  $j = 5$  contributes even less than the direct dissociation for the major part of the spectrum. The spectral characteristics of  $Y_{\text{abs}}$  and  $Y_{\text{diss}}$  are very similar. This implies that also predissociation and dissociation have very similar spectral characteristics. Figure 5 shows the absorption spectrum leading to the target state  $N_A = 63$ . For this state, which lies  $113 \text{ cm}^{-1}$  below  $T_4$  ( $N_A = 63$ ), the lifetime is 720 fs. The contribution of  $k_{\text{diss}}$  to the entire flux is about 38%, independent of the frequency. Fifty percent of the total flux goes into the predissociation through channel  $j = 3$ .

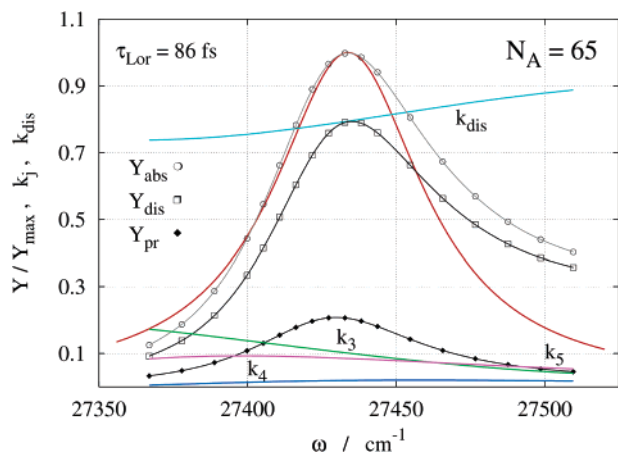


**Figure 5.** Normalized absorption spectrum of the DCI<sup>+</sup> into the  $v_A = 8$ ,  $N_A = 63$  state ( $R_1$  transition): symbols: numerical data; red line: Lorentzian function fitted to ○. Lines through  $Y_{\text{dis}}$  and  $Y_{\text{pr}}$  data are result of spline. Relative contribution of the three repulsive states to the overall predissociation process as a function of the excitation frequency (labeled  $k_3$ ,  $k_4$ , and  $k_5$ ).



**Figure 6.** Normalized absorption spectrum of the DCI<sup>+</sup> into the  $v_A = 8$ ,  $N_A = 64$  state ( $Q_1$  transition): symbols: numerical data; red line: Lorentzian function fitted to  $Y_{\text{abs}}$ . Black and gray lines through data points are the result of spline. Relative contribution of the three repulsive states to the overall predissociation process as a function of the excitation frequency (labeled  $k_3$ ,  $k_4$ , and  $k_5$ ).

Channel  $j = 5$  contributes about 10%, and  $k_4$  is negligible again with weak frequency dependence. In a later section we will discuss the influence of the laser pulse intensity and duration on the product yield for this target state. Figure 6 shows the absorption spectrum leading to the target state  $N_A = 64$ . In 0th order this state lies  $29 \text{ cm}^{-1}$  below  $T_4$  ( $N_A = 64$ ). This vicinity to  $T_4$  leads to significant deviation from Lorentzian line shape for all yield data ( $Y_{\text{abs}} = Y_{\text{dis}} + Y_{\text{pr}}$ ). To get at least a rough figure of the order of magnitude of  $\tau$ , we have included a Lorentzian line fitted to the FWHM of the data  $Y_{\text{abs}}$  (red line). This leads to  $\tau = 210$  fs. Evidently  $k_{\text{dis}}$  increases from 0.6 on the low-frequency wing of the spectrum to above 0.9 at  $150 \text{ cm}^{-1}$  above the center frequency. This increase in  $k_{\text{dis}}$  is most likely connected to the vicinity to the top of the centrifugal barrier to direct

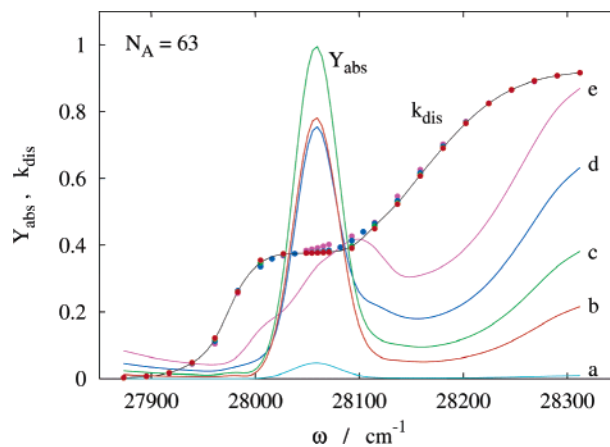


**Figure 7.** Normalized absorption spectrum of the DCI<sup>+</sup> into the  $v_A=8$ ,  $N_A=65$  state ( $R_1$  transition): symbols: numerical data; red line: Lorentzian function fitted to  $Y_{\text{abs}}$ . Black lines through data points: spline. Relative contribution of the three repulsive states to the overall predissociation process as a function of the excitation frequency (labeled  $k_3$ ,  $k_4$ , and  $k_5$ ).

dissociation. It is automatically (due to normalization) reflected in a decrease of  $k_3$  and also  $k_4$ . However, on the high-frequency wing of the spectrum  $k_4$  slightly increases again. Channel  $j=5$  shows a resonancelike variation with  $\omega$ . The increase in  $k_{\text{dis}}$  is associated with the deviation from Lorentzian line shape on the high frequency wing of the spectrum. Most likely the reason for both is that the lifetime for direct dissociation changes much faster than that for predissociation in this region. This effect appears to be smooth, indicating that there are no sharp steps in the corresponding lifetimes. Figure 7 shows the absorption spectrum leading to the target state  $N_A=65$ . This is the first state which lies above the top of the centrifugal barrier ( $4\text{ cm}^{-1}$ ). Therefore direct dissociation leading to the formation of  $D^+$  dominates. Yet,  $k_{\text{dis}}$  still increases from 0.75 to 0.9 over a range of  $100\text{ cm}^{-1}$ . States  $j=3$  and  $j=5$  each contribute on the order of 10% at the center frequency, slightly decreasing with  $\omega$ , the contribution of channel  $j=4$  is almost negligible.

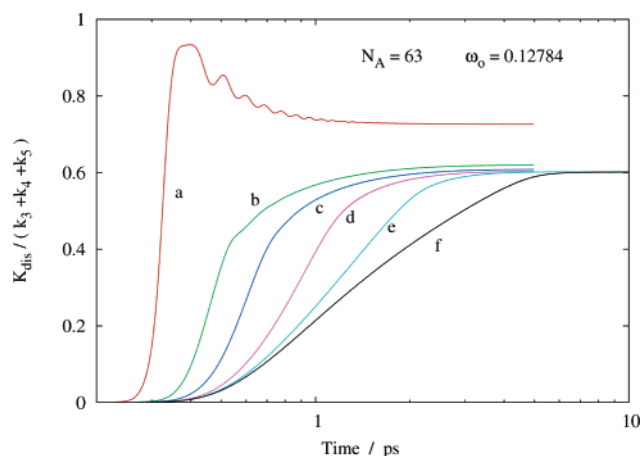
The numerical data for  $Y_{\text{abs}}$  are very non-Lorentzian on the high  $\omega$  wing of the spectrum. Obviously, in this situation the derivation of a lifetime is problematic. To get an estimate of the magnitude of  $\tau$  we have fitted a Lorentzian to the low  $\omega$  wing of the data. From this analysis we obtain a lifetime  $\tau=86\text{ fs}$ . This is considered to be a very short lifetime. In previous work<sup>27</sup> we showed that the lifetime between the rotational islands of stability increased with the RIS index from about 90 fs to 200 fs. The lifetime of the (8;65) state is even shorter than the shortest lifetimes calculated between the threshold and the first RIS (cf. Figure 2 of ref 27). The asymmetry discussed above for  $Y_{\text{abs}}$  is even more pronounced in  $Y_{\text{dis}}$  but less pronounced for  $Y_{\text{pr}}$ . This is indicative of significant multichannel interactions. In classical kinetics one would expect the same line shape for directly competing processes. We will come back to this point below.

**III.2. Results for Different Laser Pulse Intensities and Duration.** All the calculations described above utilized very low laser fields of maximum amplitude  $E=0.000025\text{ au}$  ( $1$



**Figure 8.** Absolute absorption yield spectrum of the DCI<sup>+</sup> into the  $v_A=8$ ,  $N_A=63$  state for various laser fields: a: 0.0001 au, b: 0.005 au, c: 0.007 au, d: 0.01 au, e: 0.014 au.

$au = E_{\text{hartree}}/(e \cdot a_0) = 5142\text{ MV/cm}$ ), such that the calculated absolute absorbance was only on the order of a few %, i.e., in the linear regime. At the same time the laser pulse width were always chosen significantly larger than the total dissociation lifetime. This ensured that any spectral result was unaffected by laser pulse properties. In the following we discuss results for various laser field strength and also for shorter laser pulse width. Most of these calculations were performed for the  $N_A=63$  state. Figure 8 shows the absolute absorbance yield spectrum for various laser fields between 0.005 au to 0.014 au. The laser pulse width is in each case 1 ps. As it turns out, the absorbance, in particular the absorbance at the center transition frequency, is a strongly varying function of the laser field strength. For low fields the absorbance increases with the power, as expected for the linear regime. At high laser fields the absorbance in general exhibits a maximum followed by oscillatory behavior. The precise laser field for which the absorbance becomes maximum depends on the target state and also the excitation frequency and the pulse duration. For the current condition the highest peak absorbance is observed for an intermediate laser field of 0.007 au (trace c in Figure 8). The corresponding intensity is  $I_0 = 0.5 \cdot \epsilon_0 \cdot c \cdot E_0^2 = 3.5 \cdot 10^{16}\text{ W/cm}^2 \cdot E_0^2$ , where  $E$  is given in atomic units. For this field, which is a factor 280 higher than that in Figure 5, the maximum absorbance is  $Y_{\text{abs}}=1$ . For laser fields above and below this special value the maximum absorbance at the center frequency is smaller. Even more interesting is the spectral characteristics of the absorbance and the ratio of direct dissociation to predissociation on the high-frequency wing of the excitation spectrum. Here, the contribution of direct dissociation to the total flux,  $k_{\text{dis}}$ , increases between 28 100 and 28 300  $\text{cm}^{-1}$  from about 0.4 to 0.9. It is important to emphasize that the increase in  $Y_{\text{abs}}$  does not represent a transition to the next higher rotational state, since only a single final rotational state is included in these calculations. Rather, the increase in  $k_{\text{dis}}$  is mostly likely related to surpassing the top of the barrier for direct dissociation. While the fractional yield of direct dissociation at a given excitation frequency does not depend on the laser field, the absolute absorbance does increase with increasing laser field strength amplitude in the region of the top of this barrier (around 28 300  $\text{cm}^{-1}$ ). Since



**Figure 9.** Time dependence of the branching ratio  $k_{\text{dis}}/k_{\text{prediss}}$  for various laser pulse duration: a: 0.1 ps, b: 0.3 ps, c: 0.5 ps, d: 1 ps, e: 2 ps, f: 5 ps.

the square of the laser field amplitude scales with the laser intensity (see above), this implies an increasing absolute yield of absorbance with increasing intensity. It is interesting to note that a similar result for the spectral absorbance is obtained in calculations, where the laser pulse duration is varied, but the laser field amplitude is kept constant.

Another interesting aspect is the time dependence of the individual channel contributions. In classical kinetics the branching ratio of directly competing chemical processes is time independent. We may now ask the question: down to what time regime is this picture valid. At zero time all molecules are in their ground state. With propagation in time the flux through the available channels increases. For laser pulse duration larger than 0.3 ps the yield of direct dissociation increases smoothly with time, approaching a limiting value of  $k_{\text{dis}} = 0.38$ . However, for pulse duration of 0.1 ps  $k_3$  first increases steeply to about 0.48 and decreases then with time, approaching a limiting value 0.42. This is a first indication of multichannel dynamics. The time dependence of the predissociation yield exhibits nonmonotonic behavior with transient maxima for all laser pulses between 0.1 and 5 ps. Again, this result would not be expected in classical kinetics. Finally in Figure 9 we present the time dependence of the branching ratio between direct dissociation and predissociation for laser pulses with durations between 0.1 and 5 ps. For all these laser pulse durations considered the branching ratio reaches a time independent value after several picoseconds. However, this limiting value is not approached in a step function but either by a monotonic increase for long pulses or by a nonmonotonic increase for pulse duration of 0.1 ps. We note that the limiting value for 0.1 ps differs from that for longer laser pulses. The time dependence of the branching ratio  $k_{\text{dis}}/(k_3+k_4+k_5)$  directly reflects the time dependence of the interaction between the different electronic states involved. While this is in fact not surprising from the dynamical point of view, the interesting aspect is that the fingerprint of this interaction ceases after several picoseconds and the response of the system changes from pure dynamical to classical kinetic behavior.

## IV. Summary

We have investigated the competition between direct dissociation (formation of  $\text{D}^+$ ) and predissociation (formation of  $\text{Cl}^+$ ) for high lying rovibrational states  $v_A = 8$ ,  $N_A = 60-65$  in the region of the top of a centrifugal barrier in the  $\text{A}^2\Sigma^+$  state of  $\text{DCI}^+$ . Direct dissociation becomes important starting from about  $100 \text{ cm}^{-1}$  below the top of this barrier via tunneling. Analysis of the omega dependence of the channel yields reveals significant multichannel interactions. The individual channel yields vary strongly with omega but also with time. In classical kinetics directly competing chemical reaction channels lead to a time independent product ratio. In the fragmentation of  $\text{DCI}^+$  multichannel interaction leads to a pronounced time dependence of product branching ratios. This can be rationalized as being due to a time dependence of the effective cross section for population transfer between individual levels. In principle this time dependence could in part originate from the time dependence of the laser field. We note, however, that multichannel interactions are also predicted in calculations which do not take into account the laser field.<sup>22</sup> More likely the multichannel dynamics reflect the microreversibility of population flow between the different electronic states involved. In the long time limit—reached after several picoseconds—a transition to classical kinetics is observed. For the future it would be interesting to perform similar calculations also for  $v_A = 9$  and 10, since for these vibrational levels the corresponding  $N_A$  quantum numbers would be significantly lower and therefore more easy to access experimentally.

The photochemical processes considered in the current work have been induced by one photon absorption from the ion X state into the electronic excited A state. An alternative access to the competition between dissociation and predissociation is provided by femtosecond multiphoton excitation in the IR or mid-IR. As an example we mention the competition between direct dissociation of  $\text{NaI}$ , leading to the formation of  $\text{Na}^*$  and  $\text{I}^*$ , and predissociation, leading to  $\text{Na}^+$  and  $\text{I}^-$ , as investigated theoretically by Engel and co-worker.<sup>34</sup> For the title molecule of the current work we have reported evidence for the possibility to control the competing product channels  $\text{D}^+$  vs  $\text{Cl}^+$  in a fs experiment employing 800 nm pulses.<sup>35</sup>

**Acknowledgment.** Support of this work from the DFG (436 WER 17/4/02 and 17/4/03) and from INTAS (grant no. 03-50-5765) is gratefully acknowledged. Support by J. Manz is gratefully acknowledged.

## References

- (1) Knowles, W. S. Asymmetric Hydrogenations (Nobel Lecture). *Angew. Chem. Int. Ed.* **2002**, *41*, 1999–2007.
- (2) Noyori, R. Asymmetric Catalysis: Science and Opportunities (Nobel Lecture). *Angew. Chem. Int. Ed.* **2002**, *41*, 2008–2022.
- (3) Sharpless, K. B. Searching for New Reactivity (Nobel Lecture). *Angew. Chem. Int. Ed.* **2002**, *41*, 2024–2032.
- (4) Rice, S. A.; Zhao, M. *Optimal Control of Molecular Dynamics*; Wiley-VCH: Weinheim, 2001.
- (5) Shapiro, M.; Brumer, P. *Principles of Quantum Control of Molecular Processes*; Wiley: Hoboken, NJ, 2003.

- (6) Brixner, T.; Gerber, G. Quantum Control of Gas-Phase and Liquid-Phase Femtochemistry. *ChemPhysChem* **2003**, *4*, 418–438.
- (7) Brumer, P.; Shapiro, M. Control of Unimolecular Reactions Using Coherent Light. *Chem. Phys. Lett.* **1986**, *126*, 541–546.
- (8) Tannor, D. J.; Kosloff, R.; Rice, S. A. Coherent Pulse Sequence Induced Control of Selectivity of Reactions: Exact Quantum Mechanical Calculations. *J. Chem. Phys.* **1986**, *85*, 5805–5820.
- (9) Green, D. S.; Bickel, G. A.; Wallace, S. C. (2+1) Resonance Enhanced Multiphoton Ionization of Hydrogen Chloride in a Pulsed Supersonic Jet: Spectroscopic Survey. *J. Mol. Spectrosc.* **1991**, *150*, 303–353.
- (10) Green, D. S.; Bickel, G. A.; Wallace, S. C. (2+1) Resonance Enhanced Multiphoton Ionization of Hydrogen Chloride in a Pulsed Supersonic Jet: Spectroscopy and Rydberg  $\sim$  Valence Interactions of the  $1^1\Sigma^+(0^+)$  and  $3^3\Sigma^-(1, 0^+)$  States. *J. Mol. Spectrosc.* **1991**, *150*, 354–387.
- (11) Kvaran, A.; Wang, H. S.; Waage, B. G. Three- and Two-Photon Absorption Spectroscopy: REMPI of HCl and HBr. *Can. J. Phys.* **2001**, *79*, 197–210.
- (12) Kvaran, A.; Wang, H. S. Three-Photon Absorption Spectroscopy: the  $L(1^1\Phi_3)$  and  $m(3^1\Pi_1)$  States of HCl and DCl. *Mol. Phys.* **2002**, *100*, 3513–3519.
- (13) Yench, A. J.; McConkey, A. G.; Dawber, G.; Avaldi, L.; MacDonald, M. A.; King, G. C.; Hall, R. I. Threshold Photoelectron Spectroscopy of HCl up to 40 eV. *J. Electron Spectrosc.* **1995**, *73*, 217–229.
- (14) Yench, A. J.; Cormack, A. J.; Donovan, R. J.; Hopkirk, A.; King, G. C. Threshold Photoelectron Spectroscopy of HCl and DCl. *Chem. Phys.* **1998**, *238*, 109–131.
- (15) Edvardsson, D.; Baltzer, P.; Karlsson, L.; Lundqvist, M.; Wannberg, B. Rotational Fine Structure in the UV Photoelectron Spectra of HF and HCl. *J. Electron Spectrosc.* **1995**, *73*, 105–124.
- (16) Burmeister, F.; Andersson, L. M.; Ohrwall, G.; Richter, T.; Zimmermann, P.; Godehusen, K.; Martins, M.; Karlsson, H. O.; Sorensen, S. L.; Bjorneholm, O.; Feifel, R.; Wiesner, K.; Goscinski, O.; Karlsson, L.; Svensson, S.; Yench, A. J. A Study of the Inner-Valence Ionization Region in HCl and DCl. *J. Phys. B: At., Mol. Opt. Phys.* **2004**, *37*, 1173–1183.
- (17) de Beer, E.; Koenders, B. G.; Koopmans, M. P.; de Lange, C. A. Multiphoton Ionization Processes in Hydrogen Chloride Studied by Photoelectron Spectroscopy. *J. Chem. Soc., Faraday Trans.* **1990**, *86*, 2035–2041.
- (18) de Beer, E.; Buma, W. J.; de Lange, C. A. Resonance Enhanced Multiphoton Ionization Photoelectron Spectroscopy and Pulsed Field Ionization Via the  $F\ 1^1\Delta_2(v' = 0)$  and  $f\ 3^1\Delta_2(v' = 0)$  Rydberg States of Hydrogen Chloride. *J. Chem. Phys.* **1993**, *99*, 3252–3261.
- (19) Michel, M.; Korolkov, M. V.; Weitzel, K. M. State-Selective Predissociation Spectroscopy of  $\text{HCl}^+$  and  $\text{DCl}^+$  Ions. *J. Phys. Chem. A* **2004**, *108*, 9924–9930.
- (20) Penno, M.; Holzwarth, A.; Weitzel, K. M. State Selective Predissociation Spectroscopy of Hydrogen Chloride Ions ( $\text{HCl}^+$ ) Via the  $A^2\Sigma^+ \leftarrow 2^1\Pi_{3/2}$  Transition. *Mol. Phys.* **1999**, *97*, 43–52.
- (21) Saenger, K. L.; Zare, R. N.; Mathews, C. W. A Reexamination of the Spin-Rotation Constant for  $2^1\Pi$  States: The A–X Band System of  $\text{HCl}^+$ . *J. Mol. Spectrosc.* **1976**, *61*, 216–230.
- (22) Korolkov, M. V.; Weitzel, K. M. The Predissociation Dynamics of Vibrational Eigenstates in the  $A^2\Sigma^+$  State of  $\text{HBr}^+$  Ions: Numerical Solution of Coupled Time-Dependent Schrödinger Equations. *Chem. Phys.* **2000**, *252*, 209–219.
- (23) Korolkov, M. V.; Weitzel, K. M.; Peyerimhoff, S. D. Spin–Orbit Induced Predissociation Dynamics of  $\text{HCl}^+$  and  $\text{HBr}^+$  Ions: Temporal and Spectral Representations. *Int. J. Mass Spectrom.* **2000**, *201*, 109–120.
- (24) Korolkov, M. V.; Weitzel, K. M. The Spin–Orbit Induced Predissociation Dynamics of  $\text{HCl}^+$  Ions: Rotational Islands of Stability. *Chem. Phys. Lett.* **2001**, *336*, 303–310.
- (25) Pradhan, A. D.; Kirby, K. P.; Dalgarno, A. Theoretical Study of  $\text{HCl}^+$ : Potential Curves, Radiative Lifetimes, and Photodissociation Cross Sections. *J. Chem. Phys.* **1991**, *95*, 9009–9023.
- (26) Michel, M.; Korolkov, M. V.; Weitzel, K. M. A New Route to the Dissociation Energy of Ionic and Neutral HCl Via Lineshape Analysis of Single Rotational Transitions. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4083–4086.
- (27) Korolkov, M. V.; Weitzel, K. M. Kinetic and Dynamic Aspects of Lifetime Oscillations in the Predissociation of Hydrogen Chloride Ions. *J. Phys. Chem. A* **2006**, *110*, 2924–2932.
- (28) Penno, M.; Holzwarth, A.; Weitzel, K. M. State Selective Predissociation Spectroscopy of Hydrogen Bromide Ions ( $\text{HBr}^+$ ) Via the  $2^2\Sigma^+ \leftarrow 2^1\Pi_i(i=1/2, 3/2)$  Transition. *J. Phys. Chem. A* **1998**, *102*, 1927–1934.
- (29) Kosloff, R. Propagation methods for quantum molecular dynamics. *Annu. Rev. Phys. Chem.* **1994**, *45*, 145.
- (30) Feit, M. D.; Fleck, J. A.; Steiger, A. Solution of the Schrodinger equation by a spectral method. *J. Comput. Phys.* **1982**, *47*, 412.
- (31) Korolkov, M. V.; Schmidt, B. Spin–Orbit Induced Association Under Ultrafast Laser Pulse Control. *Chem. Phys. Lett.* **2002**, *361*, 432–438.
- (32) Korolkov, M. V.; Paramonov, G. K. Vibrationally State-Selective Electronic Excitation of Diatomic Molecules by Ultrashort Laser Pulses. *Phys. Rev. A: At., Mol., Opt. Phys.* **1998**, *57*, 4998–5001.
- (33) Bisseling, R. N.; Kosloff, R.; Manz, J. Dynamics of hyperspherical and local mode resonance decay studied by time dependent wave packet propagation. *J. Chem. Phys.* **1985**, *83*, 993.
- (34) Marquetand, P.; Engel, V. Predissociation and Dissociation Dynamics in Quantum Control Fields. *Chem. Phys. Lett.* **2005**, *407*, 471–476.
- (35) Breunig, H. G.; Lauer, A.; Weitzel, K. M. Control of Branching Ratios in the Dissociative Ionization of Deuterium Chloride. *J. Phys. Chem. A* **2006**, *110*, 6395–6398.

## Halide, Ammonium, and Alkali Metal Ion Parameters for Modeling Aqueous Solutions

Kasper P. Jensen and William L. Jorgensen\*

*Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107*

Received August 4, 2006

**Abstract:** A complete set of Lennard Jones parameters for the halide ions,  $F^-$ ,  $Cl^-$ ,  $Br^-$ , and  $I^-$ , ammonium ion, and the alkali metal ions is reported. The parameters have been optimized using Monte Carlo simulations and free energy perturbation theory with the TIP4P water model to reproduce experimental free energies of hydration and locations of the first maxima of the ion-oxygen radial distribution functions, to provide water coordination numbers consistent with experimental ranges, and to exhibit gas-phase monohydrate energies in reasonable agreement with ab initio values. Average errors for absolute and relative free energies of hydration for the ions are ca. 1 kcal/mol. For the halides, this is the first self-consistent set of parameters that has been optimized for aqueous-phase performance. The good results for relative free energies of hydration are particularly auspicious for use of the new parameters in a wide variety of liquid-phase simulations where halide and alkali cations are systematically varied.

### Introduction

The study of the hydration of ions is of major importance in chemistry and biochemistry. When ions are part of a system to be studied by classical simulations, a set of parameters describing the interactions of the ions with other atoms is necessary. These parameters are usually those of a Lennard-Jones (LJ) potential, which is an empirical, effective two-atom potential. Electrostatic interactions are described by Coulomb's law and fixed atomic charges. The LJ parameters of the ions can be fitted to reproduce thermodynamic and structural data, whereas the atomic charges are assigned a priori as the formal charges of the ions. The properties are derived from statistical-mechanical simulations of model systems, which are intended to resemble an infinitely dilute solution.

A main problem with simulating hydration of ions is the electrostatic interactions. Because of their range, it is necessary to correct for the use of a finite model system. Doing this can strongly affect both structure and free energy of hydrated ions.<sup>1–3</sup> For an ion in an aqueous model system, both the ion–water (IW) and water–water (WW) interactions contribute to the hydration free energy:<sup>2</sup> WW interactions disfavor ion hydration, whereas IW interactions favor

hydration. The problem of long-range electrostatics in general remains unsolved, although progress is taking place.<sup>4–6</sup>

When the simulation system is truncated, corrections for the absent bulk are needed. Born theory can be applied to correct to the asymptotic limit of cavity sizes, as this limit is determined by the long-range behavior of Coulomb's law. For normal ions with small cavities given by ionic or covalent radii, the hydration enthalpies are inversely proportional to the radii as implied by Born theory.<sup>7</sup> However, the contributions to the free energy of solvation from the region closest to even a spherical ion are not well modeled by Born theory, since deviations from continuum theory are largest at short range.<sup>3,8</sup> Notably, the hydration of anions and cations with the same cavity sizes should differ, owing to the different structure of the water molecules around an anion and a cation, but Born theory only depends on the squared ion charge. However, a Born correction is justified at longer distances and can be used to correct for hydration energies outside a cavity radius defined by the cutoff procedure.

There are two main approaches to modeling ion hydration: One is to use space-filling replicas of the simulation cell with periodic boundary conditions (PBC), typically boxes of water with a central ion in each cell. Another is to simulate only one cell without replicas and include nonperiodic boundary conditions (NPBD). Examples of the latter include

\* Corresponding author e-mail: william.jorgensen@yale.edu.

a pure vacuum calculation, a truncated physical system without periodicity, or possibly a truncated system with constraints to the surface region, as in a droplet with spherical boundary conditions (SBC). There are advantages and disadvantages to both approaches.<sup>9</sup>

For PBC, the use of lattice techniques such as Ewald summation provides an opportunity to calculate the electrostatics of the system, as a periodic solution to the Poisson equation. However, Ewald summation does not describe the asymptotically decaying electrostatic field, because of the periodicity, i.e., the finite concentration. A self-energy correction can be used to remove the artificial self-interaction, but the model system is still made up of replicas of finite systems. In charged systems, the use of a neutralizing background charge density, usually chosen to be uniform, is deemed necessary, even though this does not model a real counterion density.<sup>10</sup> Another issue is the boundary conditions: It has been found that tin foil boundary conditions give substantial errors.<sup>11</sup> However, this artifact can be remedied by using vacuum boundary conditions instead.<sup>11,12</sup>

Even with these approximations, calculations with PBC on small clusters give reasonable and size independent hydration energies when using Ewald summation with a self-energy correction.<sup>13</sup> A cancellation of self-energies and water polarization contributes to this success, but this effect is not yet well understood.<sup>9</sup> As an alternative, a Born correction may be applied in PBC with a cavity radius equal to the IW cutoff radius, but the resulting free energies will depend critically, even arbitrarily, on the value of the WW cutoff.<sup>3</sup>

The alternative to PBC is typically to use a droplet of water with the ion situated in the center. To remove finite-size effects, several suggestions for SBC are available. Constraints on the density of the droplet and on the polarization of surface water molecules to fit bulk properties are common.<sup>14</sup> The surface potential, the work required to move a unit charge through the surface, was suggested to make up for the difference between PBC and SBC.<sup>15,16</sup> It varies less than 0.5 kcal/mol with increasing system size beyond a cavity radius of 10 Å,<sup>15</sup> though estimates of its magnitude vary, e.g., 2.3 kcal/mol,<sup>17,18</sup> 4 kcal/mol,<sup>19</sup> and 12 kcal/mol,<sup>15</sup> depending on the model.<sup>20</sup> A value of 3 kcal/mol was calculated specifically for the TIP4P water model used in the present work.<sup>21</sup> Interestingly, experiments face very similar problems, with pros and cons on the sides of cluster-ion experiments and bulk solutions.<sup>19</sup>

With PBC, a decision on truncation of the IW and WW interactions is necessary. There is still no consensus on the best way to truncate interactions. While people often use exactly half the box length as the IW cutoff,<sup>22,23</sup> it is also advised to use IW cutoffs “significantly shorter than half the box length”.<sup>24</sup> Cutoff corrections from continuum theory have been made for both IW and WW interactions, rendering the earlier problematic calculations<sup>3</sup> size independent to within  $\pm 0.7$  kcal/mol.<sup>25</sup> A comparison was done between PBC and SBC, in the form of the surface constrained all-atom solvent model by Warshel and co-workers (SCAAS).<sup>26</sup> It was pointed out that when applying the corrections, the discrepancy between the methods was 5 kcal/mol for the  $\text{Ne} \rightarrow \text{Na}^+$  mutation. This suggests that at least one method is

**Table 1.** Experimental Ionic Radii, First Maxima of Radial Distribution Functions, and Free Energies and Entropies of Hydration

ion	$r_1$ (Å) <sup>a</sup>	$R_1$ (Å) <sup>b</sup>	$-\Delta G_{\text{hyd}}^c$			$-T\Delta S_{\text{hyd}}^c$
			Marcus <sup>d</sup>	Noyes <sup>e</sup>	Tissandier <sup>f</sup>	Marcus <sup>d</sup>
$\text{Li}^+$	0.69	2.08	113.5	122.1	126.4	13.1
$\text{Na}^+$	1.02	2.36	87.2	98.4	101.3	12.0
$\text{K}^+$	1.38	2.80	70.5	80.6	84.1	8.4
$\text{Rb}^+$	1.49	2.89	65.7	75.5	78.6	7.2
$\text{Cs}^+$	1.70	3.14	59.8	67.8	73.1	7.2
$\text{NH}_4^+$	1.48		68.1			9.6
$\text{F}^-$	1.33	2.63	111.1	89.5	102.5	10.8
$\text{Cl}^-$	1.81	3.19	81.3	76.1	72.7	6.0
$\text{Br}^-$	1.96	3.37	75.3	69.2	66.2	4.8
$\text{I}^-$	2.20	3.65	65.7	60.3	57.4	3.6

<sup>a</sup> Ionic radii from ref 27. <sup>b</sup> First maximum of ion–water oxygen radial distribution functions from ref 29. <sup>c</sup>  $\Delta G$  and  $T\Delta S$  for ion(gas)  $\rightarrow$  ion(aq) in kcal/mol. <sup>d</sup> Reference 27. <sup>e</sup> Reference 28. <sup>f</sup> Reference 19.

less accurate, or both have uncertainties of that magnitude.<sup>25</sup> However, similar numbers (−109 and −108 kcal/mol) were obtained for PBC and SCAAS using 10 Å cutoffs, suggesting that the methods yield similar results when the cutoffs are identical.<sup>25</sup> Instead, the infinite cutoffs that can be applied in the SCAAS models seem to be responsible for the difference. Since this effect mainly applies to WW interactions, these calculations are more repulsive than the solvent–solvent cutoff correction would imply, thereby giving less favorable hydration energies by up to 5 kcal/mol. Other corrections for WW cutoffs have been studied, but the results were only reasonably independent of system size when differences in IW and WW cutoffs were 2 Å or less.<sup>24</sup>

Some compilations of experimental data on ion hydration are shown in Table 1. Since experimental hydration free energies range over at least 10 kcal/mol,<sup>19,27,28</sup> reproducing absolute free energies of hydration is not as important as obtaining *size independent* models of ion hydration, i.e., models that give similar answers for various cell sizes, and therefore can be corrected with a simple constant independent of system size.

Importantly, relative free energies of hydration within a series remain in quantitative agreement, except for the value for  $\text{F}^-$  computed by Noyes, which is off by 15 kcal/mol. Another problem is the relative free energies between the alkali metal ion and halide series. In the Marcus compilation,<sup>27</sup>  $\text{Li}^+ \approx \text{F}^-$ , whereas in the Tissandier compilation,<sup>19</sup>  $\text{Li}^+$  is better hydrated by 24 kcal/mol. In particular, the Tissandier values are similar for  $\text{Cs}^+$  and  $\text{Cl}^-$ , which have similar ionic radii ( $r_1$ ) and first maxima of their radial density distributions ( $R_1$ ) (these two properties are linearly correlated). However, it is well-known that anions have more negative  $\Delta G_{\text{hyd}}$  for similar radii,<sup>30</sup> so Tissandier’s cluster-pair approximation seems less adequate for separating hydration contributions for anions and cations. Because of these considerations, the Marcus compilation has been used as the target data for the present work.

A number of additional computational studies of ions in solution should be noted. For the particular case of halide and alkali metal ions, early Monte Carlo simulations of  $\text{Li}^+$ ,



$\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{F}^-$ , and  $\text{Cl}^-$  in water were carried out by Mezei and Beveridge,<sup>31</sup> and simulations of  $\text{Na}^+$  in water, methanol, and tetrahydrofuran were performed by Chandrasekhar and Jorgensen.<sup>32</sup> Another detailed study described the volumes, coordination numbers, hydration energies, and energy components for  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{F}^-$ , and  $\text{Cl}^-$  based on a parameter set fitted to HF/3-21G\* or HF/3-21+G geometries and interaction energies.<sup>33</sup> For obtaining hydration free energies, Brooks studied  $\text{Cl}^-$  and  $\text{Br}^-$  ions by thermodynamic perturbation and temperature derivatives.<sup>23</sup> Free energy perturbations (FEPs) have been applied by Lybrand et al. to compute the difference in  $\Delta G_{\text{hyd}}$  for  $\text{Cl}^-$  and  $\text{Br}^-$ ,<sup>34</sup> by Jorgensen et al. to compute the absolute  $\Delta G_{\text{hyd}}$  of  $\text{Cl}^-$ ,<sup>35</sup> and by Åqvist to obtain absolute free energies for alkaline and alkaline earth metal ions.<sup>36</sup> These studies showed that relative free energies of hydration can be obtained with good accuracy. The study by Åqvist demonstrated that it is also meaningful to aim for *absolute* free energies of hydration, and from there, relative free energies follow. Another study by Migliore et al.<sup>37</sup> used a similar perturbation approach to calculate the relative hydration free energies of  $\text{Li}^+$ ,  $\text{Na}^+$ , and  $\text{K}^+$  as well as  $\text{F}^-$  and  $\text{Cl}^-$ .

Ion parameters used in standard software packages differ substantially and give rise to different structures and energetics for ions in water.<sup>38</sup> The present focus is on development of parameters optimized to reproduce liquid-phase properties, as in the philosophy of the OPLS force fields and the approach by Åqvist.<sup>36</sup> In addition to Åqvist's widely used parameters, other LJ parameters were developed for  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{F}^-$ , and  $\text{Cl}^-$  by Chandrasekhar et al.,<sup>33</sup> for these ions and  $\text{K}^+$  by Pettitt et al.,<sup>39</sup> and for  $\text{Na}^+$  and  $\text{K}^+$  by Marrone and Merz.<sup>40</sup> All alkali metal ions and halide ions were parametrized by Peng et al. based on experimental crystal data.<sup>41</sup> They also included a compilation and discussion of other alkali metal and halide ion parameters. Some negative conclusions<sup>40</sup> regarding transferability of ion models were found to be invalid.<sup>42,43</sup> The treatment of long-range interactions remains a major reason for discrepancies between ion hydration studies.

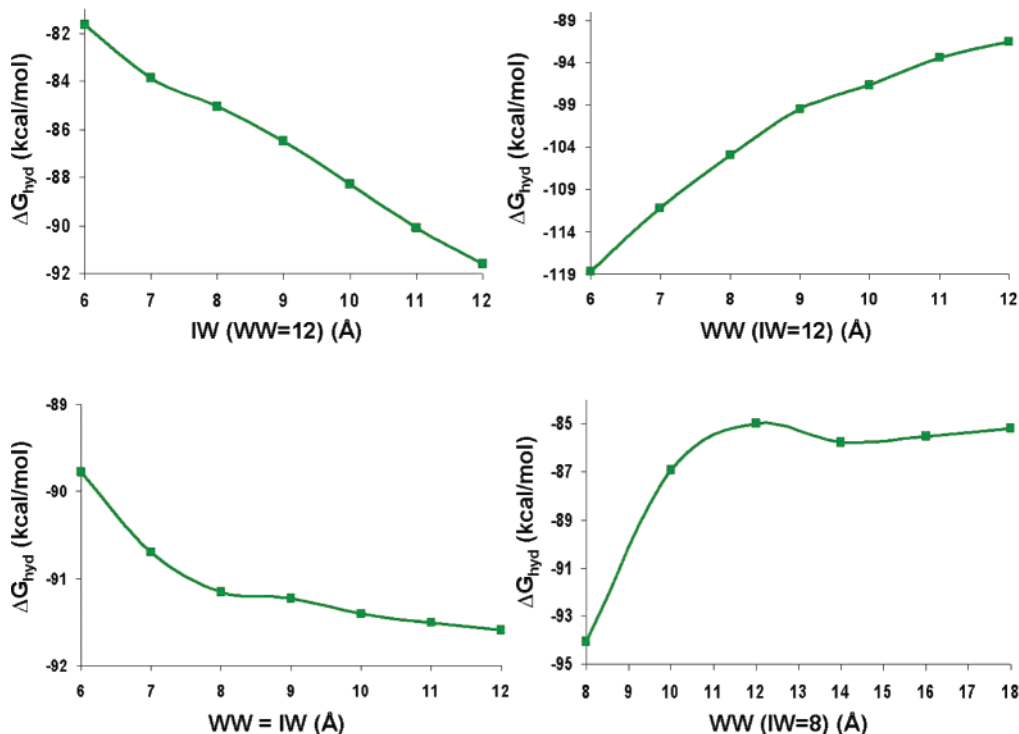
Despite these efforts, no consistent OPLS-style (based on fitting to liquid-phase data) set of parameters has so far been developed for the halide ion series. By consistent, it is meant that relative free energies and structures between the ions should be well reproduced. The purpose of the present work is to develop such a consistent set of force field parameters for the halide ions. In addition, it would be highly desirable to have one set of force field parameters giving free energies and structures consistent for *both* cations and anions, since an important goal is to obtain reliable energies and structures also in situations containing multiple ions, such as for a protein with counterions and added salt. Therefore, parameters have also been obtained for the alkali metal ions and united-atom  $\text{NH}_4^+$ , using the same set of experimental target data.

It has been argued that the effective, additive LJ potential is suitable for simulations as long as parameters are obtained in a consistent manner;<sup>42</sup> the limited adoption of polarizable force fields indicates that effective pair potentials will continue to play an important role in molecular simulations.

## Methods

**Computational Details.** Monte Carlo (MC) statistical mechanics simulations were performed to obtain averaged thermodynamic and structural results. All calculations were carried out with the BOSS software, version 4.6, following standard protocols including use of the TIP4P water model<sup>44</sup> and preferential sampling for water molecules near the ion.<sup>45</sup> As described below, most simulations were carried out for a single ion in water droplets with 10- or 15-Å radii containing 137 or 472 water molecules. Free energy perturbations (FEPs) were carried out to obtain ensemble-averaged free energies of hydration by annihilating the ions. In one set of FEPs, charges were neutralized, and subsequently the Lennard-Jones parameters were perturbed to zero. Absolute free energies were obtained in this way for all ions. Generally, 10 windows with double-wide sampling were used for an FEP. Each MC simulation was run for 5–10M configurations of equilibration and 10–20M configurations of averaging, depending on system size. The typical statistical uncertainty for the computed free-energy changes was ca. 0.1 kcal/mol. The temperature for all simulations was 298 K and the pressure was 1 atm when periodic boundary conditions were used. In the cutoff procedure, the nonbonded energy between the entire water molecule and the ion was evaluated if any of the three water atoms was closer than the cutoff distance to the ion. The simulations with PBC and TIP4P boxes were subject to a variety of both WW and IW cutoff distances, whereas in the droplet simulations all interactions were evaluated within the droplets using cutoffs of 100 Å. The Lennard-Jones parameters for the ions were developed to simultaneously yield good agreement with the experimental free energies of hydration and locations of the first peaks in ion–water radial distribution functions. Consistent with the practice for the OPLS-AA force field, geometric combining rules are used here for the Lennard-Jones parameters, i.e.,  $\epsilon_{ij} = (\epsilon_{ii} \epsilon_{jj})^{1/2}$ ,  $\sigma_{ij} = (\sigma_{ii} \sigma_{jj})^{1/2}$ . As noted previously,<sup>43</sup> lack of attention to this detail can lead to significant discrepancies.

**Calibration of the Model.** The following describes the choice of simulation model used here. Due to problems mentioned already with Ewald summation (boundary conditions, finite ion concentration, and the use of uniform background charges), such an approach was deemed impractical for the present purpose. As an alternative, using a Born correction in PBC without Ewald summation makes little sense insofar as the problem of WW interactions is not well accounted for. Ideally, all WW interactions within the “cavity” formed by the IW cutoff should be included, so that the remaining hydration free energy outside the cavity is effectively that predicted by continuum theory. But with the use of PBC, IW cutoffs have to be smaller than or equal to half the box length. As described earlier, the magnitude of the hydration free energy then depends critically on the choice of WW cutoff.<sup>3</sup> To get a clearer picture of these effects, FEP annihilations were performed for Åqvist's  $\text{Na}^+$  ion in boxes of TIP4P water with PBC and various WW and IW cutoffs. The results are shown in Figure 1. When the IW and WW cutoffs are equal,  $\Delta G_{\text{hyd}}$  varies little beyond a cutoff of 9 Å, implying cancellation of errors. However,

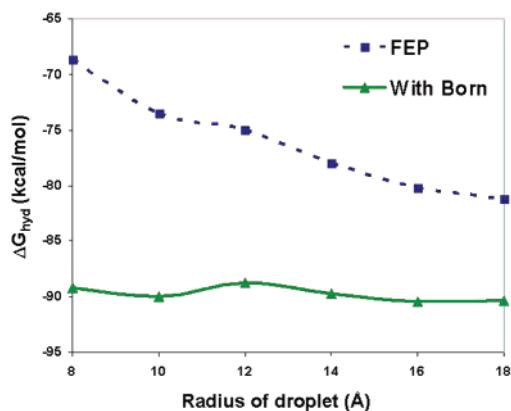


**Figure 1.** Simulations of Åqvist's  $\text{Na}^+$  in TIP4P boxes with PBC. Effect of IW and WW cutoffs.

the additional results are not consistent with continuum theory, implying that this model is unbalanced. Interestingly, size consistency is obtained as the WW cutoff increases to twice the IW cutoff. The box used to study the dependence on WW cutoff from 8 to 18 Å contained 2500 water molecules, using 10 M configurations for equilibration and 20 M for averaging. However, these calculations are very time-consuming, in particular for parametrization work, where large numbers of FEP calculations are required.

As an alternative, use of a droplet with a Born correction should give size-independent results within the error of the surface potential and the SBC.<sup>15</sup> These errors amount to 3–4 kcal/mol for pure water<sup>19</sup> independent of system size and 1 kcal/mol for the surface polarization (as calculated with SCAAS),<sup>25</sup> i.e., less than the errors in estimates of absolute hydration free energies for ions from experiment. This was tested in FEP annihilations of  $\text{Na}^+$  in TIP4P water droplets of varying size. A half-harmonic restoring force of 1 kcal/mol-Å<sup>2</sup> was used to inhibit potential evaporation of water molecules from the surface of the droplet, though this is a very rare event in the absence of the restoring potential. The results are shown in Figure 2. When the hydration free energy is given as the simulated free energy plus the Born correction, with the Born radius equal to the droplet radius, the free energies of hydration become size independent and give a free energy of hydration of -90 kcal/mol. Åqvist's result with the same ion parameters, SCAAS model, and SPC water was -98 kcal/mol.<sup>36</sup>

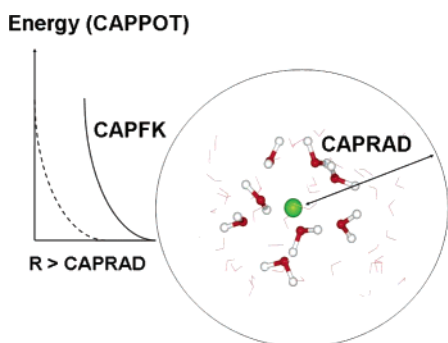
Because of these considerations, the present work has been carried out on water droplets using infinite IW and WW cutoffs. Use of a droplet with a 10-Å radius for parametrization was justified by the preliminary results. However, all optimized parameter sets were also validated in droplets with 15-Å radii. Larger droplets do not affect the free energy of



**Figure 2.** Calculated free energy of hydration for Åqvist's  $\text{Na}^+$  in TIP4P water droplets with infinite (100 Å) cutoffs for both water–water and ion–water interactions.

hydration beyond the uncertainty of the simulations ( $\sim 1$  kcal/mol), as long as infinite cutoffs and a Born correction are applied. The Born model was shown to be consistent with the scaling of size corrections in large water clusters containing an ion.<sup>46</sup>

The present model is depicted in Figure 3. The ion is positioned at the CAP origin; it is not moved so it remains at the center of the droplet throughout the simulation. Otherwise, as the ion is discharged and becomes more hydrophobic it would tend to migrate to the surface. Three setups were tried to test the stability of the model using different boundary treatments: (i) droplets without any constraints (a restoring force constant  $\text{CAPFK} = 0$ ); (ii) droplets with a potential CAPPOT starting at the droplet radius  $\text{CAPRAD}$ , with  $\text{CAPFK} = 1$  kcal/mol-Å<sup>2</sup>; and (iii) droplets with  $\text{CAPFK} = 1$  kcal/mol-Å<sup>2</sup>, but starting at  $\text{CAPRAD} + 5$  Å. The small values of CAPFK reflect a



**Figure 3.** The spherical model used in the present work. CAPRAD is the radius of the droplet in Å, CAPPOT is the half-harmonic restraint potential beyond CAPRAD (full line) or CAPRAD + 5 Å (dashed line), and CAPFK is the force constant for the half-harmonic potential.

**Table 2.** Optimized Lennard-Jones Parameters for Ions

ion	$\sigma$	$\epsilon$
F <sup>-</sup>	3.05	0.71
Cl <sup>-</sup>	4.02	0.71
Br <sup>-</sup>	4.28	0.71
I <sup>-</sup>	4.81	0.71
Li <sup>+</sup>	2.87	0.0005
Na <sup>+</sup>	4.07	0.0005
K <sup>+</sup>	5.17	0.0005
Rb <sup>+</sup>	5.60	0.0005
Cs <sup>+</sup>	6.20	0.0005
NH <sub>4</sub> <sup>+</sup>	5.34	0.0005

compromise to minimize artificial constraints on the density of the droplets, while still preventing significant evaporation of water molecules, which could compromise the validity of the Born correction. As shown below, all three setups give very similar results. The free energies of hydration are insensitive to these mild boundary constraints, mainly due to the larger radial force that the ion and neighboring water molecules already exert on the water molecules near the surface.

## Results and Discussion

**Halide Parameters.** The optimized Lennard-Jones parameters are shown in Table 2. There exists a positive correlation between both  $\sigma$  and  $\epsilon$  and  $R_1$  and a negative correlation between both  $\sigma$  and  $\epsilon$  and the magnitude of  $\Delta G_{\text{hyd}}$ . Because of this, freedom is limited during parametrization when fitting to  $R_1$  and  $\Delta G_{\text{hyd}}$ . Even then, it was possible to find one value of  $\epsilon$ , which could be used to fit all halide ions and all cations, respectively. It is not obvious that there should be a particular trend in  $\epsilon$  parameters for ions when going down a period. The trends in existing force field parameter sets<sup>41</sup> are mainly due to predefined end points of the free energy curves used during fitting. In fact, for the noble gases, the trend in  $\epsilon$  parameters is opposite of that in most ion sets, i.e.,  $\epsilon$  increases going down the period.<sup>47</sup> This is reasonable, as dispersion interactions are expected to scale with the size of the neutral atoms. However, for ions, the LJ potential must include polarization effects indirectly through fitting of parameters and not with an explicit physical term, so no particular trend

**Table 3.** Results for Halide Ions in Aqueous Solution

ion	$-\Delta G_{\text{hyd}}(10)^a$	$-\Delta G_{\text{hyd}}(15)$	$R_1(15)^b$	$-\Delta G_{\text{hyd}}(\text{exp})$	$R_1(\text{exp})$
CAPFK = 0 kcal/mol-Å <sup>2</sup>					
F <sup>-</sup>	111.8	110.3	2.75	111.1	2.63
Cl <sup>-</sup>	81.0	80.3	3.25	81.3	3.19
Br <sup>-</sup>	75.3	77.5	3.35	75.3	3.37
I <sup>-</sup>	66.0	67.2	3.55	65.7	3.65
RMS	0.4	1.5	0.08		
CAPFK = 1 kcal/mol-Å <sup>2</sup> , $R_0 = \text{CAPRAD} + 5 \text{ \AA}$					
F <sup>-</sup>	110.5	109.5	2.75	111.1	2.63
Cl <sup>-</sup>	81.5	81.2	3.25	81.3	3.19
Br <sup>-</sup>	76.5	76.2	3.35	75.3	3.37
I <sup>-</sup>	67.6	65.0	3.55	65.7	3.65
RMS	1.2	1.0	0.08		
CAPFK = 1 kcal/mol-Å <sup>2</sup>					
F <sup>-</sup>	111.6	110.4	2.75	111.1	2.63
Cl <sup>-</sup>	81.4	79.6	3.25	81.3	3.19
Br <sup>-</sup>	75.9	76.3	3.35	75.3	3.37
I <sup>-</sup>	66.1	66.9	3.60	65.7	3.65
RMS	0.4	1.2	0.07		

<sup>a</sup> Free energies of hydration  $\Delta G_{\text{hyd}}$  in kcal/mol  $\pm$  0.5 kcal/mol.

<sup>b</sup> First maxima of radial distribution functions,  $R_1$  in Å  $\pm$  0.025 Å. Parentheses refer to the radius of the water droplet, CAPRAD.

is trivially anticipated. Thus, the minimal model of identical  $\epsilon$  parameters for all halides has been adopted here. A value of  $\epsilon = 0.71$  kcal/mol gave the best fit to both structure and energies for all halide ions. On the other hand, the  $\sigma$  parameters follow the clear trend of increasing ion sizes, which is consistent with the physical picture of increasing ionic radii.

Table 3 shows the obtained free energies of hydration with the three different models: no restraint force, a force constant CAPFK of 1 kcal/mol Å<sup>2</sup> turned on at CAPRAD, and shifted to CAPRAD + 5 Å. All three models were applied with a CAPRAD of both 10 and 15 Å to test for size effects. Thus, for each final set of halide parameters, six values of  $\Delta G_{\text{hyd}}$  are presented. While optimization of the parameters was done with the smaller droplets, it is seen from Table 2 that all six models provide good absolute free energies, with a RMS deviation of 0.4–1.5 kcal/mol. This is the limiting accuracy given the approximations in the setup. Notably, the system size independence for all three CAP models is excellent.

Furthermore, the first maxima of the ion-oxygen radial distribution functions (RDFs),  $R_1$ , display a RMS deviation from experiment of 0.07–0.08 Å. The structure of the first solvation shell is very insensitive to the choice of model, again confirming the validity of the approach. Only the maxima for the models with a 15-Å radius are listed since the results are essentially identical for the smaller droplets. The largest discrepancy is for F<sup>-</sup>; the peak position could not be fit to better than 0.12 Å without losing accuracy for the free energy result. This conclusion was drawn from an extensive search that involved simultaneously varying both  $\epsilon$  and  $\sigma$ , with  $\epsilon$  varying from 0.0001 to 1.0 kcal/mol and  $\sigma$  from 2 to 4 Å. Larger values of  $\epsilon$  worked well for the halide ions, whereas smaller values were optimal for the cations. The difference again arises from the variations in the hydration of anions and cations for a given ionic radius.

**Cation Parameters.** The optimized Lennard-Jones parameters for the alkali metal ions and a united-atom model

**Table 4.** Optimized Parameters and Results for Cations in Aqueous Solution<sup>c</sup>

ion	$-\Delta G_{\text{hyd}}(10)^a$	$-\Delta G_{\text{hyd}}(15)$	$R_1(15)^b$	$-\Delta G_{\text{hyd}}(\text{exp})$	$R_1(\text{exp})$
CAPFK = 0 kcal/mol-Å <sup>2</sup>					
Li <sup>+</sup>	114.4	111.9	2.05	113.5	2.08
Na <sup>+</sup>	87.6	87.4	2.45	87.2	2.36
K <sup>+</sup>	69.4	72.9	2.85	70.5	2.80
Rb <sup>+</sup>	65.7	65.4	2.95	65.7	2.89
Cs <sup>+</sup>	59.1	60.5	3.20	59.8	3.14
NH <sub>4</sub> <sup>+</sup>	68.8	68.4	2.90	68.1	2.85
RMS	0.7	1.2	0.06		
CAPFK = 1 kcal/mol-Å <sup>2</sup> , $R_0 = \text{CAPRAD} + 5 \text{ \AA}$					
Li <sup>+</sup>	113.8	112.8	2.05	113.5	2.08
Na <sup>+</sup>	88.8	87.4	2.50	87.2	2.36
K <sup>+</sup>	70.8	71.6	2.85	70.5	2.80
Rb <sup>+</sup>	66.6	66.5	3.00	65.7	2.89
Cs <sup>+</sup>	59.3	60.0	3.20	59.8	3.14
NH <sub>4</sub> <sup>+</sup>	70.3	68.4	2.90	68.1	2.85
RMS	1.2	0.7	0.08		
CAPFK = 1 kcal/mol-Å <sup>2</sup>					
Li <sup>+</sup>	114.5	113.3	2.05	113.5	2.08
Na <sup>+</sup>	87.9	86.4	2.50	87.2	2.36
K <sup>+</sup>	71.1	69.7	2.85	70.5	2.80
Rb <sup>+</sup>	66.1	64.4	3.00	65.7	2.89
Cs <sup>+</sup>	59.7	58.9	3.20	59.8	3.14
NH <sub>4</sub> <sup>+</sup>	68.8	68.3	2.90	68.1	2.85
RMS	0.7	0.8	0.08		

<sup>a</sup> Free energies of hydration  $\Delta G_{\text{hyd}}$  in kcal/mol. <sup>b</sup> First maxima of radial distribution functions,  $R_1$  in Å. <sup>c</sup> Parentheses refer to the radius of the water droplet, CAPRAD.

of ammonium ion are displayed in Table 2. The alkali metal parameters have been reparametrized from Åqvist's original parameters and fitted to the  $\Delta G_{\text{hyd}}$  and  $R_1$  of Marcus' compilation, with the aim of providing parameters that reproduce experimental data from the same data set. Again, a minimal model with a single choice for  $\epsilon$  was successfully applied. The value of 0.0005 kcal/mol for  $\epsilon$  was best-suited to reproduce both structure and free energies simultaneously. The  $\sigma$  parameters again followed the trend of increasing ion size.

The results obtained with the new cation parameters are displayed in Table 4. As before, all six models provide good absolute free energies of hydration, this time with a RMS deviation of 0.7–1.2 kcal/mol, a similar range as for the halide ions. As seen, system-size independence of the results is obtained with both anions and cations.

The values for  $R_1$  are reproduced with a similar accuracy as before, i.e., with a RMS deviation from experiment of 0.06–0.08 Å. These deviations are similar to the experimental uncertainty, which is estimated at 0.06–0.08 Å.<sup>29</sup> Except for Li<sup>+</sup>, all peak positions are slightly larger than from experiment, but other choices of  $\epsilon$  did not provide as good overall results. As a rough conclusion, it can be said that the new parameters reproduce absolute free energies of hydration to about 1 kcal/mol and absolute and relative first maxima of ion-oxygen RDFs to within 0.1 Å.

**Comparison with Results using Prior Parameters.** To evaluate the quality of the new parameter set in more detail, its performance is compared to that using previous param-

**Table 5.** (a) Free Energies of Hydration  $\Delta G_{\text{hyd}}$  (kcal/mol) with Old Halide Parameters and (b) Comparison of Relative Free Energies of Hydration  $\Delta\Delta G_{\text{hyd}}$  (kcal/mol) Obtained with Old and New Halide Parameters

a					
ion	$\sigma$	$\epsilon$	$-\Delta G_{\text{hyd}}(10)^a$	$-\Delta G_{\text{hyd}}(15)^a$	$-\Delta G_{\text{hyd}}(\text{exp})$
F <sup>-b</sup>	2.73295	0.72	126.8	126.5	111.1
Cl <sup>-b</sup>	4.41724	0.11779	77.7	79.6	81.3
Br <sup>-c</sup>	4.62376	0.09	73.8	76.1	75.3
I <sup>-d</sup>	5.40	0.07	61.3	62.9	65.7
RMS			8.4	7.9	
b					
ion	$\Delta\Delta G_{\text{hyd}} \text{ old}^e$	$\Delta\Delta G_{\text{hyd}} \text{ new}^e$	$\Delta\Delta G_{\text{hyd}}(\text{exp})$		
F <sup>-</sup> – Cl <sup>-</sup>	48.0	30.4	29.8		
Cl <sup>-</sup> – Br <sup>-</sup>	3.7	4.3	6.0		
Br <sup>-</sup> – I <sup>-</sup>	14.1	9.8	9.6		
RMS	10.9	1.1			

<sup>a</sup> Parentheses refer to the radius of the water droplet. CAPFK = 0 kcal/mol Å<sup>2</sup>. <sup>b</sup> Reference 33. <sup>c</sup> Reference 34. <sup>d</sup> Reference 48. <sup>e</sup> Average of models with CAPRAD = 10 and 15. CAPFK = 0 kcal/mol Å<sup>2</sup>.

**Table 6.** (a) Free Energies of Hydration  $\Delta G_{\text{hyd}}$  (kcal/mol) with Old Cation Parameters and (b) Comparison of Relative Free Energies of Hydration  $\Delta\Delta G_{\text{hyd}}$  (kcal/mol) Obtained with Old and New Cation Parameters

a					
ion	$\sigma$	$\epsilon$	$-\Delta G_{\text{hyd}}(10)^a$	$-\Delta G_{\text{hyd}}(15)^a$	$-\Delta G_{\text{hyd}}(\text{exp})^a$
Li <sup>+b</sup>	2.12645	0.01828	116.5	115.0	113.5
Na <sup>+b</sup>	3.33045	0.00277	88.9	91.9	87.2
K <sup>+b</sup>	4.93463	0.00033	73.9	75.5	70.5
Rb <sup>+b</sup>	5.62177	0.00017	69.2	71.2	65.7
Cs <sup>+b</sup>	6.71700	0.00008	66.5	64.1	59.8
NH <sub>4</sub> <sup>+c</sup>	3.25	0.17	76.6	78.2	68.1
RMS			5.1	5.8	
RMS <sup>d</sup>			4.0	4.4	
b					
ions	$\Delta\Delta G_{\text{hyd}} \text{ old}^e$	$\Delta\Delta G_{\text{hyd}} \text{ new}^e$	$\Delta\Delta G_{\text{hyd}}(\text{exp})$		
Li <sup>+</sup> – Na <sup>+</sup>	25.4	25.7	26.3		
Na <sup>+</sup> – K <sup>+</sup>	15.7	16.4	16.7		
K <sup>+</sup> – Rb <sup>+</sup>	4.5	5.6	4.8		
Rb <sup>+</sup> – Cs <sup>+</sup>	4.9	5.8	5.9		
K <sup>+</sup> – NH <sub>4</sub> <sup>+</sup>	-2.7	2.6	2.4		
RMS	2.4	0.5	---		
RMS <sup>f</sup>	0.9	0.6	---		

<sup>a</sup> Parentheses refer to the radius of the water droplet. CAPFK = 0 kcal/mol Å<sup>2</sup>. <sup>b</sup> Reference 36. <sup>c</sup> Reference 49. Charges: -0.40 (N), +0.35 (H). <sup>d</sup> Excluding NH<sub>4</sub><sup>+</sup>. <sup>e</sup> Average of models with CAPRAD 10 and 15. CAPFK = 0 kcal/mol Å<sup>2</sup>. <sup>f</sup> Excluding NH<sub>4</sub><sup>+</sup>.

eters in Tables 5 and 6. Table 5a shows the absolute free energies of hydration for the halide ions obtained with the small and large droplets with CAPFK = 0 kcal/mol-Å<sup>2</sup> and using the old parameters for F<sup>-</sup>,<sup>33</sup> Cl<sup>-</sup>,<sup>33</sup> Br<sup>-</sup>,<sup>34</sup> and I<sup>-</sup>.<sup>48</sup> These parameters also yield good system-size independence, confirming the general quality of the droplet model. However, the old halide parameters yield greater deviations from the absolute free energies of hydration, which is expected

since their derivation was different. The absolute free energies for  $\text{Cl}^-$ ,  $\text{Br}^-$ , and  $\text{I}^-$  are not far from experiment; the main error is due to a large overestimation of the magnitude of the free energy of hydration for  $\text{F}^-$ . For the halide ions taken together, the RMS deviation from the Marcus free energies is 8.4 kcal/mol for the smaller droplet and 7.9 kcal/mol for the larger droplet.

The corresponding relative free energies obtained with the old parameters are compared to the results from the new ones in Table 5b; the averages of the results obtained with the small and large droplets are used. Whereas the differential hydration free energy of  $\text{F}^-$  and  $\text{Cl}^-$  is overestimated by 18 kcal/mol with the old parameter set, the error is reduced to 0.6 kcal/mol with the new set. Because the parameters for the last three halide ions are reasonable, they also yield better relative free energies. However, the overall RMS deviation from the experimental relative free energies remains high, 10.9 kcal/mol, for the old halide parameters. This was one of the main problems to be corrected in the present work. With the new parameters, the RMS deviation from experimental relative free energies is 1.1 kcal/mol.

Table 6a,b provides similar comparisons with results using the earlier parameters for cations, i.e., the alkali metal ion parameters of Åqvist<sup>36</sup> and the all-atom OPLS parameters for ammonium ion.<sup>49</sup> The old alkali cation parameters provide free energies that are uniformly larger in magnitude by just a few kcal/mol than the experimental numbers, even though they were parametrized toward slightly different target data. This shows that the models used by Åqvist and here in fact provide similar results. Nevertheless, the reparametrization was carried for both the halide ions and cations for consistency and to achieve lower overall errors in comparison to the Marcus data.

The largest deviation is for the old ammonium ion, which was parametrized independently.<sup>49</sup> The Åqvist set alone has a RMS deviation from our target data of only 4.0 and 4.4 kcal/mol for the small and large droplets, respectively. As shown in Table 6b, the Åqvist parameters perform very well for relative free energies with a RMS deviation of 0.9 kcal/mol. Thus, it appears that relative free energies can be well reproduced with a variety of models if the parameter set is well parametrized to fit absolute free energies.<sup>36</sup> The new cation set performs similarly for the alkali metal ions with a RMS deviation of 0.6 kcal/mol. Therefore, the only effect of reparametrizing the alkali metal ion parameters has been to put anions and cations on a common scale and make the parameters simpler. Larger errors are seen for ammonium, for which the old OPLS all-atom model actually gives the wrong relative free energy of hydration compared to  $\text{K}^+$ . With the new united-atom parameters, this artifact is removed, and the complete cation set including  $\text{NH}_4^+$  has a RMS deviation of 0.5 kcal/mol from the experimental relative free energies.

**Gas-Phase Ion–Water Interactions.** To test the reasonability of the developed force-field parameters, it is common to calculate the interaction energies for monohydrates (complexes of the ions with a single water molecule) and compare them to ab initio or experimental data. Indeed, many force fields have been parametrized to fit such energies.

**Table 7.** Ion–Oxygen Distances and Interaction Energies for Ion–Water Complexes in the Gas Phase

ion	$R_{\text{O}}$ (Å) OPLS TIP4P	$R_{\text{O}}$ (Å) OPLS TIP3P	$R_{\text{O}}$ (Å) ab initio	$\Delta E$ OPLS TIP4P	$\Delta E$ OPLS TIP3P	$\Delta E_0$ ab initio	$\Delta H_{\text{hyd}}$ exp <sup>d</sup>
$\text{F}^-$	2.71	2.73	2.44 <sup>a</sup>	-19.5	-19.3	-25.1 <sup>a</sup>	-23.3
$\text{Cl}^-$	3.24	3.26	3.11 <sup>a</sup>	-13.0	-13.5	-13.1 <sup>a</sup>	-13.1
$\text{Br}^-$	3.38	3.38	3.32 <sup>a</sup>	-11.9	-12.5	-11.5 <sup>a</sup>	-12.6
$\text{I}^-$	3.62	3.63	3.63 <sup>a</sup>	-10.3	-11.0	-9.6 <sup>a</sup>	-10.2
$\text{Li}^+$	1.99	1.96	1.85 <sup>b</sup>	-28.1	-31.6	-33.5 <sup>b</sup>	-34.0
$\text{Na}^+$	2.43	2.41	2.28 <sup>b</sup>	-19.4	-21.6	-22.4 <sup>b</sup>	-24.0
$\text{K}^+$	2.80	2.77	2.59	-15.0	-16.6	-17.9 <sup>b</sup>	-16.9
$\text{Rb}^+$	2.93	2.90	2.86	-13.7	-15.2	-15.1 <sup>c</sup>	-16.0
$\text{Cs}^+$	3.11	3.08	3.07	-12.3	-13.6	-13.8 <sup>c</sup>	-14.0
$\text{NH}_4^+$	2.85	2.82		-14.5	-16.0		-17.2

<sup>a</sup> MP2/6-311++G\*\* from ref 50. <sup>b</sup> MP2/cc-pVQZ values from ref 51. <sup>c</sup> MP2/cc-pVTZ values from ref 52. <sup>d</sup> Reference 52.

However, in the case of force-field parameters fit to liquid-state data, it is not expected that a perfect fit is also obtained for the interaction energies. Deviations are expected owing to polarization effects, as effective two-body potentials are biased toward the *average* many-body situation in the liquid. Nevertheless, reasonable results for complexes are expected in view of the dominant Coulombic component of the interaction energy.

Table 7 provides a comparison of the interaction energies for monohydrates obtained with ab initio methods as well as with the new force field parameters and both a TIP4P and TIP3P water molecule. The force-field results are from energy minimizations, while the ab initio values are the electronic energy difference plus a correction for the change in zero-point energy, which is quite steady near +1 kcal/mol.<sup>49,50</sup> As can be seen, the interaction energies obtained with the liquid-phase force field are mostly a little less negative than the ab initio values. Concomitantly, the force field yields slightly larger ion-oxygen distances than the ab initio calculations. The differences are greatest for the smallest ions,  $\text{F}^-$  and  $\text{Li}^+$ , which are the most bound and most polarizing. Overall, the results appear reasonable and follow the expected trends with increasing ion size. Furthermore, the differences for the ion-oxygen distances between the TIP3P and TIP4P water models are negligible, while the interaction energies are generally more favorable with TIP3P by 1–2 kcal/mol. However, the optimal water–water interaction is also stronger for the TIP3P water dimer (-6.50 kcal/mol) than the TIP4P dimer (-6.24 kcal/mol),<sup>44</sup> so the free energies of hydration for the ions with either water model should be similar.

**Structure for Ions in TIP4P Water.** Structural information for the ions in aqueous solution is reflected in the ion–water radial distribution functions and coordination numbers. For the first solvent shell, the coordination number is estimated by integrating the computed RDF to the first minimum. The results described here come from the MC simulations with the 15-Å droplets and with the restoring potential turned on 5 Å beyond CAPRAD.

Table 8 compares the performance of the new and old parameters for the first-shell coordination numbers. First of all, experimental estimates of coordination numbers vary widely,<sup>29</sup> indicating that different techniques and counterions

**Table 8.** First-Shell Coordination Numbers for Ions in Water

ion	CN (new)	CN (old)	CN (lit.)
F <sup>-</sup>	6.6	6.2	4.0–6.8 <sup>a</sup>
Cl <sup>-</sup>	7.6	7.3	5.9–8.5 <sup>a</sup>
Br <sup>-</sup>	7.7	7.4	4.2–8.9 <sup>a</sup>
I <sup>-</sup>	7.9	7.7	6.0–8.9 <sup>a</sup>
Li <sup>+</sup>	4.8	5.0	4.0–6.1 <sup>a</sup>
Na <sup>+</sup>	6.2	5.9	4.0–8.0 <sup>a</sup>
K <sup>+</sup>	7.0	7.6	5.0–8.0 <sup>a</sup>
Rb <sup>+</sup>	7.2	7.7	5, <sup>b</sup> 7.1, <sup>c</sup> 7.9 <sup>d</sup>
Cs <sup>+</sup>	7.9	7.5	6.0–8.0 <sup>a</sup>
NH <sub>4</sub> <sup>+</sup>	7.3	6.6	8.1–10.0 <sup>a</sup>

<sup>a</sup> Literature values from ref 29. <sup>b</sup> Reference 53. <sup>c</sup> From QM/MM simulations, ref 54. <sup>d</sup> Reference 55.

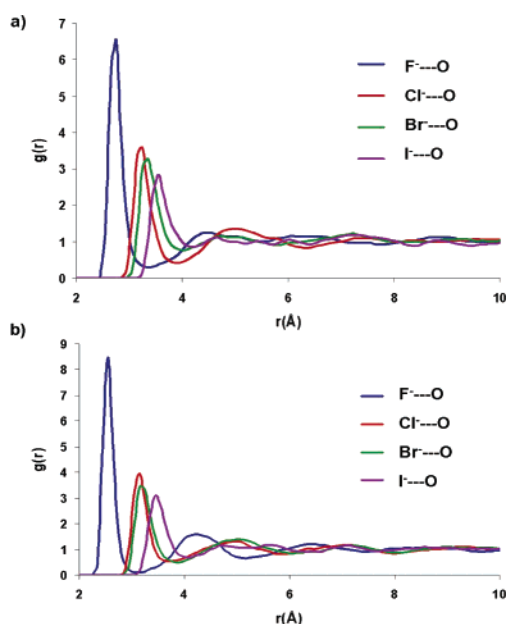
give different results. In addition, the computational estimate has an uncertainty of ca. 0.5 owing to the flatness of the RDFs near the first minima. Thus, differences in the coordination numbers should not be overinterpreted.

Looking first at the halide ions, both old and new parameters give similar results, which all fall comfortably within the experimental ranges. In addition, the expected increase in coordination number going down the periods is observed, with values of 6.6, 7.6, 7.7, and 7.9 for F<sup>-</sup>, Cl<sup>-</sup>, Br<sup>-</sup>, and I<sup>-</sup>. Qualitatively, the results confirm that the structures of the hydrated ions are consistent with experiment, in terms of how many water molecules on average are located within the first hydration sphere.

For the cations, a similar pattern is seen, with coordination numbers well within the experimental ranges and the expected trend of increasing coordination numbers with size, 4.8 for Li<sup>+</sup>, 6.2 for Na<sup>+</sup>, 7.0 for K<sup>+</sup>, 7.2 for Rb<sup>+</sup>, and 7.9 for Cs<sup>+</sup>. The coordination number of 7.3 for NH<sub>4</sub><sup>+</sup> falls below the experimental range of 8.1–10.0, even though the value is what could be expected from the trends in size. With the new parameters, NH<sub>4</sub><sup>+</sup> is between K<sup>+</sup> and Rb<sup>+</sup> in size, and all three have a coordination number of about 7. It is unclear how a model with the hydrogens of NH<sub>4</sub><sup>+</sup> explicit would lead to a higher coordination number. In fact, the old model for NH<sub>4</sub><sup>+</sup> is an all-atom one and yields a coordination number of 6.6. Also, again both the free energy of hydration and position of the first maximum in the N–O RDF can be well reproduced with the present united-atom model for ammonium ion.

The computed RDFs are shown in Figure 4 for the halide ions, and the computed positions of the first maxima for all ion–oxygen RDFs are listed in Table 9 using both the old and new parameters. For the halide ions, substantial improvement is made with the new parameters in the first-peak positions with a better trend and a lower RMS deviation from experiment, 0.08 Å vs 0.16 Å for the old halide parameter set. The experimental uncertainty is reported as 0.03–0.07 Å for halide ions.<sup>29</sup> A better fit that was also accurate for the free energies of hydration could not be obtained with any other choice of  $\epsilon$ , nor with freely varying  $\epsilon$ -values, which was initially attempted.

The second part of Table 9 compares the coordination data for the cations. The Åqvist parameter set<sup>36</sup> displays a RMS deviation from experiment of 0.08 Å, similar to that from

**Figure 4.** Radial distribution functions for halide ions with the (a) new and (b) old parameters in TIP4P water droplets with 15-Å radii at 298 K.**Table 9.** Calculated and Experimental Positions for the First Maxima of Ion–Water Oxygen Radial Distribution Functions

ion	$R_1$ (new)	$R_1$ (old)	$R_1$ (lit.) <sup>a</sup>
F <sup>-</sup>	2.75	2.55	2.63
Cl <sup>-</sup>	3.25	3.15	3.19
Br <sup>-</sup>	3.35	3.15	3.37
I <sup>-</sup>	3.55	3.45	3.65
RMS	0.08	0.16	
Li <sup>+</sup>	2.05	2.05	2.08
Na <sup>+</sup>	2.50	2.35	2.36
K <sup>+</sup>	2.85	2.65	2.80
Rb <sup>+</sup>	2.95	2.85	2.89
Cs <sup>+</sup>	3.20	3.05	3.14
NH <sub>4</sub> <sup>+</sup>	2.90	2.75	2.85
RMS	0.07	0.08	
RMS <sup>b</sup>	0.08	0.08	

<sup>a</sup> Reference 29. <sup>b</sup> Excluding NH<sub>4</sub><sup>+</sup>.

both the new alkali metal and halide parameters. The experimental uncertainty is reported as 0.06–0.08 Å for alkali metal ions.<sup>29</sup> Thus, structural accuracy similar to the Åqvist set was obtained now for both the cations and halide ions.

**Transferability to a PBC Setup.** Fluid simulations are often carried out in periodic cells. Therefore, it is important to test the validity of the relative hydration free energies and structures under such conditions. Tables 10–13 show the results calculated with a typical setup of 749 TIP4P water molecules in a periodic box with the ion in the center. Importantly, a choice of cutoff has to be made, as discussed in the Introduction. Commonly, cutoffs of 10 Å for both WW and IW interactions are chosen. As illustrated in Figure 1, the resultant absolute free energies are somewhat arbitrary. However, reasonably accurate relative free energies should be obtained since they should be dominated by the differ-

**Table 10.** Absolute and Relative Free Energies of Hydration (kcal/mol) Obtained with New Halide Parameters under Periodic Boundary Conditions

ion	$-\Delta G_{\text{hyd}}$ (calc)	$-\Delta G_{\text{hyd}}$ (exp)	diff
F <sup>-</sup>	107.0	111.1	4.1
Cl <sup>-</sup>	75.2	81.3	6.1
Br <sup>-</sup>	72.8	75.3	2.5
I <sup>-</sup>	61.9	65.7	3.8

ion	$\Delta\Delta G_{\text{hyd}}$	$\Delta\Delta G_{\text{hyd}}$	diff
F <sup>-</sup> – Cl <sup>-</sup>	31.8	29.8	2.0
Cl <sup>-</sup> – Br <sup>-</sup>	2.4	6.0	-3.6
Br <sup>-</sup> – I <sup>-</sup>	10.9	9.6	1.3
RMS			2.5

**Table 11.** Location of First Maxima of Radial Distribution Functions (Å) Obtained with New Halide Parameters under Periodic Boundary Conditions

ion	$R_1$ (calc)	$R_1$ (exp)	diff
F <sup>-</sup>	2.75	2.63	0.12
Cl <sup>-</sup>	3.25	3.19	0.06
Br <sup>-</sup>	3.35	3.37	0.02
I <sup>-</sup>	3.55	3.65	0.10

**Table 12.** Absolute and Relative Free Energies of Hydration (kcal/mol) Obtained with New Cation Parameters under Periodic Boundary Conditions

ion	$-\Delta G_{\text{hyd}}$ (calc)	$-\Delta G_{\text{hyd}}$ (exp)	diff
Li <sup>+</sup>	127.8	113.5	14.3
Na <sup>+</sup>	102.5	87.2	15.3
K <sup>+</sup>	86.6	70.5	16.1
Rb <sup>+</sup>	80.5	65.7	14.8
Cs <sup>+</sup>	76.2	59.8	16.4
NH <sub>4</sub> <sup>+</sup>	85.0	68.1	16.9

ion	$\Delta\Delta G_{\text{hyd}}$	$\Delta\Delta G_{\text{hyd}}$	diff
Li <sup>+</sup> – Na <sup>+</sup>	25.3	26.3	-1.0
Na <sup>+</sup> – K <sup>+</sup>	15.9	16.7	-0.8
K <sup>+</sup> – Rb <sup>+</sup>	6.1	4.8	1.3
Rb <sup>+</sup> – Cs <sup>+</sup>	4.3	5.9	-1.6
K <sup>+</sup> – NH <sub>4</sub> <sup>+</sup>	1.6	2.4	-0.8
RMS			1.1

**Table 13.** Location of First Maxima of Radial Distribution Functions (Å) for Cations under Periodic Boundary Conditions

ion	$R_1$ (calc)	$R_1$ (exp)	diff
Li <sup>+</sup>	2.05	2.08	0.03
Na <sup>+</sup>	2.50	2.36	0.14
K <sup>+</sup>	2.85	2.80	0.05
Rb <sup>+</sup>	3.05	2.89	0.16
Cs <sup>+</sup>	3.20	3.14	0.06
NH <sub>4</sub> <sup>+</sup>	2.90	2.85	0.05

ences in the electrostatic interactions with the water molecules nearest to the ions.

Table 10 lists the free energy results obtained with the new halide parameters and the PBC setup. The absolute free

energies are systematically less negative than experiment by 2.5–6.1 kcal/mol. The relative free energies computed with the PBC setup have a RMS deviation from experiment of 2.5 kcal/mol. This is 1.5 kcal/mol less accurate than for the spherical setups used to parametrize the force field (Table 3), but a substantial improvement over the results with the old parameters in Table 5b.

Table 11 shows the computed locations of the first maxima for the halide ion-oxygen RDFs with the new parameters under PBC. This property is insensitive to the choice of system geometry, since it is controlled by the sampling of water molecules close to the ions. The RMS deviation from experiment is again 0.08 Å, as in Table 9.

Table 12 shows the corresponding absolute and relative free energies of hydration obtained with the new cation parameters under PBC. The absolute free energies of hydration are systematically 14.3–16.9 kcal/mol too negative. However, the relative free energies have a RMS deviation from experiment of only 1.1 kcal/mol, an accuracy similar to that obtained with the droplet models. Table 13 gives the positions of the first maxima of the RDFs; the RMS deviation from experiment is 0.1 Å. Thus, for computing structures and relative free energies of hydration, the current parameter sets also work well for periodic boundary conditions.

As a last test of the transferability of the developed parameters, the absolute and relative free energies for all ions were computed with the TIP3P water model instead of TIP4P,<sup>44</sup> using a CAPRAD of 15 Å and the modified CAPPOT. Interestingly, the TIP3P water model hydrates the cations (Table 14a) more strongly, whereas it hydrates the anions (Table 14b) less than with TIP4P, in both cases by approximately 5 kcal/mol. This means that the consistency between the cation and anion series only applies when using the TIP4P water model. However, the *relative* free energies within each series are well reproduced in both TIP3P and TIP4P water. The RMS deviations from experiment using TIP3P are 1.2 and 1.7 kcal/mol for the cations and anions, respectively.

## Conclusions

Free energy perturbation calculations in the context of Monte Carlo simulations have been used to obtain a mutually consistent set of Lennard-Jones parameters for the halide ions, F<sup>-</sup>, Cl<sup>-</sup>, Br<sup>-</sup>, and I<sup>-</sup>, and for the alkali metal ions, Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Rb<sup>+</sup>, Cs<sup>+</sup>, as well as united-atom ammonium ion. The system setup for the calculations featured droplet models with different boundary restraints and radii. The alternatives gave very similar and system-size independent results when Born corrections were applied to correct for the contribution to the free energy of hydration from outside the droplet. The parameters were optimized for use with TIP4P water and yield absolute and relative free energies of hydration for the ions within about 1 kcal/mol of experiment. Structural characteristics are also well reproduced with average errors for ion–water oxygen distances of 0.1 Å or less, and the first-shell water coordination numbers fall within experimental ranges. Interaction energies and ion-oxygen distances for complexes of the ions with a single water molecule in

**Table 14.** Absolute and Relative Free Energies of Hydration for (a) Cations and (b) Anions in TIP3P Water

a. Cations			
ion	$-\Delta G_{\text{hyd}}$ (calc)	$-\Delta G_{\text{hyd}}$ (exp)	diff
Li <sup>+</sup>	120.8	113.5	7.3
Na <sup>+</sup>	92.3	87.2	5.1
K <sup>+</sup>	75.5	70.5	5.0
Rb <sup>+</sup>	70.1	65.7	4.4
Cs <sup>+</sup>	63.0	59.8	3.2
NH <sub>4</sub> <sup>+</sup>	73.1	68.1	5.0
ion	$\Delta\Delta G_{\text{hyd}}$	$\Delta\Delta G_{\text{hyd}}$	diff
Li <sup>+</sup> - Na <sup>+</sup>	28.5	26.3	2.2
Na <sup>+</sup> - K <sup>+</sup>	16.8	16.7	0.1
K <sup>+</sup> - Rb <sup>+</sup>	5.4	4.8	0.6
Rb <sup>+</sup> - Cs <sup>+</sup>	7.1	5.9	1.2
K <sup>+</sup> - NH <sub>4</sub> <sup>+</sup>	2.4	2.4	0.0
RMS			1.2
b. Anions			
ion	$-\Delta G_{\text{hyd}}$ (calc)	$-\Delta G_{\text{hyd}}$ (exp)	diff
F <sup>-</sup>	104.8	111.1	6.3
Cl <sup>-</sup>	77.6	81.3	3.7
Br <sup>-</sup>	72.5	75.3	2.8
I <sup>-</sup>	64.1	65.7	1.6
ion	$\Delta\Delta G_{\text{hyd}}$	$\Delta\Delta G_{\text{hyd}}$	diff
F <sup>-</sup> - Cl <sup>-</sup>	27.2	29.8	2.6
Cl <sup>-</sup> - Br <sup>-</sup>	5.1	6.0	0.9
Br <sup>-</sup> - I <sup>-</sup>	8.4	9.6	1.2
RMS			1.7

the gas phase were also found to be in good accord with ab initio results with the expected differences for the most polarizing ions, Li<sup>+</sup> and F<sup>-</sup>. Additional investigations included simulations under periodic boundary conditions and in TIP3P water; the high accuracy for relative free energies of hydration is maintained.

The new ion parameters are notable for their performance in these tests and for their simplicity with just a single value of the Lennard-Jones  $\epsilon$  for the anions and one for the cations. The good results for relative free energies of hydration are particularly auspicious for use of the new parameters in a wide variety of liquid-phase simulations, especially when halide and alkali cations are systematically varied.

**Acknowledgment.** Gratitude is expressed to the National Institutes of Health (GM032136) for support and to Prof. Johan Åqvist and Dr. Julian Tirado-Rives for helpful discussions. K.P.J. also acknowledges support from the European Molecular Biology Organization (long-term fellowship).

### References

- Brooks, C. L., III. *J. Chem. Phys.* **1987**, *86*, 5156–5162.
- Madura, J. D.; Pettitt, B. M. *Chem. Phys. Lett.* **1988**, *150*, 105–108.
- Straatsma, T. P.; Berendsen, H. J. C. *J. Chem. Phys.* **1988**, *89*, 5876–5886.
- Åqvist, J.; Hansson, T. *J. Phys. Chem. B* **1998**, *102*, 3837–3840.
- Heinz, T. N.; Hunenberger, P. H. *J. Chem. Phys.* **2005**, *123*, 034107.
- Kastenholz, M. A.; Hunenberger, P. H. *J. Chem. Phys.* **2006**, *124*, 124106.
- Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588–5593.
- Baker, N. A.; Hunenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 10679–10692.
- Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531–567.
- Hunenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 1856–1872.
- Vorobjev, Y. N.; Hermans, J. *J. Phys. Chem. B* **1999**, *103*, 10234–10242.
- Åqvist, J.; Hansson, T. *J. Phys. Chem.* **1994**, *98*, 3837–3840.
- Hummer, G.; Pratt, L. R.; Garcia, A. E. *J. Phys. Chem.* **1996**, *100*, 1206–1215.
- Essex, J. W.; Jorgensen, W. L. *J. Comput. Chem.* **1995**, *16*, 951–972.
- Darden, T.; Pearlman, D.; Pedersen, L. G. *J. Chem. Phys.* **1998**, *109*, 10921–10935.
- Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155–179.
- Parfenyuk, V. I. *Colloid J.* **2002**, *64*, 651–659.
- Zakharov, V. V.; Brodskaya, E. N.; Laaksonen, A. *J. Chem. Phys.* **1997**, *107*, 10675–10683.
- Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V. *J. Phys. Chem. A* **1998**, *102*, 7787–7794.
- Herce, D. H.; Darden, T.; Sagui, C. *J. Chem. Phys.* **2003**, *119*, 7621–7632.
- Wilson, M. A.; Pohorille, A.; Pratt, L. R. *J. Chem. Phys.* **1988**, *88*, 3281–3285.
- Hummer, G.; Pratt, L. R.; Garcia, A. E. *J. Phys. Chem.* **1998**, *102*, 7885–7895.
- Brooks, C. L., III. *J. Phys. Chem.* **1986**, *90*, 6680–6684.
- Resat, H.; McCammon, J. A. *J. Chem. Phys.* **1998**, *108*, 9617–9623.
- Wood, R. H. *J. Chem. Phys.* **1995**, *103*, 6177–6187.
- Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124–129.
- Marcus, Y. *Biophys. Chem.* **1994**, *51*, 111–127.
- Noyes, R. M. *J. Am. Chem. Soc.* **1962**, *84*, 513–522.
- Marcus, Y. *Chem. Rev.* **1988**, *88*, 1475–1498.
- Grossfield, A. *J. Chem. Phys.* **2005**, *122*, 024506.
- Mezei, M.; Beveridge, D. L. *J. Chem. Phys.* **1981**, *74*, 6902–6910.
- Chandrasekhar, J.; Jorgensen, W. L. *J. Chem. Phys.* **1982**, *77*, 5080–5089.
- Chandrasekhar, J.; Spellmeyer, D. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 903–910.



- (34) Lybrand, T. P.; Ghosh, I.; McCammon, J. A. *J. Am. Chem. Soc.* **1985**, *107*, 7793–7794.
- (35) Jorgensen, W. L.; Blake, J. F.; Buckner, J. K. *Chem. Phys.* **1989**, *129*, 193–200.
- (36) Åqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021–8024.
- (37) Migliore, M.; Corongiu, G.; Clementi, E.; Lie, G. C. *J. Chem. Phys.* **1988**, *88*, 7766–7771.
- (38) Patra, M.; Karttunen, M. *J. Comput. Chem.* **2004**, *25*, 678–689.
- (39) Pettitt, B. M.; Rosicky, P. J. *J. Chem. Phys.* **1986**, *84*, 5836–5844.
- (40) Marrone, T. J.; Merz, K. M., Jr. *J. Phys. Chem.* **1993**, *97*, 6524–6529.
- (41) Peng, Z.; Ewig, C. S.; Hwang, M.-J.; Waldman, M.; Hagler, A. T. *J. Phys. Chem. A* **1997**, *101*, 7243–7252.
- (42) Jorgensen, W. L.; Severance, D. L. *J. Chem. Phys.* **1993**, *99*, 4233–4235.
- (43) Åqvist, J. *J. Phys. Chem.* **1994**, *98*, 8253–8255.
- (44) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (45) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689–1700.
- (46) Rips, I.; Jortner, J. *J. Chem. Phys.* **1992**, *97*, 536–546.
- (47) Barrow, D. A.; Aziz, R. A. *J. Chem. Phys.* **1988**, *89*, 6189–6194.
- (48) McDonald, N. A.; Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1998**, *120*, 5104–5111.
- (49) Jorgensen, W. L.; Gao, J. *J. Phys. Chem.* **1986**, *90*, 2174–2182.
- (50) Kim, J.; Lee, H. M.; Suh, S. B.; Majumdar, D.; Kim, K. S. *J. Chem. Phys.* **2000**, *113*, 5259–5272.
- (51) Feller, D.; Glendening, E. D.; Woon, D. E.; Feyereisen, M. W. *J. Chem. Phys.* **1995**, *103*, 3526–3542.
- (52) Kebarle, P. *Annu. Rev. Phys. Chem.* **1977**, *28*, 445–476.
- (53) Briand, C. L.; Burton, J. J. *J. Chem. Phys.* **1976**, *64*, 2888–2895.
- (54) Hofer, T. S.; Randolph, B. R.; Rode, B. M. *J. Comput. Chem.* **2005**, *26*, 949–956.
- (55) Koneshan, S.; Rasaiah, J. C.; Lynden-Bell, R. M.; Lee, S. H. *J. Phys. Chem. B* **1998**, *102*, 4193–4204.

CT600252R

# JCTC

Journal of Chemical Theory and Computation

## Performance of the Effective Core Potentials of Ca, Hg, and Pb in Complexes with Ligands Containing N and O Donor Atoms

José-Zeferino Ramírez,<sup>†</sup> Rubicelia Vargas,<sup>\*,†</sup> Jorge Garza,<sup>†</sup> and Benjamin P. Hay<sup>‡</sup>

*Departamento de Química, Universidad Autónoma Metropolitana Iztapalapa, Apartado Postal 55-534, 09340, México, D. F., México, and Chemical Sciences Division, Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352*

Received December 11, 2005

**Abstract:** This paper presents a systematic study of the performance of the relativistic effective core potentials (RECPs) proposed by Dolg-Stoll-Preuss, Christiansen-Ermler, and Hay-Wadt for  $\text{Ca}^{2+}$ ,  $\text{Hg}^{2+}$ , and  $\text{Pb}^{2+}$ . The RECPs performance is studied when these cations are combined with ethyleneglycol, 2-aminoethanol, and ethylenediamine to form bidentate complexes. First, the description of the bidentate ligands is analyzed with the Kohn–Sham method by using SVWN, BLYP, and B3LYP exchange–correlation functionals, and they are compared with the Møller–Plesset perturbation theory (MP2); for all these methods the TZVP basis set was used. We found that the BLYP exchange–correlation functional gives similar results like those obtained by the B3LYP and MP2 methods. Thus, the bidentate metal complexes were studied with the BLYP method combined with the RECPs. To compare RECPs performance, all the systems considered in this work were studied with the relativistic all-electron Douglas-Kroll (DK3) method. We observed that the Christiansen-Ermler and Dolg-Stoll-Preuss RECPs give the best energetic and geometrical description for Ca and Hg complexes when compared with the all-electron method. For Pb complexes the spin–orbit effect and the basis set superposition error must be taken into account in the RECP calculations. In general, the trend showed in the complexation energies with the all-electron method is followed by the complexation energies computed with all the pseudopotentials tested in this work.

### I. Introduction

Coordination chemistry plays a crucial role in chemical science. Aqueous solvation, liquid–liquid partitioning, binding to proteins and other biomolecules, and catalysis are just a few of many examples. The study of the interactions that occur between ligands and metal ions is useful to establish the relationship between structure and the stability of a complex and the selectivity of a sequestering agent.<sup>1</sup> It is well-known that ligands with many donor atoms give thermodynamically more stable complexes than their analogous containing unidentate ligands.<sup>2</sup>

It is common to find neutral nitrogen and oxygen atoms as donor binding sites in multidentate ligands interacting with metal ions.<sup>3</sup> Nucleic acids, crown ethers, and macrocycles with antibiotic action are all examples where this kind of interaction is observed. From a quantum chemistry point of view, many of these systems have sizes that are prohibitive for an all-electron ab initio treatment. Additionally, some metal cations show relativistic effects, which make the calculations more expensive computationally.

These characteristics have motivated many authors to combine ab initio calculations with effective core potentials (ECPs) for modeling inner electrons in an atom.<sup>4</sup> For methods based on localized basis set functions, different ECPs approaches have been developed to describe the interaction between metal cations and several ligands. Two different

\* Corresponding author e-mail: ruvf@xanum.uam.mx.

<sup>†</sup> Universidad Autónoma Metropolitana Iztapalapa.

<sup>‡</sup> Pacific Northwest National Laboratory.

groups of ECPs may be distinguished in the literature according to the procedure to obtain them. One group considers that the atomic ECP should reproduce physical properties such as the first ionization potentials, the first excitation energies, and electron affinities. This approach has been proposed by the Stuttgart-Dresden group.<sup>5</sup> The second type of ECPs comes from the shape-consistent extraction method; in this, the ECP should give a nodeless pseudo-orbital, and it is constructed in such a way that the pseudo-orbital has the same shape of the valence part as the atomic orbital and its energy must be the same as that obtained from the all-electron atomic calculation. To this second group of ECPs belong Christiansen-Ermler<sup>6</sup> and Hay-Wadt ECPs,<sup>7</sup> for example. In the design of all these ECPs the Hartree–Fock method is used, and they are represented in terms of Gaussian functions. Dolg-Stoll-Preuss,<sup>5</sup> Christiansen-Ermler,<sup>6</sup> and Hay-Wadt<sup>7</sup> approaches are available in electronic structure codes and commonly used for the description of ligand–metal ion interactions. Some of these ECPs have been used in combination with Hartree–Fock and Møller–Plesset methods<sup>8</sup> for theoretical studies of organometallic compounds,<sup>9</sup> photochemical reactions,<sup>10</sup> metalation reactions in organic chemistry,<sup>11</sup> etc. Even when, in our knowledge, these ECPs have not been optimized or designed in the density functional theory (DFT)<sup>12</sup> context, DFT calculations have been combined with these ECPs to study adsorption phenomena onto catalyst surfaces,<sup>13</sup> defects in semiconductors,<sup>14</sup> metal–ligand bonding<sup>15</sup> and bonding energetics of zerovalent and cationic metals,<sup>16</sup> the O<sup>17</sup> and C<sup>13</sup> NMR chemical shifts in metal complexes,<sup>17</sup> and so on. All these cited examples solve a particular problem on a particular system, and there have been no systematic studies to show the performance of these ECPs on the structural and energetic of complexes where, specifically, neutral oxygen and nitrogen donor atoms interact with cations.

The aim of this paper is to study the performance of relativistic standard ECPs in combination with DFT, in the description of complexes between the metal ions Ca<sup>2+</sup>, Hg<sup>2+</sup>, and Pb<sup>2+</sup> with ligands where neutral nitrogen and oxygen are donor atoms (ethyleneglycol, 2-aminoethanol, and ethylenediamine). We chose these ligands for our research because they are analogues of common structural elements that occur in multidentate ligands. Thus, our conclusions regarding the performance of the models on these systems may be extrapolated to more complex ligands interacting with divalent cations. We have divided our work into four sections. In section II, the methodology applied in this study is described, in section III the results and discussion are presented, and in section IV the conclusions of this work are pointed out.

## II. Methodology

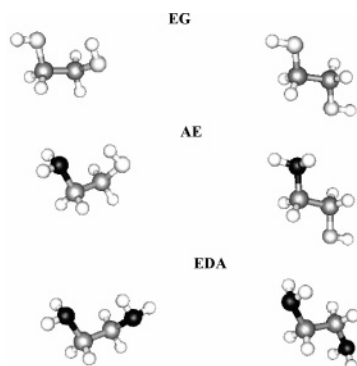
To choose the appropriate level of theory to describe the structure and conformational changes of free bidentate ligands ethyleneglycol (EG), 2-aminoethanol (AE), and ethylenediamine (EDA), within the Kohn–Sham (KS) model,<sup>12</sup> which is the most popular DFT implementation, three exchange-correlation functionals were tested. These include a local density approximation (SVWN),<sup>18</sup> a generalized gradient approximation (BLYP),<sup>19</sup> and a hybrid approach

(B3LYP).<sup>20</sup> We use, in all of these calculations, the polarized triple- $\zeta$  basis set (TZVP).<sup>21</sup> For comparison we carried out second-order Møller–Plesset (MP2) calculations with the same basis set on EG, AE, and EDA conformations. The results of both methodologies (DFT and MP2) for the most stable conformer of each molecule are contrasted with experimental data and previous results reported in the literature. From this comparison, we have selected the method to describe free bidentate ligands with the best ratio between the structural and energetic description and a reasonable computational effort.

To determine the ECPs performance on the energetic description of Ca, Hg, and Pb atoms, the first and second ionization potentials were estimated using each type of relativistic pseudopotentials tested in this work. For Ca, 10 core electrons were replaced with a relativistic effective core potential (RECP). We used three RECPs approaches: Dolg-Stoll-Preuss (Stuttgart),<sup>5</sup> Christiansen-Ermler (CRENBL),<sup>6</sup> and Hay-Wadt (LANL2DZ).<sup>7</sup> For the Hg case, 60 core electrons were replaced in Stuttgart, CRENBL and 68 for LANL2DZ RECPs. And for Pb, 60 core electrons were replaced with Stuttgart RECP, 68 with the CRENBL approach, and 78 with LANL2DZ. To determine the effect of using a different valence on the description of the energetics, Stuttgart RECPs with a different number of electrons in the core (large core, LC) were used: 18 core electrons were replaced in calcium, 78 for Hg, and 78 for Pb. To estimate spin–orbit effects in the Pb, we used the CRENBL RECP corrected by spin–orbit (CRENBL+SO) as was proposed by Christiansen<sup>6</sup> and that designed by Dolg–Stoll–Preuss for the Stuttgart RECPs (Stuttgart+SO and Stuttgart+SO+LC).<sup>5</sup> In all of them, we are using the coefficients of primitive Gaussians of RECPs and basis sets as were reported in the references.

The RECP results for ionization potentials were compared with experimental values. To provide additional points for comparison, ionization potentials were determined with relativistic all-electron (Douglas-Kroll, DK3)<sup>22a</sup> calculations in combination with the BLYP exchange correlation functional. The DK3 method is a higher-order scheme of the DK transformation extensively studied by Hess and co-workers.<sup>23</sup> The basis sets used in all calculations done with this method are as follows: (20s15p)/[4s2p] for Ca, (23s23p15d10f)/[6s4p3d1f] for Hg, and (23s23p15d10f)/[6s5p3d1f] for Pb; these were optimized for the DK3 method.<sup>22b</sup> However, this number of functions represents a minimal basis set for these metals. This is important because, for example, the basis set for Ca does not contain *d* orbitals that are important if we want to describe a bond in the complexes. For comparison, the performance of the DK3 method using a TZVP basis set (DK3/TZVP method) was tested, due to the size, just for Ca.<sup>21</sup> It is important to mention that the authors of the DK3 basis set functions suggest spherical coordinates for the calculations with this kind of functions.<sup>22c</sup> Thus, in this work we used this representation for all the calculations for metals, ligands, and complexes.

Complexes between EG, AE, and EDA and the metal cations (Ca<sup>2+</sup>, Hg<sup>2+</sup>, and Pb<sup>2+</sup>) were fully optimized with BLYP/TZVP for the ligands and each RECP described above



**Figure 1.** The most stable gauche,  $G_{\min}$ , and trans,  $T_{\min}$ , conformers for ethylenediamine (EDA), ethyleneglycol (EG), and 2-aminoethanol (AE), optimized at the BLYP/TZVP level. The black color is for N, gray for C, and white for O atoms; the H atoms are the smallest white spheres.

**Table 1.** Relative Energies (kcal mol<sup>-1</sup>) between the Most Stable Gauche and Trans Conformers of Ethyleneglycol (EG), 2-Aminoethanol (AE), and Ethylenediamine (EDA)<sup>a</sup>

ab initio level	EG	AE	EDA
SVWN	3.85	4.13	2.42
BLYP	2.78	2.71	1.54
B3LYP	2.79	2.73	1.57
MP2	2.84	2.89	1.67

<sup>a</sup> All computations were made with the TZVP basis set.

for metals. Also, the relativistic all-electron DK3 method was used in the optimization of the complexes to contrast with the RECP results, as in the case of ionization potentials; also the TZVP combined with DK3 for calcium complexes was tested. The complexation energies were calculated with each RECP as  $\Delta E = E(\text{complex}) - E(\text{ligand}) - E(\text{metal})$ . Again, these energies were compared with the relativistic all-electron (DK3) energies. Additionally, correction to the formation energies due to the basis set superposition error (BSSE) using the counterpoise method proposed by Boys and Bernardi are presented.<sup>24</sup> All the calculations were done with the NWChem v4.5 code.<sup>25</sup>

### III. Results and Discussion

**IIIa. Free Ligands.** The most stable gauche and trans conformers for EG, AE, and EDA are showed in Figure 1. From this figure we can see that the gauche conformers present an intramolecular hydrogen bond (structures at the left in Figure 1), which make this conformer more stable than the corresponding trans conformer. This observation is consistent with other studies in the literature.<sup>26</sup>

The relative energies between gauche and trans conformers are reported in Table 1 for each method tested in this work. From this table it is clear that the SVWN method gives the largest relative energy, suggesting that the hydrogen bond in the gauche conformer is overestimated. This is consistent with prior reports showing that SVWN overestimates intermolecular hydrogen bonds.<sup>27</sup> The BLYP and B3LYP results are similar to one another, and the largest deviation from the MP2 calculations is of 6.2% for AE. The MP2 relative energies are always higher than the BLYP and B3LYP ones.

The structural parameters for the hydrogen bonds in the gauche conformers are reported in Table 2. Consistent with

**Table 2.** Structural Parameters of the Hydrogen Bond in the Most Stable Gauche Conformers of Ethyleneglycol (EG), 2-Aminoethanol (AE), and Ethylenediamine (EDA)<sup>a</sup>

structural parameter	SVWN	BLYP	B3LYP	MP2
EG				
D O—H ... O	2.11	2.42	2.39	2.31
D O - - - O	2.67	2.87	2.83	2.79
< O—H ... O	114.1	107.5	107.1	109.6
AE				
D O—H ... N	1.97	2.27	2.26	2.19
D O - - - N	2.64	2.84	2.82	2.78
< O—H ... N	122.7	115.9	115.2	117.6
EDA				
D N—H ... N	2.27	2.53	2.50	2.43
D N - - - N	2.79	2.96	2.92	2.88
< N—H ... N	109.2	104.2	104.3	106.2

<sup>a</sup> Distances are in angstroms and angles in degrees. All computations were made with the TZVP basis set.

**Table 3.** Comparison of the Relative Energies (kcal mol<sup>-1</sup>) for Some Gauche and Trans Conformers of Ethyleneglycol (EG), 2-Aminoethanol (AE), and Ethylenediamine (EDA) Calculated in This Paper (BLYP/TZVP) and Previous Results

	EG		
	$G_{\min}$	$T_{\min}$	$T_1$
this work	0.00	2.78	2.85
MP2/cc-pVTZ//MP2/cc-pVDZ <sup>a</sup>	0.00	2.80	2.85
	AE		
	$G_{\min}$	$T_{\min}$	$G_1$
this work	0.00	2.71	2.05
B3LYP/6-311++G(2p,2d) <sup>b</sup>	0.00	2.85	1.90
	EDA		
	$G_{\min}$	$T_{\min}$	$T_1$
this work	0.00	1.54	1.64
MP2/6-31++G**//HF/6-31++G** <sup>c</sup>	0.00	1.66	1.75

<sup>a</sup> Reference 28. <sup>b</sup> Reference 29. <sup>c</sup> Reference 30.

the relative energies, the structural parameters from the SVWN method suggest an overestimation of the hydrogen bond strength, the donor–acceptor angles of this method are the most linear, and the hydrogen–acceptor and donor–acceptor distances are the shortest, while the BLYP and B3LYP methods give a similar description of the hydrogen bond structural parameters. This result is very important because calculations with the B3LYP exchange correlation functional are computationally more expensive than those with BLYP, and this can be significant when bigger systems are treated. The distances and angles of both methods (BLYP and B3LYP) are close to those calculated with the MP2 method. In accord with the relative energies, MP2 predicts slightly more linear angles and shorter hydrogen bond distances in comparison to the BLYP and B3LYP methods.

Because the performance of BLYP and B3LYP is very similar in the description of structural parameters and relative energies, we present in Table 3 only the relative energies of three conformers of each molecule calculated with the BLYP/TZVP method for comparison to results obtained from prior

**Table 4.** Comparison of Geometrical Parameters for the Most Stable Gauche Conformer of Ethylenglycol (EG) Predicted by the BLYP/TZVP Method and Experimental Parameters<sup>a</sup>

geometrical parameter	this work	experimental <sup>b</sup>
D O-H	0.978	0.961 ± 0.008
D C-O	1.457	1.424 ± 0.001
D C-C	1.525	1.517 ± 0.005
D O - - - O	2.868	2.813 ± 0.006
D O-H ... O	2.419	2.365 ± 0.045
< C-C-O	106.7	109.3 ± 0.4
< C-C-H	108.7	105.8 ± 2.7
< O-C-C-O	62.7	60.7 ± 1.8
< C-C-O1-H	-52.4	-54.0 ± 2.1
< C-C-O2-H	-165.8	-169.7 ± 2.0

<sup>a</sup> Distances are in angstroms and angles in degrees. <sup>b</sup> Reference 31.**Table 5.** Comparison of Geometrical Parameters for the Most Stable Gauche Conformer of Ethylenediamine (EDA) Predicted by the BLYP/TZVP Method and Experimental Parameters<sup>a</sup>

geometrical parameter	this work	experimental <sup>b</sup>
D C-C	1.536	1.545 ± 0.008
D C1-N1	1.489	1.469 ± 0.004
D C2-N2	1.481	1.469 ± 0.004
D N-H	1.024	0.995 ± 0.005
< C2-C1-N1	110.3	110.2 ± 0.7
< C1-C2-N2	110.0	110.2 ± 0.7
< N-C-C-N	64.0	64.0 ± 0.4

<sup>a</sup> Distances are in angstroms and angles in degrees. <sup>b</sup> Reference 32.

studies. We found that the relative deviations do not exceed 5%. We note that prior results obtained from a B3LYP calculation with the 6-311++G(2p,2d) basis set is very similar to that obtained by us with the BLYP/TZVP method.

Finally in Tables 4–6, the comparison between some available gas-phase experimental structural parameters of the most stable conformer of each molecule with the BLYP/TZVP results is shown. As it can be seen, only the C–C–O–H angle for the AE has a big deviation from the experimental value. In the paper where these results were reported, the authors mentioned they had doubts about this angle.<sup>33</sup> The angle of –27.0 degrees proposed in that work is not favorable for the hydrogen bond, and we found that this geometry to be unstable.

The results presented here on the ligand conformations allow us to conclude that the BLYP functional with the TZVP basis set properly describes the structural and energetic behavior of these bidentate ligands, and this method will be used in the following sections.

**IIIb. ECP Performance in the Ionization Potentials for Ca, Hg, and Pb.** To explore the energetic description of Ca, Hg, and Pb, the first and second ionization potentials (IP) were estimated by each type of RECP and the relativistic all-electron DK3 method at the BLYP ab initio level. The results are reported in Tables 7–9. To investigate the influence of a more extended valence on the energetic description, we used two RECPs from the Stuttgart group (named Stuttgart and Stuttgart-LC in the tables) that have

**Table 6.** Comparison of Geometrical Parameters for the Most Stable Gauche Conformer of 2-Aminoethanol (AE) Predicted by the BLYP/TZVP Method and Experimental Parameters<sup>a</sup>

geometrical parameter	this work	experimental <sup>b</sup>
D C-C	1.537	1.526 ± 0.016
D C-O	1.441	1.396 ± 0.010
D C-N	1.489	1.475 ± 0.023
D O-H	0.982	1.000 ± 0.020
D N-H1	1.022	1.017 ± 0.003
D N-H2	1.024	1.017 ± 0.005
D O - - - N	2.839	2.808 ± 0.005
D O-H ... N	2.269	2.300 ± 0.040
< C-C-O	111.2	112.1 ± 1.0
< C-C-N	108.8	108.1 ± 2.0
< C-O-H	104.5	108.0 ± 2.0
< C-N-H1	111.2	111.3 ± 0.8
< C-N-H2	110.8	110.4 ± 0.8
< H-N-H	107.1	109.9 ± 0.5
< O-C-C-N	55.0	55.4 ± 2.0
< C-C-O-H	-40.3	-27.0 ± 6.0
< C-C-N-H1	31.2	32.3 ± 1.0
< C-C-N-H2	-110.5	-110.5 ± 1.0

<sup>a</sup> Distances are in angstroms and angles in degrees. <sup>b</sup> Reference 33.**Table 7.** First and Second Ionization Potentials in eV for Ca<sup>a</sup>

method	first	second
Stuttgart	6.12	12.03
Stuttgart-LC	6.02	11.34
CRENBL	6.10	12.11
LANL2DZ	6.36	11.72
DK3	6.37	11.74
experimental <sup>b</sup>	6.11	11.86

<sup>a</sup> All methods were combined with BLYP. <sup>b</sup> Reference 34.**Table 8.** First and Second Ionization Potentials in eV for Hg<sup>a</sup>

method	first	second
Stuttgart	10.46	19.14
Stuttgart-LC	9.39	17.30
CRENBL	10.42	19.07
LANL2DZ	9.32	17.37
DK3	10.59	19.00
experimental <sup>b</sup>	10.43	18.74

<sup>a</sup> All methods were combined with BLYP. <sup>b</sup> Reference 34.

different numbers of core electrons. Additional RECPs reported in the tables have been described in the methodology section.

Comparing relativistic all-electron calculations using the DK3 method and experimental values, we found good agreement for the first and second ionization potentials for Ca, and they differ just in 4.3% and 1.0%, respectively. The test of the basis set TZVP with DK3 in this atom improves only the first IP since the relative error with this basis set is 0.7% (the first IP is 6.07 and the second 12.07 eV with the DK3/TZVP method), and for the second IP the relative error is 1.8%. For the Hg ionization potentials predicted by the DK3 method, a better agreement is observed; their results differ only about 1.5% from the experimental values, using just

**Table 9.** First and Second Ionization Potentials in eV for Pb<sup>a</sup>

method	first ionization potential	second ionization potential
Stuttgart	6.83	13.66
Stuttgart + SO	7.13	14.76
Stuttgart-LC	6.99	13.70
Stuttgart + SO-LC	7.29	14.41
CRENBL	6.36	12.81
CRENBL + SO	6.60	13.87
LANL2DZ	6.89	13.90
DK3	7.13	13.88
experimental <sup>b</sup>	7.41	15.01

<sup>a</sup> All methods were combined with BLYP. <sup>b</sup> Reference 34.

the basis set optimized for the DK3 described in the methodology. We can see that the DK3-BLYP performance is good for these systems. However, the DK3 method shows a different behavior for the IPs in the Pb than that observed for Ca and Hg atoms, since the first IP has a 3.8% relative error with respect to the experimental value, and the second IP 7.5%. The Ca and Hg results suggest that these deviations can be attributed to spin-orbit effects that are not considered in the implementation of the DK3 method in the code used in this work. We will come back in this issue later when the RECPs results for Pb are discussed. But in general, we may say that the energetic description of DK3 method for the metals and their ions is reasonably good, even when the basis sets optimized for this method are limited. We may increase the basis set, or test different exchange-correlation functionals, or explore the spin-orbit effects to assess the DK3 performance in the metals considered here; however, such a study is not the target of this work.

It is also observed in Tables 7–9 that Stuttgart-LC, for Ca and Hg, gives larger relative errors than Stuttgart (one magnitude order for the first IP), with respect to experimental IPs. While for Pb, the error of LC is almost the same than the more expanded valence. The Dolg-Stoll-Preuss more expanded valence RECP results are in better agreement with respect to experimental values, for calcium where the largest relative error is 1.4% for the second IP. For Hg, we found 0.3 and 2.1% deviations from the first and second IPs, respectively. As it was mentioned before, the bigger deviations are found for Pb, where the largest difference is observed in the second ionization potential, 9.0% for the more expanded valence; however, when this RECP in the corrected spin-orbit version is used, the difference with the experimental value goes to 1.7% and 3.8% for the first IP. For the first IP the Stuttgart large core corrected by the spin-orbit effect shows the lowest difference, 1.6% and 4.0% for the second IP.

Comparing all computed RECPs results for Ca and Hg with the experimental values (Tables 7 and 8), the smaller relative errors are observed in CRENBL and Stuttgart ECPs, and the relative errors are not over 2%. Contrary to Ca and Hg, the CRENBL RECP results for Pb show the largest deviations with respect to experimental IPs, 14.2 and 14.7% for the first and second IP, respectively. When the SO effect is included in the RECP, the errors diminish to 10.9 and 7.6%. It is important to say that we tested the spin-orbit

**Figure 2.** Complexes between Pb(II) and ethyleneglycol (EG), 2-aminoethanol (AE), and ethylenediamine (EDA), optimized at the BLYP/TZVP level combined with CRENBL RECP for the metal. The black color is for N, gray for C, and white for O atoms; the H atoms were suppressed for simplicity.**Table 10.** Structural Parameters for Ca<sup>2+</sup> with Ethylenediamine (EDA)<sup>a</sup>

method	D N – Ca	D N – - - N	< N–Ca–N	< N–C–C–N
Stuttgart	2.40	2.99	77.09	60.50
Stuttgart-LC	2.57	3.09	73.88	64.30
CRENBL	2.53	3.08	74.75	63.60
LANL2DZ	2.51	3.07	74.87	63.40
DK3	2.52	3.06	74.85	62.35

<sup>a</sup> All RECPs were combined with BLYP/TZVP. Distances are in angstroms and angles in degrees.

effect in the Hg IPs, using the Christiansen RECPs, and there was not a difference when this effect is not considered.

With respect to relativistic all-electron DK3 calculations, the largest deviations are obtained for LANL2DZ RECP, 12.2% for the first ionization potential and 8.6% for the second of the Hg atom.

**IIIc. Bidentate Ligand Complexes.** The metal complexes with dications of Ca, Hg, and Pb were fully optimized, using each RECP tested in this work for metals combined with the BLYP/TZVP for ligands. For comparison, relativistic all-electron DK3 calculations for all complexes also were computed. For the Ca complexes two basis sets were tested, TZVP and the minimal basis set for the DK3 method described in the methodology, although for consistency, in the correspondent tables just the results obtained with the latter basis set are reported. The optimized geometries for Pb complexes at the BLYP/TZVP level combined with CRENBL RECP are presented in Figure 2. Structural parameters for complexes with EDA ligand are reported in Tables 10–12. We decided to report just the complexes with EDA since the differences between RECPs and the DK3 method are similar in the other complexes. However, structural parameters for complexes with EG and AE are given in the Supporting Information.

Structural parameters in these tables (11 and 12) show that the geometries from the Stuttgart RECPs are in good agreement with the DK3 geometries, with the exception of the Hg<sup>2+</sup> complex, where relative errors reach around 15.5% for the N–Hg distance and 19.8% in the N–Hg–N angle when less expanded valence (Stuttgart-LC) RECP is used; it predicts a shorter donor–metal distance. For the Ca<sup>2+</sup>–EDA complex (Table 10), all RECPs predict structural parameters with a deviation less than 5% with respect to the DK3 method, for this case when the DK3/TZVP method is used the geometrical parameters are almost the same, just the N–Ca distance (2.49 Å) and the N–Ca–N (75.81 degrees) show a small change. For the same ligand with Hg<sup>2+</sup>, the largest relative error, after Stuttgart-LC, with respect to DK3 was found for Hay-Wadt RECP, 7.2% for the N–C–C–N angle (see Table 11). Curiously, better agreement is found

**Table 11.** Structural Parameters for Hg<sup>2+</sup> with Ethylenediamine (EDA)<sup>a</sup>

method	D N – Hg	D N – – N	< N–Hg–N	< N–C–C–N
Stuttgart	2.40	3.09	80.34	58.40
Stuttgart-LC	2.12	3.03	91.26	55.90
CRENBL	2.38	3.09	80.89	58.90
LANL2DZ	2.44	3.09	78.54	60.40
DK3	2.51	3.10	76.20	56.33

<sup>a</sup> All RECPs were combined with BLYP/TZVP. Distances are in angstroms and angles in degrees.

**Table 12.** Structural Parameters for Pb<sup>2+</sup> with Ethylenediamine (EDA)<sup>a</sup>

method	D N – Pb	D N – – N	< N–Pb–N	< N–C–C–N
Stuttgart	2.46	2.95	73.58	58.40
Stuttgart+SO	2.46	2.95	73.62	58.40
Stuttgart-LC	2.47	2.95	73.34	58.10
Stuttgart+SO-LC	2.47	2.95	73.39	58.10
CRENBL	2.58	2.99	70.85	60.40
CRENBL+SO	2.58	2.99	70.88	60.50
LANL2DZ	2.41	2.92	74.67	56.40
DK3	2.51	2.97	72.57	58.79

<sup>a</sup> All RECPs were combined with BLYP/TZVP. Distances are in angstroms and angles in degrees.

**Table 13.** Formation Energy in kcal/mol for Ca<sup>2+</sup> Complexes, without (wo) Considering BSSE Correction and with (w) This Correction<sup>a</sup>

Ca <sup>2+</sup> – EG	wo/BSSE	w/BSSE
Stuttgart	–110.46	–109.56
Stuttgart-LC	–87.01	–86.54
CRENBL	–92.19	–91.89
LANL2DZ	–92.36	–91.37
DK3	–91.46	–90.90
Ca <sup>2+</sup> – AE	wo/BSSE	w/BSSE
Stuttgart	–116.33	–115.21
Stuttgart-LC	–91.64	–91.22
CRENBL	–97.80	–97.34
LANL2DZ	–97.56	–96.60
DK3	–96.65	–96.06
Ca <sup>2+</sup> – EDA	wo/BSSE	w/BSSE
Stuttgart	–123.61	–122.29
Stuttgart-LC	–98.25	–97.90
CRENBL	–105.45	–104.82
LANL2DZ	–104.77	–103.93
DK3	–103.90	–103.28

<sup>a</sup> The BLYP exchange correlation functional was used in all calculations.

for the Pb<sup>2+</sup>–ligand, where the largest deviation was 4.2% for the same angle as the Hg<sup>2+</sup> complex and also for LANL2DZ RECP, as it is shown in Table 12. For all metal–ligand complexes, there is no significant change in angles and distances for the inclusion of spin–orbit correction in the CRENBL and Stuttgart RECPs.

The complexation energies for all metals, ligands, and methods explored in this work are reported in Tables 13–15. These complexation energies were computed both with and without BSSE corrections. Although the BSSE correc-

**Table 14.** Formation Energy in kcal/mol for Hg<sup>2+</sup> Complexes, without (wo) Considering BSSE Correction and with (w) This Correction<sup>a</sup>

Hg <sup>2+</sup> – EG	wo/BSSE	w/BSSE
Stuttgart	–172.27	–170.49
Stuttgart-LC	–169.14	–168.01
CRENBL	–171.17	–170.09
LANL2DZ	–139.36	–138.29
DK3	–164.80	–162.04
Hg <sup>2+</sup> – AE	wo/BSSE	w/BSSE
Stuttgart	–202.45	–200.67
Stuttgart-LC	–190.34	–189.34
CRENBL	–201.10	–199.84
LANL2DZ	–162.71	–161.54
DK3	–193.56	–190.52
Hg <sup>2+</sup> – EDA	wo/BSSE	w/BSSE
Stuttgart	–229.74	–227.62
Stuttgart-LC	–212.43	–211.52
CRENBL	–228.49	–226.83
LANL2DZ	–186.35	–185.00
DK3	–221.80	–218.60

<sup>a</sup> The BLYP exchange correlation functional was used in all calculations.

tions are small for Ca<sup>2+</sup> and Hg<sup>2+</sup>, they are important for Pb<sup>2+</sup> complexes. In the latter case, application of the BSSE correction results in about 7% change to the complexation energy when using the DK3 method. As expected, the RECP results are not as affected as the all-electron calculations, with the largest correction found in LANL2DZ and Stuttgart-LC RECPs for Pb<sup>2+</sup> complexes where the BSSE is around 4.6%.

Comparing the computed complexation energies using RECPs versus the DK3 method, we found that the worst deviation occurs with the Stuttgart RECP in Ca<sup>2+</sup> complexes, which gives about 20% of relative error in all Ca–ligand complexes. This behavior is observed due to the limited basis set optimized for the DK3 method. However, when the DK3/TZVP method is applied, the binding energies corrected by the BSSE are as follows: –98.39 kcal/mol for the EG, –103.42 kcal/mol for the AE and, –110.23 kcal/mol for the EDA calcium complexes; in all cases when the basis set of calcium has polarization functions the binding energy is increased in about 7.5%. This makes that the relative errors of Stuttgart RECP with respect to the DK3 method diminish around 10%. With respect to the DK3/TZVP method, Stuttgart-LC RECP shows the largest error (about 12%), and for the other RECPs the relative error is increased from less than 2% to around 6%. It means that when the basis set includes polarization functions in this atom, the DK3/TZVP and Stuttgart RECP methods give similar results.

An additional test was made for the DK3 method, and just for the Ca<sup>2+</sup> complexes we performed a conventional DFT calculations BLYP/TZVP; these results were compared with those obtained with DK3/TZVP. Geometrical parameters for all complexes are very similar; the bigger difference in distances is 0.06 Å and in angles 1.2 degrees. With respect to the complexes binding energies they differ by less than

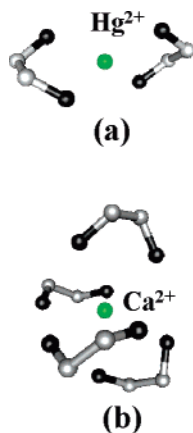
**Table 15.** Formation Energy in kcal/mol for  $\text{Pb}^{2+}$  Complexes, without (wo) Considering BSSE Correction and with (w) This Correction

$\text{Pb}^{2+} - \text{EG}$	wo/BSSE	w/BSSE
Stuttgart	-118.10	-116.39
Stuttgart+SO	-120.88	-118.20
Stuttgart-LC	-116.36	-110.97
Stuttgart+SO-LC	-117.41	-112.01
CRENBL	-101.65	-100.36
CRENBL+SO	-103.15	-101.84
LANL2DZ	-126.79	-120.96
DK3	-118.73	-110.23
$\text{Pb}^{2+} - \text{AE}$	wo/BSSE	w/BSSE
Stuttgart	-130.83	-128.91
Stuttgart+SO	-134.34	-131.54
Stuttgart-LC	-130.09	-124.28
Stuttgart+SO-LC	-131.57	-125.75
CRENBL	-112.74	-111.17
CRENBL+SO	-114.94	-113.35
LANL2DZ	-139.66	-133.40
DK3	-132.37	-123.39
$\text{Pb}^{2+} - \text{EDA}$	wo/BSSE	w/BSSE
Stuttgart	-144.90	-142.78
Stuttgart+SO	-149.12	-146.16
Stuttgart-LC	-145.22	-138.90
Stuttgart+SO-LC	-147.12	-140.79
CRENBL	-125.35	-123.49
CRENBL+SO	-128.20	-126.33
LANL2DZ	-153.64	-146.83
DK3	-147.21	-137.81

<sup>a</sup>The BLYP exchange correlation functional was used in all calculations.

1%. It means that the errors of the DK3 method occur, mainly due to the minimal basis set optimized for this method, more than the relativistic DK treatment.

With the LANL2DZ RECPs, the largest deviations with respect to DK3 were observed for the  $\text{Hg}^{2+}$  complexes, where the relative error is around 15.0%. Finally the CRENBL



**Figure 3.** Calculated structures for (a)  $[\text{Hg}(\text{EDA})_2]^{2+}$  and (b)  $[\text{Ca}(\text{EDA})_4]^{2+}$  complexes, optimized at the BLYP/TZVP level combined with CRENBL RECP for the metal. The black color is for N and gray for C; the H atoms were suppressed for simplicity.

RECPs give the largest differences with the  $\text{Pb}^{2+}$  complexes, but when the spin-orbit effect is considered in the binding energies for these complexes the relative error diminishes from 9.7% to 8% on average, after the BSSE correction.

In general, Stuttgart RECP binding energies are the values closest to the DK3 results, with the exception of the calcium complexes. It is important to note that just the Stuttgart RECPs (small core) are comparable in size to the TZVP basis set.

Regardless of the RECP, the complexation energies follow the same trend as the DK3 calculations. For all metal cations the most stable complex is that built with the EDA ligand, where two nitrogen atoms are acting as donor atoms, followed by the AE ligand, and the least stable complexes are formed with EG. It has been observed that in the gas phase the neutral nitrogen donor shows stronger metal coordinating properties than the neutral oxygen donor,<sup>1</sup> and our results agree with this observation. For a given ligand, always the complexes with  $\text{Hg}^{2+}$  are more stable than those with  $\text{Pb}^{2+}$ , and the latter more stable than complexes where  $\text{Ca}^{2+}$  is the metal ion. Thus, the most stable complex is the  $\text{Hg}^{2+}$ -EDA complex.

It is important to note that there are no experimental values to compare with our theoretical results for the simple 1:1 M:L complexes. Although there are crystal structures that contain bidentate ligands (EG, EDA, or AE) bound to the metal ion, additional bound ligands are also present. To benchmark the ability of the methodologies to describe experimental geometries, we carried out the same calculations with all RECPs in the  $[\text{Hg}(\text{EDA})_2]^{2+}$  and  $[\text{Ca}(\text{EDA})_4]^{2+}$  complexes where experimental structures are reported.<sup>35,36</sup> In Figure 3, the calculated structures of these complexes are depicted. The root-mean-square errors (RMS) of all RECPs geometries with respect to the experimental structural are on average 0.37 Å for the calcium complex and 0.31 Å for the  $[\text{Hg}(\text{EDA})_2]^{2+}$ . These errors show that all RECPs give a close description of the experimental structures, even when the calculations are performed in a vacuum.

In Tables 16 and 17 some structural parameters are listed to obtain a more detailed comparison between experimental and theoretical data. Relative errors with respect to experimental values are also given for  $[\text{Ca}(\text{EDA})_4]^{2+}$  and  $[\text{Hg}(\text{EDA})_2]^{2+}$  complexes in Tables 16 and 17. As we can see for the Ca complex all RECPs relative errors with respect to the experimental are less than 9%, the lowest errors are found when the Stuttgart more expanded valence RECP is used, and Stuttgart-LC, LANL2DZ, and CRENBL give similar errors. In the case of the Hg complex (Table 17) the lower relative errors, in general, are found with the Stuttgart and CRENBL RECPs. For this complex, the bigger relative error is found in one of the N-C-N dihedral angles; the cause of this difference is due to the presence of a  $\text{ClO}_4^-$  ion that appears in the experimental crystal structure, and it is close enough to interact with the  $\text{CH}_2$  group of one of the EDA molecules in the Hg complex. The ligand does not show that this interaction has a better agreement with the calculated dihedral angle.

In Tables 16 and 17 also DK3 structural parameters are given; in the calcium complex case the basis set used for



**Table 16.** Structural Parameters for the  $[\text{Ca}(\text{EDA})_4]^{2+}$  Complex<sup>a</sup>

geometrical parameter	Stuttgart	Stuttgart-LC	CRENBL	LANL2DZ	DK3	exp <sup>b</sup>
D N–Ca	2.70 (5.47)	2.79 (8.98)	2.78 (8.59)	2.77 (8.20)	2.76 (7.81)	2.56
D N–Ca	2.73 (4.60)	2.82 (8.05)	2.80 (7.28)	2.80 (7.28)	2.79 (6.90)	2.61
D N – – – N	2.94 (3.52)	2.98 (4.93)	2.97 (4.58)	2.97 (4.58)	2.97 (4.58)	2.84
< N–Ca–N	65.51 (1.64)	64.18 (3.63)	64.45 (3.23)	64.45 (3.23)	64.70 (2.85)	66.60
< N–C–C–N	–59.40 (3.50)	–61.10 (6.46)	–60.90 (6.12)	–60.90 (6.12)	–60.50 (5.40)	–57.39

<sup>a</sup> All RECPs were combined with BLYP/TZVP. Relative errors are reported in parentheses. Distances are in angstroms and angles in degrees.  
<sup>b</sup> Reference 35.

**Table 17.** Structural Parameters for the  $[\text{Hg}(\text{EDA})_2]^{2+}$  Complex<sup>a</sup>

geometrical parameter	Stuttgart	Stuttgart-LC	CRENBL	LANL2DZ	DK3	exp <sup>b</sup>
D N–Hg	2.41 (5.24)	2.20 (3.93)	2.40 (4.80)	2.48 (8.30)	2.43 (6.11)	2.29
D N–Hg	2.41 (3.88)	2.20 (5.17)	2.40 (3.45)	2.48 (6.90)	2.43 (4.74)	2.32
D N – – – N	2.97 (2.77)	2.92 (1.04)	2.97 (2.77)	3.02 (4.50)	2.96 (2.42)	2.89
D N – – – N	2.97 (1.02)	2.92 (0.68)	2.97 (1.02)	3.02 (2.72)	2.96 (0.68)	2.94
< N–Hg–N	76.14 (1.41)	83.09 (7.59)	76.35 (1.14)	74.83 (3.11)	75.07 (2.80)	77.23
< N–Hg–N	76.14 (2.98)	83.09 (5.87)	76.35 (2.71)	74.83 (4.65)	75.07 (4.35)	78.48
< N–C–C–N	–58.50 (0.46)	–55.90 (4.88)	–57.80 (1.65)	–60.70 (3.28)	–57.60 (1.99)	–58.77
< N–C–C–N	–58.50 (9.89)	–55.90 (13.89)	–58.10 (10.51)	–60.70 (6.50)	–57.60 (11.28)	–64.92

<sup>a</sup> All RECPs were combined with BLYP/TZVP. Relative errors are reported in parentheses. Distances are in angstroms and angles in degrees.  
<sup>b</sup> Reference 36.

**Table 18.** Structural Parameters for the  $[\text{Pb}(\text{EDA})_2]^{2+}$  Complex<sup>a</sup>

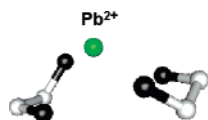
geometrical parameter	Stuttgart	Stuttgart-LC	CRENBL	LANL2DZ	DK3
D N–Pb	2.55 (1.16)	2.56 (0.78)	2.66 (3.10)	2.49 (3.49)	2.58
D N–Pb	2.66 (0.75)	2.70 (0.75)	2.75 (2.61)	2.63 (1.87)	2.68
D N – – – N	2.96 (0.34)	2.97 (0.00)	3.00 (1.01)	2.94 (1.01)	2.97
< N–Pb–N	69.23 (0.68)	68.59 (0.25)	67.42 (1.95)	70.01 (1.82)	68.76
< N–C–C–N	–59.50 (1.16)	–59.80 (0.66)	–61.50 (2.16)	–58.40 (2.99)	–60.20

<sup>a</sup> All RECPs were combined with BLYP/TZVP; numbers in parentheses are relative errors. Distances are in angstroms and angles in degrees.

the DK3 calculation was TZVP although there is not a big difference with the results obtained for structural parameters when the small basis set optimized for DK3 is used, as it was observed in the bidentate ligand complexes.

We also found a report of X-ray structure for a complex of Pb with EDA ( $[\text{Pb}(\text{EDA})_2]^{2+}$ ), but coordinates are not available. In this case we made the comparison only with respect to a DK3 all-electron calculation. For this Pb complex the RMS is 0.07 Å on average for RECPs with respect to the DK3 computation, and the largest RMS was 0.12 Å given for the overimposition with the LANL2DZ RECP. In

Table 18 a more detailed comparison between some RECPs and DK3 geometrical parameters is shown. As it can be seen, relative errors are not over 3.5%. For this complex it is important to mention that we took as a starting point, the geometry of the  $[\text{Hg}(\text{EDA})_2]^{2+}$  replacing the Hg by Pb. However all calculations give a structure where the metal ion is not at the middle of both ligands as in the Hg case; instead, the Pb is found above the center of the ligands as it is shown in Figure 4, and a similar behavior was observed previously for halogenated lead hydrides and methyl lead compounds.<sup>37</sup>



**Figure 4.** Calculated structure for the  $[\text{Pb}(\text{EDA})_2]^{2+}$  complex optimized with the relativistic all-electron DK3 method combined with BLYP/TZVP. The black color is for N and gray for C; the H atoms were suppressed for simplicity.

#### IV. Concluding Remarks

For the free ligands, both BLYP and B3LYP with the TZVP basis set reproduce experimental geometries as well as the geometries and relative energies calculated with other methods using bigger basis sets. Since BLYP and B3LYP exchange-correlation functionals both yield similar results, the BLYP method was selected for this study because it is computationally cheaper. As expected from earlier studies of intermolecular hydrogen bonds, the local SVWN exchange correlation functional overestimates hydrogen bond strength within the ligands, yielding shorter hydrogen bond distances and more linear hydrogen bonds.

The best results for Ca and Hg ionization potentials come from Dolg-Stoll-Preuss and Christiansen-Ermler RECPs. The Pb ionization potentials predicted by the RECPs show important differences with respect to the experimental ones. To obtain better results in Pb ionization potentials, it is important to use the version where spin-orbit effects are incorporated. However, a detailed study is required in this system to see which is the impact of a different exchange-correlation functional or different basis sets in the computation of IPs.

The basis sets optimized for the DK3 method reported in the literature, which were used here, are minimal for all metals. When the TZVP basis set is tested in combination with DK3 for the calcium atom, the first ionization potential improves with respect to the experimental value. However, in the description of complexes, the role of the basis set on the calcium atom is more relevant in small systems since in bigger systems the basis set functions of the ligands compensate for the limitations of the DK3 basis set of this atom. These problems are not observed for Hg and Pb complexes since, even when these metals have associated a minimal basis set, there are *d* and *f* functions involved.

For the complexes, Dolg-Stoll-Preuss RECPs show that it is necessary to use extended valence, even for Ca, to obtain geometrical parameters that are comparable to those obtained from relativistic all-electron calculations. The comparison between the Christiansen and Dolg-Stoll-Preuss RECPs with and without spin-orbit effects reveals that this correction does not significantly impact the geometrical parameters for any of the metal complexes. However, this correction is important to predict binding energies since it contributes at least 2 kcal/mol, depending on the complex.

All complexation energies calculated with RECPs follow qualitatively the same trend as those obtained from relativistic all-electron calculations. For all metal cations the most stable complex is formed with EDA. For each ligand, the complexes formed with  $\text{Hg}^{2+}$  are always more stable than those formed with  $\text{Pb}^{2+}$ , and these are always more stable than complexes with  $\text{Ca}^{2+}$ .

The Dolg-Stoll-Preuss more expanded valence and Christiansen-Ermler RECPs combined with BLYP/TZVP for the ligand are always closer to the all-electron DK3 method; we may recommend both methods since in the systems tested here the performance with bidentate ligands is good in all cases except for  $\text{Pb}^{2+}$ . In this case, the poor results can be notably improved if the BSSE is considered.

**Acknowledgment.** Financial support for J.-Z.R. was provided by CONACYT, México, through the scholarship 169487. R.V. acknowledges to CONACYT for financial support through the project 39621-F. B.P.H. was supported by the Environmental Management Science Program of the Office of Science, U.S. Department of Energy (Grant 73759). The research was performed in part at the Pacific Northwest National Laboratory (operated by Battelle for the DOE). We thank to the Laboratorio de Supercómputo y Visualización en Paralelo at UAM-Iztapalapa for the access to its computer facilities.

**Supporting Information Available:** Cartesians coordinates of optimized structures in angstroms of all free bidentate ligands and complexes and tables of structural parameters for 2-aminoethanol and ethyleneglycol complexes. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### References

- (1) Hancock, R. D.; Martell, A. E. *Chem. Rev.* **1989**, *89*, 1875–1914.
- (2) Martell, A. E.; Hancock, R. D. Introductory overview. In *Metal Complexes in Aqueous Solutions*; Fackler, J. P., Jr., Ed.; Plenum: New York, 1996; Modern Inorg. Chem. No. 7, pp 1-14.
- (3) Izatt, R. M.; Pawlak, K.; Bradshaw, J. S.; Bruening, R. L. *Chem. Rev.* **1995**, *95*, 2529–2586.
- (4) (a) Phillips, J. C.; Kleinman, L. *Phys. Rev.* **1959**, *116*, 287–294. (b) Weeks, J. D.; Rice, S. A. *J. Chem. Phys.* **1968**, *49*, 2741–2755. (c) Kahn, L.; Baybutt, P.; Truhlar, D. G. *J. Chem. Phys.* **1976**, *65*, 3826–3853. (d) Hay, P. J.; Wadt, W. R.; Kahn, L. R. *J. Chem. Phys.* **1978**, *68*, 3059–3066. (e) Preuss, H.; Stoll, H.; Wedig, U.; Krueger, T. *Int. J. Quantum Chem.* **1981**, *19*, 113–130.
- (5) (a) Fuentealba, P.; Szentpaly, L. V.; Preuss, H.; Stoll, H. *J. Phys. B* **1985**, *18*, 1287–1296. (b) Kaupp, M.; Schleyer, P. V. R.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1991**, *94*, 1360–1366. (c) Kuechle, W.; Dolg, M.; Stoll, H.; Preuss, H. *Mol. Phys.* **1991**, *74*, 1245–1263. (d) Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Mol. Phys.* **1993**, *78*, 1211–1224. (e) Metz, B.; Stoll, H.; Dolg, M. *J. Chem. Phys.* **2000**, *113*, 2563–2569.
- (6) (a) Hurley, M. M.; Fernandez Pacios, L.; Christiansen, P. A.; Ross, R. B.; Ermler, W. C. *J. Chem. Phys.* **1986**, *84*, 6840–6853. (b) Ross, R. B.; Powers, J. M.; Atashroo, T.; Ermler, W. C.; LaJohn, L. A.; Christiansen, P. A. *J. Chem. Phys.* **1990**, *93*, 6654–6670.
- (7) (a) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270–283. (b) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 284–298. (c) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299–310.

- (8) (a) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622. (b) Pople, J. A.; Seeger, R.; Krishnan, R. *Int. J. Quantum Chem. Symp.* **1977**, *11*, 149–163. (c) Szabo, A.; Ostlund, N. S. The Hartree–Fock approximation. In *Modern Quantum Chemistry*, 1st ed.; Dover Publications Inc.: Mineola, NY, 1996; pp 108–230. (d) Szabo, A.; Ostlund, N. S. Many-body perturbation theory. In *Modern Quantum Chemistry*, 1st ed.; Dover Publications Inc.: Mineola, NY, 1996; pp 320–379.
- (9) (a) Ehlers, A. W.; Dapprich, S.; Vydroshchikov, S. F.; Frenking, G. *Organometallics* **1996**, *15*, 105–117. (b) Borisov, Y. A.; Peregudov, A. S. *J. Struct. Chem.* **2003**, *44*, 567–573. (c) Shagun, V. A.; Shevchenko, S. G.; Frolov, Yu. L. *J. Struct. Chem.* **2003**, *44*, 736–740. (d) Fleischer, H.; Glang, S.; Schollmeyer, D.; Mitzel, N. W.; Buehl, M. *Dalton Trans.* **2004**, *21*, 3765–3771.
- (10) (a) Schwerdtfeger, P.; Bruce, A. E.; Bruce, M. R. M. *J. Am. Chem. Soc.* **1998**, *120*, 6587–6597. (b) Su, P.-H.; Lin, F.-W.; Yeh, C.-S. *J. Phys. Chem. A* **2001**, *105*, 9643–9648.
- (11) Kremer, T.; Junge, M.; Schleyer, P. V. *Organometallics* **1996**, *15*, 3345–3359.
- (12) Parr, R. G.; Yang, W. Density-functional theory. In *Density-Functional Theory of Atoms and Molecules*, 1st ed.; Breslow, R., Goodenough, J. B., Halpern, J., Rowlinson, J. S., Eds.; Oxford University Press: New York, 1989; International series of monographs on chemistry No. 16, pp 47–69.
- (13) (a) Haase, F.; Sauer, J. *J. Am. Chem. Soc.* **1998**, *120*, 13503–13512. (b) Cakmak, M.; Srivastava, G. P. *Surf. Sci.* **1999**, *435*, 420–424. (c) Phillips, M. A.; Besley, N. A.; Gill, P. M. W.; Moriarty, P. *Phys. Rev. B* **2003**, *67*, 035309.
- (14) (a) Devita, A.; Gillan, M. J.; Lin, J. S. *Phys. Rev. B* **1992**, *46*, 12964–12973. (b) Pesola, M.; von Boehm, J.; Sammal-korpi, V. *Phys. Rev. B* **1999**, *60*, R16267–R16270.
- (15) (a) Cosentino, U.; Moro, G.; Pitea, D.; Villa, A.; Fantucci, P. C.; Maiocchi, A.; Uggeri, F. *J. Phys. Chem. A* **1998**, *102*, 4606–4614. (b) Zheng K.; Wang J.; Shen Y.; Peng W.; Yun, F. *J. Comput. Chem.* **2002**, *23*, 436–443. (c) Kumondai, K.; Toyoda, M.; Ishihara, M.; Katakuse, I.; Takeuchi, T.; Ikeda, M.; Iwamoto, K. *J. Chem. Phys.* **2005**, *123*, 024314. (d) Jemmis, E. D.; Parameswaran, P.; Phukan, A. K. *Mol. Phys.* **2005**, *103*, 897–903.
- (16) (a) Merlic, C. A.; Miller, M. M.; Hietbrink, B. N.; Houk, K. N. *J. Am. Chem. Soc.* **2001**, *123*, 4904–4918. (b) Combariza, M. Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrosc.* **2004**, *15*, 1128–1135. (c) Danks, T. N.; Wagner, G. *Dalton Trans.* **2005**, 2933–2940.
- (17) (a) Kaupp, M.; Malkin, V. G.; Malkina, O. L.; Salahub, D. R. *J. Am. Chem. Soc.* **1995**, *117*, 1851–1852. (b) Kaupp, M.; Malkin, V. G.; Malkina, O. L.; Salahub, D. R. *Chem.-Eur. J.* **1996**, *2*, 24–30. (c) Cai, S.-H.; Chen, Z.; Wan, H.-L. *J. Phys. Chem. A* **2002**, *106*, 1060–1066. (d) Cai, S.-H.; Chen, Z.; Chen, Z.-W.; Wan, H.-L. *Chem. Phys. Lett.* **2002**, *362*, 13–18.
- (18) (a) Slater, J. C. The Self-Consistent Field for Molecular and Solids. In *Quantum Theory of Molecular and Solids*; McGraw-Hill: New York, 1974; Vol. 4. (b) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- (19) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (20) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652. (b) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- (21) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560–571.
- (22) (a) Nakajima, T.; Hirao, K. *J. Chem. Phys.* **2002**, *116*, 8270–8275. (b) The basis sets employed was provided by Dr. Nakajima in a personal communication. (c) Personal communication with Dr. Nakajima.
- (23) (a) Hess, B. A. *Phys. Rev. A* **1986**, *33*, 3742–3748. (b) Jansen, G.; Hess, B. A. *Phys. Rev. A* **1989**, *39*, 6016–6017.
- (24) Boys, S. B.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
- (25) (a) Aprà, E.; Windus, T. L.; Straatsma, T. P.; Bylaska, E. J.; de Jong, W.; Hirata, S.; Valiev, M.; Hackler, M.; Pollack, L.; Kowalski, K.; Harrison, R.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju V.; Krishnan, M.; Auer, A. A.; Brown, E.; Cisneros, G.; Fann, G.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. *NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.5*; Pacific Northwest National Laboratory: Richland, WA 99352-0999, U.S.A., 2003. (b) Kendall, R. A.; Aprà, E.; Bernholdt, D. E.; Bylaska, E. J.; Dupuis, M.; Fann, G. I.; Harrison, R. J.; Ju, J.; Nichols, J. A.; Nieplocha, J.; Straatsma, T. P.; Windus, T. L.; Wong, A. T. *Comput. Phys. Commun.* **2000**, *128*, 260–283.
- (26) (a) Radom, L.; Lathan, W. A.; Hehre, W. J.; Pople, J. A. *J. Am. Chem. Soc.* **1973**, *95*, 693–698. (b) Chang, Y.-P.; Su, T.-M.; Li, T.-W.; Chao, I. *J. Phys. Chem. A* **1997**, *101*, 6107–6117.
- (27) Vargas, R.; Garza, J.; Dixon, D. A.; Hay, B. P. *J. Am. Chem. Soc.* **2000**, *122*, 4750–4755.
- (28) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1994**, *116*, 3892–3900.
- (29) Vorobyov, I.; Yappert, M. C.; DuPré, D. B. *J. Phys. Chem. A* **2002**, *106*, 668–679.
- (30) Bultinck, P.; Goeminne, A.; Van de Vondel, D. *J. Mol. Struct. (THEOCHEM)* **1995**, *339*, 1–14.
- (31) Kazerouni, M. R.; Hedberg, L.; Hedberg, K. *J. Am. Chem. Soc.* **1997**, *119*, 8324–8331.
- (32) (a) Yokozeki, A.; Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2926. (b) Kazerouni, M. R.; Hedberg, L.; Hedberg, K. *J. Am. Chem. Soc.* **1994**, *116*, 5279–5284.
- (33) (a) Penn, R. E.; Curl, R. F., Jr. *J. Chem. Phys.* **1971**, *55*, 651–658. (b) Penn, R. E.; Olsen, R. J. *J. Mol. Spectrosc.* **1976**, *62*, 423–428.
- (34) Huheey, J. E.; Keiter, E. A.; Keiter, R. L. The structure of the atom. In *Inorg. Chem.: Principles of Structure and Reactivity*, 4th ed.; Piro, J., Ed.; Harper Collins: New York, 1990; pp 10–45.
- (35) Waters, A. F.; White, A. H. *Aust. J. Chem.* **1996**, *49*, 61–72.
- (36) Duplancic, T.; Grdenic, D.; Kamenar, B.; Matkovic, P.; Sikirica, M. *J. Chem. Soc., Dalton* **1976**, *10*, 887–890.
- (37) Kaupp, M.; Schleyer, P. R. *J. Am. Chem. Soc.* **1993**, *115*, 1061–1073.

# JCTC Journal of Chemical Theory and Computation

## Macroscopic $pK_a$ Calculations for Fluorescein and Its Derivatives

Marcin Król,<sup>\*,†,‡</sup> Marta Wrona,<sup>§</sup> Christopher S. Page,<sup>†</sup> and Paul A. Bates<sup>\*,†</sup>

*Biomolecular Modelling Laboratory, Cancer Research UK London Research Institute  
Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK,  
Department of Bioinformatics and Telemedicine, Collegium Medicum,  
Jagiellonian University, Kopernika 17, Kraków 31-501, Poland, and  
Grey Cancer Institute, Free Radicals Research Group, Mount Vernon Hospital,  
P.O. Box 100, Northwood, Middlesex HA6 2JR, UK*

Received July 18, 2006

**Abstract:** This study describes the calculation of the microscopic dissociation and tautomerization constants of fluorescein and its derivatives, 2',7'-dichlorofluorescein (DCF) and 2',7'-difluorofluorescein (DFF), in an aqueous environment. In vacuo free energies were obtained using complete basis set (CBS) and DFT-based methods, while free energies of solvation were calculated with the CPCM implicit solvation protocol using the UAHF, UAKS, and Pauling radii sets. Our results indicate that the different vacuum protocols give free energy changes upon dissociation within 1 kcal/mol of each other for a given molecule. Therefore, we suggest that the computationally less intensive PBE1PBE/6-311+G(2d,2p)//PBE1PBE/6-31+G(d) model chemistry may reasonably be used in  $pK_a$  calculations of larger molecules. The calculations also provided a rigorous test of the implicit solvation models. Relative calculations of dissociation constants gave results in good agreement with experiment; absolute values deviated from experimental data by 1–3  $pK_a$  units. Consistently better results were obtained with the Pauling radii set. The influence of geometry relaxation on going from vacuum to solvent is negligible for  $pK_{a2}$  and larger for  $pK_{a1}$  but still smaller than the variation due to the radii set. Calculation of tautomerization constants gave more variable results, with none of the solvation methods able to reproduce experimental values consistently, although certain individual constants were correctly calculated.

### Introduction

Proton-transfer reactions in an aqueous environment are of great importance in many areas of chemistry and biology,<sup>1</sup> and a knowledge of the dissociation constant,  $K_a$ —more

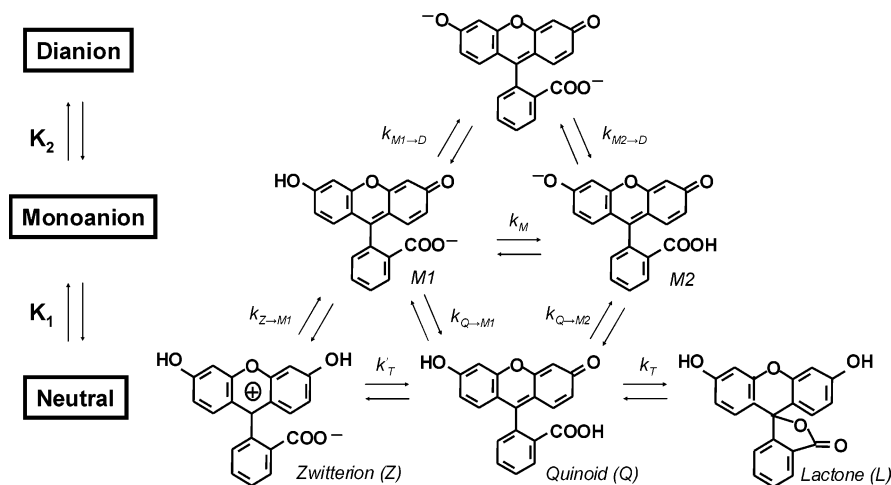
frequently expressed in terms of  $pK_a$ —allows for the prediction of the protonation states of a molecule of interest at any given pH value. Consequently, the theoretical prediction of  $pK_a$  values using quantum chemical methods has attracted a great deal of interest.<sup>2–20</sup> The various approaches used previously differ in their description of the thermodynamic cycle, the level of molecular orbital theory used for obtaining the energies in a vacuum, and in the methods used to treat solvation effects. The choice of the thermodynamic cycle will not be addressed here, as it has recently been discussed in detail elsewhere.<sup>3,21</sup> However, the level of theory used for vacuum and solvation calculations not only greatly influ-

\* Corresponding author phone: +4420 7269 3023; fax: +4420 7269 3258; e-mail: mykrol@cyf-kr.edu.pl (M.K.) or phone: +4420 7269 3223; fax: +4420 7269 3258; e-mail: paul.bates@cancer.org.uk (P.A.B.).

† Cancer Research U.K. London Research Institute Lincoln's Inn Fields Laboratories.

‡ Jagiellonian University.

§ Mount Vernon Hospital.



**Figure 1.** Prototropic equilibria for fluorescein. Tautomerization ( $k_T$ ,  $k_T$ ,  $k_M$ ) and macroscopic ( $K_1$  and  $K_2$ ) and microscopic ( $k_{Z \rightarrow M1}$ ,  $k_{Q \rightarrow M1}$ ,  $k_{Q \rightarrow M2}$ ,  $k_{M1 \rightarrow D}$ ,  $k_{M2 \rightarrow D}$ ) dissociation constants are shown.

ences the accuracy of the results but also the computational tractability and must, therefore, be optimized for a given set of molecules of interest.

It has been reported that pK<sub>a</sub> calculations are approaching chemical accuracy for certain classes of organic molecules.<sup>5,22</sup> However, previous studies have tended to focus on predicting dissociation constants for small compounds containing only one ionizable group. While efforts are underway to extend the methodology for larger solutes,<sup>23</sup> the calculation of solvation free energies of doubly charged systems presents significant computational challenges, as indicated by a recent study of acetate and bicarbonate dianions.<sup>24</sup> An assessment of how methods developed for pK<sub>a</sub> calculations of small molecules can deal with larger organic species existing in several tautomeric forms and containing multiple ionizable groups is, therefore, of great interest.

Fluorescein and its derivatives, 2',7'-dichlorofluorescein (DCF) and 2',7'-difluorofluorescein (DFP, Oregon Green 488), are the subjects of the present study. These molecules have been extensively analyzed in our laboratory<sup>25–27</sup> and are well-known fluorescent probes that have seen widespread use in biological and biochemical applications. In particular, they have been used in fluorescent protein labeling and imaging,<sup>28,29</sup> as Ca<sup>2+</sup> and pH indicators,<sup>30</sup> and in the measurement of oxidative stress.<sup>25,31,32</sup> Despite the large quantities of experimental data available, there is still a need for more detailed information about the basic chemical properties of these fluorescein-derived probes. The ability to calculate prototropic properties of these dyes is of use both in understanding the probe chemistry and in developing new probes.

Fluorescein, DCF, and DFP can exist in aqueous solution in a number of prototropic forms: cationic, neutral, monoanionic, and dianionic. It follows that three ground-state macroscopic pK<sub>a</sub>'s are involved. Furthermore, there are three different tautomers for the neutral species—a quinoid, a zwitterion, and a lactonic form—and a further two for the monoanion, with ionized carboxyl or hydroxyl groups.<sup>33,34</sup> Equilibria between the different tautomers and ionization states are described by tautomerization and microscopic dissociation constants. Possible anionic and neutral forms

of fluorescein together with microscopic ( $k$ ) and macroscopic ( $K_1$  and  $K_2$ ) equilibrium constants are shown in Figure 1. It should be noted that microscopic dissociation constants are readily obtained from theoretical calculations; however, they cannot be directly compared to the experimentally derived macroscopic constants, unless tautomerization constants are also calculated.

We established a three-species model to calculate the pK<sub>a</sub> values of the neutral-monoanion and monoanion-dianion transitions, where the neutral form was represented by a quinoid (Q) and the monoanion by the M<sub>1</sub> tautomer (see the Methods section). However, the dissociation constant of the cation was not calculated, since its pK<sub>a</sub> is very low, and so is not in the biological pH range.<sup>35</sup>

Conventionally, complete basis set (CBS) model chemistries<sup>36,37</sup> are the methods of choice to obtain vacuum energies, with CBS-QB3<sup>38</sup> being the reference method used in much of the recent work on pK<sub>a</sub> calculations. This protocol yields very good agreement with experimental data. However, it is computationally very demanding, both in terms of CPU time and disk storage, and is not practical for calculating the vacuum free energies of the systems studied here; instead, a modified version of the faster CBS-4M<sup>39</sup> was used. A new, state-of-the-art DFT-based protocol for in vacuo energy calculations has recently been proposed,<sup>40</sup> which produces results with an accuracy comparable to the CBS methods but without the computational overhead. In this paper we will compare both CBS and DFT methods for calculating energies of the molecules in a vacuum.

In principle, solvation effects may be evaluated using either continuum<sup>41,42</sup> or explicit solvation models<sup>43–45</sup> or a combination of the two (cluster-continuum methods).<sup>18,46–49</sup> It has been argued that for the proper description of the hydration inclusion of explicit water molecules around the hydrophilic groups of the solute molecule is mandatory.<sup>13</sup> Indeed, several studies have shown that addition of a few explicit water molecules substantially improved the agreement between calculated and experimental solvation energies and pK<sub>a</sub> values.<sup>48–50</sup> However, there is still debate on how many water molecules should be added to the solute<sup>18</sup> although Pliego

and Riveros have proposed a variational principle for finding the number of explicit water molecules, such that the optimum is the number of molecules which minimize the solvation free energy.<sup>47</sup> Furthermore, Kelly and co-workers showed that adding one explicit water molecule significantly improves the accuracy of solvation free energies calculated with their new SM6 implicit solvent model<sup>50</sup> but pointed out that it did not work for other models. Given the above and the fact that the molecules studied in this work are rather large and contain multiple ionizable sites, adding explicit water molecules is nontrivial. Therefore, we assess the performance of the more widely used pure implicit solvent model for the calculation of solvation effects on systems of this type, bearing in mind the known shortcomings of the description.

Many different continuum models have been developed (for a recent review see ref 42), and their accuracy depends in part upon the way in which the cavity containing the solute is generated. Since the values of the atomic radii used to generate the solute cavity have been optimized for different levels of quantum mechanical theory, it is important to choose radii corresponding to the method used. Furthermore, Takano and Houk have recently shown that some radii sets are better suited to the calculation of solvation free energies of charged species than others.<sup>51</sup> Consequently, in this work we examine the influence of several radii sets on the results obtained.

Solvation effects should, in principle, include both the solvation free energy change (defined as the energy change upon moving the molecule from vacuum to the solution phase with a fixed geometry)<sup>52</sup> and the relaxation energy (the difference in energy between the molecule optimized in a vacuum and optimized in solvent).<sup>53</sup> However, most studies often include only the former term, calculated using a vacuum-optimal or solvent-optimal geometry, and neglect the relaxation effect. While this neglect may be justified for small molecules, it may not be true for the molecules tested in this work. Therefore, we discuss the effect of this approximation on the quality of the calculated dissociation constants.

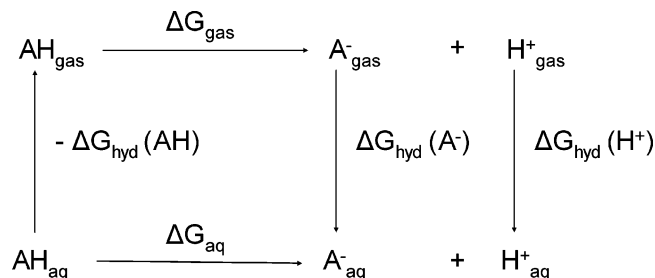
The aim of the present work is to assess the performance of implicit solvent models in the calculation of  $pK_a$  values for fluorescein and its derivatives. The protocol should be able to reproduce experimental results to within chemical accuracy but should also be computationally feasible for such large molecules. To this end, we compare results of  $pK_a$  calculations for fluorescein and its derivatives using CBS-4M and DFT-based methods for vacuum free energy estimation. Free energies of solvation are estimated by CPCM solvation methods with different radii sets.

## Methods

**Computational Design.**  $pK_a$  can be related to the free energy change  $\Delta G_{aq}$  of the dissociation reaction using the formula

$$pK_a = \frac{\Delta G_{aq}}{RT \ln 10} \quad (1)$$

We have used the following thermodynamic cycle to calculate  $\Delta G_{aq}$ :



With this cycle

$$\Delta G_{aq} = -\Delta G_{hyd}(AH) + \Delta G_{gas} + \Delta G_{hyd}(A^-) + \Delta G_{hyd}(H^+) \quad (2)$$

and

$$pK_a = \frac{G_{gas}(A^-) - G_{gas}(AH) + \Delta G_{hyd}(A^-) - \Delta G_{hyd}(AH) + G_{gas}(H^+) + \Delta G_{hyd}(H^+)}{RT \ln 10} \quad (3)$$

Here  $G_{gas}(H^+) = -6.28$  kcal/mol from the Sackur-Tetrode equation and  $\Delta G_{hyd}(H^+) = -264.0$  kcal/mol.<sup>54,55</sup> The value of  $\Delta G_{hyd}(H^+)$  was obtained for the 1 atm gas phase/1 M solution standard state and must be converted to 1 M gas phase/1 M solution standard state by adding  $-1.9$  kcal/mol.<sup>56</sup> Note that other values for  $\Delta G_{hyd}(H^+)$  have also been suggested. For example, values based on the absolute potential correction for the hydrogen standard electrode range from  $-254$  to  $-261$  kcal/mol.<sup>57</sup> Liptak and Shields used the value of  $-264.61$  kcal/mol derived from the thermodynamic cycle of acetic acid.<sup>4</sup> However, recently the value of  $-264.0$  kcal/mol has been widely accepted as the most reliable estimate of the solvation energy of the proton.<sup>56,58</sup>

When calculating  $pK_a$  values of molecules with multiple tautomeric forms, one should take into account all tautomers for a given ionization state. Therefore, both tautomerization and microscopic dissociation constants must be used to calculate macroscopic dissociation constants. From Figure 1 we can define

$$K_1 = \frac{[M][H^+]}{[N]} \quad (4)$$

where  $[M]$  is the concentration of all singly ionized tautomers and  $[N]$  is the concentration of all neutral tautomers. However,

$$[M] = [M_1] + [M_2] \quad (5)$$

$$[N] = [Z] + [Q] + [L] \quad (6)$$

where  $M_1$ ,  $M_2$ ,  $Z$ ,  $Q$ , and  $L$  are concentrations of the different tautomers. From Figure 1 we readily obtain expressions for  $k'_T$ ,  $k_T$ ,  $k_M$ , and  $k_{Q \rightarrow M_1}$ :

$$k'_T = \frac{[Q]}{[Z]} \quad (7)$$

$$k_T = \frac{[L]}{[Q]} \quad (8)$$

$$k_M = \frac{[M_2]}{[M_1]} \quad (9)$$

$$k_{Q \rightarrow M1} = \frac{[M_1][H^+]}{[Q]} \quad (10)$$

Combining all equations, we can derive an expression for  $K_1$  in terms of a microscopic constant  $k_{Q \rightarrow M1}$  and tautomerization constants  $k'_T$ ,  $k_T$ , and  $k_M$

$$K_1 = \frac{k_{Q \rightarrow M1} \times k'_T \times (1 + k_M)}{1 + k'_T + (k_T \times k'_T)} \quad (11)$$

and

$$pK_1 = pk_{Q \rightarrow M1} - \log \frac{k'_T \times (1 + k_M)}{1 + k'_T + (k_T \times k'_T)} \quad (12)$$

Similarly,

$$pK_2 = pk_{M1 \rightarrow D} + \log(1 + k_M) \quad (13)$$

However, since experimental data for fluorescein and its derivatives show that  $[M_2] \sim 0$ ,<sup>33,59</sup> eqs 12 and 13 may be simplified to

$$pK_1 = pk_{Q \rightarrow M1} - \log \frac{k'_T}{1 + k'_T + (k_T \times k'_T)} \quad (14)$$

$$pK_2 = pk_{M1 \rightarrow D} \quad (15)$$

To avoid errors in the absolute  $pK_a$  determination we have calculated microscopic  $pK_a$ s for DCF and DFF relative to the experimental  $pK_a$  values for fluorescein. By calculating the difference in  $pK_a$  values obtained from eq 3 for two molecules—AH and BH—we obtain the  $pK_a$  of AH relative to the  $pK_a$  of BH:

$$pK_a(\text{AH}) = \frac{1}{RT \ln 10} \times [G_{\text{gas}}(\text{A}^-) - G_{\text{gas}}(\text{B}^-) + G_{\text{gas}}(\text{BH}) - G_{\text{gas}}(\text{AH}) + \Delta G_{\text{hyd}}(\text{A}^-) - \Delta G_{\text{hyd}}(\text{B}^-) + \Delta G_{\text{hyd}}(\text{BH}) - \Delta G_{\text{hyd}}(\text{AH})] + pK_a(\text{BH}) \quad (16)$$

In this work AH represented DCF and DFF, and BH represented fluorescein.

The tautomerization constant  $k_T$  can be calculated from a free energy change  $\Delta G_{\text{aq}}^T$  of a tautomerization process in solution using

$$k_T = \exp\left(\frac{-\Delta G_{\text{aq}}^T}{RT}\right) \quad (17)$$

where  $\Delta G_{\text{aq}}^T$  is the difference in the total free energy in solution between two tautomers.

**Gas-Phase Free Energies and Structures.** A well-established protocol for calculating free energies in a vacuum is CBS-QB3. However, this proved excessively expensive computationally for the molecules under consideration. Alternative methods were therefore sought, which, on one hand, would be computationally less demanding, and, on the

other hand, would still retain the high quality of CBS-QB3. Consequently, we have examined four methods to calculate gas-phase free energies, which are defined below:

V1 – CBS-4M (geometry optimized at HF/6-31+G(d))

V2 – PBE1PBE<sup>60</sup>/aug-cc-pVTZ // PBE1PBE/6-31+G(d)

V3 – PBE1PBE/6-311+G(2d,2p) // PBE1PBE/6-31+G(d)

V4 – PBE1PBE/6-311+G(2d,2p) //CPCM/ HF/6-31+G(d)

In the CBS-4M model chemistry geometry optimization is usually performed at the HF/3-21G level; however, for the molecules studied here, it led to the monoanion accepting a lactonic form, which is not observed experimentally.<sup>61</sup> Since geometries optimized at the HF/6-31+G(d) level maintained a correct, quinoid form, we used this level of theory for CBS-4M energy evaluations.

In the V2 and V3 methods, geometry optimization of a DCF monoanion led to a lactone. In this case, a solvent-optimized geometry, which held the correct quinoid form, was used to perform SP calculations in a vacuum at the PBE1PBE/6-311+G(2d,2p) level. For consistency, we have also performed identical calculations (utilizing solvent-optimized geometries) for fluorescein and DFF and have termed this approach V4. In fact, in this approach all calculations were performed on the solvent geometry; the relaxation effect in going from vacuum to solvent has been neglected.

In all methods, correction for the treatment of internal rotors as harmonic oscillators has been neglected. It should be noted that in principle this might be the source of significant errors if absolute free energy values are of interest.<sup>62</sup> However, the errors would be expected to cancel when calculating relative differences in free energy values for molecules with almost identical internal rotations.

All calculations were performed using the *Gaussian03* software.<sup>63</sup>

**Solvation Free Energies and Structures.** Free energies of solvation in water were calculated using the CPCM implicit solvent method<sup>64</sup> as implemented in *Gaussian03* with three of the standard radii sets (UAHF,<sup>65</sup> UAKS, and Pauling<sup>66</sup>) using the HF/6-31+G(d) model chemistry, which led to three different solvation energies for each molecule and its corresponding protonation state:

S1 – CPCM(UAHF)/HF/6-31+G(d)

S2 – CPCM(UAKS)/HF/6-31+G(d)

S3 – CPCM(Pauling)/HF/6-31+G(d)

Although the Pauling radii set was originally parametrized for neutral molecules, Takano and Houk<sup>51</sup> have shown that they give the lowest mean absolute deviations (MADs) of solvation energies for all tested radii sets when applied to anions, and so it was of interest to examine whether this extends to dianions. They also showed that the UA0 and UFF radii yield significantly inferior results and so have not been considered here. Furthermore, the UAHF and UAKS radii sets, although closely related, perform quite differently,

especially for anions. Both of these sets are, therefore, included in our study.

The geometries of all molecules were optimized at the HF/6-31+G(d) level and CPCM/HF/6-31+G(d) level using UAHF radii. Relaxation energies were calculated as the difference in SCF energy between solvent- and vacuum-optimized geometries. These were added to the solvation energies to give the total contribution to the energy change upon moving the molecule from a vacuum to a water environment. Note that the relaxation energy can be calculated in a vacuum (as the difference between in vacuo SCF energy for solvent- and vacuum optimized geometries) or in solvent (as the difference between CPCM SCF energy for solvent- and vacuum optimized geometries). Relaxation energies were included in the dissociation constant calculations with methods V1, V2, and V3 but were not used for V4 as this method uses the geometry optimized in water both for calculation of  $\Delta G_{\text{gas}}$  and  $\Delta G_{\text{hydr}}$ .

To investigate the dependence of the optimal geometry in solvent on the radii set, we have performed additional CPCM/HF/6-31+G(d) optimizations using UAKS and Pauling radii on fluorescein. Subsequently, we compared solvation energies calculated for these geometries with energies obtained from single point (SP) calculations using UAKS or Pauling radii on geometries obtained with UAHF radii. Although the absolute values differ by up to 3 kcal/mol, the solvation free energy difference between the pair of molecules (e.g. the neutral species and its anion) is always below 1 kcal/mol (see Supporting Information Table S1). Therefore, the error introduced by not reoptimizing the geometry with different radii sets would be negligible. Consequently, for DCF and DFF, only geometries optimized with UAHF radii are considered.

**Tautomerization Constants.** To calculate tautomerization constants, the geometries of the neutral zwitterion, quinoid, and lactone forms of fluorescein and its derivatives were optimized at the CPCM/HF/6-31+G(d) level using UAHF radii. Frequency calculations were performed at the same level of theory. The free energy of solvation was calculated using UAHF and Pauling radii sets at the CPCM/HF/6-31+G(d) level, and no relaxation energy was included.  $G_{\text{aq}}^{\text{T}}$  was calculated as the sum of the SCF energy at the PBE1PBE/6-311+G(2d,2p) level, thermal correction to the Gibbs free energy from the CPCM frequency calculations and solvation free energy.

## Results and Discussion

**Relative  $pK_{\text{a}}$  Values.** As described in detail in the Methods section, the accuracy of the calculated macroscopic  $pK_{\text{a}}$  values depends on two factors: the correct calculation of both the microscopic dissociation constants and tautomerization constants. To dissect the two contributions we have compared calculated microscopic  $pK_{\text{Q-M1}}$  values for fluorescein and its derivatives with the value of  $pK_{\text{Q-M1}}$  obtained from  $pK_{\text{a1}}$  and tautomerization constants measured experimentally.  $pK_{\text{M1-D}}$  is compared directly to  $pK_{\text{a2}}$  since  $k_{\text{M}} \sim 0$ . The comparison of calculated tautomerization constants with experimental values is discussed in a later section.

**Table 1.** Absolute Values of  $pK_{\text{Q-M1}}$  and  $pK_{\text{M1-D}}$  for DCF Calculated Relative to Fluorescein<sup>a</sup>

	$pK_{\text{Q-M1}}$		$pK_{\text{M1-D}}$	
	V1	V4	V1	V4
S1	3.53	4.11	3.18	2.14
S2	3.52	4.09	3.18	2.14
S3	3.80	4.38	4.88	3.84
exp.	3.50		5.19	

<sup>a</sup> Experimental values were obtained from  $pK_{\text{a1}} = 4.00$ ,  $pK_{\text{a2}} = 5.19$ ,  $k_{\text{T}} = 2.00$ ,  $K_{\text{T}} = 7.14$ , and  $k_{\text{M}} = 0$ .<sup>33</sup>

**Table 2.** Absolute Values of  $pK_{\text{Q-M1}}$  and  $pK_{\text{M1-D}}$  for DFF Calculated Relative to Fluorescein<sup>a</sup>

	$pK_{\text{Q-M1}}$				$pK_{\text{M1-D}}$			
	V1	V2	V3	V4	V1	V2	V3	V4
S1	3.37	3.36	3.28	4.49	3.55	3.71	3.75	2.74
S2	3.39	3.38	3.30	4.49	3.55	3.71	3.75	2.74
S3	3.47	3.46	3.37	4.59	4.63	4.80	4.84	3.82
exp.	3.37				4.69			

<sup>a</sup> Experimental values were obtained from  $pK_{\text{a1}} = 3.61$  and  $pK_{\text{a2}} = 4.69$ .<sup>67</sup>  $k_{\text{T}} = 0.68$ ,  $K_{\text{T}} = 16.94$  (calculated from the ratio of different tautomers<sup>67</sup>), and  $k_{\text{M}} = 0$ .

Tables 1 and 2 show values of  $pK_{\text{Q-M}}$  and  $pK_{\text{a2}}$  obtained from eq 16. Note that by using eq 16 hydration and vacuum free energies of a proton are not needed, and differences in corresponding free energies, employed in the equation, are likely to lead to the cancellation of errors in solvation and gas-phase free energies.

Analysis of the data shown in Tables 1 and 2 reveals that there is a perfect agreement between the calculated and experimental values for the  $pK_{\text{Q-M1}}$  for all three radii sets and vacuum methods V1–V3. In the case of  $pK_{\text{M1-D}}$  the quality of the results depends on the radii set used. Results obtained with the united atom (UA) radii deviate from experimental data by  $\sim 1$  (for DFF) and  $\sim 2$  (for DCF)  $pK_{\text{a}}$  units, while the Pauling all-atom radii set reproduces the values correctly. This is surprising since Pauling radii were not optimized for solvation energies of anionic species and should give inferior results compared to the UAHF and UAKS. On the other hand, as Takano and Houk observed, Pauling radii gave the lowest MADs for anionic species from all the radii sets tested. The calculation of  $pK_{\text{M1-D}}$  involves estimation of solvation energies for singly and doubly charged anions, and Pauling radii, according to Takano and Houk's data, are likely to give better estimates of these energies.

It is interesting to note that the protocol V4, which involved calculating vacuum free energies on geometries optimized in water, does not yield correct predictions, with the overestimation of the first dissociation constants and underestimation of the second. This is most likely caused by the incorrect description of the proton dissociation in a vacuum, if the solvent-optimized geometries are used for this process, and by the neglect of the relaxation energy.

**Absolute  $pK_{\text{a}}$  Values.** Given the encouraging results obtained in the relative  $pK_{\text{a}}$  calculations, we further estimate microscopic dissociation constants for all molecules using eq 3. This is much more demanding because the absolute values of solvation and gas-phase free energies must be



**Table 3.** Absolute Values of pK<sub>Q→M1</sub> and pK<sub>M1→D</sub> for Fluorescein<sup>a</sup>

	pK <sub>Q→M1</sub>				pK <sub>M1→D</sub> (exp. 6.80)			
	V1	V2	V3	V4	V1	V2	V3	V4
S1	5.08	5.43	5.17	5.94	10.65	10.09	9.62	10.01
S2	5.44	5.80	5.54	6.33	10.65	10.09	9.62	10.01
S3	2.12	2.46	2.21	2.98	8.40	7.85	7.37	7.76
exp.	3.49				6.80			

<sup>a</sup> Experimental values were obtained from pK<sub>a1</sub> = 4.45, pK<sub>a2</sub> = 6.80, K<sub>T</sub> = 6.09, K<sub>T</sub> = 0.5, and K<sub>M</sub> = 0.<sup>33</sup>

**Table 4.** Absolute Values of pK<sub>Q→M1</sub> and pK<sub>M1→D</sub> for DCF<sup>a</sup>

	pK <sub>Q→M1</sub>		pK <sub>M1→D</sub>	
	V1	V4	V1	V4
S1	5.11	6.56	7.16	5.47
S2	5.49	6.94	7.16	5.47
S3	2.42	3.87	6.61	4.92
exp.	3.50		5.19	

<sup>a</sup> Experimental values were obtained from pK<sub>a1</sub> = 4.00, pK<sub>a2</sub> = 5.19, K<sub>T</sub> = 2.00, K<sub>T</sub> = 7.14, and K<sub>M</sub> = 0.<sup>33</sup>

accurately calculated, and the correct value of the proton solvation free energy must be obtained from experimental data. The latter value is critical for the accuracy of absolute pK<sub>a</sub> estimates. As discussed briefly in the Methods section, several different values of the proton solvation energy were used in the past. However, currently, a value of -269 kcal/mol is accepted as the most reliable estimate of the proton solvation energy.<sup>56,58</sup>

Tables 3–5 show the results of the pK<sub>Q→M1</sub> and pK<sub>M1→D</sub> calculations, together with the corresponding experimental data, for fluorescein, DCF, and DFF, respectively. Solvation and vacuum free energies calculated with the methods S1–S3 and V1–V4 are given in the Supporting Information Tables S2–S4 for fluorescein, DCF, and DFF, respectively. For DCF results obtained from the V1 and V4 methods only are shown.

For a given solvation method the differences in the pK<sub>a</sub> values produced by the V1–V3 vacuum methods are all below 0.5 pK<sub>a</sub> units for the first dissociation constant and below one pK<sub>a</sub> unit for the second dissociation constant. All three methods gave convergent results; therefore, the computationally costly methods V1 and V2 may be successfully substituted by the substantially less demanding V3 method. This was further confirmed by the test calculations performed on 3-chloro-4-hydroxyphenol molecule, which showed that methods V1–V3 yield very similar results for Δ(G(A<sup>-</sup><sub>gas</sub>) – G(AH<sub>gas</sub>)) compared to CBS-QB3 but with a fraction of the computational cost (see Supporting Information Table S5).

As with the relative pK<sub>a</sub> calculations, the V4 method does not produce satisfactory results and should be avoided. Good performance of the method in a few cases (such as for the Pauling radii for fluorescein and pK<sub>M1→D</sub> for DCF and DFF) is most probably due to a fortuitous cancellation of errors.

pK<sub>a</sub> values presented in Tables 3–5 for the first dissociation constant differ by ~2 pK<sub>a</sub> units for UA radii sets and by ~1 pK<sub>a</sub> unit for Pauling radii. Mean deviations (MDs), given in Table 6, show that the deviation from experimental values for pK<sub>Q→M1</sub> is very similar for all three molecules and indicate

**Table 5.** Absolute Values of pK<sub>Q→M1</sub> and pK<sub>M1→D</sub> for DFF<sup>a</sup>

	pK <sub>Q→M1</sub> (exp. 3.37)				pK <sub>M1→D</sub> (exp. 4.69)			
	V1	V2	V3	V4	V1	V2	V3	V4
S1	4.95	5.30	4.96	6.94	7.52	7.12	6.69	6.06
S2	5.35	5.69	5.35	7.34	7.52	7.12	6.69	6.06
S3	2.09	2.43	2.09	4.08	6.37	5.97	5.53	4.90
exp.	3.37				4.69			

<sup>a</sup> Experimental values were obtained from pK<sub>a1</sub> = 3.61 and pK<sub>a2</sub> = 4.69.<sup>67</sup> K<sub>T</sub> = 0.68 and K<sub>T</sub> = 16.94 were calculated from the ratio of different tautomers,<sup>67</sup> and K<sub>M</sub> = 0.

**Table 6.** Mean Deviations (MDs) in pK<sub>a</sub> Units of the pK<sub>Q→M1</sub> and pK<sub>M1→D</sub> with Respect to Experimental Values for Different Solvation Methods and Molecules<sup>a</sup>

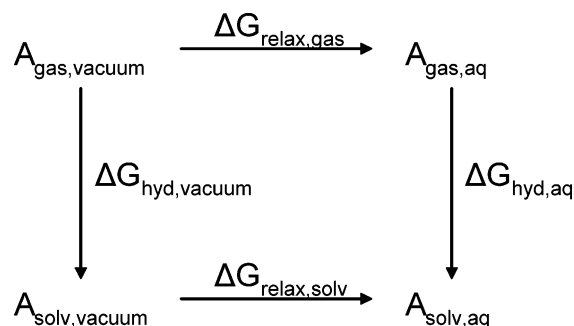
	fluorescein		DCF		DFF	
	pK <sub>Q→M1</sub>	pK <sub>M1→D</sub>	pK <sub>Q→M1</sub>	pK <sub>M1→D</sub>	pK <sub>Q→M1</sub>	pK <sub>M1→D</sub>
S1	1.73	3.32	1.61	1.97	1.70	2.43
S2	2.11	3.32	1.99	1.97	2.09	2.43
S3	-1.23	1.07	-1.08	1.42	-1.17	1.26

<sup>a</sup> Methods S1–S3 employ the UAHF, UAKS, and Pauling radii sets, respectively. MDs for fluorescein and DFF include pK<sub>a</sub> values calculated for the V1–V3 vacuum methods; MDs for DCF are for pK<sub>a</sub> values calculated for V1 only.

that there is a systematic error in the calculation of either absolute solvation energies or gas-phase acidities. Indeed, this is the reason why relative pK<sub>Q→M1</sub> values were in such good agreement with experiment. However, as different vacuum protocols yield very similar pK<sub>a</sub> values, it must be the solvation energies which deviate from the correct ones and incur the error. This was to be expected, as the correct calculation of solvation free energies for anions would require adding explicit solvent molecules. Bearing this in mind, implicit solvent models are still capable of obtaining pK<sub>Q→M1</sub> values which are within 1 pK<sub>a</sub> unit of experimental values for fluorescein-like molecules, if the Pauling radii set is used. It should be noted that UA radii sets cause overestimation of the pK<sub>Q→M1</sub> values, while the Pauling radii set underestimates them.

In the case of the second dissociation constant, the MDs for the values obtained with the UA radii sets differ among test molecules, which indicates that the implicit solvent model is unable to obtain the correct values of solvation energies but, more importantly, is not able to treat the molecules in a consistent way. The difference between calculated and experimental values of pK<sub>M1→D</sub> differ from ~3 pK<sub>a</sub> units for fluorescein to ~2 pK<sub>a</sub> units for DCF. On the other hand, if the Pauling radii set is used, the MDs are similar for all three molecules with an average value of ~1.2 pK<sub>a</sub> units. Again, the fact that the Pauling radii perform better, both in terms of MDs and the ability to sustain a similar level of error, is somewhat surprising given the fact that the Pauling radii set was not optimized for the ionized molecules. On the other hand, UA radii, although optimized on ions, were shown to perform well only if used for the molecules present in the training set.<sup>48</sup>

It should be noted that we have taken reference experimental data from papers by Mchedlov-Petrosyan<sup>33,59</sup> and co-workers (for fluorescein and DCF) and by Orte et al.<sup>67</sup> (for DFF), but other values can be found in the literature.



**Figure 2.** Two ways of solvating a molecule A. The first suffix (gas, solv) describes the environment of the molecule; the second suffix describes the geometry of the molecule (vacuum, aq), e.g.  $A_{\text{gas,aq}}$  is the molecule in a solvent-optimized geometry being placed in a vacuum.  $\Delta G_{\text{hyd}}$  is the change in free energy upon solvating the molecule, and  $\Delta G_{\text{relax}}$  is the change in free energy upon changing the geometry. Note that generally  $\Delta G_{\text{hyd,vacuum}} \neq \Delta G_{\text{hyd,aq}}$  and hence  $\Delta G_{\text{relax,gas}} \neq \Delta G_{\text{relax,solv}}$ .

For example, different authors measured  $\text{p}K_{\text{a}1}$  of fluorescein in the range 4.23–4.39 and  $\text{p}K_{\text{a}2} = 6.31$ –6.72.<sup>34,35,68</sup> Klonis and Sawyer also obtained slightly different tautomerization constants ( $k_{\text{T}} = 4.3$ –6.6,  $k'_{\text{T}} = 0.7$ –1.1 depending on the dissociation reaction model). For DCF, Leonhardt et al. established  $\text{p}K_{\text{a}1} = 3.5$  and  $\text{p}K_{\text{a}2} = 4.95$ .<sup>69</sup> Although there is still some disparity within the experimental data, it is much smaller than the error associated with the calculations of dissociation and tautomerization constants, and its impact on the performance of the methods presented in this work is, therefore, minor.

**The Influence of Relaxation Energy on the Calculated Dissociation Constants.** There is an ongoing debate in the literature on the use of vacuum- or solvent-optimized geometries for the calculation of solvation free energies used in the calculations of dissociation constants. Some investigators have shown that geometry optimization in an aqueous environment is crucial for the correct prediction of  $\text{p}K_{\text{a}}$  values in cases where substantial charge delocalization is possible, e.g. in aromatic systems.<sup>4,5</sup> Others, however, have not observed such a dependence,<sup>10,24</sup> and geometry optimization in water did not alter the results; in one case it actually made them worse,<sup>11</sup> even though charge delocalization was possible.

To understand fully the significance of the geometry used for the calculation of solvation free energies for  $\text{p}K_{\text{a}}$  calculations, we should consider the energy difference between the molecule in a vacuum and in solvent, optimized on their respective energy surfaces. As can be seen in Figure 2, this difference comprises two components: the energy change on solvating the molecule ( $\Delta G_{\text{hyd}}$ ) and the energy change on going from the vacuum to the solvent geometry ( $\Delta G_{\text{relax}}$ ). If  $\Delta G_{\text{relax}}$  is not included, using  $\Delta G_{\text{hyd}}$  calculated on the vacuum-optimized geometry leads to the total free energy of a solvated molecule  $G(A_{\text{solv,vacuum}})$  being higher than the correct free energy of the molecule A in solvent. This happens because one only takes into account solvation of the vacuum geometry (in terms of  $\Delta G_{\text{hyd,vacuum}}$ ) but does not allow the molecule to relax in water, which would lower

**Table 7.** Absolute Values of  $\Delta G_{\text{relax,gas}}$  and  $\Delta G_{\text{relax,solv}}$  for Fluorescein, DCF, and DFF Molecules [kcal/mol]<sup>a</sup>

		neutral	mono-anion	dianion	$\Delta \text{p}K_{\text{Q-M}1}$	$\Delta \text{p}K_{\text{M}1-\text{D}}$
fluorescein	$\Delta G_{\text{relax,gas}}$	1.48	2.18	1.75	0.51	-0.31
	$\Delta G_{\text{relax,solv}}$	-1.36	-3.10	-3.13	-1.28	-0.02
DCF	$\Delta G_{\text{relax,gas}}$	1.34	2.32	1.58	0.72	-0.54
	$\Delta G_{\text{relax,solv}}$	-1.23	-2.86	-2.81	-0.77	0.02
DFF	$\Delta G_{\text{relax,gas}}$	1.50	2.39	1.78	0.65	0.23
	$\Delta G_{\text{relax,solv}}$	-1.36	-3.07	-3.04	-1.23	0.02

<sup>a</sup>  $\Delta \text{p}K_{\text{Q-M}1}$  and  $\Delta \text{p}K_{\text{M}1-\text{D}}$  show the error in a respective  $\text{p}K_{\text{a}}$  due to the neglect of relaxation.

the total free energy of the system. Conversely, using  $\Delta G_{\text{hyd,aq}}$  leads to  $G(A_{\text{solv,aq}})$  being too low. In this case  $\Delta G_{\text{relax,gas}}$  is neglected. Since this change is made in vacuo the neglected term is always unfavorable. Therefore, it does not matter which geometry is used for the calculation of  $\Delta G_{\text{hyd}}$ , provided that the corresponding correction for the geometry relaxation  $\Delta G_{\text{relax}}$  is taken into account as well. To assess how important the relaxation effect is for our test molecules, we have calculated absolute values of  $\Delta G_{\text{relax,gas}}$  and  $\Delta G_{\text{relax,solv}}$  with HF/6-31+G(d) and the Pauling radii set (data shown in Table 7, relaxation energy values for all radii used are shown in Supporting Information Tables S6–S8). The net effect of the relaxation energies on the values of dissociation constants is calculated according to eq 3 as  $\Delta \text{p}K_{\text{Q-M}1} = \Delta(M-N)/RT \ln 10$  and  $\Delta \text{p}K_{\text{M}1-\text{D}} = \Delta(D-M)/RT \ln 10$ , where  $N$ ,  $M$ , and  $D$  are the relaxation energy values of neutral, mono-, and dianions of respective molecules.

Analysis of the data in Table 7 reveals that although absolute values of relaxation energies are up to 3 kcal/mol, the net effect of the relaxation energy on the values of dissociation constants is smaller due to the equivalent magnitude of relaxation effects for the pair of molecules (e.g. the neutral species and its anion). Overall, the neglect of relaxation incurs a larger error for the first dissociation constant. If the solvation free energy was calculated on the vacuum-optimized geometry, not including the relaxation energy would lead to the value of  $\text{p}K_{\text{Q-M}1}$  being too high by over 1  $\text{p}K_{\text{a}}$  unit for fluorescein and DFF. This error would appear to be smaller if the solvation free energy was calculated on the solvent-optimized geometry. On the other hand, for  $\text{p}K_{\text{M}1-\text{D}}$ , the neglect of the relaxation causes only minor errors not exceeding 0.5  $\text{p}K_{\text{a}}$  unit. The fact that the relaxation effect is more profound for the first dissociation constant may be explained on the grounds that the solvent has a different influence on the geometry of the neutral and charged molecules; this results in different relaxation energies, which do not fully compensate.

Even though the errors incurred by the neglect of relaxation are on average smaller than errors caused by the inaccurate description of solvent effects, we should be aware that the use of the geometry optimized in a vacuum for the calculation of  $\Delta G_{\text{hyd}}$  will yield higher  $\text{p}K_{\text{a}}$  values, while the use of geometry optimized in solvent will lead to lower  $\text{p}K_{\text{a}}$  values. This is, indeed, in line with the findings of earlier studies.<sup>5</sup>

**Tautomerization Constants for Fluorescein and Its Derivatives.** Tautomerization constants calculated using

**Table 8.** Experimental<sup>33,67</sup> and Calculated Tautomerization Constants and the Correction to the pK<sub>a1</sub> (Corr, in pK<sub>a</sub> Units) Given by  $\log(k_T \times (1 + k_M)/1 + k_T + (k_T \times k'_T))$ 

constant	fluorescein			DCF			DFF		
	UAHF	Pauling	exp.	UAHF	Pauling	exp.	UAHF	Pauling	exp.
$k_T$	7144	5.86	6.09	452	2.14	2.00	133	0.13	0.68
$k'_T$	0.18	5E-05	0.50	794	0.007	7.14	2468	0.06	16.94
Corr	-3.85	-4.46	-0.96	-2.66	-2.14	-0.50	-2.13	-1.26	-0.24

different radii sets are shown in Table 8, together with experimental data. Total free energies of all tautomers calculated with both UAHF and Pauling radii are given in the Supporting Information Table S9. We have used only two different radii sets (UAHF and Pauling), given the difference they produced for the pK<sub>a</sub> value calculations, and not considered UAKS radii, as they gave solvation energies close to those derived with UAHF.

Experimental data show that the free energy differences between tautomers in aqueous solution are within 1 kcal/mol for most cases, which sets up a challenging task for the computational methods to reproduce the differences correctly: the accuracy of free energy estimates both in vacuo and in solvent must be within 1 kcal/mol.

First, it should be noted that neither of the two radii sets used to obtain free energies is able to predict both tautomerization constants correctly. On the other hand,  $k_T$  is very well estimated if the Pauling radii set is used to calculate solvation free energies, although it fails to predict  $k'_T$  with reasonable accuracy (errors of 4 orders of magnitude in  $k$  or approximately 5 kcal/mol), and for the DCF and DFF molecules, the relative stability of the Q and Z tautomers is reversed. Conversely, UAHF radii are able to reproduce  $k'_T$  values for a fluorescein molecule and, within 2 orders of magnitude, for the other molecules and maintain the correct relative stabilities of the Q and Z tautomers. This radii set is also able to describe the correct relative stabilities of the L and Q tautomers of both fluorescein and DCF; however, the errors in  $k_T$  are up to 3 orders of magnitude (up to approximately 4 kcal/mol).

Additional contributions to the total free energies of the tautomers in solvent comprise the energy of each tautomer in a vacuum and the thermal correction to the free energy, which includes the zero point energy (ZPE) and entropic contributions. While there are errors also associated with both of these, it has been shown previously that ZPEs converge in the 6-31+G(d) basis set<sup>40</sup> and are, therefore, unlikely to be a significant source of error. To check if the vacuum energy estimation makes a major contribution to the overall error, we have calculated vacuum SCF energies for the fluorescein tautomers using the aug-cc-pVTZ basis set (data not shown), but this has not improved the values of the tautomerization constants. This is also consistent with the pK<sub>a</sub> calculations, which show that differences between the V2 and V3 methods are negligible.

As mentioned in the Methods section, tautomerization constants contribute to the values of the macroscopic pK<sub>as</sub>; therefore, errors in the former will have a negative impact on the quality of the latter. Table 8 shows the value of the

correction to the pK<sub>a1</sub> value which is due to the existence of different tautomeric forms of the molecule. For all of the molecules under consideration, the fact that the neutral forms of the dyes exist as multiple tautomers increases the pK<sub>a1</sub> value (see eq 14). This increase is roughly inversely proportional to the stability of the Q tautomer, being the largest for fluorescein and the smallest for DFF and is never larger than 1 pK<sub>a</sub> unit. Conversely, corrections calculated with both the UAHF and Pauling radii sets are substantially larger. This is caused by the incorrect prediction of the stabilities of various tautomers and either over-stabilization of the zwitterions (Pauling radii) or lactone (UAHF radii). In both cases, the stability of the quinoid form is severely underestimated, leading to an increase in the absolute value of the correction. The use of the calculated tautomerization constants (Table 8) together with microscopic dissociation constants (Tables 3–5) would lead to large discrepancies between calculated and experimental macroscopic dissociation constants, and these discrepancies would be mainly due to the incorrect estimation of tautomerization constants.

Theoretical investigations into the tautomerization constants of fluorescein have been published previously.<sup>70</sup> Those authors used a different solvation method (Poisson–Boltzmann continuum) and were able to reproduce the  $k_T$  value correctly but did not calculate  $k'_T$  since they failed to obtain an optimized geometry of the zwitterion. Indeed, it has since been shown that implicit solvent models are unable to describe the zwitterionic tautomer of carboxylic monoacids correctly,<sup>15</sup> which may be the reason for the poor prediction of  $k'_T$ . It should be noted that the proper calculation of the solvation free energy of the zwitterionic tautomer would require the use of explicit water molecules.

## Conclusions

We have used several different calculation protocols to estimate microscopic dissociation and tautomerization constants for fluorescein and two derivatives, DCF and DFF. Evaluation of free energies in a vacuum employed CBS-4M and DFT-based protocols, while solvation free energies were calculated with the CPCM method and several radii sets. We found that all three in vacuo protocols performed similarly. From this we conclude that the least computationally demanding PBE1PBE/6-311+G(2d,2p)//PBE1PBE/6-31+G(d) may be the method of choice for systems which are too large in practice to be treated with the CBS-QB3 model chemistry. Relative calculations of microscopic dissociation constants produced good agreement with experiment for pK<sub>Q-M1</sub> and pK<sub>M1-D</sub> (in the latter case only for Pauling radii). As expected, results of absolute calculations

deviated more from the experimental data with mean deviations between 1 and 3 pK<sub>a</sub> units. Better results were obtained if the solvation free energy was obtained with the Pauling radii set.

We have also investigated the contribution of the geometry relaxation on the calculated values of dissociation constants. It turned out that relaxation effects are negligible for pK<sub>M1→D</sub>, since geometries of both mono- and dianions are changed similarly, and the net effect is insignificant. For pK<sub>Q→M1</sub> the effect is larger but still smaller than the accuracy of the implicit solvent model. In the case of tautomerization constants, the Pauling radii set correctly predicted *k*<sub>T</sub>, whereas UAHF performed better for *k*'<sub>T</sub>; however, neither set gave accurate estimates for both. The results described in this paper indicate that implicit solvent models are able to produce satisfactory results; however, they should be carefully tested for the set of molecules of interest, and appropriate parameters best describing solvation effects should be chosen. This task may be simplified in the future with the development of implicit solvation methods, which are less dependent on a correct choice of parameters to define the solute cavity. Such methods are gradually being developed and include isodensity surface approaches<sup>6,71</sup> or the use of a mixed implicit-explicit aqueous environment.<sup>18,47</sup>

**Acknowledgment.** We thank Professor Peter Wardman and Dr. Tomasz Borowski for careful reading of the manuscript. This work was funded by Cancer Research U.K. All calculations were performed at the Academic Computer Centre Cyfronet of Stanislaw Staszic University of Mining and Metallurgy in Krakow, Poland (grant KBN/SGI2800/UJ/031/2002).

**Supporting Information Available:** Solvation free energies calculated on geometries optimized with different radii sets (Table S1), solvation and vacuum free energies calculated with the methods S1–S3 and V1–V4 (Tables S2–S4), vacuum free energies of 3-chloro-4-hydroxyphenol (Table S5), relaxation energy values for all radii sets (Tables S6–S8), and total free energies of all tautomers (Table S9). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- Bell, R. P. *The Proton in Chemistry*; Chapman and Hall: London, 1973.
- Chen, I.-J.; MacKerell, A. D., Jr. *Theor. Chem. Acc.* **2000**, *103*, 483–494.
- Liptak, M. D.; Shields, G. C. *Int. J. Quantum Chem.* **2001**, *85*, 727–741.
- Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314–7319.
- Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, *124*, 6421–6427.
- Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413–7422.
- Adam, K. R. *J. Phys. Chem. A* **2002**, *106*, 11963–11972.
- Saracino, G. A. A.; Improta, R.; Barone, V. *Chem. Phys. Lett.* **2003**, *373*, 411–415.
- Dahlke, E. E.; Cramer, C. J. *J. Phys. Org. Chem.* **2003**, *16*, 336–347.
- Magill, A. M.; Yates, B. F. *Aust. J. Chem.* **2004**, *57*, 1205–1210.
- Murłowska, K.; Sadlej-Sosnowska, N. *J. Phys. Chem. A* **2004**, *109*, 5590–5595.
- Fu, Y.; Liu, L.; Li, R.-Q.; Liu, R.; Guo, Q.-X. *J. Am. Chem. Soc.* **2004**, *126*, 814–822.
- Almerindo, G. I.; Tondo, D. W.; Pliego, J. R., Jr. *J. Phys. Chem. A* **2004**, *108*, 166–171.
- Busch, M. S. A.; Knapp, E. W. *ChemPhysChem* **2004**, *5*, 1513–1522.
- Soriano, E.; Cerdán, S.; Ballesteros, P. *J. Mol. Struct. – Theochem* **2004**, *684*, 121–128.
- Han, J.; Tao, F.-M. *J. Phys. Chem. A* **2005**, *110*, 257–263.
- Han, J.; Deming, R. L.; Tao, F.-M. *J. Phys. Chem. A* **2005**, *109*, 1159–1167.
- Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 2493–2499.
- Namazian, M.; Kalantary-Fotooh, F.; Noorbala, M. R.; Searles, D. J.; Coote, M. L. *J. Mol. Struct. – THEOCHEM* **2006**, *758*, 275–278.
- Eckert, F.; Klamt, A. *J. Comput. Chem.* **2006**, *27*, 11–19.
- Pliego, J. R., Jr. *Chem. Phys. Lett.* **2003**, *367*, 145–149.
- Magill, A. M.; Cavell, K. J.; Yates, B. F. *J. Am. Chem. Soc.* **2004**, *126*, 8717–8724.
- Barone, V.; Improta, R.; Rega, N. *Theor. Chem. Acc.* **2004**, *111*, 237–245.
- Gao, D.; Svoronos, P.; Wong, P. K.; Maddalena, D.; Hwang, J.; Walker, H. *J. Phys. Chem. A* **2005**, *109*, 10776–10785.
- Burkitt, M. J.; Wardman, P. *Biochem. Biophys. Res. Commun.* **2001**, *282*, 329–333.
- Wrona, M.; Wardman, P. *Free Radical Biol. Med.* **2005**, *38*, 262–270.
- Wrona, M.; Wardman, P. *Free Radical Biol. Med.* **2006**, in press.
- Royall, J. A.; Ischiropoulos, H. *Arch. Biochem. Biophys.* **1993**, *302*, 348–355.
- Haugland, R. P. *Handbook of Fluorescent Probes and Research products*, 9th ed.; Molecular Probes, Inc.: Eugene, 2002.
- Minta, A.; Kao, J. P.; Tsien, R. Y. *J. Biol. Chem.* **1989**, *264*, 8171–8178.
- LeBel, C. P.; Ischiropoulos, H.; Bondy, S. C. *Chem. Res. Toxicol.* **1992**, *5*, 227–231.
- Kane, D. J.; Sarafian, T. A.; Anton, R.; Hahn, H.; Gralla, E. B.; Valentine, J. S.; Ord, T.; Bredesen, D. E. *Science* **1993**, *262*, 1274–1277.
- Mchedlov-Petrosyan, N. O.; Rubtsov, M. I.; Lukatskaya, L. L. *Dyes Pigm.* **1992**, *18*, 179–198.
- Klonis, N.; Sawyer, W. H. *J. Fluoresc.* **1996**, *6*, 147–157.
- Sjöback, R.; Nygren, J.; Kubista, M. *Spectrochim. Acta A* **1995**, *51*, L7–L21.
- Nyden, M. R.; Petersson, G. A. *J. Chem. Phys.* **1981**, *75*, 1843–1862.

- (37) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532–6542.
- (38) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822–2827.
- (39) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A., Jr. *J. Chem. Phys.* **1996**, *104*, 2598–2619.
- (40) Liptak, M. D.; Shields, G. C. *Int. J. Quantum Chem.* **2005**, *105*, 580–587.
- (41) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (42) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.
- (43) Sicinska, D.; Paneth, P.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *106*, 2708–2713.
- (44) Rivelino, R.; Coutinho, K.; Canuto, S. *J. Phys. Chem. B* **2002**, *106*, 12317–12322.
- (45) Rode, B. M.; Schwenka, C. F.; Tongraar, A. *J. Mol. Liq.* **2004**, *110*, 105–122.
- (46) Tuñón, I.; Rinaldi, D.; Ruiz-López, M. F.; Rivail, J. L. *J. Phys. Chem.* **1995**, *99*, 3798–3805.
- (47) Pliego, J. R., Jr.; Riveros, J. M. *J. Phys. Chem. A* **2001**, *105*, 7241–7247.
- (48) Pliego, J. R., Jr.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434–7439.
- (49) Cao, Z.; Lin, M.; Zhang, Q.; Mo, Y. *J. Phys. Chem. A* **2004**, *108*, 4277–4282.
- (50) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133–1152.
- (51) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70–77.
- (52) Ben-Naim, A. *Solvation thermodynamics*; Plenum Press: New York, 1987.
- (53) Silva, C. O.; Silva, E. C. D.; Nascimento, M. A. C. *J. Phys. Chem. A* **2000**, *104*, 2402–2409.
- (54) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R., Jr. *J. Phys. Chem. A* **1998**, *102*, 7787–7794.
- (55) Tuttle, T. R.; Malaxos, S.; Coe, J. V. *J. Phys. Chem. A* **2002**, *106*, 925–932.
- (56) Camaioni, D. M.; Schwerdtfeger, C. A. *J. Phys. Chem. A* **2005**, *109*, 10795–10797.
- (57) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610–5620.
- (58) Palascak, M. W.; Shields, G. C. *J. Phys. Chem. A* **2004**, *108*, 3692–3694.
- (59) Mchedlov-Petrosyan, N. O.; Kleschchevnikova, V. N. *J. Chem. Soc. Faraday Trans.* **1994**, *90*, 629–640.
- (60) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (61) Samoilov, D. V.; Mchedlov-Petrosyan, N. O.; Martynova, V. P.; El'tsov, A. V. *Russ. J. Gen. Chem.* **2000**, *70*, 1259–1271.
- (62) Ayala, P. Y.; Schlegel, H. B. *J. Chem. Phys.* **1998**, *108*, 2314–2325.
- (63) 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, 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.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02 ed.*; Gaussian, Inc.: Wallingford, CT, 2004.
- (64) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669–681.
- (65) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210–3221.
- (66) Lide, D. R. *CRC Handbook of Chemistry and Physics*, 86th edition (June 23, 2005); CRC: 2005.
- (67) Orte, A.; Crovetto, L.; Talavera, E. M.; Boens, N.; Alvarez-Pez, J. M. *J. Phys. Chem. A* **2005**, *109*, 734–747.
- (68) Smith, S. A.; Pretorius, W. A. *Water SA* **2002**, *28*, 395–402.
- (69) Leonhardt, H.; Gordon, L.; Livingston, R. *J. Phys. Chem.* **1971**, *75*, 245–249.
- (70) Jang, Y. H.; Hwang, S.; Chung, D. S. *Chem. Lett.* **2001**, *30*, 1316–1317.
- (71) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098–16104.

## 2-Phenylpyridine: To Twist or Not To Twist?

Kerwin D. Dobbs\*<sup>†</sup> and Karl Sohlberg<sup>‡</sup>

*DuPont Central Research & Development, Experimental Station, P.O. Box 80320, Wilmington, Delaware 19880-0320, and Department of Chemistry, Drexel University, 3141 Chestnut Street, Philadelphia, Pennsylvania 19104*

Received July 26, 2006

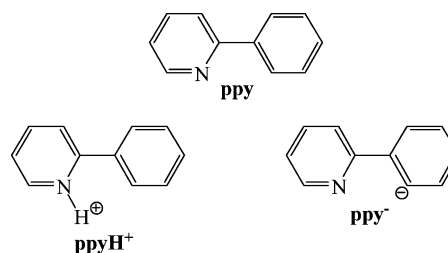
**Abstract:** Density functional theory methods were used to investigate the structures associated with 2-phenylpyridine, **ppy**, and several of its electronic states. The structure of **ppy** has the aromatic rings twisted with respect to one another by  $\sim 21^\circ$ , which is about half the value found for biphenyl. In comparison with **ppy**, both the isoelectronic cation, **ppyH<sup>+</sup>**, and anion, **ppy<sup>-</sup>**, have larger twist angles. The extent of twisting is governed by the delicate balance between  $\pi$  conjugation and repulsive orbital/steric interactions, and the magnitudes of these interactions were investigated by examining the torsional energy barriers for all three molecular species. In contrast, every one of the investigated open-shell structures—**ppy<sup>+•</sup>**, **ppy<sup>-•</sup>**, **ppy<sup>\*•</sup>**, **ppyH<sup>+•</sup>**, and **ppy<sup>-•\*</sup>**—has coplanar aromatic rings, that is, no twist angle. Frontier molecular orbital analyses reveal that the  $\pi$ -type bonding between the bridging carbons becomes dominant over any repulsive orbital and steric interactions, thereby leading to coplanar rings. Also, the energetics associated with **ppy** and its various electronic states were investigated and reported.

### Introduction

Materials with conjugated organic moieties, whether they be molecular, oligomeric, or polymeric in nature, are being used in devices as semiconductors or light-emitting diodes and have captured the imaginations of researchers investigating the associated structural, electronic, and optical properties.<sup>1–3</sup> These conjugated moieties are usually olefins (e.g., ethylene and acetylene), freely rotating aromatic rings (e.g., benzene and thiophene), rigidly linked biphenyl-like units (fluorene and carbazole), or fused aromatic rings (e.g., tetracene and pentacene). In conjunction with experimental work, the past decade has seen the fundamental properties of these conjugated materials being ascertained through quantum mechanical modeling of the actual molecular system or the repeat unit of an oligomer/polymer system.<sup>4</sup> By using first-principles computational chemistry methods such as density functional theory (DFT), one will obtain computational results which are accurate and reliable, thereby

extending limited experimental data and allowing the confident study of systems which may be difficult to do experimentally. One such conjugated molecular system with a paucity of experimental molecular property data is 2-phenylpyridine, **ppy**.

The only existing structural information for **ppy** is obtained indirectly through the X-ray crystal structures of organometallic complexes containing **ppy** as the formally anionic ligand, **ppy<sup>-</sup>**. This ligand and its derivatives are commonly found in organometallic species which are electroluminescent.<sup>5,6</sup> In all of these structures, the **ppy<sup>-</sup>** ligand was shown



\* Corresponding author e-mail: Kerwin.D.Dobbs@usa.dupont.com.

<sup>†</sup> DuPont Central Research & Development (this paper is DuPont contribution #8747).

<sup>‡</sup> Drexel University.

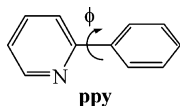
to have coplanar, or very nearly coplanar, aromatic rings. Previous computations on the neutral **ppy** molecule, however, have found the aromatic rings to be twisted with respect to

each other in the range of 21–32°. <sup>7,8</sup> In acidic aqueous media, **ppy** is in its protonated form, **ppyH<sup>+</sup>**, and may exhibit either fluorescent or phosphorescent characteristics. <sup>9–11</sup> Also, this experimental work resulted in qualitative arguments which suggest a twisted structure for the ground state of **ppyH<sup>+</sup>** but a planar structure for the first excited triplet state. To date, no first-principles computational chemistry data exist either for the singlet states of **ppyH<sup>+</sup>** and **ppy<sup>-</sup>** or for the first excited triplet states of all three molecular species—**ppy\***, **ppyH<sup>+</sup>\***, and **ppy<sup>-\*</sup>**.

The ionizing (electron removal/addition) of conjugated polymer chains can have profound effects on their conductivity. Oxidation/reduction creates a localized charge defect in the polymer chain known as a positive/negative polaron. <sup>12</sup> The radical cation, **ppy<sup>•+</sup>**, and radical anion, **ppy<sup>•-</sup>**, may be considered as models for single-charged defects in polymers which contain **ppy** as a comonomer. What is unknown is the effect on the molecular structure of **ppy** upon ionization.

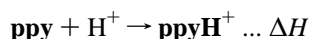
### Computational Model

DFT methods are used to understand the structure and energetics of **ppy** and its various electronic states. The first task is to firmly establish the equilibrium structure of **ppy** and the energy maxima (at  $\phi = 0^\circ$  and  $90^\circ$ ) for rotation about the CC bond linking the two aromatic rings.

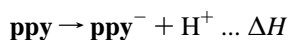


A full torsional energy profile is not necessary because it has been investigated previously. <sup>8</sup>

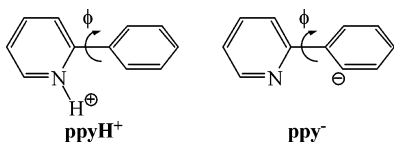
The second task is to investigate the structural and energetic properties of the isoelectronic cationic and anionic states of **ppy**. For the protonation of **ppy** to form **ppyH<sup>+</sup>**, a proton is attached to the pyridine nitrogen



The release of energy is known as the proton affinity of **ppy** and is defined as the negative enthalpy,  $-\Delta H$ , for the above reaction. The isoelectronic anion, **ppy<sup>-</sup>**, of interest is the one in which  $\text{H}^+$  is removed from the phenyl carbon “ortho” to the CC bridge between the two rings of **ppy**, and the enthalpy for this reaction



is known as the deprotonation energy of **ppy**. In addition, the torsional barriers of **ppyH<sup>+</sup>** and **ppy<sup>-</sup>** at  $\phi$  values of  $0^\circ$  and  $90^\circ$  are computed and compared with the **ppy** results.



Third, an electron is subtracted and added to **ppy** to form, respectively, the radical cation, **ppy<sup>•+</sup>**, and the radical anion, **ppy<sup>•-</sup>**, and the computed structures of these two charged, radical species are compared to that of **ppy**. In addition, the

**Table 1.** Inter-Ring Distances (CC in Angstroms) and Twist Angles ( $\phi$  in Degrees) for **ppy<sup>a</sup>**

basis sets	CC		$\phi$	
	BP86	B3LYP	BP86	B3LYP
6-31G(d)	1.492	1.488	15.0 (16.0)	18.2 (19.3)
6-311G(d)	1.490	1.488	18.1 (19.4)	20.4 (21.8)
6-311G(2d,2p)	1.489	1.486	18.2 (19.4)	20.8 (21.9)

<sup>a</sup> The numbers in parentheses are the alternate dihedral angles between the aromatic rings (see text).

ionization potential (IP) and electron affinity (EA) of **ppy** to form **ppy<sup>•+</sup>** and **ppy<sup>•-</sup>**, respectively,

$$\text{IP} = E(\text{ppy}^{\bullet+}) - E(\text{ppy})$$

may be determined as “adiabatic” or “vertical” values. The

$$\text{EA} = -[E(\text{ppy}^{\bullet-}) - E(\text{ppy})]$$

adiabatic value is the energy difference between the charged, radical species in its relaxed (optimized) geometry and the respective neutral, singlet species in its relaxed geometry. For the vertical value, the energy difference is between the charged, radical species at the geometry of its neutral state and the neutral species in its relaxed geometry. Both the adiabatic and vertical IPs and EAs are computed in order to determine the effects of structure on these energy values.

As part of the third task, the focus on open-shell analogues continues with the comparison of the first excited triplet states, **ppy\***, **ppyH<sup>+</sup>\***, and **ppy<sup>-\*</sup>**, with respect to their singlet ground states, **ppy**, **ppyH<sup>+</sup>**, and **ppy<sup>-</sup>**. In addition to examining the structural differences between the singlet and triplet molecular species, the singlet–triplet energy differences are also investigated. The adiabatic energy difference,  $\Delta E_{\text{ad}}(\text{S}_0 \rightarrow \text{T}_1)$ , is the energy needed to excite a molecule from its singlet ground-state geometry to the relaxed geometry of its first excited triplet state. An alternative method to use is time-dependent DFT (TDDFT), <sup>13–16</sup> which has proven to be most successful for low-energy excitations. <sup>17,18</sup>

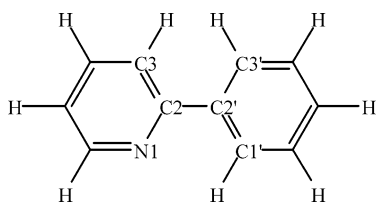
### Computational Details

All calculations were performed with the Gaussian 03 suite of programs. <sup>19</sup> The neutral, singlet **ppy** structure was optimized by DFT with the density functionals BP86 <sup>20,21</sup> and B3LYP <sup>22,23</sup> in conjunction with a few selected basis sets: 6-31G(d), <sup>24</sup> 6-311G(d), <sup>25</sup> and 6-311G(2d,2p). <sup>26</sup> The hybrid functional B3LYP has proven to be good for the rotational barriers of biphenyl <sup>27</sup> as well as for relative energetics in general. The pure BP86 functional was chosen mainly because of the enhanced performance gains of the calculations using this functional compared to B3LYP. Pure functionals are able to take advantage of using density fitting basis sets which expand the density in a set of atom-centered functions when computing the Coulomb interaction instead of computing all of the two-electron integrals. <sup>28,29</sup> Table 1 presents selected metric parameters for the equilibrium structure of **ppy** at several different levels of theory. For the purposes of this study, the metric parameters of interest

**Table 2.** Inter-Ring Distances (CC in Angstroms) and Torsional Barrier Heights (kcal/mol) for **ppy** at Twist Angle  $\phi = 0.0^\circ$  and  $90.0^\circ$  for Different Levels of Computation

$\phi$	basis sets	BP86			B3LYP		
		CC	$\Delta H$	$\Delta G$	CC	$\Delta H$	$\Delta G$
0.0	6-31G(d)	1.493	0	1	1.491	0	1
	6-311G(d)	1.492	0	1	1.490	0	1
	6-311G(2d,2p)	1.491	0	1	1.489	0	1
90.0	6-31G(d)	1.504	4	6	1.498	4	5
	6-311G(d)	1.501	4	5	1.497	4	5
	6-311G(2d,2p)	1.500	4	5	1.496	4	5

are the inter-ring CC distance and the twist angle,  $\phi$ , which is the dihedral angle defined as N1–C2–C2'–C1':



Values for the dihedral angle C3–C2–C2'–C3' are also given in Table 1 in parentheses and turn out to be noticeably different from the defined twist angle because of the nonplanarity of the two aromatic rings.<sup>8</sup> Because each optimized geometry was characterized by a harmonic vibrational frequency analysis at the same level of computation, the calculated frequencies were used to determine zero-point energy and vibrational thermal corrections (at 298.15 K) for use in determining energetic properties unless otherwise stated. To determine rotational barriers of **ppy**, two structures of **ppy** were fully optimized with the twist angle,  $\phi$ , being constrained to  $0^\circ$  and  $90^\circ$ . The enthalpic and free energy torsional barrier heights as well as the inter-ring CC distances for **ppy** at  $\phi = 0^\circ$  and  $90^\circ$  are presented in Table 2.

The remaining calculations were all done with the B3LYP functional, and any energetic properties which involved **ppy** utilized the B3LYP/6-311G(2d,2p) geometry of **ppy**. To investigate the structures of the isoelectronic cationic (**ppyH**<sup>+</sup>) and anionic (**ppy**<sup>−</sup>) singlet states of **ppy**, the polarized triple- $\zeta$  basis sets, 6-311G(2d,2p), and the diffuse basis sets, 6-311+G(2d,2p) and 6-311++G(2d,2p),<sup>30</sup> were used. Two energy minima exist for **ppy**<sup>−</sup>: a local minimum at  $\phi < 50^\circ$  and a global minimum at  $\phi = 180^\circ$ . For this study, the former structure is of greater interest because this rotational isomer closely resembles the **ppy** anionic ligand in organometallic structures. The inter-ring CC distances and twist angles at the different levels of computation are presented in Table 3. A computational procedure, similar to the one employed for the neutral **ppy** species, was used to determine the inter-ring CC bond lengths and the torsional barrier heights (at  $\phi = 0^\circ$  and  $90^\circ$ ) of **ppyH**<sup>+</sup> and **ppy**<sup>−</sup> but only at the B3LYP/6-311+G(2d,2p) level (see Table 4). As can be seen in Table 3, the extra set of diffuse functions (for H) does not alter the structures of these ionic species. The proton affinity and deprotonation energy of **ppy** were computed at the B3LYP/6-311++G(2d,2p) level and are presented in Table 5 as the energy property  $\Delta H(\text{H}^+)$ .

**Table 3.** Inter-Ring Distances (CC in Angstroms) and Twist Angles ( $\phi$  in Degrees) for **ppyH**<sup>+</sup> and **ppy**<sup>−</sup>, Using the B3LYP Hybrid Functional and the Indicated Basis Sets

basis sets	<b>ppyH</b> <sup>+</sup>		<b>ppy</b> <sup>−</sup>	
	CC	$\phi$	CC	$\phi$
6-311G(2d,2p)	1.463	33.0 (33.6)	1.488	43.1 (41.8)
6-311+G(2d,2p)	1.464	33.1 (33.7)	1.489	48.1 (46.4)
6-311++G(2d,2p)	1.464	33.1 (33.7)	1.489	48.2 (46.4)

**Table 4.** Inter-Ring Distances (CC in Angstroms) and Torsional Barrier Heights (kcal/mol) for **ppyH**<sup>+</sup> and **ppy**<sup>−</sup> at Twist Angle  $\phi = 0.0^\circ$  and  $90.0^\circ$  for the B3LYP/6-311+G(2d,2p) Level of Theory

$\phi$	<b>ppyH</b> <sup>+</sup>			<b>ppy</b> <sup>−</sup>		
	CC	$\Delta H$	$\Delta G$	CC	$\Delta H$	$\Delta G$
0.0	1.467	1	2	1.497	2	3
90.0	1.483	3	4	1.500	0	1

**Table 5.** Energy Properties for **ppyH**<sup>+</sup> and **ppy**<sup>−</sup> at the B3LYP/6-311++G(2d,2p) Level of Theory<sup>a</sup>

property	<b>ppyH</b> <sup>+</sup>	<b>ppy</b> <sup>−</sup>
$\Delta H(\text{H}^+)$	10 (230)	17 (398)
$\Delta E_{\text{ad}}(\text{S}_0 \rightarrow \text{T}_1)$	2.6 (60)	0.8 (19)
$\Delta E_{\text{TDDFT}}(\text{S}_0 \rightarrow \text{T}_1)$	2.9 (67)	1.4 (32)

<sup>a</sup> Energies are given in eV with numbers in parentheses given in kcal/mol.

**Table 6.** Inter-Ring Distances (CC in Angstroms) for **ppy**, **ppy**<sup>+</sup>, **ppy**<sup>−</sup>, **ppy**<sup>\*</sup>, **ppyH**<sup>+</sup>, and **ppy**<sup>−\*</sup> Using the B3LYP Hybrid Functional

basis sets	<b>ppy</b>	<b>ppy</b> <sup>+</sup>	<b>ppy</b> <sup>−</sup>	<b>ppy</b> <sup>*</sup>	<b>ppyH</b> <sup>+</sup>	<b>ppy</b> <sup>−*</sup>
6-311G(2d,2p)	1.486	1.443	1.434	1.387	1.398 <sup>a</sup>	1.425
6-311+G(2d,2p)		1.444	1.435		1.399	1.426
6-311++G(2d,2p)		1.444	1.435		1.399	1.426

<sup>a</sup> The aromatic rings are coplanar at all levels of theory for all open-shell molecules with the exception of this one, in which the optimized twist angle is actually  $1.1^\circ$  at this particular level of theory.

For the neutral, triplet state **ppy**<sup>\*</sup>, the structure and frequencies were computed using the 6-311G(2d,2p) basis sets. These same basis sets plus the two diffuse versions were used to compute the structures and frequencies of the remaining open-shell molecules: the ionic radical doublets, **ppy**<sup>+</sup> and **ppy**<sup>−</sup>, as well as the ionic triplets, **ppyH**<sup>+</sup> and **ppy**<sup>−\*</sup>. Because the two aromatic rings in each of these optimized structures are coplanar, only the inter-ring CC bond lengths are presented in Table 6.

The energetic properties of the neutral, singlet state of **ppy** with respect to the three open-shell species, **ppy**<sup>\*</sup>, **ppy**<sup>+</sup>, and **ppy**<sup>−</sup>, were obtained at the B3LYP/6-311++G(2d,2p) level of computation and are presented in Table 7. Each of the adiabatic IP, EA, and  $\Delta E_{\text{ad}}(\text{S}_0 \rightarrow \text{T}_1)$  values are determined from the energy difference between the **ppy** structure and each of the three structure-optimized open-shell states. To determine the vertical IP and EA values, single-point energy calculations were first carried out at the designated computational level for each of the open-shell



**Table 7.** Energy Properties for **ppy** Computed at the B3LYP/6-311++G(2d,2p) Level of Theory<sup>a</sup>

property	energy
IP <sub>ad</sub>	8.0 (185)
IP <sub>vert</sub>	8.2 (189)
EA <sub>ad</sub>	0.3 (6)
EA <sub>vert</sub>	-0.1 (-2)
$\Delta E_{\text{ad}}(\text{S}_0 \rightarrow \text{T}_1)$	2.8 (65)
$\Delta E_{\text{TDDFT}}(\text{S}_0 \rightarrow \text{T}_1)$	3.2 (73)

<sup>a</sup> Energies are given in eV with numbers in parentheses given in kcal/mol.

species at the singlet **ppy** geometry. Then, the vertical energy is determined by subtracting the energy of the singlet **ppy** from each of these new single-point energies [e.g., IP<sub>vert</sub> =  $E(\text{ppy}^{\bullet+} @ \text{ppy geom.}) - E(\text{ppy})$ ]. By their nature, vertical energies do not have corrections for zero-point energy and vibrational thermal effects. To complete the energetic picture for the triplet energy surface relative to the singlet one, TDDFT computations were carried out for the first seven excited triplet states on the singlet structure of **ppy**. Only the excitation energy to the first excited triplet state,  $\Delta E_{\text{TDDFT}}(\text{S}_0 \rightarrow \text{T}_1)$ , is presented in Table 7. In a manner similar to the one just described for **ppy**, the energetics for **ppyH**<sup>+</sup> and **ppy**<sup>-</sup> relative to their respective triplet states have been completed and are reported as  $\Delta E_{\text{ad}}(\text{S}_0 \rightarrow \text{T}_1)$  and  $\Delta E_{\text{TDDFT}}(\text{S}_0 \rightarrow \text{T}_1)$  values in Table 5.

## Results and Discussion

**ppy**. The equilibrium gas-phase structure of **ppy** is known only through computational results, and the highest-level calculations to date may be found in two recent reports in the literature. In a thermochemical study of phenylpyridine isomers, da Silva et al.<sup>7</sup> obtained an inter-ring CC distance of 1.488 Å and a twist angle of 20.7° at the B3LYP/6-31G(d) level of theory. In examining the torsional barrier of **ppy**, Göller and Grummt<sup>8</sup> used four different methods: HF/6-31G(d), MP2/6-31G(d), BPW91/6-31+G(d), and B3LYP/6-31+G(d). Respectively, these methods yielded inter-ring CC bond lengths of 1.491, 1.479, 1.491, and 1.489 Å as well as twist angles of 27.9°, 31.5°, 26.7°, and 21.9°. We wanted to see if the inter-ring CC distance and the twist angle of **ppy** would greatly change either as the basis set became larger or in switching from a pure density functional (BP86) to a hybrid one (B3LYP). An exhaustive study with a variety of basis sets and density functionals was unnecessary because experimental structure information is not available for judging the accuracy of the computed methods. As indicated in Table 1, our best DFT values, (BP86;B3LYP)/6-311G(2d,2p), for these metric parameters essentially agree with previous computational studies in that the CC bond length is about 1.49 Å and the twist angle is uncertain (ours is 18–21°). This latter uncertainty is probably due to the shallowness of the torsional potential energy surface where the aromatic rings are nearly coplanar (vide infra and ref 8). The MP2 results of Göller and Grummt indicate a CC bond length which is shorter by 0.01 Å compared to all of the DFT results. A similar result was found in a definitive computational study of biphenyl by Arulmozhiraja and

Fujii.<sup>27</sup> Also, the highest-level DFT computations in this latter study yielded a twist angle of ~40° for biphenyl. The second dihedral angle between the aromatic rings (numbers in parentheses in Table 1) is about 1° larger than the defined twist angle and is indicative of nonplanar rings. In fact, all of the computational methods in the current study agree that the largest distortion from planarity in each ring is at the bridging carbon atom.

Just as in the isoelectronic biphenyl molecule, the twist angle between the two aromatic rings is a delicate balance between electronic and steric effects. The two major electronic effects are the degree of  $\pi$  conjugation between the two rings, leading to coplanar rings, and the repulsive orbital interactions between the  $\pi$  systems of each ring, leading to a twisted structure. Repulsive steric interactions between the ortho hydrogens on different rings tend to direct the rings to be perpendicular to one another.

One way to gauge the magnitude of these effects is to compute the torsional barriers of **ppy** at  $\phi = 0^\circ$  (rings are coplanar) and at  $\phi = 90^\circ$  (rings are perpendicular), and the corresponding enthalpic and free energy values are presented in Table 2. Within the reliability of the methods being used, the torsional barriers turned out to be independent of the basis set and density functional. There is no enthalpic barrier for making the rings coplanar, and the corresponding free energy barrier is only 1 kcal/mol. Concurrently, the CC bond length has lengthened by no more than 0.003 Å. Coplanarity would maximize the repulsive interactions between the two rings, but these interactions must already be well-balanced by  $\pi$  conjugation because of the small rotational barrier and minor effect on the inter-ring CC distance. This small energy barrier for coplanarity not only explains the wide range in optimized twist angles from different computational methods but also would predict difficulty in obtaining a small uncertainty for the twist angle in an experimental structure.

On the other hand, the enthalpic (4 kcal/mol) and free energy (5 kcal/mol) barriers for making the rings perpendicular are more substantial. Also, there is a noticeable lengthening of the inter-ring CC distance by about 0.01 Å toward a typical CC single bond length. Maximizing the twist angle essentially eliminates any repulsive orbital and steric interactions between the aromatic rings, but it also disrupts any  $\pi$ -type bonding between the two rings. Göller and Grummt<sup>8</sup> had obtained similar magnitudes for the torsional barriers of **ppy**, but they had claimed that a three-center–four-electron interaction between the nitrogen and ortho C–H bond on the neighboring ring (N1...H1'C1') contributed to the very low barrier for  $\phi = 0^\circ$ . We cannot support such a claim, which was based solely on geometrical arguments. Overall, our computed results point to weaker repulsive orbital and steric interactions compared to the  $\pi$ -electron delocalization between the rings. The BP86 functional has proven itself to provide structural and torsional energy results comparable to those with B3LYP. For the rest of this investigation, however, the B3LYP functional was used, mainly because of its proven success for reaction energetics.<sup>31</sup>

**ppyH**<sup>+</sup> and **ppy**<sup>-</sup>. To date, there have been no computational chemistry reports in the literature for either the protonated or anionic forms of **ppy** in order to determine

if they have twist angles and, if so, the energy barriers for coplanar as well as perpendicular rings. We have high confidence in the structure results for **ppy** with the 6-311G(2d,2p) basis sets. To gauge basis set convergence for the ionic forms of **ppy**, however, diffuse functions were initially added to the heavy atoms of this basis set [6-311+G(2d,2p)], and then to the hydrogens [6-311++G(2d,2p)], and selected optimized metric parameters with the B3LYP functional are presented in Table 3. With the first set of diffuse functions, the inter-ring CC bond length in either ionic molecule increases by only 0.001 Å while the twist angle for **ppyH**<sup>+</sup> increases by only 0.1°. But, the **ppy**<sup>-</sup> twist angle is more sensitive to the addition of the first set of diffuse functions, increasing by 5°. No significant changes in metric parameters occur with the addition of diffuse functions for hydrogens. Just like for **ppy**, the rings are nonplanar for either ionic species. This distortion in planarity (see numbers in parentheses in Table 3) is smaller for **ppyH**<sup>+</sup> than for **ppy**, but it is larger for **ppy**<sup>-</sup>. In fact, the second dihedral angle is actually *smaller* than the defined twist angle by nearly 2°, which may be indicative of a repulsive steric interaction between the ortho hydrogens being less than the repulsive interaction between the lone pairs on the nitrogen and the anionic carbon atom.

Compared to **ppy** at the same level of theory, B3LYP/6-311G(2d,2p), the inter-ring CC bond length is 0.023 Å shorter and the twist angle is about 12° larger for **ppyH**<sup>+</sup>. The shorter bond length points to a weaker orbital repulsion compared to the  $\pi$  conjugation, while the larger twist angle is the result of two sets of ortho-hydrogen steric interactions. For **ppy**<sup>-</sup>, the inter-ring CC bond length is slightly longer by 0.002 Å, while the twist angle is noticeably larger by 22° compared to that for **ppy**. Because **ppy** and **ppy**<sup>-</sup> both have the same number of ortho-hydrogen steric interactions, the much larger twist angle in **ppy**<sup>-</sup> must be the result of orbital repulsions, most likely, the repulsive interaction between the lone pairs on the nitrogen and the anionic carbon atom.

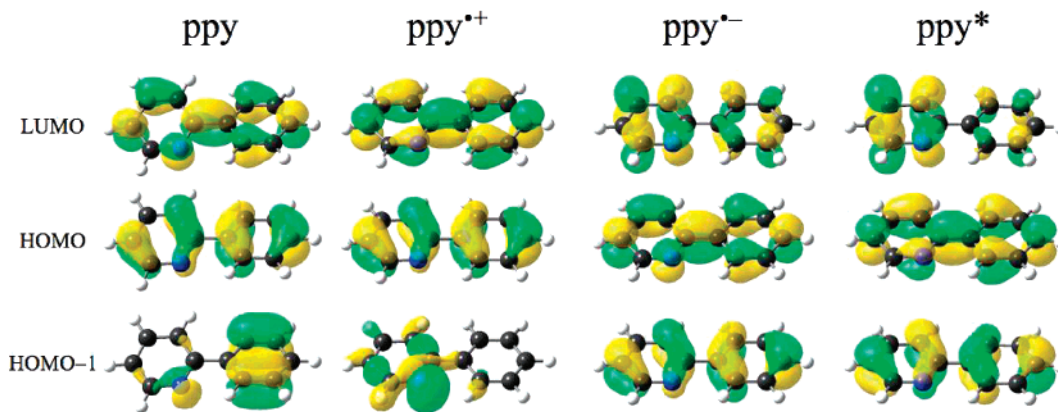
The torsional energy barriers and inter-ring CC distances of **ppyH**<sup>+</sup> and **ppy**<sup>-</sup> at the B3LYP/6-311+G(2d,2p) level are presented in Table 4. When compared to **ppy**, the resistance to ring coplanarity for these ions is somewhat greater. Just like for **ppy**, the ortho-hydrogen steric interactions are at a maximum for coplanar rings in **ppyH**<sup>+</sup>, but two sets of these steric interactions (compared to only one set in **ppy**) have yielded a slightly larger barrier (by 1 kcal/mol) and a slightly longer (by 0.003 Å) CC bond compared to the relaxed structure of **ppyH**<sup>+</sup>. For **ppy**<sup>-</sup>, the repulsive lone-pair interaction and  $\pi$  overlap are both maximized for coplanar rings, but the apparently stronger repulsive interaction results in a rotational barrier 2 kcal/mol larger than that in **ppy**, while the CC bond has lengthened by 0.008 Å compared to that in the relaxed structure of **ppy**<sup>-</sup>. The rotational barrier at 90° for **ppyH**<sup>+</sup> is similar to that of **ppy**, and the CC bond is 0.019 Å longer. These results are a reflection of disrupting the  $\pi$  conjugation (again, the stronger interaction) and eliminating any steric repulsions between the rings. In contrast, making the rings perpendicular for **ppy**<sup>-</sup> has no enthalpic barrier, a free energy barrier of 1 kcal/

mol, and a longer (by 0.011 Å) CC bond between the rings. Although the  $\pi$  conjugation has been disrupted, it has been almost equally balanced by the easing of the repulsive interaction between the lone pairs on the nitrogen and the anionic carbon atom.

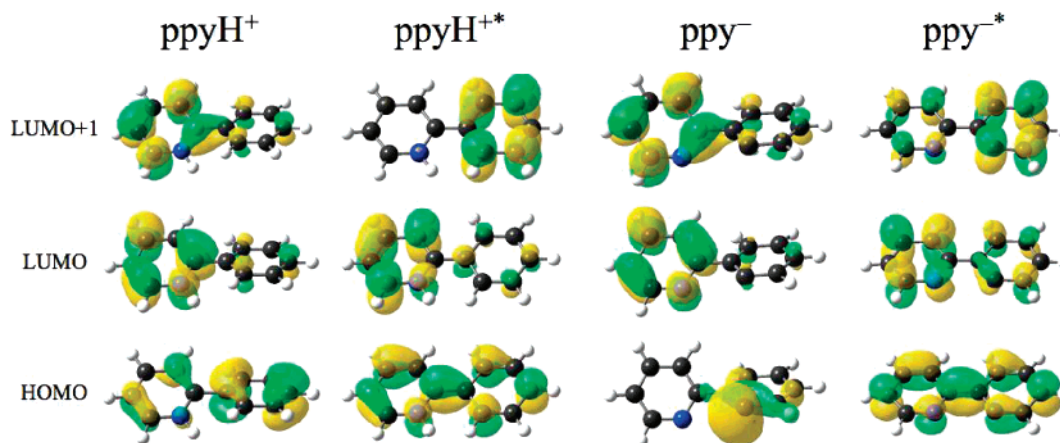
**ppy**<sup>+</sup>, **ppy**<sup>-</sup>, **ppy**<sup>\*</sup>, **ppyH**<sup>+</sup>, and **ppy**<sup>-</sup>. To our knowledge, neither the two radical ions nor the three triplet species have been studied computationally. Using the B3LYP hybrid functional, the optimized geometries for these open-shell molecules were determined with the 6-311G(2d,2p) basis sets, and the effects to the inter-ring CC bond lengths of adding diffuse functions to the heavy atoms, and then to hydrogen, for the charged species is presented in Table 6. As was found for **ppyH**<sup>+</sup> and **ppy**<sup>-</sup>, the inter-ring distances for all of the ionic molecules increase by 0.001 Å with the addition of diffuse functions for carbon and nitrogen, but no further change in this metric parameter occurs when also adding diffuse functions for hydrogen.

The structures of all of these open-shell species are rather interesting because each of the inter-ring CC bond lengths is much shorter than that computed for the corresponding closed-shell species and the aromatic rings in each structure are coplanar. As a point of reference, DFT computations for the isoelectronic biphenyl open-shell molecules<sup>27,32</sup> have yielded similar conclusions with the exception that the biphenyl radical cation has a twist angle<sup>27</sup> which is in agreement with experimental resonance Raman spectra.<sup>33</sup> Compared to **ppy**, the B3LYP/6-311G(2d,2p) values for the inter-ring bond lengths of **ppy**<sup>+</sup> and **ppy**<sup>-</sup> are shorter by 0.043 and 0.052 Å, respectively, whereas this bond length is still even shorter, by 0.099 Å, for **ppy**<sup>\*</sup>. When compared to the closed-shell molecules **ppyH**<sup>+</sup> and **ppy**<sup>-</sup>, the shortening of the inter-ring CC bond for the corresponding open-shell molecules, **ppyH**<sup>+</sup> and **ppy**<sup>-</sup>, is similar (0.065 and 0.063 Å, respectively). At first glance, the shorter inter-ring distances and coplanar aromatic rings seem to be counter-intuitive. These structural preferences, however, are easily explained by examining the respective frontier molecular orbitals.

Figure 1 compares the highest occupied molecular orbitals (HOMOs) and lowest unoccupied molecular orbital (LUMOs) of **ppy**, **ppy**<sup>+</sup>, **ppy**<sup>-</sup>, and **ppy**<sup>\*</sup>.<sup>34,35</sup> The HOMO of **ppy** represents a  $\pi^*$ -like repulsive interaction between the two rings, contributing to the twist angle of ~21°, whereas the corresponding LUMO displays  $\pi$ -like bonding between the two bridging carbon atoms. Although the HOMO and LUMO of **ppy**<sup>+</sup> are identical to those of **ppy**, there is now one less electron in the HOMO of **ppy**<sup>+</sup>, which leads to orbital repulsion being reduced and  $\pi$  conjugation becoming dominant. The concomitant structural effects are a shorter distance between the rings and a twist angle of zero.  $\pi$  bonding between the bridging carbons exists for the LUMO of **ppy**. Therefore, adding an electron to this LUMO to create **ppy**<sup>-</sup> results in an increase in the  $\pi$  conjugation between the aromatic rings (see the HOMO of **ppy**<sup>-</sup>), which overwhelms the existing repulsive  $\pi^*$ -like interaction (see the HOMO-1 of **ppy**<sup>-</sup>). Again, this dominant  $\pi$ -bonding interaction leads to a shorter inter-ring CC bond and coplanar rings when compared to the **ppy** structure. Now, what would



**Figure 1.** Frontier molecular orbitals for **ppy**, **ppy<sup>+</sup>**, **ppy<sup>-</sup>**, and **ppy<sup>\*</sup>** which explain the shorter inter-ring CC bond distances and coplanar aromatic rings of the open-shell molecules compared to **ppy**.



**Figure 2.** Frontier molecular orbitals for **ppyH<sup>+</sup>**, **ppyH<sup>+</sup><sup>\*</sup>**, **ppy<sup>-</sup>**, and **ppy<sup>-</sup><sup>\*</sup>** which explain the shorter inter-ring CC bond distances and coplanar aromatic rings of the triplet molecules compared to their respective singlet analogues.

the result be of combining the effects seen in **ppy<sup>+</sup>** and in **ppy<sup>-</sup>**, that is, to reduce the orbital repulsion, as in **ppy<sup>+</sup>**, and increase the  $\pi$  conjugation, as in **ppy<sup>-</sup>**? As it turns out, this is exactly what happens with the triplet state of **ppy**. For **ppy<sup>\*</sup>**, the frontier orbitals are nearly the same as those for **ppy<sup>-</sup>**, but the electron occupancy is different. One electron is in the  $\pi$ -like HOMO of **ppy<sup>\*</sup>**, but now, there is only one electron in the  $\pi^*$ -like HOMO-1. This combined effect is reflected in the much shorter CC bridge bond compared to that in **ppy**, which is essentially the sum of the bond-shortening effects found in the radical ions ( $0.043 \text{ \AA} + 0.052 \text{ \AA} = 0.097 \text{ \AA} \approx 0.99 \text{ \AA}$ ).

The explanation for the structural preferences found in **ppyH<sup>+</sup><sup>\*</sup>** and **ppy<sup>-</sup><sup>\*</sup>** relies on comparing the frontier orbitals of these triplet molecules with their respective singlet analogues (see Figure 2). By examining only the LUMOs of **ppyH<sup>+</sup>** and **ppy<sup>-</sup>**, it is difficult to imagine how the  $\pi$ -bonding HOMOs came about in forming **ppyH<sup>+</sup><sup>\*</sup>** and **ppy<sup>-</sup><sup>\*</sup>**. A closer look at the LUMO+1 orbitals of the singlet species, however, reveals the beginning of  $\pi$  overlap between the bridging carbons. The obvious  $\pi$ -bonding in the HOMOs of the triplet molecules is actually the result of the linear combination of the LUMO and LUMO+1 of the respective singlet species. As expected, this  $\pi$  bonding between the two aromatic rings explains a twist angle of zero for **ppyH<sup>+</sup><sup>\*</sup>** and **ppy<sup>-</sup><sup>\*</sup>**.

**Energetic Properties.** The energetic properties of the singlet state of **ppy** with respect to the open-shell species, **ppy<sup>+</sup>**, **ppy<sup>-</sup>**, and **ppy<sup>\*</sup>**, were computed at the B3LYP/6-31++G(2d,2p) level and may be found in Table 7. The adiabatic ionization potential,  $IP_{\text{ad}}$ , of **ppy** to form the radical cation, **ppy<sup>+</sup>**, is 8.2 eV and is not significantly different from the experimental ionization potential of 8.16 eV for the isoelectronic biphenyl molecule.<sup>36</sup> The vertical ionization potential,  $IP_{\text{vert}}$ , of **ppy** is 0.2 eV higher than the  $IP_{\text{ad}}$  value, and this difference is a measure of the structural relaxation energy for the ionization process. To form the radical anion, **ppy<sup>-</sup>**, the computed adiabatic electron affinity,  $EA_{\text{ad}}$ , of **ppy** is 0.3 eV. As indicated by the vertical electron affinity,  $EA_{\text{vert}}$ , the structural relaxation energy for creating this anion is 0.4 eV. Several benchmark computational investigations of adiabatic ionization potentials and electron affinities have concluded that DFT methods, similar to the one in this study, have an average error in accuracy of 0.1–0.2 eV.<sup>37–40</sup>

The adiabatic energy difference,  $\Delta E_{\text{ad}}$ , between the ground state of **ppy** and its corresponding triplet state is 2.8 eV. The larger value of 3.2 eV for  $\Delta E_{\text{TDDFT}}$  provides an upper bound for the  $S_0 \rightarrow T_1$  excitation energy and is indicative of the **ppy<sup>\*</sup>** structure being very different from the corresponding singlet structure (vide supra).

The energetic properties of the two charged, isoelectronic states of **ppy**, **ppyH<sup>+</sup>**, and **ppy<sup>-</sup>** at the B3LYP/6-311++G-

(2d,2p) level are presented in Table 5. Protonation of **ppy** is essentially the addition of  $H^+$  to the nitrogen of the pyridine ring. The computed proton affinity for creating **ppyH**<sup>+</sup> is 230 kcal/mol, which is very similar to the accepted experimental proton affinity of 222 kcal/mol for pyridine.<sup>41</sup> Deprotonation of **ppy** is analogous to deprotonation of benzene:  $H^+$  is removed from the phenyl ring, leaving behind a negatively charged carbon. To form **ppy**<sup>-</sup>, the computed enthalpy of deprotonation is 398 kcal/mol, which very closely agrees with the experimental enthalpy of deprotonation of 401.7 kcal/mol for benzene.<sup>42</sup>

From the data in Table 5, the energetic picture of the triplet energy surface for **ppyH**<sup>+</sup> relative to the singlet one is similar to that of **ppy** but very different from that of **ppy**<sup>-</sup>. Compared to **ppy**, the singlet–triplet energy difference,  $\Delta E_{ad}$ , for **ppyH**<sup>+</sup> is 0.2 eV lower at 2.6 eV. The upper bound for the  $S_0 \rightarrow T_1$  excitation energy of **ppyH**<sup>+</sup> is 2.9 eV and is 0.3 eV lower than the corresponding value for **ppy**. When **ppy** is in acidic media as **ppyH**<sup>+</sup>, the lowest phosphorescent peak is experimentally determined to be 2.89 eV.<sup>10</sup> This experimental energy value is closest to the computed  $\Delta E_{TDDFT}$  value of 2.9 eV, which may indicate that the molecular structure of the excited triplet state more closely resembles that of the singlet state before losing its energy in the emission process. The  $\Delta E_{ad}$  value of 0.8 eV between the singlet ground state of **ppy**<sup>-</sup> and its first excited triplet state is about 3 times smaller than the corresponding values for **ppy** and **ppyH**<sup>+</sup>.

## Conclusion

Density functional theory methods were used to investigate the structures and energetics associated with 2-phenylpyridine, **ppy**, and several of its electronic states. The structure of **ppy** has an inter-ring CC bond length of 1.49 Å, and the aromatic rings are twisted with respect to one another by  $\sim 21^\circ$ , which is about half the value found for biphenyl. The extent of twisting is governed by the delicate balance between  $\pi$  conjugation and repulsive orbital/steric interactions, and the magnitudes of these interactions were investigated by examining the torsional energy barriers. For coplanar rings, the torsional free energy barrier is only 1 kcal/mol, indicating that any repulsive interaction between the two rings is well-balanced by  $\pi$  conjugation. Whereas for perpendicular rings, the torsional free energy barrier is larger at 5 kcal/mol, most likely the result of rupturing any  $\pi$ -type bonding between the rings. The aforementioned metric parameters and energetics were computed at the B3LYP/6-311G(2d,2p) level, whereas the metric parameters and energetics summarized in the following paragraphs were computed at the B3LYP/6-311++G(2d,2p) level.

In comparison with **ppy**, the isoelectronic cation, **ppyH**<sup>+</sup>, has a larger twist angle ( $33^\circ$ ) which is explained by the extra set of ortho-hydrogen steric interactions. These repulsive interactions, however, are weak compared to  $\pi$  conjugation, as evidenced by the torsional free energy barriers at twist angles of  $0^\circ$  and  $90^\circ$  being 2 and 4 kcal/mol, respectively. The isoelectronic anion, **ppy**<sup>-</sup>, has an even larger twist angle of  $48^\circ$ , which is primarily due to the repulsive interaction between the lone pairs on the pyridine nitrogen and on the anionic carbon of the phenyl ring. This lone-pair repulsion

is significant enough to yield a torsional barrier of 3 kcal/mol for coplanar rings and to nearly counterbalance  $\pi$  conjugation as evidenced by a torsional barrier of 1 kcal/mol for perpendicular rings. When **ppy**<sup>-</sup> is used as an organometallic ligand, the lone-pair repulsion is obviously of no consequence because these electrons are now involved in bonding with the metal and do not interfere with coplanarization of the rings. As expected, the proton affinity (230 kcal/mol) and deprotonation energy (398 kcal/mol) of **ppy** are very similar to the experimental values for pyridine and benzene, respectively.

Every one of the investigated open-shell structures—**ppy**<sup>•+</sup>, **ppy**<sup>•-</sup>, **ppy**<sup>\*</sup>, **ppyH**<sup>•+</sup>, and **ppy**<sup>•-</sup>—has no twist angle between the aromatic rings. A frontier orbital analysis which compares each open-shell molecule with its closed-shell counterpart reveals that the  $\pi$ -type bonding between the bridging carbons becomes dominant over any repulsive orbital and steric interactions, thereby leading to coplanar rings. The predicted ionization potential (8.2 eV), electron affinity (0.3 eV), and singlet–triplet energy difference (2.8 eV) of **ppy** are of high enough quality to present a challenge for experimental work to reproduce. The energy required to form the triplet species, **ppyH**<sup>•+</sup>, from the ground-state singlet, **ppyH**<sup>+</sup>, is computed to be 2.6 eV, yet experimental phosphorescence studies have found the lowest triplet energy to be 2.89 eV, which is closer to the  $\Delta E_{TDDFT}$  value of 2.9 eV. Because UV–vis data are sensitive to molecular structure, the good agreement between  $\Delta E_{TDDFT}$  and experimental results may indicate that the molecular structure of the excited triplet state in the experiment more closely resembles that of the singlet state before losing its energy in the emission process. The very small singlet–triplet energy gap for **ppy**<sup>-</sup> (0.8 eV) indicates an easy transformation to **ppy**<sup>•-</sup> under applied voltage. When **ppy**<sup>-</sup> is used as an organometallic ligand, the overall planar system of the triplet species may be connected to the electroluminescent properties of the organometallic species, but further computational work is necessary to establish a firm connection.

**Supporting Information Available:** Optimized coordinates and corresponding total energies for all of the molecules (26 pages). This information is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Jenekhe, S. A. *Chem. Mater.* **2004**, *16*, 4381.
- (2) Bendikov, M.; Wudl, F.; Perepichka, D. F. *Chem. Rev.* **2004**, *104*, 4891.
- (3) Kottas, G. S.; Clarke, L. I.; Horinek, D.; Michl, J. *Chem. Rev.* **2005**, *105*, 1281.
- (4) Bredas, J.-L.; Beljonne, D.; Coropceanu, V.; Cornil, J. *Chem. Rev.* **2004**, *104*, 4971.
- (5) Grushin, V. V.; Herron, N.; LeCloux, D. D.; Marshall, W. J.; Petrov, V. A.; Wang, Y. *Chem. Commun.* **2001**, 1494.
- (6) Brooks, J.; Babayan, Y.; Lamansky, S.; Djurovich, P. I.; Tsyba, I.; Bau, R.; Thompson, M. E. *Inorg. Chem.* **2002**, *41*, 3055.

- (7) Silva, M. A. V. R. d.; Matos, M. A. R.; Rio, C. A.; Morais, V. M. F.; Wang, J.; Nichols, G.; Chickos, J. S. *J. Phys. Chem. A* **2000**, *104*, 1774.
- (8) Göller, A.; Grummt, U.-W. *Chem. Phys. Lett.* **2000**, *321*, 399.
- (9) Kubin, J.; Testa, A. C. *J. Photochem. Photobiol., A* **1994**, *83*, 91.
- (10) Sarkar, A.; Chakravorti, S. *J. Lumin.* **1995**, *65*, 163.
- (11) Pohlert, G.; Virdee, S.; Scaiano, J. C.; Sinta, R. *Chem. Mater.* **1996**, *8*, 2654.
- (12) André, J.-M.; Delhalle, J.; Brédas, J.-L. *Quantum Chemistry Aided Design of Organic Polymers: An Introduction to the Quantum Chemistry of Polymers and its Applications*; World Scientific: New Jersey, 1991; Vol. 2.
- (13) Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *52*, 997.
- (14) Bauernschmitt, R.; Ahlrichs, R. *Chem. Phys. Lett.* **1996**, *256*, 454.
- (15) Casida, M. E.; Jamorski, C.; Casida, K. C.; Salahub, D. R. *J. Chem. Phys.* **1998**, *108*, 4439.
- (16) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1998**, *109*, 8218.
- (17) Wiberg, K. B.; Oliveira, A. E. d.; Trucks, G. *J. Phys. Chem. A* **2002**, *106*, 4192.
- (18) Jacquemin, D.; Preat, J.; Wathelet, V.; Fontaine, M.; Perpete, E. A. *J. Am. Chem. Soc.* **2006**, *128*, 2072.
- (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 C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (20) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098.
- (21) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822.
- (22) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
- (23) Hertwig, R. H.; Koch, W. *Chem. Phys. Lett.* **1997**, *268*, 345.
- (24) Hariharan, P. C.; Pople, J. A. *Chem. Phys. Lett.* **1972**, *16*, 217.
- (25) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.
- (26) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.
- (27) Arulmozhiraja, S.; Fujii, T. *J. Chem. Phys.* **2001**, *115*, 10589.
- (28) Dunlap, B. I. *J. Chem. Phys.* **1983**, *78*, 3140.
- (29) Dunlap, B. I. *THEOCHEM* **2000**, *529*, 37.
- (30) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294.
- (31) Dobbs, K. D.; Dixon, D. A. *J. Phys. Chem.* **1996**, *100*, 3965.
- (32) Lee, S. Y. *Bull. Korean Chem. Soc.* **1998**, *19*, 93.
- (33) Lapouge, C.; Buntinx, G.; Poizat, O. *J. Mol. Struct.* **2003**, *651–653*, 747.
- (34) Figures 1 and 2 display only the  $\alpha$  orbitals of the open-shell molecules. The molecular orbitals were created with the GaussView program (see ref 35).
- (35) Dennington, R., II; Keith, T.; Millam, J.; Eppinnett, K.; Hovell, W. L.; Gilliland, R. *GaussView*, revision 3.09; Semichem, Inc.: Shawnee Mission, KS, 2003.
- (36) Lias, S. G. Ionization Energy Evaluation in NIST Chemistry WebBook. In *NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, 2005. <http://webbook.nist.gov> (accessed May 2006).
- (37) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42.
- (38) Oliveira, G. d.; Martin, J. M. L.; Proft, F. d.; Geerlings, P. *Phys. Rev. A: At., Mol., Opt. Phys.* **1999**, *60*, 1034.
- (39) Rienstra-Kiracofe, J. C.; Tschumper, G. S.; Schaefer, H. F., III; Nandi, S.; Ellison, G. B. *Chem. Rev.* **2002**, *102*, 231.
- (40) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2005**, *122*, 184107.
- (41) Hunter, E. P. L.; Lias, S. G. *J. Phys. Chem. Ref. Data* **1998**, *27*, 413.
- (42) Davico, G. E.; Bierbaum, V. M.; DePuy, C. H.; Ellison, G. B.; Squires, R. R. *J. Am. Chem. Soc.* **1995**, *117*, 2590.

## Nonlinearity of the Bifunctional of the Nonadditive Kinetic Energy: Numerical Consequences in Orbital-Free Embedding Calculations

Marcin Dułak and Tomasz A. Wesolowski\*

 Département de Chimie Physique, Université de Genève, 30, quai Ernest-Ansermet,  
CH-1211 Genève 4, Switzerland

Received July 24, 2006

**Abstract:** The bifunctional of the nonadditive kinetic energy in the reference system of noninteracting electrons ( $T_s^{\text{nad}}[\rho_A, \rho_B] = T_s[\rho_A + \rho_B] - T_s[\rho_A] - T_s[\rho_B]$ ) is the key quantity in orbital-free embedding calculations because they hinge on approximations to  $T_s^{\text{nad}}[\rho_A, \rho_B]$ . Since  $T_s^{\text{nad}}[\rho_A, \rho_B]$  is not linear in  $\rho_A$ , the associated potential (functional derivative)  $\delta T_s^{\text{nad}}[\rho, \rho_B]/\delta \rho|_{\rho=\rho_A}(\vec{r})$  changes if  $\rho_A$  varies. In this work, for two approximations to  $T_s^{\text{nad}}[\rho_A, \rho_B]$ , which are nonlinear in  $\rho_A$  (gradient-free and gradient-dependent), their linearized versions are constructed, and the resulting changes (*linearization errors*) in various properties of embedded systems (orbital energies, dipole moments, interaction energies, and electron densities) are analyzed. The considered model embedded systems represent typical nonbonding interactions: van der Waals contacts, hydrogen bonds, complexes involving charged species, and intermolecular complexes of the charge-transfer character. For van der Waals and hydrogen bonded complexes, the linearization of  $T_s^{\text{nad}}[\rho_A, \rho_B]$  affects negligibly the calculated properties. Even for complexes, for which large complexation induced changes of the electron density can be expected, such as the water molecule in the field of a cation, the linearization errors are about 2 orders of magnitude smaller than the interaction induced shifts of the corresponding properties. Linearization of  $T_s^{\text{nad}}[\rho_A, \rho_B]$  is shown to be inadequate for the complexes of a strong charge-transfer character. Compared to gradient-free approximation to  $T_s^{\text{nad}}[\rho_A, \rho_B]$ , introduction of gradients increases the linearization error.

### Introduction

The key quantity in orbital-free embedding calculations,<sup>1</sup> in which the subsystem of primary interest (*subsystem A*) is described at the orbital-level, whereas the environment of this subsystem is described using only its electron density ( $\rho_B$ ), is the bifunctional of the nonadditive kinetic energy

$$T_s^{\text{nad}}[\rho_A, \rho_B] = T_s[\rho_A + \rho_B] - T_s[\rho_A] - T_s[\rho_B] \quad (1)$$

where  $T_s[\rho]$  denotes the kinetic energy in the reference system of noninteracting electrons as defined in the Levy's constrained search.<sup>2</sup>

The functional derivative of  $T_s^{\text{nad}}[\rho_A, \rho_B]$  with respect to  $\rho_A$  is a component of the effective potential in Kohn–Sham-like one-electron equations<sup>1</sup> for embedded orbitals ( $\phi_i^A$ ), which are used to construct the embedded electron density ( $\rho_A = 2\sum_{i=1}^{N^A} |\phi_i^A|^2$ )

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{eff}}^{\text{KSCED}}[\rho_A, \rho_B; \vec{r}]\right]\phi_i^A = \epsilon_i^A \phi_i^A \quad i = 1, N^A \quad (2)$$

where  $2N^A$  is the number of electrons in the embedded subsystem. The label KSCED stands for Kohn–Sham Equations with Constrained Electron Density and is used here to indicate that ( $\{\phi_i^A\}$ ) are not the Kohn–Sham orbitals<sup>3</sup> and that the effective multiplicative potential in these equations is not the corresponding Kohn–Sham effective potential for neither the whole system nor the isolated subsystem A

\* Corresponding author e-mail: tomasz.wesolowski@chiphys.unige.ch.

( $V^{\text{KS}}[\rho_A; \vec{r}]$ ). All formulas are given in atomic units for the closed-shell case in this work. These units are also used in the discussion of numerical results except for dipole moments (in Debye, 1 D = 0.39343 e·Bohr) and orbital energies (in eV, 1 eV = 0.00367493 Hartree).

The effective potential in eq 2 has the following form

$$V_{\text{eff}}^{\text{KSCEd}}[\rho_A, \rho_B; \vec{r}] = V_{\text{eff}}^{\text{KS}}[\rho_A; \vec{r}] + V_{\text{eff}}^{\text{emb}}[\rho_A, \rho_B; \vec{r}], \quad (3)$$

where the part representing the environment reads

$$V_{\text{eff}}^{\text{emb}}[\rho_A, \rho_B; \vec{r}] = \sum_{i_B}^{N_{\text{nuc}}^B} -\frac{Z_{i_B}}{|\vec{r} - \vec{R}_{i_B}|} + \int \frac{\rho_B(\vec{r}')}{|\vec{r}' - \vec{r}|} d\vec{r}' + \left. \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho} \right|_{\rho=\rho_A+\rho_B} - \left. \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho} \right|_{\rho=\rho_A} + \left. \frac{\delta T_s^{\text{nad}}[\rho, \rho_B]}{\delta \rho} \right|_{\rho=\rho_A} \quad (4)$$

Various computational studies based on eq 2 have been reported recently.<sup>4–11</sup> The functional derivative of  $T_s^{\text{nad}}[\rho_A, \rho_B]$  is also used in the Cortona's formulation of density functional theory (DFT)<sup>12</sup> as well as in the wave function-in-DFT embedding approach by Carter, Wang, and collaborators.<sup>13</sup>

In the context of embedding, it is worthwhile to underline the qualitative difference between the electrostatic components (the first two terms) and the last three terms of the embedding potential in eq 4. The electrostatic components correspond to functionals which are linear in  $\rho_A$ . (The functional  $F[f]$  is linear if  $F[\alpha p + \beta q] = \alpha F[p] + \beta F[q]$ .) The terms depending on exchange-correlation- and kinetic energies, however, correspond to functionals which are not linear in  $\rho_A$ . For a recent discussion of deviation from linearity of the exact functional  $T[\rho]$ , which lies at the origin of nonlinearity of  $T_s^{\text{nad}}[\rho_A, \rho_B]$ , see refs 14 and 15 for instance. The common approximations to  $T[\rho]$  also lead to such an analytic expression for  $T_s^{\text{nad}}[\rho_A, \rho_B]$ , which is nonlinear in  $\rho_A$ . For instance, the regular gradient expansion truncated to zeroth order (Thomas-Fermi functional,<sup>16,17</sup> known also as the local density approximation (LDA)), leads to the following expression

$$\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B] = (3/10)(3\pi^2)^{2/3} \int ((\rho_A + \rho_B)^{5/3} - \rho_A^{5/3} - \rho_B^{5/3}) d\vec{r} \quad (5)$$

which is obviously nonlinear in  $\rho_A$ . Similarly, using the generalized gradient approximation (GGA) to  $T_s[\rho]$

$$T_s^{\text{GGA}}[\rho] = (3/10)(3\pi^2)^{2/3} \int \rho^{5/3}(\vec{r}) F(s(\vec{r})) d\vec{r}$$

where  $s = |\nabla \rho|/(2\rho k_F)$  with  $k_F = (3\pi^2\rho)^{1/3}$ , and  $F(s)$  is an analytic function of  $s$  which might originate from various types of considerations, leads to an analytic expression for  $T_s^{\text{nad}}[\rho_A, \rho_B]$  which is also nonlinear in  $\rho_A$ :

$$\tilde{T}_s^{\text{nad(GGA)}}[\rho_A, \rho_B] = (3/10)(3\pi^2)^{2/3} \int ((\rho_A + \rho_B)^{5/3} F(s_{AB}) - \rho_A^{5/3} F(s_A) - \rho_B^{5/3} F(s_B)) d\vec{r} \quad (6)$$

As a result, in any practical calculations the associated effective potential ( $\delta T_s^{\text{nad}}[\rho_A, \rho_B]/\delta \rho_A$ ) changes, if  $\rho_A$  varies.

From the practical point of view, it would be desirable to replace the  $\rho_A$ -dependent potential by a quantity which is not  $\rho_A$ -dependent. Such a simplification would make it possible to avoid updating the relevant component of the embedding potential during the self-consistent cycle and/or in cases where the geometry of the embedded subsystem changes. Neglecting such changes has been recognized in ref 13 as a possible additional approximation to the embedding potential of the eq 4 form.

Linearizing  $T_s^{\text{nad}}[\rho_A, \rho_B]$  in  $\rho_A$  around some reference ( $\rho_A^0$ ) provides a way to eliminate the dependency of  $\delta T_s^{\text{nad}}[\rho, \rho_B]/\delta \rho|_{\rho=\rho_A}(\vec{r})$  on  $\rho_A$ :

$$T_s^{\text{nad}}[\rho_A, \rho_B] \approx T_s^{\text{nad}}[\rho_A^0, \rho_B] + \int \left. \frac{\delta T_s^{\text{nad}}[\rho, \rho_B]}{\delta \rho} \right|_{\rho=\rho_A^0}(\vec{r})(\rho_A - \rho_A^0) d\vec{r}. \quad (7)$$

$\rho_A^0$  can be chosen to be the electron density of the isolated subsystem A.

The potential corresponding to the linearized  $T_s^{\text{nad}}[\rho_A, \rho_B]$  reads

$$\left. \frac{\delta T_s^{\text{nad}}[\rho, \rho_B]}{\delta \rho} \right|_{\rho=\rho_A}(\vec{r}) \approx \left. \frac{\delta T_s^{\text{nad}}[\rho, \rho_B]}{\delta \rho} \right|_{\rho=\rho_A^0}(\vec{r}) \quad (8)$$

The approximation of eq 8 applied in eq 2 affects obviously the embedded orbitals. As a consequence, molecular properties are affected by this approximation. For any observable, the difference between the results obtained from the linearized and nonlinearized versions of eq 2 are referred to as *linearization error* (LE) in this work. The approximation of eq 8 is referred to as *linearization approximation*.  $\rho_A^0$  and  $\rho_B$  are chosen as the Kohn–Sham ground-state electron densities of the isolated subsystems A and B. The local density approximation<sup>18–20</sup> to the exchange-correlation energy functional ( $E_{\text{xc}}[\rho]$ ) is used in all calculations.

The set of model embedded subsystems considered here (see Table 1) comprises molecules involved in common nonbonding interactions occurring in soft-condensed matter: van der Waals contacts, hydrogen bonded complexes, complexes involving charged species, and a representative intermolecular complex of a charge-transfer character.<sup>21</sup> They form a series, in which formation of the complex involves electron density deformation of increasing magnitude. In particular, the electron density deformation in the  $\text{NH}_3\text{--ClF}$  complex is known to be strong.<sup>21</sup> The adequacy of the *linearization approximation*, in which the electron density of the isolated molecule is used as  $\rho_A^0$  in eq 11, can be expected to decrease along this series. Note that homomolecular dimers occur twice in the tables because the two monomers are not equivalent.

The *linearization errors* of the following complexation induced properties: energies of interactions between the subsystems, dipole moments, orbital energies, and deformations of electron density are analyzed.

Unless specified, the discussed results are obtained using the aug-cc-pVTZ<sup>22,23</sup> basis set for all elements except for lithium (cc-pVTZ<sup>24,23</sup>). The atomic basis sets are centered on all atoms in the whole investigated complex including

**Table 1:** Errors in Orbital Energies ( $\delta\Delta\epsilon^A$ ) Arising from the Linearization of  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$  for the Highest Occupied Embedded Orbital (HOEO) and the Lowest Unoccupied Embedded Orbital (LUEO)<sup>c</sup>

subsystem		$\delta\Delta\epsilon_{\text{HOEO}}^A$	$\delta\Delta\epsilon_{\text{LUEO}}^A$	$\Delta\epsilon_{\text{HOEO}}^A$	$\Delta\epsilon_{\text{LUEO}}^A$
A	B				
Ne	Ne	0	-0.1	-10.9	-42.8
CH <sub>4</sub>	CH <sub>4</sub>	0.1	1.5	-2.1	-293.8
CH <sub>4</sub>	CH <sub>4</sub>	0.1	1.5	-2.1	-293.8
C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>	0.4	2.9	-58.1	-169.9
C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>	0.4	2.9	-58.1	-169.9
H <sub>2</sub> O	H <sub>2</sub> O <sup>a</sup>	-1.5	-2.2	508.6	168.4
H <sub>2</sub> O <sup>a</sup>	H <sub>2</sub> O	3.9	1.9	-644.9	-252.9
NH <sub>3</sub> <sup>a</sup>	H <sub>2</sub> O	7.0	1.9	-746.4	-223.0
H <sub>2</sub> O	NH <sub>3</sub> <sup>a</sup>	-2.2	-2.8	710.6	134.0
HF <sup>a</sup>	HF	3.6	3.3	-1037.7	-506.6
HF	HF <sup>a</sup>	-1.2	-2.1	361.5	141.5
NH <sub>3</sub>	CIF	183.7	596.0	-1253.7	-1129.2
CIF	NH <sub>3</sub>	-3.2	-19.0	817.1	1207.9
Li <sup>+</sup>	H <sub>2</sub> O	-0.5	0.6	1449.2	270.4
H <sub>2</sub> O	Li <sup>+</sup>	23.3	112.7	-7127.2	-6547.8

<sup>a</sup> Acceptor of the hydrogen bond. <sup>b</sup> Geometries of the complexes taken from refs 26 and 4 for Li<sup>+</sup>-H<sub>2</sub>O. <sup>c</sup> The complexation induced shifts of HOEO ( $\Delta\epsilon_{\text{HOEO}}^A$ ) and LUEO ( $\Delta\epsilon_{\text{LUEO}}^A$ ) are given for reference. All values in meV.

the embedded subsystem and its environment (KSCED(s) calculations according to the nomenclature of ref 25) for equilibrium geometries of each dimer.<sup>26</sup> In a dedicated section concerning the role of the size of the atomic basis sets on the evaluated errors the following series of basis sets is used: aug-cc-pVDZ, aug-cc-pVTZ, aug-cc-pVQZ, and aug-cc-pV5Z.<sup>22,23</sup> Additionally, errors due to linearization of  $T_s^{\text{nad}}[\rho_A, \rho_B]$  in  $\rho_A$  have been analyzed for exchange-correlation energy functional of the GGA type (Perdew and Wang<sup>27,28</sup>). All the calculations are performed on the “(99,590)p” pruned grid using 10<sup>-9</sup> self-consistent cycle convergence criterion and the GEN-A4\* auxiliary basis set.<sup>29</sup>

Two levels of approximation to  $T_s^{\text{nad}}[\rho_A, \rho_B]$  and the associated functional derivative are considered: (a) local density approximation  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$ <sup>1</sup>, defined in eq 8, and (b)  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$ <sup>25</sup>, defined in eq 9, in which the function  $F(s)$  has the Lembarki-Chermette (LC94) form<sup>30</sup>

$$F^{\text{LC94}}(s) = \frac{1 + 0.093907s \operatorname{arcsinh}(76.32s) + (0.26608 - 0.0809615e^{-100s^2})s^2}{1 + 0.093907s \operatorname{arcsinh}(76.32s) + 0.57767 \cdot 10^{-4}s^4}$$

Tildas used above indicate that the considered functionals are not exact.

The GGA97 approximation to  $T_s^{\text{nad}}[\rho_A, \rho_B]$  is considered here in addition to  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$  because of its use in most of our own studies applying eq 2 for embedded subsystems where  $\rho_A$  and  $\rho_B$  do not overlap strongly. The GGA97 choice is motivated by the fact that the associated functional derivative was shown to be the most accurate among several gradient-dependent approximations<sup>25,31</sup> in the case of small overlaps between  $\rho_A$  and  $\rho_B$ .

## Results and Discussions

**Orbital Energies.** Table 1 collects the linearization errors in the energies of the highest occupied (HOEO) and

**Table 2:** Errors in Orbital Energies ( $\delta\Delta\epsilon^A$ ) Arising from the Linearization of  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$  for the Highest Occupied Embedded Orbital (HOEO) and the Lowest Unoccupied Embedded Orbital (LUEO)<sup>c</sup>

subsystem		$\delta\Delta\epsilon_{\text{HOEO}}^A$	$\delta\Delta\epsilon_{\text{LUEO}}^A$	$\Delta\epsilon_{\text{HOEO}}^A$	$\Delta\epsilon_{\text{LUEO}}^A$
A	B				
Ne	Ne	0	0	-2.9	82.2
CH <sub>4</sub>	CH <sub>4</sub>	0	0.7	16.6	-252.0
CH <sub>4</sub>	CH <sub>4</sub>	0	0.7	16.6	-252.0
C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>	0.2	1.9	-39.4	-143.7
C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>	0.2	1.9	-39.4	-143.7
H <sub>2</sub> O	H <sub>2</sub> O <sup>a</sup>	-2.9	-3.7	551.4	216.4
H <sub>2</sub> O <sup>a</sup>	H <sub>2</sub> O	3.5	1.4	-589.1	-220.0
NH <sub>3</sub> <sup>a</sup>	H <sub>2</sub> O	6.0	1.1	-669.8	-196.6
H <sub>2</sub> O	NH <sub>3</sub> <sup>a</sup>	-4.1	-4.2	756.5	174.1
HF <sup>a</sup>	HF	3.8	2.7	-993.4	-468.8
HF	HF <sup>a</sup>	-2.4	-4.3	410.2	212.5
NH <sub>3</sub>	CIF	89.8	173.6	-979.7	-603.7
CIF	NH <sub>3</sub>	-6.3	-25.0	884.3	1333.5
Li <sup>+</sup>	H <sub>2</sub> O	-1.8	-0.8	1535.4	398.8
H <sub>2</sub> O	Li <sup>+</sup>	20.4	82.9	-7077.2	-6568.9

<sup>a</sup> Acceptor of the hydrogen bond. <sup>b</sup> Geometries of the complexes taken from refs 26 and 4 for Li<sup>+</sup>-H<sub>2</sub>O. <sup>c</sup> The complexation induced shifts of HOEO ( $\Delta\epsilon_{\text{HOEO}}^A$ ) and LUEO ( $\Delta\epsilon_{\text{LUEO}}^A$ ) are given for reference. All values in meV.

lowest unoccupied embedded orbitals (LUEO) in the  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$  case. For molecules involved in van der Waals complexes or hydrogen bonds, LE in orbital energies is so small (meV range) that it can be neglected in almost any discussion of chemical relevance. Typically, LEs are 2 orders of magnitude smaller than the complexation induced shifts of orbital energies. The complexation induced shift of the orbital energy for the subsystem A ( $\Delta\epsilon^A$ ) is defined as the difference between the energies of the corresponding orbitals in the embedded- and free subsystem A ( $\Delta\epsilon^A = \epsilon^A - \epsilon^{A0}$ ). For H<sub>2</sub>O in Li<sup>+</sup>-H<sub>2</sub>O, numerical values of LE are significantly larger (0.1 eV range); they are, however, still about 2 orders of magnitude smaller than the total complexation induced shifts of this property. Such small relative errors in this case might even seem surprising, taking into account the fact that the cation polarizes strongly the water molecule in the complex and the adequacy of using  $\rho_A^0$  in eq 8 is not evident. LEs amount to less than 10% of the complexation induced shifts in the orbital energies in this case. For NH<sub>3</sub> in the NH<sub>3</sub>-CIF complex, however, LE in the energy of LUEO reaches 0.6 eV which is unacceptable because this error represents about 50% of the magnitude of complexation induced shifts of HOEO or LUEO. This failure of the *linearization approximation* of eq 8 could be expected because the NH<sub>3</sub>-CIF complex is known for its charge-transfer character.<sup>21</sup>

It is worthwhile to notice that LEs in the energies of unoccupied orbitals are larger than the ones of occupied orbitals in almost all cases. The fact that *linearization approximation* affects more unoccupied orbitals than occupied ones, indicates that linearizing  $\tilde{T}_s^{\text{nad}}[\rho_A, \rho_B]$  should be applied with a proper care in such applications of the orbital-free embedding potential which aim at the energies of the electronic excitations.<sup>32,33</sup>



**Table 3:** Relative Errors in the Interaction Energy ( $\delta E_{\text{int}}^A$ / $E_{\text{int}}^A$  in %), Arising from the Linearization of  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$  and  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]^c$ 

subsystem		GGA97		LDA	
A	B	$\delta E_{\text{int}}^A/E_{\text{int}}^A$	$E_{\text{int}}^A$	$\delta E_{\text{int}}^A/E_{\text{int}}^A$	$E_{\text{int}}^A$
Ne	Ne	0	-518	0	-130
CH <sub>4</sub>	CH <sub>4</sub>	0	-2524	0	-683
CH <sub>4</sub>	CH <sub>4</sub>	0	-2524	-0.01	-683
C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>	-0.01	-4521	-0.01	-2445
C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>	-0.01	-4521	-0.01	-2445
H <sub>2</sub> O	H <sub>2</sub> O <sup>a</sup>	-0.04	-11487	-0.11	-6492
H <sub>2</sub> O <sup>a</sup>	H <sub>2</sub> O	-0.09	-11620	-0.10	-6012
NH <sub>3</sub> <sup>a</sup>	H <sub>2</sub> O	-0.15	-14662	-0.14	-8053
H <sub>2</sub> O	NH <sub>3</sub> <sup>a</sup>	-0.05	-14530	-0.15	-8672
HF <sup>a</sup>	HF	-0.14	-10142	-0.19	-5112
HF	HF <sup>a</sup>	-0.02	-9010	-0.08	-4558
NH <sub>3</sub>	CIF	-28.18	-23024	-122.35	-1423
CIF	NH <sub>3</sub>	-0.25	-19437	-6.42	-1003
Li <sup>+</sup>	H <sub>2</sub> O	0	-46752	0	-41423
H <sub>2</sub> O	Li <sup>+</sup>	-0.41	-72880	-0.31	-65132

<sup>a</sup> Acceptor of the hydrogen bond. <sup>b</sup> Geometries of the complexes taken from refs 26 and 4 for Li<sup>+</sup>-H<sub>2</sub>O. <sup>c</sup> The interaction energies ( $E_{\text{int}}^A$  in  $\mu$ Hartree) are given for reference.

Usually, LEs in orbital energies are smaller for  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$  than for  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$  (compare Tables 1 and 2).

**Interaction Energies.** Once the embedded orbitals ( $\{\phi_i^A\}$ ) are obtained from eq 2, the interaction energy between subsystem A and its environment (subsystem B) reads

$$\begin{aligned}
 E_{\text{int}}^A &= E^{\text{sc}}[\{\phi_i^A\}, \rho_B] - E^{\text{KS}}[\rho_A^0] - E^{\text{KS}}[\rho_B] \\
 &= V[\rho_A + \rho_B] + J[\rho_A + \rho_B] + E_{\text{xc}}[\rho_A + \rho_B] \\
 &\quad + T_s^{\text{nad}}[\rho_A, \rho_B] + 2 \sum_{i=1}^{N_A} \left\langle \phi_i^A \left| -\frac{1}{2} \nabla^2 \right| \phi_i^A \right\rangle + T_s[\rho_B] \\
 &\quad + E_{\text{NN}}^{\text{AB}} - E^{\text{KS}}[\rho_A^0] - E^{\text{KS}}[\rho_B] \quad (9)
 \end{aligned}$$

where  $E^{\text{KS}}[\rho_A^0]$  and  $E^{\text{KS}}[\rho_B]$  denote the Kohn–Sham energy functionals for isolated subsystems A and B, whereas  $E_{\text{NN}}^{\text{AB}}$  is the energy of repulsion between nuclei in the total system,  $V[\rho_A + \rho_B]$  is the functional of the nuclear attraction energy (total), and  $J[\rho_A + \rho_B]$  is the functional of the electron–electron Coulomb repulsion (total). Note that in the numerical evaluation of  $E_{\text{int}}^A$ , some components of the energy are not calculated at all because they cancel each other (for instance the kinetic energy contribution  $T_s$  to  $E^{\text{KS}}[\rho_B]$  cancels the  $T_s[\rho_B]$  term).

Table 3 collects the linearization errors in the energy of interaction between the embedded subsystem and its environment for  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$  and  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$ . Linearization of  $\tilde{T}_s^{\text{nad}}[\rho_A, \rho_B]$  is very adequate as it affects the interaction energies by a small amount (range of  $\mu$ Hartrees) in all cases except for the charge-transfer complex where it reaches 649  $\mu$ Hartree (for the GGA97 functional). As discussed previously, the charge-transfer case lies outside of the domain of applicability of the *linearization approximation*. Among other systems, the largest LE occurs for water interacting with Li<sup>+</sup> where it reaches 30  $\mu$ Hartree (again for the GGA97 functional). The GGA97 interaction

**Table 4:** Errors in the Complexation Induced Dipole Moments ( $\delta \Delta \mu^A$ ), Arising from the Linearization of  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]^c$ 

subsystem		$\delta \Delta \mu_x^A$	$\delta \Delta \mu_y^A$	$\delta \Delta \mu_z^A$	$\Delta \mu_x^A$	$\Delta \mu_y^A$	$\Delta \mu_z^A$
Ne	Ne	0	0	0	0	0	5.0
Ne	Ne	0	0	0	0	0	-5.0
CH <sub>4</sub>	CH <sub>4</sub>	0	0	-0.3	0	0	30.7
CH <sub>4</sub>	CH <sub>4</sub>	0	0	0.3	0	0	-30.7
C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>	0.4	-1.0	0	-122.0	55.1	0
C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>	-0.4	1.0	0	122.0	-55.1	0
H <sub>2</sub> O	H <sub>2</sub> O <sup>a</sup>	2.5	0	-0.2	-209.7	-3.2	41.5
H <sub>2</sub> O <sup>a</sup>	H <sub>2</sub> O	4.3	0	0.2	-258.4	4.7	-64.8
NH <sub>3</sub> <sup>a</sup>	H <sub>2</sub> O	7.9	-0.6	0	-337.1	75.0	0.3
H <sub>2</sub> O	NH <sub>3</sub> <sup>a</sup>	3.7	0.2	0	-289.5	-19.3	-0.1
HF <sup>a</sup>	HF	-4.0	-0.3	0	276.0	58.9	0
HF	HF <sup>a</sup>	-1.3	0.1	0	101.9	-37.0	0
NH <sub>3</sub>	CIF	0	0	287.6	0	0	-983.1
CIF	NH <sub>3</sub>	0	0	12.4	0	0	-675.1
Li <sup>+</sup>	H <sub>2</sub> O	0	0	0	0	0	-13.1
H <sub>2</sub> O	Li <sup>+</sup>	0	0	32.5	0	0	-1657.0

<sup>a</sup> Acceptor of the hydrogen bond. <sup>b</sup> Geometries of the complexes taken from refs 26 and 4 for Li<sup>+</sup>-H<sub>2</sub>O. <sup>c</sup> The total complexation induced dipole moments ( $\Delta \mu^A$ ) are given for reference. All values in mDebye.

energies are larger in absolute values than the LDA ones, and the effect of linearization of  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$  is systematically slightly larger than it is for  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$ .

For practical purposes, it is important that LE in energy does not vary with the changing geometry. To this end, LEs were analyzed along the LDA dissociation energy curve for Li<sup>+</sup>-H<sub>2</sub>O. As indicated in previous sections the *linearization approximation* is still applicable for Li<sup>+</sup>-H<sub>2</sub>O despite the strong polarization of water due to the electric field of the cation. A larger basis set (aug-cc-pVQZ for O and H, and cc-pVQZ for Li) than the one used in calculations discussed so far and  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$  were applied. LE remain small in the whole range of intermolecular distances. At an equilibrium intermolecular distance of 1.80 Å, LE in the interaction energy amounts to 25  $\mu$ Hartree, and it does not exceed 80  $\mu$ Hartree even at intermolecular distances as short as 0.7 Å. LEs are decreasing with increasing intermolecular distance, which reflects the fact that the exact  $T_s^{\text{nad}}[\rho_A, \rho_B]$  functional disappears for nonoverlapping  $\rho_A$  and  $\rho_B$ , and this asymptotic condition is satisfied also by  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$  and  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$  considered in this work. Therefore, the contribution of the kinetic energy terms decreases with increasing intersystem distance.

**Dipole Moments.** The *linearization errors* in dipole moments are collected in Tables 4 and 5. Except for the charge-transfer complex, linearization of  $\tilde{T}_s^{\text{nad}}[\rho_A, \rho_B]$  is very adequate in all embedded systems. For  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$ , LEs in dipole moments are small and do not exceed 6% of the total complexation induced dipole moments. Opposite, to the previously analyzed observables, the complexation induced dipole moments depend strongly on the choice of the approximation for  $T_s^{\text{nad}}[\rho_A, \rho_B]$ .

**Complexation Induced Density Deformations.** Orbital energies, interaction energies, and dipole moments are global

**Table 5:** Errors in the Complexation Induced Dipole Moments ( $\delta\Delta\mu^A$ ), Arising from the Linearization of  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]^c$ 

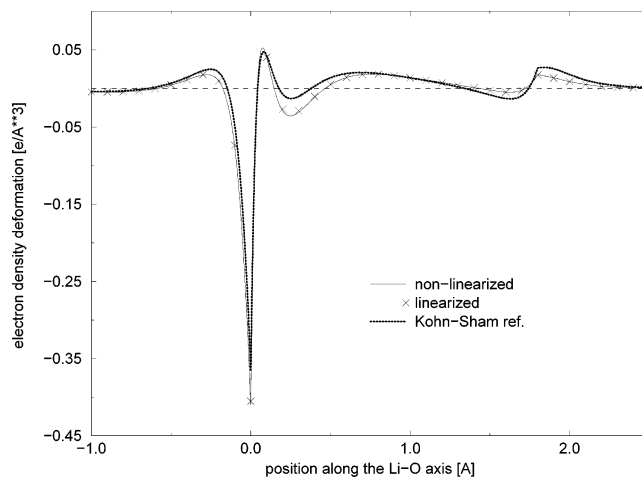
subsystem		$\delta\Delta\mu_x^A$	$\delta\Delta\mu_y^A$	$\delta\Delta\mu_z^A$	$\Delta\mu_x^A$	$\Delta\mu_y^A$	$\Delta\mu_z^A$
Ne	Ne	0	0	0	0	0	0.7
Ne	Ne	0	0	0	0	0	-0.7
CH <sub>4</sub>	CH <sub>4</sub>	0	0	0	0	0	-6.3
CH <sub>4</sub>	CH <sub>4</sub>	0	0	0	0	0	6.3
C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>	0.4	-0.6	0	-117.2	10.9	0
C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>	-0.4	0.6	0	117.2	-10.9	0
H <sub>2</sub> O	H <sub>2</sub> O <sup>a</sup>	4.3	0	-0.4	-262.7	-3.3	43.9
H <sub>2</sub> O <sup>a</sup>	H <sub>2</sub> O	3.8	0	0.3	-206.4	4.6	-65.4
NH <sub>3</sub> <sup>a</sup>	H <sub>2</sub> O	6.3	-0.6	0	-264.9	72.3	0.3
H <sub>2</sub> O	NH <sub>3</sub> <sup>a</sup>	6.1	0.4	0	-344.1	-21.6	-0.1
HF <sup>a</sup>	HF	-3.9	-0.3	0	238.1	58.1	0.0
HF	HF <sup>a</sup>	-2.5	0.2	0	145.9	-38.0	0.1
NH <sub>3</sub>	CIF	0	0	134.2	0	0	-675.8
CIF	NH <sub>3</sub>	0	0	18.1	0	0	-795.8
Li <sup>+</sup>	H <sub>2</sub> O	0	0	0.2	0	0	-19.7
H <sub>2</sub> O	Li <sup>+</sup>	0	0	27.9	0	0	-1596.0

<sup>a</sup> Acceptor of the hydrogen bond. <sup>b</sup> Geometries of the complexes taken from refs 26 and 4 for Li<sup>+</sup>-H<sub>2</sub>O. <sup>c</sup> The total complexation induced dipole moments ( $\Delta\mu^A$ ) are given for reference. All values in mDebye.

quantities for the embedded system. Any change in the effective potential such as linearization of  $T_s^{\text{nad}}[\rho_A, \rho_B]$  might affect them less than local quantities such as electron density perturbation:  $\Delta\rho_A = \rho_A - \rho_A^0$ . Analyses made in previous sections indicate that Li<sup>+</sup>-H<sub>2</sub>O is the one among the analyzed systems, for which LEs are the largest being sufficiently small that the *linearization approximation* is fully applicable. For this system, large changes of electron density can be expected due to the polarization of the water molecule (subsystem A) by the positive charge of the cation. Indeed, electron density deformation derived from conventional Kohn-Sham calculations is the largest in the vicinity of the oxygen atom (see Figure 1). The embedding calculations, without linearization of  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$  lead to electron density deformation which follows closely the Kohn-Sham trends. It indicates that  $\delta\tilde{T}_s^{\text{nad(LDA)}}[\rho, \rho_B]/\delta\rho|_{\rho=\rho_A}(\vec{r})$  is a very good approximation to the exact potential  $\delta T_s^{\text{nad}}[\rho, \rho_B]/\delta\rho|_{\rho=\rho_A}(\vec{r})$ . Interestingly, linearization of  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B](\vec{r})$  does not lead to any noticeable effect on complexation induced density deformations. LE is so small in this case that it is almost not visible on the figure. In fact, the largest LEs do not exceed  $5 \times 10^{-4} \text{ e}\text{\AA}^{-3}$ , i.e., are between 1 or 2 orders of magnitude smaller than the complexation induced density deformations.

It is worthwhile to notice that LEs are much smaller than the errors of  $\delta\tilde{T}_s^{\text{nad(LDA)}}[\rho, \rho_B]/\delta\rho|_{\rho=\rho_A}(\vec{r})$  which are responsible for the deviations from the reference Kohn-Sham data.

**Nonlinearity of  $T_s^{\text{nad}}[\rho_A, \rho_B]$ : The Effect of Changing the Basis Sets and Approximations for  $E_{\text{xc}}[\rho]$ .** Increasing the flexibility of the used basis sets and changing the approximation for the exchange-correlation energy functional provides another possibility to detect flaws of the *linearization approximation* of eq 8. Similar analyses as the ones discussed in the previous sections are made here using a



**Figure 1.** Complexation induced deformation of the electron density of H<sub>2</sub>O in Li<sup>+</sup>-H<sub>2</sub>O calculated using linearized and nonlinearized  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$ . The oxygen atom is situated at  $z = 0.0 \text{ \AA}$  and the lithium at  $z = 1.8 \text{ \AA}$ . The deformation of the total electron density derived from the LDA Kohn-Sham calculations is given for reference.

series of atomic basis sets of increasing completeness starting from aug-cc-pVDZ until aug-cc-pV5Z. Five representative embedded systems for which previous calculations indicate the adequacy of the linearization of  $T_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$  are chosen for this analysis: (i) Ne in the Ne<sub>2</sub> dimer, (ii) HF (donor of the hydrogen bond) in the (HF)<sub>2</sub> dimer, (iii) HF (acceptor of the hydrogen bond) in the (HF)<sub>2</sub> dimer, (iv) H<sub>2</sub>O (donor of the hydrogen bond) in the (H<sub>2</sub>O)<sub>2</sub> dimer, and (v) H<sub>2</sub>O (acceptor of the hydrogen bond) in the (H<sub>2</sub>O)<sub>2</sub> dimer.

LEs in the interaction energy are lower or equal than 1  $\mu$ Hartree for all five systems, using all the basis sets. As far as LEs in the orbital energies are concerned, they lie well below the 5 meV threshold. It is worthwhile to note that such small errors are negligible compared to the variability of the complexation induced shifts in the orbital energies calculated using various basis sets.

The negligible variation of LEs of all investigated properties upon the changes of the basis set results probably from the fact that embedded orbitals are constructed using *all* atomic centers including the atoms of the environment. Therefore, even at the aug-cc-pVDZ level  $\delta T_s^{\text{nad}}[\rho, \rho_B]/\delta\rho|_{\rho=\rho_A}(\vec{r})$  is adequately represented around atoms of the environment.

The change of approximation for the exchange-correlation energy from LDA to GGA does not influence the observed trends concerning the errors due to linearization of  $T_s^{\text{nad}}[\rho_A, \rho_B]$  in  $\rho_A$ . For instance, for PW91 approximation the largest LE of the interaction energy is found for NH<sub>3</sub>-CIF and reaches 532  $\mu$ Hartree (for the GGA97 functional, using the aug-cc-pVTZ basis set).

## Conclusions

The linearization of  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$  and  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$  using the Kohn-Sham ground-state electron density of the isolated subsystem A in eq 8 leads to negligible numerical effects on properties for all (except for the NH<sub>3</sub>-CIF complex) studied systems. The largest linearization errors

are 2 orders of magnitude smaller than the complexation induced shifts in molecular properties in such cases. Among the analyzed properties, the energies of unoccupied orbitals are typically the most affected by the linearization of  $T_s^{\text{nad}}[\rho_A, \rho_B]$  in  $\rho_A$ . Among these cases where the *linearization approximation* is adequate the largest *linearization error* occurs for the energy of the lowest unoccupied embedded orbital in the case of the water molecule in the  $\text{Li}^+ - \text{H}_2\text{O}$  complex. Linearization of  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$  results in the change of the orbital energy in the range of 0.1 eV which is still about 60 times smaller than the whole complexation induced energy shift of this orbital.

In the  $\text{NH}_3 - \text{ClF}$  case, even larger deformations of electron density accompany the formation of the complex. For this system using the electron densities of isolated monomers to linearize  $T_s^{\text{nad}}[\rho_A, \rho_B]$  was shown not to be adequate.

The linearization in  $\rho_A$  leads usually to smaller effects in the case of  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$  than  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$ . Since neither  $\tilde{T}_s^{\text{nad(LDA)}}[\rho_A, \rho_B]$  nor  $\tilde{T}_s^{\text{nad(GGA97)}}[\rho_A, \rho_B]$  are exact, this result indicates that the adequacy of linearization should be checked if intended to be applied to any new type of approximation for  $T_s^{\text{nad}}[\rho_A, \rho_B]$ .

Our calculations show that the linearization error does not vary significantly with changing atomic basis sets within the aug-cc-pVXZ family and that similar linearization errors are obtained using the LDA and the GGA (PW91) exchange-correlation functionals.

**Acknowledgment.** This work was supported by Swiss National Science Foundation.

### References

- (1) Wesolowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 8050–8053.
- (2) Levy, M. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 6062–6065.
- (3) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- (4) Stefanovich, E. V.; Truong, T. N. *J. Chem. Phys.* **1996**, *104*, 2946–2955.
- (5) Mei, W. N.; Boyer, L. L.; Mehl, M. J.; Ossowski, M. M.; Stokes, H. T. *Phys. Rev. B* **2000**, *61*, 11425–11431.
- (6) Trail, J. R.; Bird, D. M. *Phys. Rev. B* **2000**, *62*, 16402–16411.
- (7) Shimojo, F.; Kalia, R. K.; Nakano, A.; Vashishta, P. *Comput. Phys. Commun.* **2005**, *167*, 151–164.
- (8) Neugebauer, J.; Louwerson, M. J.; Baerends, E. J.; Wesolowski, T. A. *J. Chem. Phys.* **2005**, *122*, 094115.
- (9) Olsson, M. H. M.; Hong, G. Y.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025–5039.
- (10) Cholý, N.; Lu, G.; E, W.; Kaxiras, E. *Phys. Rev. B* **2005**, *71*, 094101.
- (11) Iannuzzi, M.; Kirchner, B.; Hutter, J. *Chem. Phys. Lett.* **2006**, *421*, 16–20.
- (12) Cortona, P. *Phys. Rev. B* **1991**, *44*, 8454–8458.
- (13) Kluner, T.; Govind, N.; Wang, Y. A.; Carter, E. A. *J. Chem. Phys.* **2002**, *116*, 42–54.
- (14) Chan, G. K. L.; Handy, N. C. *Phys. Rev. A* **1999**, *59*, 2670–2679.
- (15) Gal, T. *Phys. Rev. A* **2001**, *6406*, 062503.
- (16) Thomas, L. H. *Proc. Cambridge Philos. Soc.* **1927**, *23*, 542.
- (17) Fermi, E. *Z. Phys.* **1928**, *48*, 73.
- (18) Dirac, P. A. M. *Proc. Cambridge Philos. Soc.* **1930**, *26*, 376–385.
- (19) Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.* **1980**, *45*, 566–569.
- (20) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- (21) Ruiz, E.; Salahub, D. R.; Vela, A. *J. Phys. Chem.* **1996**, *100*, 12265–12276.
- (22) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (23) Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 02/25/04, as developed and distributed by the *Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory* which is part of the *Pacific Northwest Laboratory*, P.O. Box 999, Richland, WA 99352, U.S.A., and funded by the *U.S. Department of Energy*. The *Pacific Northwest Laboratory* is a multiprogram laboratory operated by *Battelle Memorial Institute* for the *U.S. Department of Energy* under contract DE-AC06-76RLO 1830. Contact *David Feller* or *Karen Schuchardt* for further information. <http://www.emsl.pnl.gov/forms/basisform.html> (accessed May 25, 2004).
- (24) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (25) Wesolowski, T. A. *J. Chem. Phys.* **1997**, *106*, 8516–8526.
- (26) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415–432.
- (27) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671–6687.
- (28) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1993**, *48*, 4978–4978.
- (29) Köster, A. M.; Calaminici, P.; Escalante, S.; Flores-Moreno, R.; Goursot, A.; Patchkovskii, S.; Reveles, J. U.; Salahub, D. R.; Vela, A.; Heine, T. *The deMon User's Guide, Version 1.0.3, 2003–2004*. <http://www.deMon-software.com/> (accessed Sep 8, 2006).
- (30) Lembarki, A.; Chermette, H. *Phys. Rev. A* **1994**, *50*, 5328–5331.
- (31) Wesolowski, T. A.; Chermette, H.; Weber, J. *J. Chem. Phys.* **1996**, *105*, 9182–9190.
- (32) Casida, M. E.; Wesolowski, T. A. *Int. J. Quantum Chem.* **2004**, *96*, 577–588.
- (33) Wesolowski, T. A. *J. Am. Chem. Soc.* **2004**, *126*, 11444–11445.

CT600241Q

## Elimination of Translational and Rotational Motions in Nuclear Orbital Plus Molecular Orbital Theory: Contribution of the First-Order Rovibration Coupling

Kaito Miyamoto, Minoru Hoshino, and Hiromi Nakai\*

*Department of Chemistry, School of Science and Engineering, Waseda University,  
Tokyo 169-8555, Japan*

Received June 20, 2006

**Abstract:** The translation- and rotation-free nuclear orbital plus molecular orbital (TRF–NOMO) theory was developed to determine nonadiabatic nuclear and electronic wave functions. This study implemented a computational program for the TRF–NOMO method including first-order rotational terms, which corresponds to rovibronic coupling. Numerical assessments of first-order TRF–NOMO Hartree–Fock as well as second-order Møller–Plesset perturbation methods were carried out for several small molecules. The first-order contributions give small corrections in energy. Thus, we confirm that the approximate zeroth-order treatment is sufficient for eliminating the rotational contamination.

### 1. Introduction

Born–Oppenheimer (BO) approximation,<sup>1</sup> which divides nuclear and electronic motions, is a fundamental concept of modern molecular theory. Although BO treatment is suitable for accurately describing various chemical and physical phenomena, it cannot take into account coupling between the nuclear and electronic motions, that is, the non-BO effect.

Adamowicz and co-workers<sup>2–5</sup> proposed the non-BO theory by utilizing the explicit correlated Gaussian (ECG) basis functions, which involve the internal coordinates among nuclei and electrons. The ECG approach has been shown to achieve considerably high accuracy, that is, spectroscopic accuracy. The problem with this approach is that the complexity of the explicit formula increases with an increase of the number ( $N$ ) of identical particles, which requires different programming codes for different  $N$  values. Furthermore, the computational cost increases very rapidly ( $N$  factorial).

On the other hand, we have developed the nuclear orbital plus molecular orbital (NOMO) theory,<sup>6–12</sup> which determines nuclear and electronic wave functions simultaneously without BO approximation. In NOMO theory, total wave function is constructed by nuclear orbitals (NOs), one-nucleus wave functions, and molecular orbitals (MOs), one-electron ones. We have proposed that it is convenient to adopt Gaussian basis functions with the center on each atomic position in

the appropriate molecular configuration, such as equilibrium and optimized ones.<sup>6–12</sup> The use of Gaussian basis functions has been accepted by other groups<sup>13–28</sup> probably because it is consistent with conventional MO theory within BO approximation. While Gaussian basis functions can describe a vibrational state accurately, translational and rotational states are not adequately reproduced because their motions are limited in some local regions represented by the functions. Thus, we have presented translation-free (TF)– and translation- and rotation-free (TRF)–NOMO theories and clarified the importance of eliminating translational and rotational contaminations in obtaining accurate results in NOMO and similar approaches.<sup>7,10–12</sup>

Sutcliffe has pointed out that it is possible to rigorously construct a TF–NOMO Hamiltonian, but it is impossible to rigorously construct a TRF–NOMO Hamiltonian for general systems.<sup>29</sup> This distinction arises because translations are separable from rotations and vibrations, but rotations and vibrations are coupled for general systems. The essential problems Sutcliffe points out undoubtedly exist, and the TRF treatment for nonrigid rotator systems cannot rigorously succeed in general cases, nor can it do so in the NOMO theory. We have focused on the locality of the Gaussian functions, of which the orbital centers can approximately define the rigid-body rotator. It is possible to define center-of-mass (COM), angular, and internal coordinates uniquely for the rigid-body rotator, while it is impossible for the general case where rotational and vibrational motions couple. Thus, the

\* Corresponding author fax: +81-3-3205-2504; e-mail: nakai@waseda.jp.

rotational operator has been expanded in a Taylor series with respect to the displacement  $\Delta\mathbf{x}$  based on the rigid-body rotator. As a result, the nuclear wave function represented by Gaussian basis functions can be separated into zeroth-order rigid-body rotation and higher-order coupling. TRF–NOMO theory adopts this unique definition of the COM, angular, and internal coordinates for the zeroth-order rotator.<sup>10</sup>

In a previous study, however, we implemented a programming code for the TRF–NOMO method corresponding only to the zeroth-order terms of the rotational Hamiltonian and numerically tested their contributions. Therefore, the assessment of the higher-order contribution is of great importance for investigating the reliability of the zeroth-order treatment and, furthermore, the validity of the TRF–NOMO formalism itself, on the basis of the Taylor expansion of the rotational operator.

The purpose of the present study is to implement the computational program for the TRF–NOMO method involving the first-order rotational terms, which are much more complicated than the zeroth-order terms, and the numerical assessment of their contribution. The organization of this paper is as follows. First, section 2 describes the theoretical aspects of the first-order TRF–NOMO method. Section 3 indicates the implementation of this method. In section 4, we present the numerical assessments of the present treatment. Concluding remarks are summarized in section 5. Furthermore, the second-order Møller–Plesset (MP2) treatment for the first-order TRF–NOMO method is described in the Appendix.

## 2. Theory

In this section, we summarize TRF–NOMO/Hartree–Fock (HF) theory,<sup>10</sup> which determines electronic and nuclear wave functions simultaneously while eliminating translational and rotational motions. The total Hamiltonian adopted in the original NOMO theory contaminates translational and rotational motions. Thus, the Hamiltonian is called the translation- and rotation-contaminated (TRC) Hamiltonian:

$$\hat{H}_{\text{TRC}} = \hat{T}^e + \hat{T}^n + \hat{V}^{ee} + \hat{V}^{en} + \hat{V}^{nn} \quad (1)$$

where

$$\hat{T}^e = - \sum_p \frac{1}{2} \nabla(\mathbf{x}_p)^2 \equiv \sum_p \hat{t}^e(\mathbf{x}_p) \quad (2)$$

$$\hat{T}^n = - \sum_P \frac{1}{2m_P} \nabla(\mathbf{x}_P)^2 \equiv \sum_P \hat{t}^n(\mathbf{x}_P) \quad (3)$$

$$\hat{V}^{ee} = \sum_{p < q} \frac{1}{r_{pq}} \quad (4)$$

$$\hat{V}^{en} = - \sum_{p,P} \frac{Z_P}{r_{pP}} \quad (5)$$

$$\hat{V}^{nn} = \sum_{P < Q} \frac{Z_P Z_Q}{r_{PQ}} \quad (6)$$

Here,  $\hat{T}^e$  and  $\hat{T}^n$  are electronic and nuclear kinetic operators, respectively. Two-particle operators consist of e–e ( $\hat{V}^{ee}$ ), e–n ( $\hat{V}^{en}$ ), and n–n ( $\hat{V}^{nn}$ ) interactions. While the summation of  $p$

and  $q$  in eqs 2, 4, and 5 runs over electrons, that of  $P$  and  $Q$  in eqs 3, 5, and 6 runs over nuclei.

We proposed a scheme to eliminate the contribution of translational motion from the TRC Hamiltonian.<sup>7</sup> By subtracting the translational Hamiltonian  $\hat{T}_T$  from  $\hat{H}_{\text{TRC}}$ , the TF Hamiltonian is given by

$$\hat{H}_{\text{TF}} = \hat{H}_{\text{TRC}} - \hat{T}_T \quad (7)$$

where

$$\hat{T}_T = - \frac{1}{2M} \sum_{\mu} \nabla(\mathbf{x}_{\mu})^2 - \frac{1}{M} \sum_{\mu < \nu} \nabla(\mathbf{x}_{\mu}) \nabla(\mathbf{x}_{\nu}) \quad (8)$$

Here,  $M$  is the total mass of all particles. The summation runs over all particles.

We have further proposed a scheme to eliminate the effect of translational and rotational motions from the NOMO calculation.<sup>10</sup> The TRF Hamiltonian is given by subtracting the rotational Hamiltonian  $\hat{T}_R$  from  $\hat{H}_{\text{TF}}$  as follows:

$$\hat{H}_{\text{TRF}} = \hat{H}_{\text{TF}} - \hat{T}_R = \hat{H}_{\text{TRC}} - \hat{T}_T - \hat{T}_R \quad (9)$$

where

$$\hat{T}_R = \sum_{\alpha} \frac{1}{2I_{\alpha}} \left( \sum_{\mu} \hat{L}_{\alpha,\mu}^2 + 2 \sum_{\mu < \nu} \hat{L}_{\alpha,\mu} \hat{L}_{\alpha,\nu} \right) \quad (10)$$

Here,  $I_{\alpha}$  is the principle moment of inertia, which is not a constant but a function (or an operator) due to coupling with the vibration.

NOMO theory adopts Gaussian basis functions, which are localized around the centers. Then, the position of the  $\mu$ th particle  $\mathbf{x}_{\mu} = (x_{\mu}, y_{\mu}, z_{\mu})$  is described by the center  $\mathbf{x}_{\mu}^0 = (x_{\mu}^0, y_{\mu}^0, z_{\mu}^0)$  and the displacement  $\Delta\mathbf{x}_{\mu}$  as follows:

$$\mathbf{x}_{\mu} = \mathbf{x}_{\mu}^0 + \Delta\mathbf{x}_{\mu} \quad (11)$$

Because of the locality of the Gaussian function, it is possible to define a quasi-COM coordinate  $\mathbf{x}_G^0$  by

$$\mathbf{x}_G^0 = \frac{\sum_{\mu} m_{\mu} \mathbf{x}_{\mu}^0}{\sum_{\mu} m_{\mu}} \quad (12)$$

The coordinate  $\mathbf{x}_{\mu}^0$  is translated to a new coordinate  $\tilde{\mathbf{x}}_{\mu}^0$ , whose origin agrees with the COM  $\mathbf{x}_G^0$ :

$$\tilde{\mathbf{x}}_{\mu}^0 = \mathbf{x}_{\mu}^0 - \mathbf{x}_G^0 \quad (13)$$

Because  $\tilde{\mathbf{x}}_{\mu}^0$  is a constant, the atoms located on  $\{\tilde{\mathbf{x}}_{\mu}^0\}$  compose a rigid-body rotator, of which the tensor of inertia is defined by

$$\mathbf{I}^0 = \begin{bmatrix} \sum_{\mu} m_{\mu} (\tilde{y}_{\mu}^{02} + \tilde{z}_{\mu}^{02}) & - \sum_{\mu} m_{\mu} \tilde{x}_{\mu}^0 \tilde{y}_{\mu}^0 & - \sum_{\mu} m_{\mu} \tilde{x}_{\mu}^0 \tilde{z}_{\mu}^0 \\ - \sum_{\mu} m_{\mu} \tilde{x}_{\mu}^0 \tilde{y}_{\mu}^0 & \sum_{\mu} m_{\mu} (\tilde{z}_{\mu}^{02} + \tilde{x}_{\mu}^{02}) & - \sum_{\mu} m_{\mu} \tilde{y}_{\mu}^0 \tilde{z}_{\mu}^0 \\ - \sum_{\mu} m_{\mu} \tilde{x}_{\mu}^0 \tilde{z}_{\mu}^0 & - \sum_{\mu} m_{\mu} \tilde{y}_{\mu}^0 \tilde{z}_{\mu}^0 & \sum_{\mu} m_{\mu} (\tilde{x}_{\mu}^{02} + \tilde{y}_{\mu}^{02}) \end{bmatrix} \quad (14)$$

By diagonalizing the tensor, the principal moments of inertia and the transformed coordinates are determined as follows:

$${}^t\mathbf{U}^0\mathbf{U} = \begin{bmatrix} I_x^0 & 0 & 0 \\ 0 & I_y^0 & 0 \\ 0 & 0 & I_z^0 \end{bmatrix} \quad (15)$$

$$\mathbf{r}_\mu^0 = \mathbf{U}\tilde{\mathbf{x}}_\mu^0 \quad (16)$$

where  $\mathbf{U}$  is a unitary transform matrix with three dimensions. The general coordinate  $\mathbf{r}_\mu$  of the  $\mu$ th particle, which cannot be treated as a rigid-body rotator, is rewritten by  $\mathbf{r}_\mu^0 = (r_{x_\mu}^0, r_{y_\mu}^0, r_{z_\mu}^0)$  and the displacement  $\Delta\mathbf{r}_\mu = (\Delta r_{x_\mu}, \Delta r_{y_\mu}, \Delta r_{z_\mu})$ :

$$\mathbf{r}_\mu = \mathbf{r}_\mu^0 + \Delta\mathbf{r}_\mu \quad (17)$$

As a result, the rotational term can be uniquely defined using the coordinates  $\{\mathbf{r}_\mu\}$ .

Furthermore, the rotational operator is expanded in a Taylor series with respect to  $\Delta\mathbf{r}_\mu$  as follows:

$$\hat{T}_R = \hat{T}_{R_0} + \hat{T}_{R_1} + O(\Delta\mathbf{r}^2) \quad (18)$$

$$\hat{T}_{R_0} = \sum_{\alpha}^{x,y,z} \left( \sum_{\mu} \frac{1}{2I_{\alpha}^0} \hat{L}_{\alpha,\mu}^0{}^2 + \sum_{\mu < \nu} \frac{1}{I_{\alpha}^0} \hat{L}_{\alpha,\mu}^0 \hat{L}_{\alpha,\nu}^0 \right) \quad (19)$$

$$\begin{aligned} \hat{T}_{R_1} = & \sum_{\mu} \sum_{\alpha}^{x,y,z} \frac{1}{2I_{\alpha}^0} \left( \hat{L}_{\alpha,\mu}^0{}^2 + \hat{L}_{\alpha,\mu}^0 \Delta\hat{L}_{\alpha,\mu} + \Delta\hat{L}_{\alpha,\mu} \hat{L}_{\alpha,\mu}^0 - \right. \\ & \left. \frac{2}{I_{\alpha}^0} \Delta\hat{L}_{\alpha,\mu} \hat{L}_{\alpha,\mu}^0{}^2 \right) + \sum_{\mu < \nu} \sum_{\alpha}^{x,y,z} \frac{1}{I_{\alpha}^0} \left\{ \hat{L}_{\alpha,\mu}^0 \Delta\hat{L}_{\alpha,\nu} + \Delta\hat{L}_{\alpha,\mu} \hat{L}_{\alpha,\nu}^0 - \right. \\ & \left. \frac{1}{I_{\alpha}^0} (\Delta\hat{L}_{\alpha,\mu} \hat{L}_{\alpha,\nu}^0{}^2 + \Delta\hat{L}_{\alpha,\nu} \hat{L}_{\alpha,\mu}^0{}^2 + 2\Delta\hat{L}_{\alpha,\mu} \hat{L}_{\alpha,\mu}^0 \hat{L}_{\alpha,\nu}^0 + \right. \\ & \left. 2\Delta\hat{L}_{\alpha,\nu} \hat{L}_{\alpha,\mu}^0 \hat{L}_{\alpha,\nu}^0) \right\} - \sum_{\mu \neq \nu, \mu \neq \lambda, \nu < \lambda} \sum_{\alpha}^{x,y,z} \frac{2}{I_{\alpha}^0} \Delta\hat{L}_{\alpha,\mu} \hat{L}_{\alpha,\nu}^0 \hat{L}_{\alpha,\lambda}^0 \quad (20) \end{aligned}$$

where  $x$  components of  $I_{\alpha}^0$ ,  $\Delta\hat{L}_{\alpha,\mu}$ ,  $\hat{L}_{\alpha,\mu}^0$ , and  $\Delta\hat{L}_{\alpha,\mu}$  are

$$I_x^0 = \sum_{\mu} m_{\mu} (r_{y_{\mu}}^0{}^2 + r_{z_{\mu}}^0{}^2) \quad (21)$$

$$\Delta\hat{L}_{x,\mu} = m_{\mu} (r_{y_{\mu}}^0 \Delta r_{y_{\mu}} + r_{z_{\mu}}^0 \Delta r_{z_{\mu}}) \quad (22)$$

$$\hat{L}_{x,\mu}^0 = -i \left( r_{y_{\mu}}^0 \frac{\partial}{\partial r_{z_{\mu}}^0} - r_{z_{\mu}}^0 \frac{\partial}{\partial r_{y_{\mu}}^0} \right) \quad (23)$$

$$\Delta\hat{L}_{x,\mu} = -i \left( \Delta r_{y_{\mu}}^0 \frac{\partial}{\partial r_{z_{\mu}}^0} - \Delta r_{z_{\mu}}^0 \frac{\partial}{\partial r_{y_{\mu}}^0} \right) \quad (24)$$

In eq 18, the second- and higher-order terms are collected in the last term  $O(\Delta\mathbf{r}^2)$ . The HF equations for NOs and MOs in the TRF–NOMO formalism with first-order rovibrational coupling are easily derived by applying the variational procedure. The HF equations are similar to the TRC–, TF–, and the zeroth-order TRF–NOMO ones. However, the present first-order TRF–NOMO method involves additional terms, some of which require three-body integrals.

To differentiate between the zeroth- and first-order TRF–NOMO methods, we adopt the abbreviations TR<sub>0</sub>F and TR<sub>1</sub>F, respectively.

### 3. Implementation

We have implemented a computational program<sup>30</sup> for the first-order TRF–NOMO/HF method by modifying the GAMESS program package.<sup>31</sup> For simplicity, we have concentrated here on nuclear contributions of translational and rotational motions, which should be the main parts in those motions. Thus, the translational Hamiltonian is approximated as follows:

$$\hat{T}_T(\mathbf{x}) \approx \hat{T}_T^n(\mathbf{x}) = -\frac{1}{2M^n} \sum_P^{\text{nuc}} \nabla(\mathbf{x}_P)^2 - \frac{1}{M^{nP < Q}} \sum_{P < Q}^{\text{nuc}} \nabla(\mathbf{x}_P) \nabla(\mathbf{x}_Q) \quad (25)$$

where  $M^n$  is the total mass of all nuclei. The rotational Hamiltonian is also approximated as follows:

$$\hat{T}_{R_0} \approx \hat{T}_{R_0}^n = \sum_{\alpha}^{x,y,z} \left( \sum_{\alpha} \frac{1}{2I_{\alpha}^{n,0}} \hat{L}_{\alpha}^0{}^2 + \sum_{P < Q}^{\text{nuc}} \frac{1}{I_{\alpha}^{n,0}} \hat{L}_{\alpha,P}^0 \hat{L}_{\alpha,Q}^0 \right) \quad (26)$$

$$\begin{aligned} \hat{T}_{R_1} \approx \hat{T}_{R_1}^n = & \sum_P^{\text{nuc},x,y,z} \sum_{\alpha} \frac{1}{2I_{\alpha}^{n,0}} \left( \hat{L}_{\alpha,P}^0 \Delta\hat{L}_{\alpha,P} + \Delta\hat{L}_{\alpha,P} \hat{L}_{\alpha,P}^0 - \right. \\ & \left. \frac{2}{I_{\alpha}^{n,0}} \Delta\hat{L}_{\alpha,P} \hat{L}_{\alpha,P}^0{}^2 \right) + \sum_{P < Q}^{\text{nuc},x,y,z} \sum_{\alpha} \frac{1}{I_{\alpha}^{n,0}} \left\{ \hat{L}_{\alpha,P}^0 \Delta\hat{L}_{\alpha,Q} + \Delta\hat{L}_{\alpha,P} \hat{L}_{\alpha,Q}^0 - \right. \\ & \left. \frac{1}{I_{\alpha}^{n,0}} (\Delta\hat{L}_{\alpha,P} \hat{L}_{\alpha,Q}^0{}^2 + \Delta\hat{L}_{\alpha,Q} \hat{L}_{\alpha,P}^0{}^2 + 2\Delta\hat{L}_{\alpha,P} \hat{L}_{\alpha,P}^0 \hat{L}_{\alpha,Q}^0 + \right. \\ & \left. 2\Delta\hat{L}_{\alpha,P} \hat{L}_{\alpha,Q}^0 \hat{L}_{\alpha,P}^0) \right\} - \sum_{P \neq Q, P \neq R, Q < R} \sum_{\alpha}^{x,y,z} \frac{2}{I_{\alpha}^{n,0,2}} \Delta\hat{L}_{\alpha,P} \hat{L}_{\alpha,Q}^0 \hat{L}_{\alpha,R}^0 \quad (27) \end{aligned}$$

where  $I_{\alpha}^{n,0}$  is the principal moment of inertia for the rigid-body rotator, which consists of all nuclei. Because three-body integrals due to first-order rotational operators can be constructed by direct products of one-body integrals, for example, the integral of the three-body term in eq 20 is given by

$$- \sum_{\substack{I \neq J, I \neq K \\ J < K}} \sum_{\alpha}^{x,y,z} \frac{2}{I_{\alpha}^0} \langle \varphi_I | \Delta\hat{L}_{\alpha} | \varphi_J \rangle \langle \varphi_J | \hat{L}_{\alpha}^0 | \varphi_K \rangle \langle \varphi_K | \hat{L}_{\alpha}^0 | \varphi_K \rangle \quad (28)$$

Here,  $I$ ,  $J$ , and  $K$  represent nuclear occupied orbitals; we first evaluate and save the one-body integrals into memory and then use them reading from the memory. The approximation in eqs 25–27 corresponds to neglecting coupling between electronic motion and translation or rotation. It is noted, however, that the TRF–NOMO method using eqs 25–27 still involves vibronic coupling, that is, a nonadiabatic effect.

The present program is capable of performing different kinds of NOMO/HF calculations: TRC–, TF–, TR<sub>0</sub>F–, and

**Table 1.** Even-Tempered Parameters  $\alpha$  and  $\beta$  According to eq 37 from the Hydrogen to Fluorine Atom

atom	H	D	T	Li	B	C	N	F
$\alpha$	2.872	5.741	8.597	19.997	313.790	342.018	399.109	541.491
$\beta$	3.162	3.162	3.162	3.162	3.162	3.162	3.162	3.162

**Table 2.** Total ( $E_{\text{tot}}$ ), Translational ( $E_{\text{trans}}$ ), Zeroth-Order Rotational ( $E_{\text{rot0}}$ ), First-Order Rotational ( $E_{\text{rot1}}$ ), Vibrational ( $E_{\text{vib}}$ ), and Electronic ( $E_{\text{elec}}$ ) Energies (in hartree) Calculated by the TRC-, TF-, TR<sub>0</sub>F-, and TR<sub>1</sub>F-NOMO/HF Methods

		TRC	TF	$\Delta E^{\text{TF}^a}$	TR <sub>0</sub> F	$\Delta E^{\text{R}_0\text{F}^b}$	TR <sub>1</sub> F	$\Delta E^{\text{R}_1\text{F}^c}$
H <sub>2</sub>	$E_{\text{tot}}$	-1.052371	-1.074314	-21.943	-1.104088	-29.774	-1.104069	0.019
	$E_{\text{trans}}$	0.018337	(0.027092)		(0.191744)		(0.095525)	
	$E_{\text{rot0}}$	0.011396	0.016916		(0.080830)		(0.081221)	
	$E_{\text{rot1}}$	0.000422	0.000452		-0.000069		(0.000030)	
	$E_{\text{vib}}$	0.006518	0.009723		0.014310		0.014274	
	$E_{\text{elec}}$	-1.089044	-1.101406		-1.118329		-1.118343	
D <sub>2</sub>	$E_{\text{tot}}$	-1.074233	-1.090522	-16.289	-1.112184	-21.662	-1.112172	0.012
	$E_{\text{trans}}$	0.013614	(0.020054)		(0.068366)		(0.068710)	
	$E_{\text{rot0}}$	0.008534	0.012550		(0.058017)		(0.058320)	
	$E_{\text{rot1}}$	0.000202	0.000208		-0.000042		(0.000018)	
	$E_{\text{vib}}$	0.004877	0.007295		0.010391		0.010372	
	$E_{\text{elec}}$	-1.101460	-1.110576		-1.122533		-1.122544	
T <sub>2</sub>	$E_{\text{tot}}$	-1.084308	-1.097901	-13.593	-1.115902	-18.001	-1.115893	0.008
	$E_{\text{trans}}$	0.011367	(0.016734)		(0.054419)		(0.054503)	
	$E_{\text{rot0}}$	0.007143	0.010518		(0.046134)		(0.046203)	
	$E_{\text{rot1}}$	0.000139	0.000130		-0.000023		(0.000006)	
	$E_{\text{vib}}$	0.004085	0.006086		0.008308		0.008293	
	$E_{\text{elec}}$	-1.107043	-1.114635		-1.124187		-1.124187	

<sup>a</sup>  $\Delta E^{\text{TF}} = E_{\text{tot}}^{\text{TF}} - E_{\text{tot}}^{\text{TRC}}$  (in mhartree) <sup>b</sup>  $\Delta E^{\text{R}_0\text{F}} = E_{\text{tot}}^{\text{R}_0\text{F}} - E_{\text{tot}}^{\text{TF}}$  (in mhartree) <sup>c</sup>  $\Delta E^{\text{R}_1\text{F}} = E_{\text{tot}}^{\text{R}_1\text{F}} - E_{\text{tot}}^{\text{R}_0\text{F}}$  (in mhartree)

TR<sub>1</sub>F-NOMO/HF using the following energy expectations:

$$E_{\text{tot}}^{\text{TRC}} = \langle \Phi_0 | \hat{H}_{\text{TRC}} | \Phi_0 \rangle \quad (29)$$

$$E_{\text{tot}}^{\text{TF}} = \langle \Phi_0 | \hat{H}_{\text{TRC}} - \hat{T}_{\text{T}}^n | \Phi_0 \rangle \quad (30)$$

$$E_{\text{tot}}^{\text{TR}_0\text{F}} = \langle \Phi_0 | \hat{H}_{\text{TRC}} - \hat{T}_{\text{T}}^n - \hat{T}_{\text{R}_0}^n | \Phi_0 \rangle \quad (31)$$

$$E_{\text{tot}}^{\text{TR}_1\text{F}} = \langle \Phi_0 | \hat{H}_{\text{TRC}} - \hat{T}_{\text{T}}^n - \hat{T}_{\text{R}_0}^n - \hat{T}_{\text{R}_1}^n | \Phi_0 \rangle \quad (32)$$

Furthermore, translational ( $E_{\text{trans}}$ ), zeroth-order rotational ( $E_{\text{rot0}}$ ), first-order rotational ( $E_{\text{rot1}}$ ), and vibrational ( $E_{\text{vib}}$ ) energy components are estimated as follows:

$$E_{\text{trans}} = \langle \Phi_0 | \hat{T}_{\text{T}}^n | \Phi_0 \rangle \quad (33)$$

$$E_{\text{rot0}} = \langle \Phi_0 | \hat{T}_{\text{R}_0}^n | \Phi_0 \rangle \quad (34)$$

$$E_{\text{rot1}} = \langle \Phi_0 | \hat{T}_{\text{R}_1}^n | \Phi_0 \rangle \quad (35)$$

Because the rest of the nuclear motion corresponds to vibration, the vibrational energy is evaluated by

$$E_{\text{vib}} = \langle \Phi_0 | \hat{T}^n - \hat{T}_{\text{T}}^n - \hat{T}_{\text{R}_0}^n - \hat{T}_{\text{R}_1}^n | \Phi_0 \rangle \quad (36)$$

The electronic energy including nuclear repulsion is given by

$$E_{\text{elec}} = \langle \Phi_0 | \hat{T}^c + \hat{V}^{\text{ce}} + \hat{V}^{\text{en}} + \hat{V}^{\text{nn}} | \Phi_0 \rangle \quad (37)$$

#### 4. Numerical Assessment

Numerical assessment on the effect of the first-order rotational terms in the TRF-NOMO treatment was per-

formed for small molecules: H<sub>2</sub>, D<sub>2</sub>, T<sub>2</sub>, Li<sub>2</sub>, B<sub>2</sub>, N<sub>2</sub>, F<sub>2</sub>, H<sub>3</sub><sup>+</sup>, BH<sub>3</sub>, CH<sub>4</sub>, and C<sub>2</sub>H<sub>4</sub>. We have carried out HF calculations for the TRC-, TF-, TR<sub>0</sub>F-, and TR<sub>1</sub>F-NOMO methods. Primitive functions, of which exponents correspond to correlation-consistent polarization plus valence triple- $\zeta$  (cc-pVTZ) bases of Dunning,<sup>32</sup> were adopted as electronic basis functions (EBFs): (5s2p1d), (11s5p2d1f), and (10s5p2d1f) functions for {H, D, T}, Li, and {B, C, N, F}, respectively. For nuclear basis functions (NBFs), we used (7s7p7d) primitive Gaussian functions. The exponents  $\zeta_n$  of the NBFs are determined by the even-tempered scheme.<sup>7,33</sup>

$$\zeta_n = \alpha\beta^n \quad (38)$$

Table 1 shows the values of  $\alpha$  and  $\beta$  of each atom. Orbital centers of both EBFs and NBFs in H<sub>2</sub>, D<sub>2</sub>, and T<sub>2</sub> molecules were settled at the experimental geometries,<sup>34</sup> and those of others are the optimized geometry at the conventional MO/HF level.

Table 2 summarizes the results of the TRC-, TF-, TR<sub>0</sub>F-, and TR<sub>1</sub>F-NOMO/HF calculations for H<sub>2</sub>, D<sub>2</sub>, and T<sub>2</sub> molecules. The total energies decrease in the order of the TRC-, TF-, and TR<sub>0</sub>F-NOMO/HF methods. The energy difference between  $E_{\text{tot}}^{\text{TRC}}$  and  $E_{\text{tot}}^{\text{TF}}$ , which is represented by  $\Delta E^{\text{TF}}$ , corresponds to the energetic improvement achieved by eliminating the contamination of the translational motion.  $\Delta E^{\text{TF}}$  values for H<sub>2</sub>, D<sub>2</sub>, and T<sub>2</sub> are -21.9, -16.3, and -13.6 mhartree, respectively. The energy difference between  $E_{\text{tot}}^{\text{TF}}$  and  $E_{\text{tot}}^{\text{TR}_0\text{F}}$ , given by  $\Delta E^{\text{R}_0\text{F}}$ , corresponds to the elimination of zeroth-order rotational motion, that is, the rigid-body rotator.  $\Delta E^{\text{R}_0\text{F}}$  values for H<sub>2</sub>, D<sub>2</sub>, and T<sub>2</sub> are -29.8, -21.6, and -18.0 mhartree, respectively. Thus, the elimina-

**Table 3.** Total Energies (in hartree) of Several Small Molecules Calculated by the TRC-, TF-, TR<sub>0</sub>F-, and TR<sub>1</sub>F-NOMO/HF Methods

	TRC	TF	$\Delta E^{\text{TF}a}$	TR <sub>0</sub> F	$\Delta E^{\text{R}_0\text{F}b}$	TR <sub>1</sub> F	$\Delta E^{\text{R}_1\text{F}c}$
Li <sub>2</sub>	-14.553448	-14.643200	-89.753	-14.763556	-120.356	-14.763556	0.000
B <sub>2</sub>	-48.170671	-48.382923	-212.252	-48.671256	-288.333	-48.671258	-0.003
N <sub>2</sub>	-107.637234	-108.020683	-383.449	-108.541786	-521.104	-108.541786	0.000
F <sub>2</sub>	-196.803609	-197.361041	-557.432	-198.087631	-726.590	-198.087631	0.000
H <sub>3</sub> <sup>+</sup>	-1.192826	-1.210845	-18.019	-1.239881	-29.036	-1.239819	0.061
BH <sub>3</sub>	-25.898990	-26.099861	-200.871	-26.123998	-24.138	-26.124007	-0.008
CH <sub>4</sub>	-39.518843	-39.779765	-260.922	-39.801179	-21.414	-39.801338	-0.159
C <sub>2</sub> H <sub>4</sub>	-76.842228	-77.093110	-250.882	-77.260050	-166.940	-77.259630	0.421

<sup>a</sup>  $\Delta E^{\text{TF}} = E_{\text{tot}}^{\text{TF}} - E_{\text{tot}}^{\text{TRC}}$  (in mhartree) <sup>b</sup>  $\Delta E^{\text{R}_0\text{F}} = E_{\text{tot}}^{\text{TR}_0\text{F}} - E_{\text{tot}}^{\text{TF}}$  (in mhartree) <sup>c</sup>  $\Delta E^{\text{R}_1\text{F}} = E_{\text{tot}}^{\text{TR}_1\text{F}} - E_{\text{tot}}^{\text{TR}_0\text{F}}$  (in mhartree)

tion of these contaminations is of great importance to improving accuracy.

On the other hand, the energy changes from the TR<sub>0</sub>F treatment to the TR<sub>1</sub>F one are extremely small. Moreover, they increase slightly.  $\Delta E^{\text{R}_1\text{F}}$  values, which are defined by the difference between  $E_{\text{tot}}^{\text{TR}_0\text{F}}$  and  $E_{\text{tot}}^{\text{TR}_1\text{F}}$ , for H<sub>2</sub>, D<sub>2</sub>, and T<sub>2</sub> are less than 0.1 mhartree. Therefore, it is confirmed that TR<sub>0</sub>F treatment is a good approximation with chemical accuracy.

In Table 2, the energy components,  $E_{\text{tot}}$ ,  $E_{\text{trans}}$ ,  $E_{\text{rot}0}$ ,  $E_{\text{rot}1}$ ,  $E_{\text{vib}}$ , and  $E_{\text{elec}}$  are also given. In TRC treatment, the translational and zeroth-order rotational contaminations, that is,  $E_{\text{trans}}^{\text{TRC}}$  and  $E_{\text{rot}0}^{\text{TRC}}$ , are less than  $|\Delta E^{\text{TF}}|$  and  $|\Delta E^{\text{R}_0\text{F}}|$ , respectively. Similarly,  $E_{\text{rot}0}^{\text{TF}}$  in the TF treatment is less than  $|\Delta E^{\text{R}_0\text{F}}|$ . Thus,  $E_{\text{tot}}^{\text{TRC}} - E_{\text{trans}}^{\text{TRC}}$  and  $E_{\text{tot}}^{\text{TF}} - E_{\text{rot}0}^{\text{TF}}$  are higher than  $E_{\text{tot}}^{\text{TF}}$  and  $E_{\text{tot}}^{\text{TR}_0\text{F}}$ . This means that the self-consistent-field (SCF) procedure is important for the adequate removal of the translational and rotational contaminations.

$E_{\text{vib}}$  is related to the zero-point energy (ZPE). Experimental ZPEs of H<sub>2</sub>, D<sub>2</sub>, and T<sub>2</sub> are 0.010, 0.007, and 0.006 hartree, respectively. While  $E_{\text{vib}}^{\text{TRC}}$  of H<sub>2</sub> is about 0.003 hartree smaller than the experimental value,  $E_{\text{vib}}^{\text{TR}_0\text{F}}$  and  $E_{\text{vib}}^{\text{TR}_1\text{F}}$  are about 0.004 hartree larger.  $E_{\text{vib}}^{\text{TF}}$  is comparatively close to the experimental value. Similar trends are seen for D<sub>2</sub> and T<sub>2</sub>. However, this does not mean that TF treatment is the most suitable for evaluating ZPEs. The agreement brought about by TF treatment may be due to cancellation between the effect of the rotational elimination and the many-body effect, that is, e-n, n-n, and e-e correlations. A related discussion is mentioned in the Appendix.

Table 3 shows the results of the TRC-, TF-, TR<sub>0</sub>F-, and TR<sub>1</sub>F-NOMO/HF calculations for Li<sub>2</sub>, B<sub>2</sub>, N<sub>2</sub>, F<sub>2</sub>, H<sub>3</sub><sup>+</sup>, BH<sub>3</sub>, CH<sub>4</sub>,<sup>34</sup> and C<sub>2</sub>H<sub>4</sub> molecules. As in the cases of H<sub>2</sub>, D<sub>2</sub>, and T<sub>2</sub>, the total energies decrease considerably in the order of TRC-, TF-, and TR<sub>0</sub>F-NOMO/HF for all molecules. On the other hand, the energy changes from the TR<sub>0</sub>F treatment to the TR<sub>1</sub>F one are considerably smaller. While  $\Delta E^{\text{R}_1\text{F}}$  values, corresponding to the first-order rovibration coupling of the diatomic molecules, are on the order of microhartrees,  $\Delta E^{\text{R}_1\text{F}}$  values of the polyatomic molecules are from tens to hundreds of microhartrees. The maximum ratio  $|\Delta E^{\text{R}_1\text{F}}/\Delta E^{\text{R}_0\text{F}}|$  is 0.742% for CH<sub>4</sub>. This indicates that the contribution of first-order rovibration coupling is significantly smaller than zeroth-order rotational contamination. Furthermore, it shows the validity of the TRF-NOMO formalism, based on the Taylor expansion of the rotational operator.

## 5. Conclusions

In the present study, we have implemented the computational code for the TR<sub>1</sub>F-NOMO/HF method. Numerical assessments of the TRC-, TF-, TR<sub>0</sub>F-, and TR<sub>1</sub>F-NOMO/HF methods, which were performed for 11 small molecules, have clarified the importance of the elimination of translational and rotational contaminations. The results demonstrated that the zeroth-order rigid-body term for rotational motion makes the main contribution, and the first-order rovibrational coupling term is significantly smaller in energy. In other words, it is true that the rovibrational coupling exists in the TR<sub>0</sub>F and TR<sub>1</sub>F Hamiltonian, although the effect is shown to be negligibly small in the present assessment. For rigid molecules, we confirm that the approximate zeroth-order treatment is sufficient for eliminating rotational contamination and that the Taylor expansion with respect to displacement  $\Delta \mathbf{x}$  for the rotational operator is reasonable to achieve chemical accuracy (millihartrees) in NOMO calculations. In the case of estimating the nonadiabatic effect, which often requires spectroscopic accuracy (microhartrees), it might be necessary to take into account the first- and higher-order rovibrational coupling. Furthermore, the estimation of the first- and higher-order rovibrational coupling contribution in nonrigid molecules or the excited state is an interesting subject and will be examined in the near future.

**Acknowledgment.** Some of the calculations were performed at the Research Center for Computational Science (RCCS) of the Okazaki National Research Institutes. This study was partially supported by a Grant-in-Aid for Exploratory Research “KAKENHI 16655010” from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT), by a NAREGI Nano-Science Project of MEXT, by the 21st-Century Center Of Excellence (21COE) “Practical Nano-Chemistry” from MEXT, and the “Development of High-Performance Computational Environment for Quantum Chemical Calculation and Its Assessment” from the Advanced Research Institute for Science and Engineering (RISE), Waseda University.

## Appendix: TRF-NOMO/MP2 Theory

We have formulated TRF-NOMO/HF and MP2 methods in our previous work.<sup>10,12</sup> However, numerical assessments of TRF-NOMO/HF and MP2 methods were limited to zeroth-order rotational terms. The main body in this paper shows the implementation of the TR<sub>1</sub>F-NOMO/HF calcula-



**Table 4.** Total ( $E_{\text{tot}}$ ), Translational ( $E_{\text{trans}}$ ), Zeroth-Order Rotational ( $E_{\text{rot0}}$ ), First-Order Rotational ( $E_{\text{rot1}}$ ), Vibrational ( $E_{\text{vib}}$ ), and Electronic ( $E_{\text{elec}}$ ) Energies (in hartree) Calculated by the TRC-, TF-, TR<sub>0</sub>F-, and TR<sub>1</sub>F-NOMO/MP2 Methods

		TRC	TF	$\Delta E^{\text{TF}a}$	TR <sub>0</sub> F	$\Delta E^{\text{R}_0^{\text{F}b}$	TR <sub>1</sub> F	$\Delta E^{\text{R}_1^{\text{F}c}$
H <sub>2</sub>	$E_{\text{tot}}$	-1.110118	-1.131626	-21.508	-1.143040	-11.414	-1.143494	-0.454
	$E_{\text{trans}}$	0.018500	(0.033951)		(0.194822)		(0.098315)	
	$E_{\text{rot0}}$	0.010485	0.011443		(0.078776)		(0.079352)	
	$E_{\text{rot1}}$	0.000375	0.000416		-0.000092		(0.000602)	
	$E_{\text{vib}}$	0.007315	0.008372		0.013309		0.013353	
	$E_{\text{elec}}$	-1.146791	-1.151858		-1.156258		-1.156847	
D <sub>2</sub>	$E_{\text{tot}}$	-1.124192	-1.140331	-16.138	-1.149150	-8.820	-1.149418	-0.267
	$E_{\text{trans}}$	0.013718	(0.025071)		(0.070578)		(0.070842)	
	$E_{\text{rot0}}$	0.007927	0.008606		(0.057963)		(0.056900)	
	$E_{\text{rot1}}$	0.000180	0.000190		-0.000054		(0.000357)	
	$E_{\text{vib}}$	0.005402	0.006239		0.009633		0.009659	
	$E_{\text{elec}}$	-1.151419	-1.155366		-1.158729		-1.159076	
T <sub>2</sub>	$E_{\text{tot}}$	-1.130901	-1.144174	-13.273	-1.152012	-7.838	-1.152189	-0.177
	$E_{\text{trans}}$	0.011457	(0.020742)		(0.056194)		(0.056207)	
	$E_{\text{rot0}}$	0.006646	0.007375		(0.044988)		(0.045094)	
	$E_{\text{rot1}}$	0.000125	0.000119		-0.000031		(0.000225)	
	$E_{\text{vib}}$	0.004508	0.005233		0.007687		0.007698	
	$E_{\text{elec}}$	-1.153635	-1.156900		-1.159668		-1.159887	

<sup>a</sup>  $\Delta E^{\text{TF}} = E_{\text{tot}}^{\text{TF}} - E_{\text{tot}}^{\text{TRC}}$  (in mhartree) <sup>b</sup>  $\Delta E^{\text{R}_0^{\text{F}}} = E_{\text{tot}}^{\text{TR}_0^{\text{F}}} - E_{\text{tot}}^{\text{TF}}$  (in mhartree) <sup>c</sup>  $\Delta E^{\text{R}_1^{\text{F}}} = E_{\text{tot}}^{\text{TR}_1^{\text{F}}} - E_{\text{tot}}^{\text{TR}_0^{\text{F}}}$  (in mhartree)

**Table 5.** Total Energies (in hartree) of Several Small Molecules Calculated by the TRC-, TF-, TR<sub>0</sub>F-, and TR<sub>1</sub>F-NOMO/MP2 Methods

	TRC	TF	$\Delta E^{\text{TF}a}$	TR <sub>0</sub> F	$\Delta E^{\text{R}_0^{\text{F}b}$	TR <sub>1</sub> F	$\Delta E^{\text{R}_1^{\text{F}c}$
Li <sub>2</sub>	-14.697667 (-144.220)	-14.794092 (-150.892)	-96.425	-14.872552 (-108.996)	-78.460	-14.872556 (-109.000)	-0.004
B <sub>2</sub>	-48.585915 (-415.244)	-48.796114 (-413.191)	-210.198	-48.972819 (-301.563)	-176.705	-48.972831 (-301.572)	-0.012
N <sub>2</sub>	-108.412959 (-775.726)	-108.809830 (-789.147)	-396.870	-109.107783 (-565.997)	-297.953	-109.107798 (-566.012)	-0.015
F <sub>2</sub>	-197.860037 (-1056.428)	-198.462909 (-1101.868)	-602.872	-198.867987 (-780.356)	-405.077	-198.867997 (-780.366)	-0.010
H <sub>3</sub> <sup>+</sup>	-1.261926 (-69.100)	-1.278850 (-68.004)	-16.923	-1.292210 (-52.329)	-13.360	-1.292606 (-52.787)	-0.396
BH <sub>3</sub>	-26.174954 (-275.965)	-26.337320 (-237.460)	-162.366	-26.363152 (-239.154)	-25.832	-26.363193 (-239.187)	-0.041
CH <sub>4</sub>	-39.937333 (-418.490)	-40.147101 (-367.335)	-209.767	-40.168661 (-367.482)	-21.560	-40.168842 (-367.504)	-0.181
C <sub>2</sub> H <sub>4</sub>	-77.563113 (-720.885)	-77.806160 (-713.050)	-243.047	-77.919316 (-659.266)	-113.156	-77.918563 (-658.934)	0.753

<sup>a</sup>  $\Delta E^{\text{TF}} = E_{\text{tot}}^{\text{TF}} - E_{\text{tot}}^{\text{TRC}}$  (in mhartree) <sup>b</sup>  $\Delta E^{\text{R}_0^{\text{F}}} = E_{\text{tot}}^{\text{TR}_0^{\text{F}}} - E_{\text{tot}}^{\text{TF}}$  (in mhartree) <sup>c</sup>  $\Delta E^{\text{R}_1^{\text{F}}} = E_{\text{tot}}^{\text{TR}_1^{\text{F}}} - E_{\text{tot}}^{\text{TR}_0^{\text{F}}}$  (in mhartree)

tion and the numerical assessments. We examine the contribution of the first-order rotational term to the TR<sub>1</sub>F-NOMO/MP2 calculation.

Table 4 shows the energy components,  $E_{\text{tot}}$ ,  $E_{\text{trans}}$ ,  $E_{\text{rot0}}$ ,  $E_{\text{rot1}}$ ,  $E_{\text{vib}}$ , and  $E_{\text{elec}}$ , of H<sub>2</sub>, D<sub>2</sub>, and T<sub>2</sub> calculated by the TRC-, TF-, TR<sub>0</sub>F-, and TR<sub>1</sub>F-NOMO/MP2 methods. The electronic and nuclear basis functions are the same as the NOMO/HF calculations presented in section 3. For the energy components, similar trends to those in Table 2 were seen in Table 4. All  $\Delta E^{\text{R}_1^{\text{F}}}$  values in Table 4 are on the order of 0.1 mhartree. In Table 2,  $E_{\text{vib}}^{\text{TF}}$  at the HF level is close to the experimental ZPE rather than  $E_{\text{vib}}^{\text{TRC}}$ ,  $E_{\text{vib}}^{\text{TR}_0^{\text{F}}}$ , and  $E_{\text{vib}}^{\text{TR}_1^{\text{F}}}$ . However, the inclusion of a many-body effect decreases  $E_{\text{vib}}^{\text{TF}}$ ,  $E_{\text{vib}}^{\text{TR}_0^{\text{F}}}$ , and  $E_{\text{vib}}^{\text{TR}_1^{\text{F}}}$ . As a result, the discrepancy of  $E_{\text{vib}}^{\text{TF}}$  from the experimental ZPE increases with the MP2 treatment, although those of  $E_{\text{vib}}^{\text{TR}_0^{\text{F}}}$  and  $E_{\text{vib}}^{\text{TR}_1^{\text{F}}}$  decrease. Therefore, we think that the agreement brought

about by TF treatment at the HF level is due to cancellation between the effect of the rotational elimination and the many-body effect.

Table 5 shows the results of the TRC-, TF-, TR<sub>0</sub>F-, and TR<sub>1</sub>F-NOMO/MP2 calculations for several small molecules except for H<sub>2</sub>, D<sub>2</sub>, and T<sub>2</sub>. The electronic and nuclear basis functions are the same as the above calculations. The correlation energies, which involve the e-e, e-n, and n-n contributions, are given in parentheses. The many-body effect becomes more important as the nuclear charges or numbers of nuclei and electrons increase. The correlation energies are approximately comparable to the effect of the elimination of the translational and rotational contaminations. The total energies decrease considerably in the order of TRC-, TF-, and TR<sub>0</sub>F-NOMO/MP2 for all molecules. On the other hand, the total energy differences between the TR<sub>0</sub>F and TR<sub>1</sub>F treatments are significantly smaller, that is, less

than 1 mhartree. This trend is the same as that of the HF treatment. Therefore, approximate zeroth-order treatment is sufficient for eliminating the rotational contamination, and the Taylor expansion with respect to displacement  $\Delta\mathbf{x}$  for the rotational operator is reasonable in the MP2-level calculation.

### References

- (1) Born, M.; Oppenheimer, R. *Ann. Phys.* **1927**, *84*, 457–488.
- (2) Cafiero, M.; Bubin, S.; Adamowicz, L. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1491–1501.
- (3) Kinghorn, D. B.; Adamowicz, L. *J. Chem. Phys.* **2000**, *113*, 4203–4205.
- (4) Bubin, S.; Adamowicz, L. *J. Chem. Phys.* **2004**, *121*, 6249–6253.
- (5) Cafiero, M.; Adamowicz, L. *J. Chem. Phys.* **2005**, *122*, 184305.
- (6) Tachikawa, M.; Mori, K.; Nakai, H.; Iguchi, K. *Chem. Phys. Lett.* **1998**, *290*, 437–442.
- (7) Nakai, H. *Int. J. Quantum Chem.* **2002**, *86*, 511–517.
- (8) Nakai, H.; Sodeyama, K.; Hoshino, M. *Chem. Phys. Lett.* **2001**, *345*, 118–124.
- (9) Nakai, H.; Sodeyama, K. *J. Chem. Phys.* **2003**, *118*, 1119–1127.
- (10) Nakai, H.; Hoshino, M.; Miyamoto, K.; Hyodo, S. *J. Chem. Phys.* **2005**, *122*, 164101.
- (11) Sodeyama, K.; Miyamoto, K.; Nakai, H. *Chem. Phys. Lett.* **2006**, *421*, 72–76.
- (12) Hoshino, M.; Nakai, H. *J. Chem. Phys.* **2006**, *124*, 194110.
- (13) Tachikawa, M. *Chem. Phys. Lett.* **2002**, *360*, 494–500.
- (14) Tachikawa, M.; Taneda, K.; Mori, K. *Int. J. Quantum Chem.* **1999**, *75*, 497–510.
- (15) Tachikawa, M.; Osamura, Y. *Theor. Chem. Acc.* **2000**, *104*, 29–39.
- (16) Tachikawa, M.; Shiga, M. *J. Am. Chem. Soc.* **2005**, *127*, 11908–11909.
- (17) Shibl, M. F.; Tachikawa, M.; Kühn, O. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1368–1373.
- (18) Udagawa, T.; Ishimoto, T.; Tokiwa, H.; Tachikawa, M.; Nagashima, U. *Chem. Phys. Lett.* **2004**, *389*, 236–240.
- (19) Tachikawa, M. *THEOCHEM* **2003**, *630*, 75–79.
- (20) Tachikawa, M.; Buenker, R. J.; Kimura, M. *J. Chem. Phys.* **2003**, *119*, 5005–5009.
- (21) Shigeta, Y.; Ozaki, Y.; Kodama, K.; Nagao, H.; Kawabe, H.; Nishikawa, K. *Int. J. Quantum Chem.* **1998**, *69*, 629–637.
- (22) Shigeta, Y.; Nagao, H.; Nishikawa, K.; Yamaguchi, K. *Int. J. Quantum Chem.* **1999**, *75*, 875–883.
- (23) Shigeta, Y.; Takahashi, H.; Yamanaka, S.; Mitani, M.; Nagao, H.; Yamaguchi, K. *Int. J. Quantum Chem.* **1998**, *70*, 659–669.
- (24) Webb, S. P.; Iordanov, T.; Hammes-Schiffer, S. *J. Chem. Phys.* **2002**, *117*, 4106–4118.
- (25) Pak, M. V.; Hammes-Schiffer, S. *Phys. Rev. Lett.* **2004**, *92*, 103002.
- (26) Swalina, C.; Pak, M. V.; Hammes-Schiffer, S. *J. Chem. Phys.* **2005**, *123*, 014303.
- (27) Reyes, A.; Pak, M. V.; Hammes-Schiffer, S. *J. Chem. Phys.* **2005**, *123*, 064104.
- (28) Bochevarov, A. D.; Valeev, E. F.; Sherrill, C. D. *Mol. Phys.* **2004**, *102*, 111–123.
- (29) Sutcliffe, B. *J. Chem. Phys.* **2005**, *123*, 237101.
- (30) The present program is still a pilot code, not the result of years of optimization and efficiency improvements. Nonetheless, some idea of the time required to compute the NOMO terms can be gained from the following: for CH<sub>4</sub>, the traditional SCF time in seconds (SCF cycle) is 11 (9), while TRC, TF, TR<sub>0</sub>F, and TR<sub>1</sub>F are 183 (19), 184 (19), 175 (18), and 321 (35) using a Xeon 3.2 GHz workstation.
- (31) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (32) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007–1023. Note that we found that uncontracted basis functions for the electronic bases produce considerably more accurate results than those of contracted ones. For example, the improvement of total energy is about 0.3 hartree for B<sub>2</sub> at the TR<sub>0</sub>F-NOMO/MP2 level. Thus, we adopted the uncontracted bases in this paper as in the case of our previous papers: refs 10 and 12.
- (33) Bardo, R. D.; Ruedenberg, K. *J. Chem. Phys.* **1974**, *60*, 918–931.
- (34) Huber, K. P.; Herzberg, G. *MOLECULAR SPECTRA and MOLECULAR STRUCTURE IV. Constants of Diatomic Molecules*; Van Nostrand Reinhold: New York, 1979; pp 250, 266, 270.

CT6002065

## A Barrier-Free Atomic Radical-Molecule Reaction: F + Propene

Ji-Lai Li, Cai-Yun Geng, Xu-Ri Huang,\* and Chia-Chung Sun

State Key Laboratory of Theoretical and Computational Chemistry,  
Institute of Theoretical Chemistry, Jilin University, Changchun 130023,  
People's Republic of China

Received September 18, 2005

**Abstract:** The possible reaction mechanism of atomic radical F with propene is investigated theoretically by a detailed potential energy surface (PES) calculation at the UMP2/6-311++G(d,p) and CCSD(T)/cc-pVTZ (single-point) levels using ab initio quantum chemistry methods and transition-state theory. Various possible reaction paths including addition–isomerization–elimination reactions and direct H-atom abstraction reactions are considered. Among them, the most feasible pathway should be the atomic radical F ( $^2F$ ) attacking on the C=C double bond in propene ( $CH_3CH=CH_2$ ) to form a weakly bound complex **I1** with no barrier, followed by atomic radical F addition to the C=C double bond to form the low-lying intermediate isomer **3** barrierlessly. Starting from intermediate isomer **3**, the most competitive reaction pathway is the dissociation of the C2–C3 single bond via transition state **TS3–P5**, leading to the product **P5**,  $CH_3 + CHF=CH_2$ . However, in the direct H-atom abstraction reactions, the atomic radical F picking up the *b*-allylic hydrogen of propene barrierlessly is the most feasible pathway from thermodynamic consideration. The other reaction pathways on the doublet PES are less competitive because of thermodynamical or kinetic factors. No addition–elimination mechanism exists on the potential energy surface. Because the intermediates and transition states involved in the major pathways are all lower than the reactants in energy, the title reaction is expected to be rapid. Furthermore, on the basis of the analysis of the kinetics of all channels through which the addition and abstraction reactions proceed, we expect that the competitive power of reaction channels may vary with experimental conditions for the title reaction. The present study may be helpful for probing the mechanisms of the title reaction and understanding the halogen chemistry.

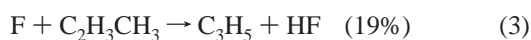
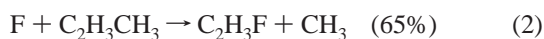
### 1. Introduction

The reactivity of the fluorine atom is an interesting research topic because of the complicated nature of the reaction mechanism involved in these reactions and its importance in many industrial applications. A variety of reaction mechanisms are involved in its reaction with different molecules. The reactions of fluorine atoms with unsaturated hydrocarbons have been studied, both theoretically and experimentally.<sup>1–16</sup> It was found that this class of reactions

proceeds primarily by the addition of fluorine atoms to the double bond of the hydrocarbon molecules to form long-lived chemical-activated radical complexes, which further decompose unimolecularly to give predominantly hydrogen atoms or methyl radicals. In contrast, in the F atom reaction with methane,<sup>17–20</sup> the reaction mainly occurs through an abstraction mechanism to produce a HF molecule and a methyl radical ( $CH_3$ ) in which the  $CH_3$  part is largely a spectator during the reaction process. Recently, Lee and co-workers<sup>5</sup> carried out the crossed-beam study on the reaction dynamics of the F atom reaction with the propene molecule with both single and double bonds, in which reactions occur

\* Corresponding author fax: (+86) 431-894-5942; e-mail: lijilai2008@yahoo.com.

through both direct abstraction and long-lived complex formation mechanisms. As shown in eqs 1–3, three different reaction channels of the  $F + C_2H_3CH_3$  reaction have been identified: (1) the H atom formation channel takes place through some direct reaction mechanism and is likely to be a long-lived complex formation process; (2) the  $CH_3$  formation process mainly proceeds through a long-lived complex formation mechanism, and (3) HF formation shows clearly that a direct pickup-type reaction mechanism is responsible for this channel. Experimental results show that the  $CH_3$  formation is the most important reaction pathway for the title reaction, while both H and HF formation channels are also significant.



However, Lee et al.<sup>5</sup> did not further provide unambiguous information on the site-specific effect in the H-atom elimination channels and the HF formation channels although there are three chemically different hydrogen sites in the propene molecule. This information may be important in the site specificity of the sequential chain processes. To the best of our knowledge, the nature of different reaction channels in the atomic radical **F** reactions with hydrocarbon molecules has not been investigated in detail in all previous experiments, and no report is found about the theoretical study on the title reaction. In addition, recently, Braña and Sordo have built up the potential energy surfaces (PESs) of the analogous **Cl** + **propene** reaction.<sup>21</sup> Now, the sequential question is whether the mechanism of the title reaction **F** + propene is similar to the **Cl** + **propene** reaction or not.

In view of their potential importance and the rather limited information on them, a detailedly theoretical study on the doublet potential energy surfaces (DPESs) of the title reactions is therefore very desirable. The objective of the present article is threefold: (1) provide the elaborated isomerization and dissociation channels on the  $C_3H_6F$  DPESs, (2) investigate the site-specific effect of the title reactions to assist further experiment identification, and (3) provide some useful insight into the mechanism of the **F** + hydrocarbons reaction. Some conclusions that are drawn in this article may be helpful to resolve the question mentioned above and for further experimental studies of this reaction.

## 2. Computational Method Details

All of the calculations reported in this work were carried out on SGI O3800 servers using the Gaussian 98 and 03 program packages.<sup>22</sup> Unless otherwise stated, all geometries of the reactants, complexes, intermediates, transition states (TSs), and products have been fully optimized with the unrestricted Møller–Plesset second-order perturbation UMP2-(FULL)<sup>23</sup> method using the 6-311++G(d,p) basis set. The presence of diffuse functions in the basis set allows for an appropriate representation of the dispersion forces that should play an important role in the stabilization of the weakly

bound structures considered in this work.<sup>24</sup> Vibrational frequencies, also calculated at the same level of theory, have been used to characterize stationary points, zero-point energy (ZPE) corrections calculations. The number of imaginary frequencies for intermediates and transition states are 0 and 1, respectively. The ZPE and vibrational frequencies were scaled by a factor of 0.95 for anharmonicity correction.<sup>25</sup> To confirm that the transition states connect between designated intermediates, intrinsic reaction coordinate (IRC)<sup>26</sup> calculations were performed at the UMP2(FULL)/6-311++G-(d,p) level of theory.

For the purpose of obtaining more reliable energies of various structures, the coupled-cluster CCSD(T) method with single, double, and perturbative treatment of triple excitations<sup>27</sup> in conjunction with the correlation-consistent polarized valence triple- $\zeta$  basis sets cc-pVTZ<sup>28</sup> was used. The UMP2(FULL)/6-311++G(d,p) optimized geometries were used for the single-point coupled cluster calculations without reoptimization at the CCSD(T)/cc-pVTZ levels.

The major problem in the application of unrestricted single-determinant reference wave functions is that of contamination with higher spin states. The severe spin contamination could lead to a deteriorated estimation of the barrier height.<sup>29,30</sup> We have examined the spin contamination before and after annihilation for the radical species and transition states involved in the  $F + C_2H_3CH_3$  reaction. For doublet systems, the expectation values of  $\langle S^2 \rangle$  range from 0.98 to 0.75 before annihilation, and after annihilation,  $\langle S^2 \rangle$  is 0.75~0.76 (the exact value for a pure doublet is 0.75). This suggests that the wave function is not severely contaminated by states of higher multiplicity.<sup>31, 32</sup>

Recent studies on systems similar to the ones considered in this article concluded that the use of spin annihilation techniques to generate spin-projected energies is mandatory for calculations of both reaction energies and barrier heights.<sup>33</sup> All of the UMP2 energy values reported in this work correspond to spin-projected calculations (PMP2) to correct for spin contamination using Schlegel's algorithm.<sup>34</sup> On the other hand, Stanton has shown that all spin contamination is essentially removed from a coupled cluster wave function.<sup>35</sup> Considering that, even with modest spin contamination, the position and height of the barriers may be affected,<sup>36</sup> several points along the reaction path (IRC) were optimized at the UMP2 level, followed by single-point calculations at the CCSD(T) level to verify the negative barriers at the reactant entrance, because the reaction path should be less sensitive to spin contamination.<sup>37</sup> Therefore, we expect that the CCSD(T) calculations reported in this work are, in this regard, reliable.

The thermodynamic functions ( $\Delta H$ ,  $\Delta S$ , and  $\Delta G$ ) were estimated within the ideal gas, rigid-rotor, and harmonic oscillator approximations. A temperature of 298.15 K and a pressure of 1 atm were assumed.

## 3. Results and Discussion

For the present  $C_3H_6F$  system, there are enantiomorphs. For convenient and clear discussion, we only present moieties of reaction channels that have enantiomorphic varieties and

some important enantiomorphous structures. Six intermediate complexes (loose structure), six intermediate isomers, 11 products, and 21 transition states are obtained at the UMP2-(FULL)/6-311++G(d,p) or UMP2(FULL)/6-311G level of theory. The structures of the reactants, intermediate complexes and isomers, transition states, and products are depicted in Figure 1. The symbol **TS $x$ - $y$**  is used to denote a transition state, where  $x$  and  $y$  are the corresponding isomers or products. By means of the interrelation among the reactants, complexes, isomers, transition states, and products as well as the corresponding relative energies, the schematic profiles of the potential energy surface are depicted as shown in Figure 2. The schematic map of reaction mechanisms and activation energies of the title reaction is shown in Scheme 1. The direct H-abstraction manners are shown in Scheme 2. The calculated energetic data ( $\Delta E$ ,  $\Delta H$ ,  $\Delta S$ , and  $\Delta G$ ) of various products, complexes, isomers, and transition states and values obtained experimentally are listed in Table S1 as Supporting Information. It should be noted that the energy of **F** + propene is set at zero as a reference for other species for convenient discussion. As shown in Table S1, it is clear that the results calculated at the CCSD(T)/cc-pVTZ//UMP2-(FULL)/6-311++G(d,p) level of theory are in good agreement with the experimental values. Some apparent inconsistencies appearing in Figure 2 and in Table S1 in this work arise from the use of single-point spin-projected energies<sup>34</sup> (see, e.g., in Figure 2a, where **TS1-3** lies 2.27 kcal/mol below **I1**, etc.). We checked that all of the unprojected energy profiles are topologically consistent. In addition, the harmonic vibrational frequencies of some important species are given in Table S2 as Supporting Information. A comparison between the theoretical results and available experimental data for the reactant propene, product HF, C<sub>2</sub>H<sub>3</sub>F, and C<sub>2</sub>H<sub>4</sub> is also presented in Table S2 (Supporting Information).

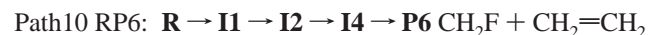
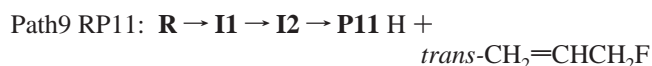
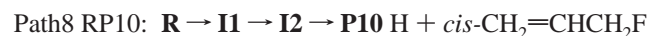
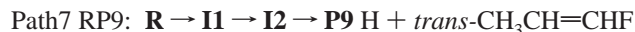
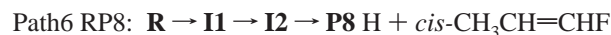
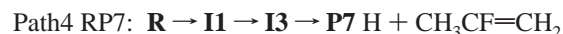
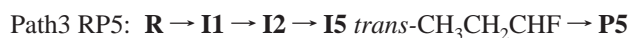
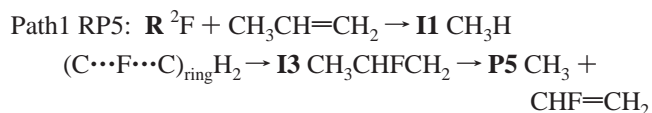
**3.1. Analysis of Reaction Mechanism of Atomic Radical F with Propene.** As the atomic radical **F** (“•” denotes the unpaired single electron) can have either a direct abstraction or an addition mechanism, two distinguishable types of initial attacks have been revealed for the radical-molecule reaction **F** + propene, namely, the picking-up on hydrogen (direct hydrogen abstraction) and the attack on the C=C double bond (addition–isomerization–elimination). A detailed discussion on the title reaction mechanisms as shown in Figure 2 is given as follows.

**3.1.A. Addition–Isomerization–Elimination Reaction Mechanism.** First, the atomic radical **F** (<sup>2</sup>F) and propene (CH<sub>3</sub>CH=CH<sub>2</sub>) approach each other forming a weakly bound complex **1** (**I1**). Complex **1** contains two C•••F bonds (see Figure 1), which stabilize the complex by 9.30 kcal/mol relative to the reactant **R**, F + CH<sub>3</sub>CH=CH<sub>2</sub> (0.0 kcal/mol). It is obvious that this addition process is a barrier-free association.<sup>15,21</sup> Then, complex **1** can undergo two concerted three-center F shifts; one of the F shifts is the C1•••F bond being shortened and the C2•••F bond being elongated (**TS1-2**) to produce the low-lying intermediate isomer **2**, CH<sub>3</sub>CHCHF (**I2**, -42.65), with large exothermicity, while the other is the C1•••F bond being elongated and the C2•••F bond being shortened, leading to another low-lying intermediate isomer **3**, CH<sub>3</sub>CFCH<sub>2</sub> (**I3**, -43.96), with larger exothermic-

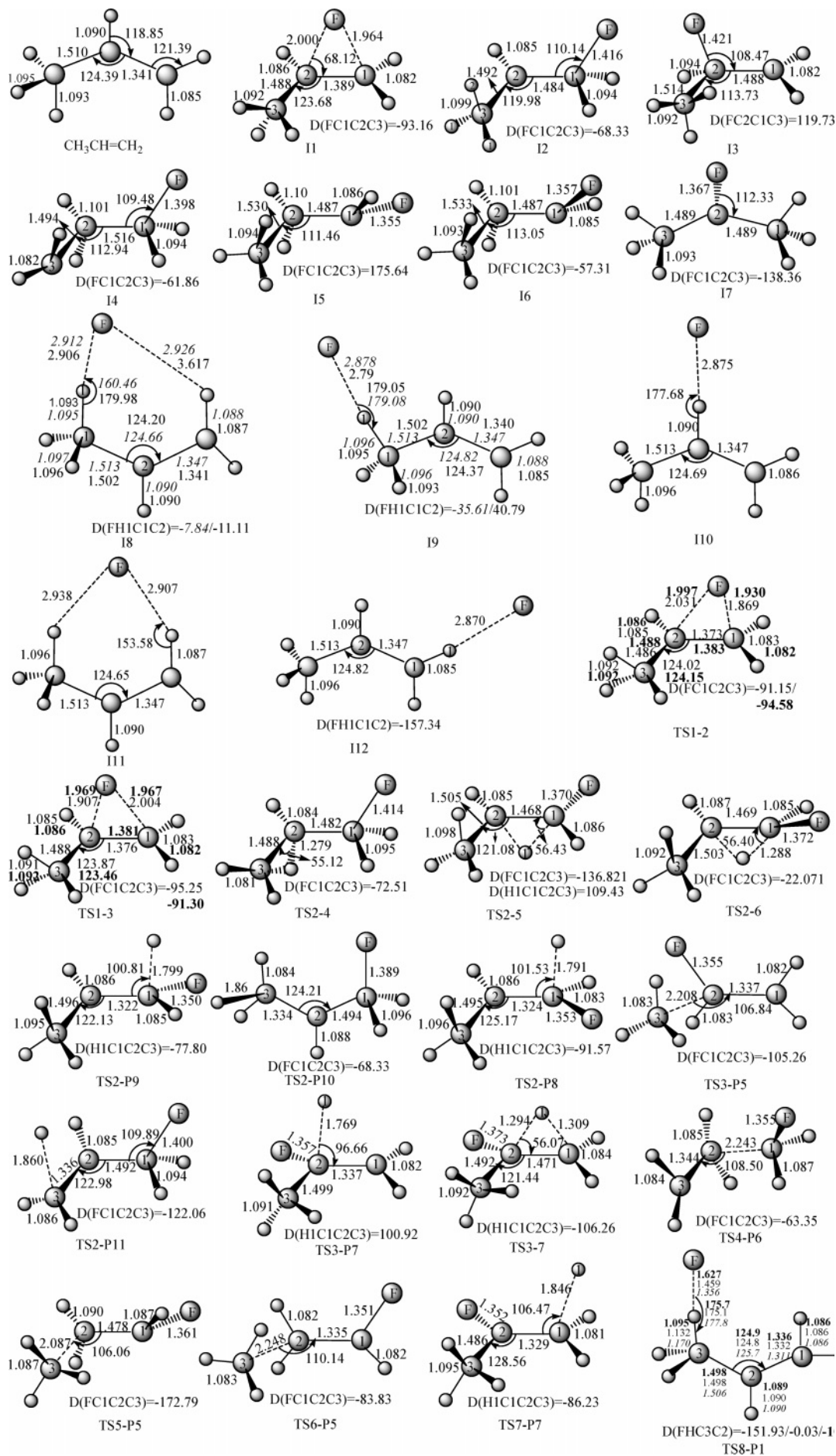
ity. The **I1** → **I2** and **I1** → **I3** energy barriers are negative, -2.11 and -2.27 kcal/mol, respectively, at the CCSD(T)/cc-pVTZ//UMP2/6-311++G(d,p) level of theory. IRCMax-(CCSD(T)//UMP2)<sup>37</sup> calculations confirmed these connections, in which **TS1-3** and **TS1-2** lie -1.13 and -1.28 kcal/mol lower in energy than **I1**, respectively. As was indicated earlier, most radical additions to the unsaturated carbon–carbon double bond have negative activation barriers.<sup>15,21</sup> Thus, the presence of **I1** on the PES is mandatory, for the topological consistency of the PES of the title reaction. To our knowledge, it is the first time that the existence of a prereactive complex in the title reactions has been reported.

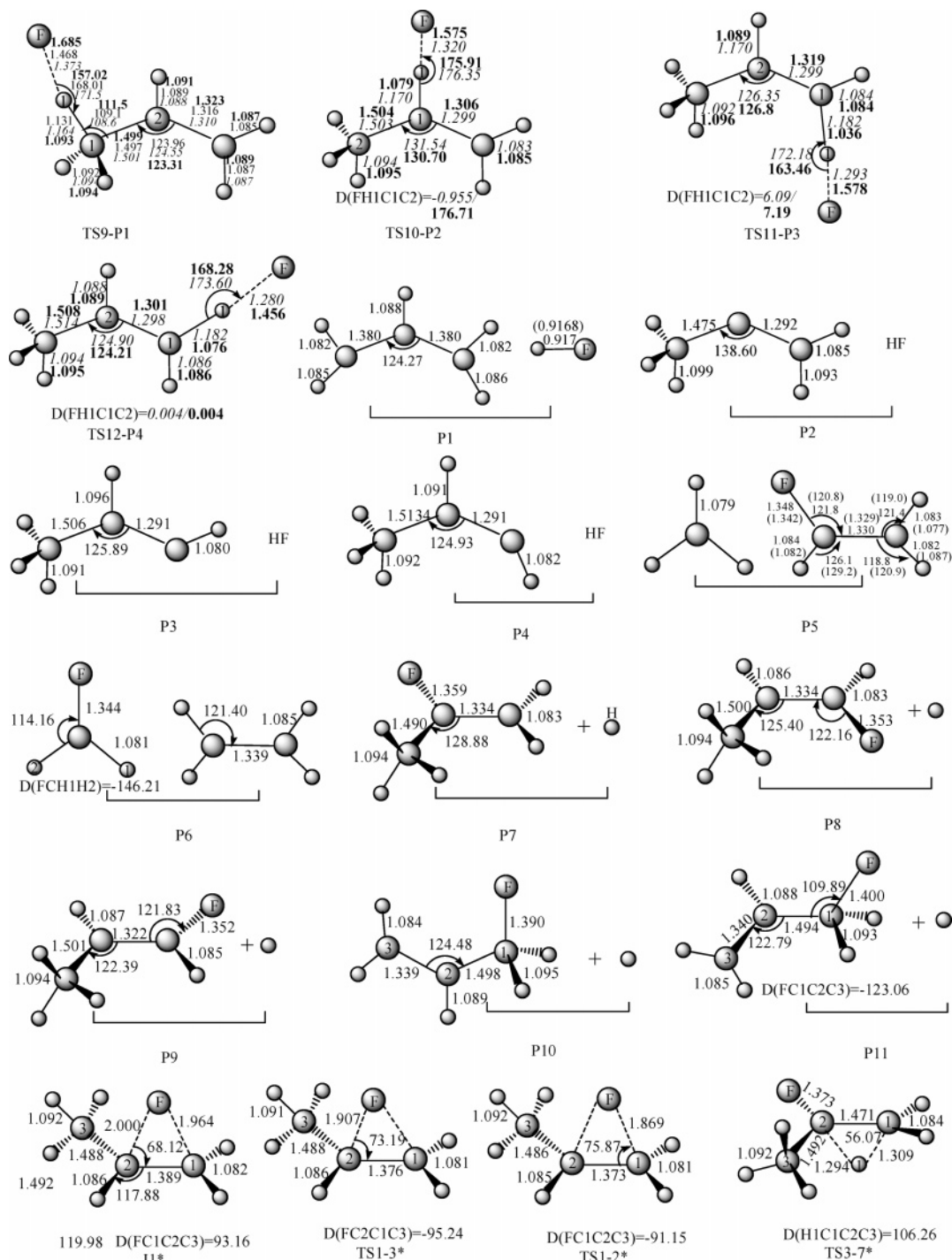
The overall processes **R** → **I1** → **I2** and **R** → **I1** → **I3** are both associated with the addition of the atomic radical **F** to the C=C double bond of CH<sub>3</sub>CH=CH<sub>2</sub>. They are typical **F**-addition mechanisms due to the electron-rich double bond in propene and the strong electronegativity of the F atom. As shown in Figure 1, the structural features of the addition transition states **TS1-2** and **TS1-3** are more similar to those of complex **1** than those of isomers **2** and **3**, respectively. It is consistent with the Hammond postulate, which states that a transition state will be structurally and energetically similar to the species (reactant, intermediate, or product) nearest to it on the reaction path. This is also in good agreement with the conclusion obtained by Yang and Zhang,<sup>15</sup> who have employed the molecular intrinsic characteristic contour and the electron density map methods.

With a large amount of heat released from the addition processes, isomers **2** and **3** can easily undergo 10 isomerization and dissociation pathways that can be expressed as follows:



In terms of small species in products, the title reaction has three different channels in the addition–isomerization–elimination reaction mechanism, that is, (I) the CH<sub>3</sub> radical formation channels, paths 1–3; (II) the H-atom formation channels, paths 4–9; and (III) the CH<sub>2</sub>F radical formation channels, that is, path 10.

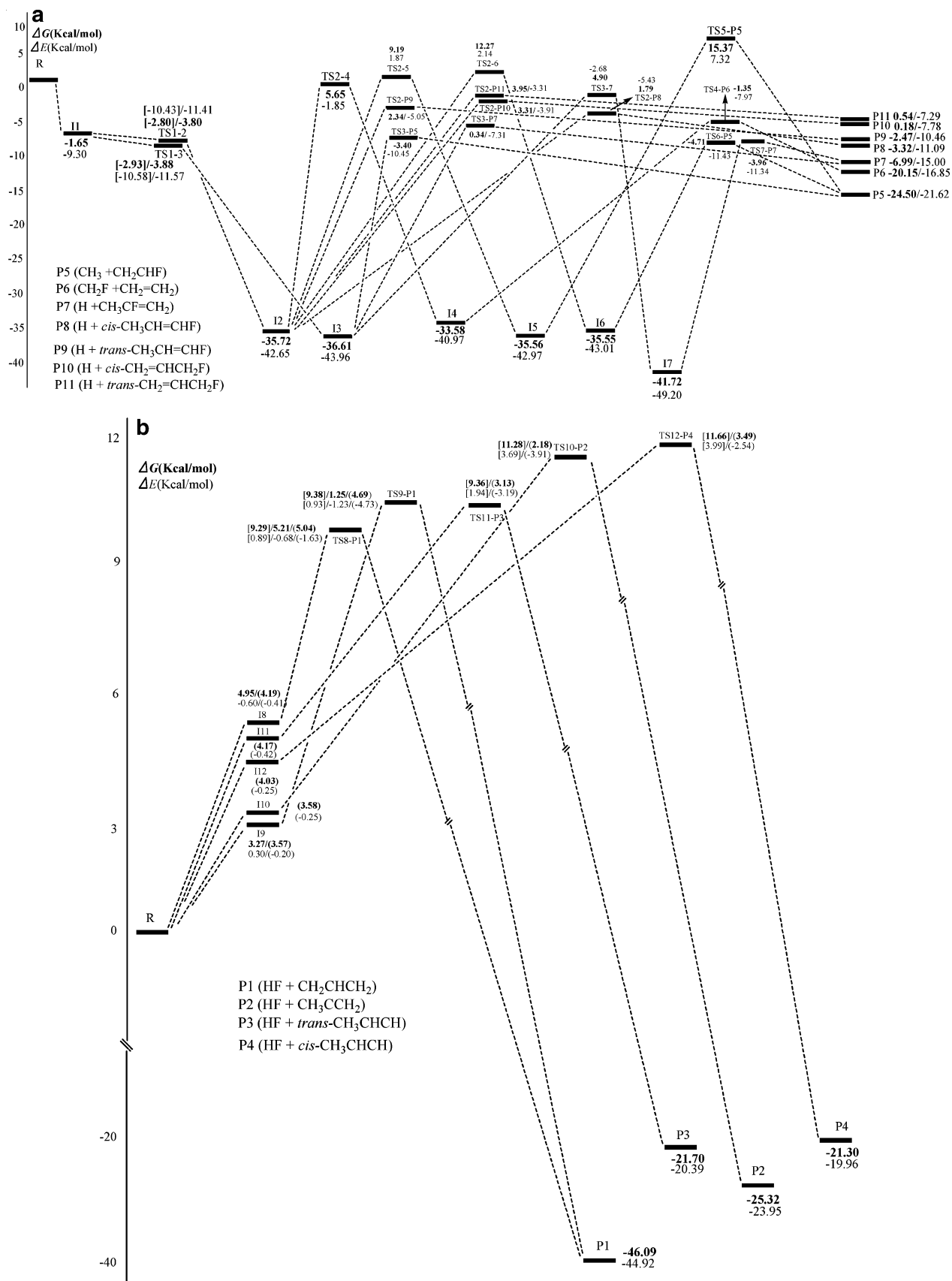




**Figure 1.** Optimized geometries of the reactants, the complexes, the doublet intermediate isomers, the products, and the transition states for the F + propene reaction. Bond distances are in angstroms, and angles are in degrees. Numbers in roman type show the structures at the UMP2/6-311++G(d,p) level of theory. Numbers in italics show the structures at the UMP2/6-311G level of theory. Numbers in bold show the structures at the Max{CCSD(T)}/IRC{UMP2} + ZPE level of theory. Numbers in parentheses are the experimental values (ref 38 for HF, ref 39 for CH<sub>2</sub>=CHF). Symbols with asterisks indicate enantiomorphs corresponding to the same symbols.

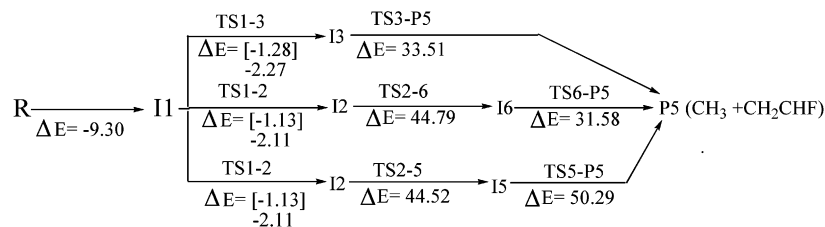
3.1.A.I. The CH<sub>3</sub> Radical Formation Channel. Clearly, from complex **1**, CH<sub>3</sub>H(C···F···C)<sub>ring</sub>H<sub>2</sub> (-9.30), the pathway Path1 RP5 can reach the products **P5**, CH<sub>3</sub> + CHF=CH<sub>2</sub>, easily by going through transition state **TS1-3** (-10.58), intermediate isomer **I3**, CH<sub>3</sub>CHFCH<sub>2</sub> (-43.96), and **TS3-P5** (-10.45). As shown in Figure 1, the geometry of transition state **TS3-P5** calculated at the UMP2/6-311++G(d,p) level is found to have an elongated C2···C3 bond length

( $r_{C2-C3} = 2.208 \text{ \AA}$ ). Most importantly, this geometry is indicative of a "late" transition state, corresponding to a C-C distance significantly longer than the equilibrium bond length (1.510 Å in propene). Because all of the energies of the transition states and isomers in pathway Path1 RP5 are lower than that of the reactants, the rate of this pathway should be very fast. Furthermore, because the pathway Path1 RP5 has no transition state in the first reaction step and tight structure

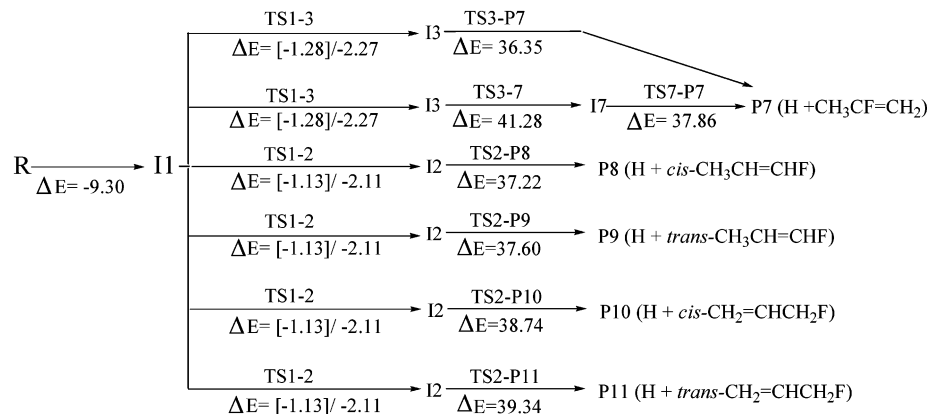
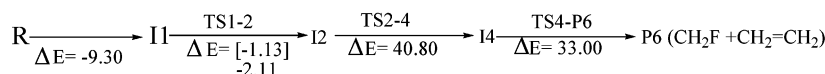


**Figure 2.** (a) Schematic potential energy surfaces of the addition–isomerization–elimination channels in the F + propene reaction. (b) Schematic potential energy surfaces of the hydrogen abstraction channels in the F + propene reaction. Numbers in roman type are relative energies at the CCSD(T)/cc-pVTZ//UMP2/6-311++G(d,p) + ZPE level of theory. Numbers in bold are relative Gibbs energies at the CCSD(T)/cc-pVTZ//UMP2/6-311++G(d,p) + ZPE level of theory. Numbers in parentheses are values at the CCSD(T)/cc-pVTZ//UMP2/6-311G + ZPE level of theory. Numbers in square brackets are values at the Max{CCSD(T)}/IRC{UMP2} + ZPE level of theory.

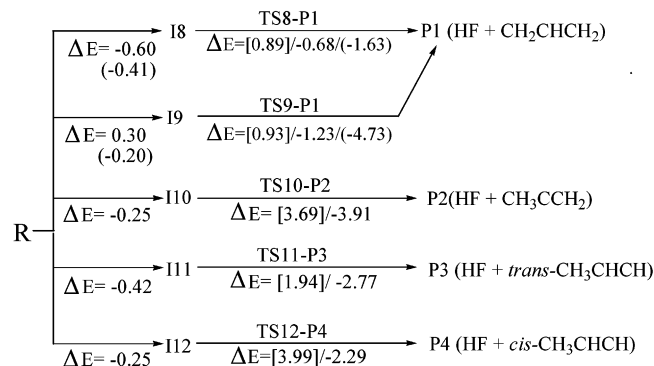


**Scheme 1.** Reaction Mechanism and Activation Energies (kcal/mol) of the Reaction of F with Propene<sup>a</sup>The CH<sub>3</sub> Formation Channels

The H Formation Channels

The CH<sub>2</sub>F Formation Channel

The HF Formation Channels



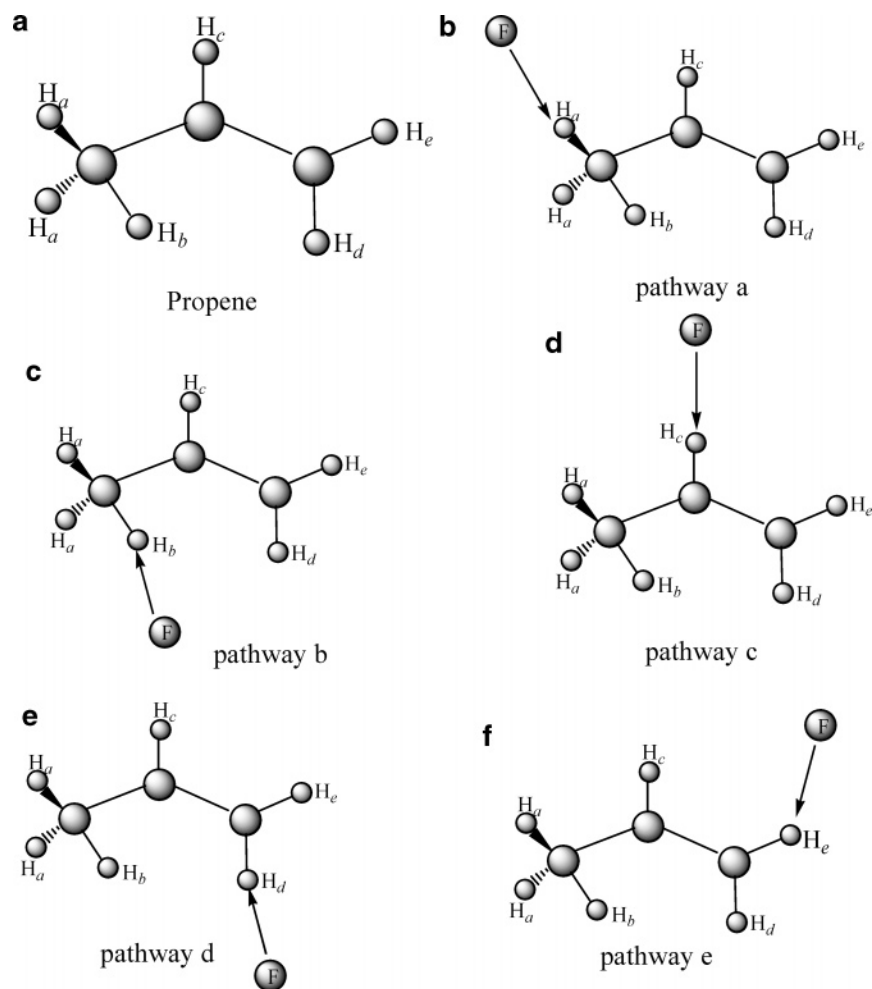
<sup>a</sup> Activation energies are calculated at the CCSD(T)/cc-pVTZ//UMP2/6-311++G(d,p) + ZPE (numbers in roman type), CCSD(T)/cc-pVTZ//UMP2/6-311G + ZPE (numbers in parentheses), and Max{CCSD(T)}//IRC{UMP2} + ZPE (numbers in square brackets) levels of theory.

isomers and transition states to the products, this pathway should possess the character of negative temperature dependence.

In pathway Path2 RP5, a successive 1,2-H-shift and dissociation steps **I2** → **I6** → **P5** can lead to product **P5**. It is an H-shift–methyl-elimination mechanism. Because the rate-determining energy barrier (44.79) of Path2 RP5 is higher than the ones of Path1 RP5 (33.51), pathway Path2 RP5 should be less competitive than pathway Path1 RP5. For Path3 RP5, similar successive 1,2-H-shifts and dissociation steps **I2** → **I5** → **P5** can also lead to product **P5**. As shown in Figure 1, the geometries of transition states **TS2–6** (2.14) and **TS2–5** (1.87) belong to a cis–trans configuration relationship and are energetically close and high-lying. Yet, the rather high energy barrier (50.29) of **TS5–P5** (7.32)

makes the pathway Path3 RP5 process thermodynamically prohibited and far from competitive with pathways Path1 RP5 and Path2 RP5 under normal conditions.

3.1.A.II. The H-Atom Formation Channel. There are two pathways (Path4 and Path5) through which complex **1** can transform to the product **P7**. As shown in Figure 1, the transition structure **TS3–7** in Path5 RP7 is a hydrogen migration transition state. The relative energy of **TS3–7** is 4.63 kcal/mol higher than that of **TS3–P7**. Both the relative energies of rate-determining transition states and the simplicity considerations support the viability of pathway Path4 RP7. On the other hand, it is obvious that pathway Path4 RP7 is similar to pathway Path1 RP5 because pathway Path4 RP7 has the same steps of **R** → **I1** → **I3** as pathway Path1 RP5. The difference of the two pathways is how intermediate

**Scheme 2.** The Five Different Chemical H Atoms in Propene and the Five Possible Direct Abstraction Manners in the HF Formation Channel

isomer **3** dissociates to relevant educts. In pathway Path1 RP5, **I3** can reach the product **P5** via surmounting the transition state **TS3–P5** (–10.45), a 33.51 kcal/mol energy barrier, while in pathway Path4 RP7, **I3** dissociates via overcoming the transition state **TS3–P7** (–7.31), a 36.35 kcal/mol energy barrier, leading to the product **P7**. The difference of the two pathways may lead us to the fact that pathway Path4 RP7 is less competitive than pathway Path1 RP5.

For the other four pathways, Path6 through Path9, which have the same steps of **R** → **I1** → **I2**, complex **1** (–9.30) transforms to intermediate isomer **2** (–42.65) via transition state **TS1–2** (–10.43) jointly, and subsequently, isomer **2** transforms to four different couples of educts **P8**, H + *cis*-CH<sub>3</sub>CH=CHF (–11.09); **P9**, H + *trans*-CH<sub>3</sub>CH=CHF; **P10**, H + *cis*-CH<sub>2</sub>=CHCH<sub>2</sub>F; and **P11** H + *trans*-CH<sub>2</sub>=CHCH<sub>2</sub>F via **TS2–P8** (–5.43), **TS2–P9** (–5.05), **TS2–P10** (–3.91), and **TS2–P11** (–3.31) via surmounting 37.22, 37.60, 38.74, and 39.34 kcal/mol energy barriers, respectively. As shown in Figure 1, their difference is how intermediate isomer **2** transforms to four different couples of educts, that is, the H elimination of the four different chemical sites of the H atom from intermediate isomer **2** (site specificity). From the thermodynamic stability of products and the rate-determining reaction steps, we can safely draw the conclusion that, for the six pathways in the H-atom formation channel mentioned

above, the order of reaction competition abilities is Path4 RP7 > Path6 RP8 > Path7 RP9 > Path8 RP10 > Path9 RP11 > Path5 RP7.

According to our calculations, an unambiguous conclusion can be drawn to give a rational interpretation of the site-specific effect in the H-atom formation channel. As shown in Figure 1, **TS3–P7**, **TS2–P8**, **TS2–P9**, **TS2–P10**, and **TS2–P11** are the C2–H elimination of **I3**, CH<sub>3</sub>CHF=CH<sub>2</sub>; the C1–H<sub>t</sub> elimination of **I2**, CH<sub>3</sub>CH=CH<sub>2</sub>F; the C1–H<sub>c</sub> elimination of **I2**; the C3–H1 elimination of **I2**; and the C3–H2 elimination of **I2**, respectively, where the subscripts “t” and “c” denote the trans and cis H atoms of C1–H of **I2**, respectively. Because the pathway Path4 RP7 is the most feasible pathway in the H-atom formation channel, CH<sub>3</sub>CF=CH<sub>2</sub> may be the main contribution in the C<sub>3</sub>H<sub>5</sub>F product. The contributions of *cis*-CH<sub>3</sub>CH=CHF, *trans*-CH<sub>3</sub>CH=CHF, *cis*-CH<sub>2</sub>=CHCH<sub>2</sub>F, and *trans*-CH<sub>2</sub>=CHCH<sub>2</sub>F decrease orderly in terms of the competition abilities of their pathways.

3.1.A.III. The CH<sub>2</sub>F Radical Formation Channel. As shown in Figure 2a, the pathway Path10 RP6 is the only pathway leading to product **P6**, CH<sub>2</sub>F + CH<sub>2</sub>=CH<sub>2</sub>, in the CH<sub>2</sub>F radical formation channel. In pathway Path10 RP6, the characteristic steps are **I2** → **I6** → **P6**, CH<sub>2</sub>F + CH<sub>2</sub>=CH<sub>2</sub>; a three-center H migration; and a C–C bond rupture process. We should mention that the C–C rupture process in the CH<sub>2</sub>F radical formation channel is obviously different from

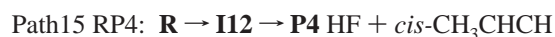
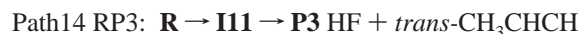
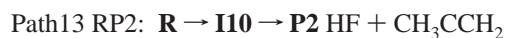
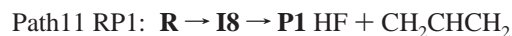
the C–C rupture processes in the CH<sub>3</sub> radical formation channel. The former process occurs between C2 and C1, while all of the latter processes occur between C2 and C3 in the C<sub>3</sub>H<sub>6</sub>F system as shown in Figure 1. In terms of the formation heat of the CH<sub>2</sub>F radical formation channel being a few kilocalories per mole more than that of the H-atom formation channel and the relative energy of rate-determining transition state **TS2–4** (–1.85) lying above those of **TS3–P7** (–7.31), **TS2–P8** (–5.43), **TS2–P9** (–5.05), **TS2–P10** (–3.91), **TS2–P11** (–3.31), and **TS4–P6** (–7.97), we conjecture that the CH<sub>2</sub>F radical formation channel is less favorable than the H-atom formation channel in low-temperature ranges, while the latter should be competitive with the former under high-temperature conditions. Thus, the actual feasibilities of these two channels may depend on the reaction conditions in the experiments.

To briefly summarize, the CH<sub>3</sub> radical formation channel is the most competitive reaction channel, while the CH<sub>2</sub>F radical formation channel may be prohibited in low-temperature ranges among the three channels. The pathway Path1 RP5 is the most competitive pathway among all 10 of the pathways. The CH<sub>2</sub>F radical formation pathway Path10 RP6 may be more favorable than the H-atom formation pathway Path6 RP8 and Path7 RP9 in high-temperature ranges on the doublet PESs.

Now, it is requisite to compare the addition–isomerization–elimination PES feature of the title reaction with that of the analogous reaction F + ethylene, a detailed theoretical investigation on the formation of the intermediate complex, and a subsequent unimolecular decomposition.<sup>15</sup> Theoretical calculations show that there exist both similarity and discrepancy. In general, the addition–isomerization–elimination PES features of the two analogous reactions are very similar. First, both reactions involve the same initial association pathway; that is, the fluorine atom approaches ethylene/propene, perpendicular to the C=C double bond, forming the prereactive complex directly at the reaction entrance. The prereactive complex transforms to the fluoroethyl/fluoropropyl radicals via the addition transition structures with no barrier. Second, both radicals, fluoroethyl and fluoropropyl, can eliminate the H atom directly via surmounting 37~39 kcal/mol barriers or indirectly by intramolecular H shifts with 41~45 kcal/mol barriers. Furthermore, by comparison of the minima and transition structures in F + propene with the counterpoints in F + ethylene, we can find that their relative energies as well as the barrier heights vary little. For instance, according to Zhang and Yang's calculations,<sup>15</sup> the relative energies of the prereaction complex (**I<sub>add</sub>**), the addition transition state (**TS<sub>add</sub>**), and the fluoroethyl radical (CH<sub>2</sub>FCH<sub>2</sub>) in F + ethylene reaction are –9.30, –11.68, and –41.49 kcal/mol at the CCSD(T)/aug-cc-pVDZ//CCSD/6-31G(d,p) + ZPE level, which are in good agreement with that of **I1** (–9.30), **TS1-3** (–11.57), **TS1-2** (–11.41), **I3** (–43.96), and **I2** (–42.65) in the C<sub>3</sub>H<sub>6</sub> + F reaction at the CCSD(T)/cc-pVTZ//UMP2/6-311++G(d,p) + ZPE level. On the other hand, the main discrepancies lie in the two folds: (a) an F-shift transition structure exists on the PES of F + ethylene, while it cannot be located on that of F + propene, and (b) the reaction pathways and possible products of F + propene

are more than that of F + ethylene. It is obvious that the reaction of F + propene involves all of the main features of the F + ethylene reaction. However, from the discussion above, one can find that the addition–isomerization–elimination PES of F + propene is more complicated than that of F + ethylene because a hydrogen atom of ethylene is substituted by methyl in propene.

*3.1.B. Direct Hydrogen Abstraction Reaction Mechanism.* Scheme 2 shows the five possible pathways through which the abstraction reaction proceeds. For CH<sub>3</sub>CH=CH<sub>2</sub>, there are five different chemical environments of the H atom, namely, allylic hydrogen (H<sub>a</sub> and H<sub>b</sub>), vinyl inner hydrogen (H<sub>c</sub>), and vinyl terminal hydrogen (H<sub>d</sub> and H<sub>e</sub>) as shown in Scheme 2. Pathways a and b correspond to the abstraction of the two nonequivalent allylic hydrogens. In pathway c, the vinyl inner hydrogen is the one involved in the abstraction reaction, whereas pathways d and e represent the abstraction of the vinyl terminal hydrogen. For convenient discussion, we list them as follows:



As shown in Figure 1, five loosely bonded intermediate complexes (**I8–I12**) prior to the formation of the transition states along the reaction coordinate are located in the first step of this mechanism involved in the atomic radical **F** attacking the H atom of propene. However, as shown in Figure 2b, all of the complex formation processes are endothermic at the CCSD(T)//MP2 level of theory. This implies that such complexes might be short-lived and not important on the PES<sup>41</sup> and can be safely ruled out. On the other hand, one can find that all five of the abstraction hydrogen processes involve negative barriers at CCSD(T)//MP2 levels. In contrast, the IRCMax method consistently gives positive barriers that differ from the CCSD(T)//MP2 levels by 3~8 kcal/mol with sizable displacements along the reaction path ( $0.6 < S < 1.2$ , see Table S3, Supporting Information). We expect that large errors are due to poor transition state geometries and positions along the IRCs at a low level of theory and thus have considerable effects on the calculated energy barriers. Furthermore, the barriers involved are moderate ( $\Delta E < 4$  kcal/mol;  $\Delta G < 12$  kcal/mol), and the formations of the final products are exergonic ( $\Delta G > -21$  kcal/mol). Therefore, both kinetic and thermodynamic considerations support the viability of such channels.

Let us now analyze the low-temperature limiting behavior. Figure 2b shows that the rate-determining barriers of the five hydrogen abstraction channels are less than 4 kcal/mol at 0 K, especially for allylic hydrogen abstraction (**TS8–P1** and **TS9–P1**,  $\Delta E < 1$  kcal/mol). Therefore, such processes may proceed in low-temperature ranges (the existence of a small barrier cannot be discarded as the errors in the calculation

ZPE because, for complexes with very small barriers, ZPE corrections are problematic, where the harmonic approximation does not work properly).

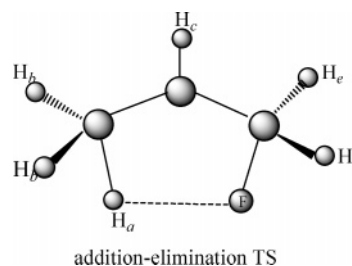
In pathway Path11 RP1, corresponding to the direct abstraction from the *b*-allylic hydrogen of  $\text{CH}_3\text{CH}=\text{CH}_2$  as shown in Figure 1 and Scheme 2, the geometry of transition state **TS8-P1** calculated at the UMP2/6-311++G(d,p) level of theory is found to have a bent structure ( $\angle_{\text{F-H-C}} = 175.1^\circ$ ) and an elongated H–F bond length ( $r_{\text{F-H}} = 1.459 \text{ \AA}$ ;  $r_{\text{C-H}} = 1.132 \text{ \AA}$ ). Most importantly, this geometry is indicative of an “early” transition state, corresponding to a H–F distance significantly longer than the equilibrium bond length (0.917 Å), specifically falling between the HF classical outer turning points of  $\nu = 4$  (1.318 Å) and  $\nu = 5$  (1.381 Å).<sup>42</sup> The calculated heat of formation at 0 K (–44.92 kcal/mol) is in good agreement with the experimental value (–48.4 kcal/mol).<sup>5</sup> As shown in Figure 2b, this hydrogen abstraction channel is an exergonic process globally ( $\Delta E = -44.92 \text{ kcal/mol}$ ), involving a small barrier ( $\Delta E = 0.89 \text{ kcal/mol}$ ) at the Max{CCSD(T)/cc-pVTZ}/IRC{UMP2/6-311++G(d,p)} + ZPE level of theory. Therefore, both kinetic and thermodynamic considerations support the viability of this channel.

The *a*-allylic hydrogen abstraction pathway Path12 RP1 is totally similar to that described above for *b*-allylic hydrogen abstraction. And the relative energy of **TS9-P1** (0.93) is only 0.04 kcal/mol higher than that of **TS8-P1** (0.89) at the Max{CCSD(T)/IRC{UMP2}} level of theory. The hairlike difference of the two pathways may lead us to the fact that pathway Path12 RP1 is less favorable than pathway Path11 RP1 appreciably.

Pathways Path13 through Path15 (see Figure 1 and Scheme 2) correspond to the possible abstraction of the vinyl hydrogens of propene. We should mention that the three transition states **TS10-P2**, **TS11-P3**, and **TS12-P4** are located at the UMP2/6-311G level of theory. Despite numerous attempts, we cannot locate these transition states at the higher levels of theory UMP2/6-311G(d,p) and UMP2/6-311++G(d,p). As shown in Table S1 (Supporting Information), the energy barriers and products in pathways Path13 through Path15 are higher than those in pathways Path11 through Path12 RP1. Therefore, both kinetic and thermodynamic considerations may rule out the significance of pathways Path13 through Path15.

We should mention that we also reoptimized **TS8-P1** and **TS9-P1** at the UMP2/6-311G level. It is worth noting that the comparison of geometries obtained at the UMP2/6-311G level with those at the UMP2/6-311++G(d,p) level shows that the discrepancies are considerably large. For example, both of the forming  $\text{F}\cdots\text{H}$  bond lengths in the two transition structures obtained at the UMP2/6-311G level are about 0.1 Å shorter than that at the UMP2/6-311++G(d,p) level. Likewise, both of the breaking  $\text{C}\cdots\text{H}$  bond lengths in them at a low level are 0.03–0.04 Å shorter than that at a high level. Fallaciously, the transition state **TS8-P1** is energetically high-lying among the five abstraction hydrogen transition states at the CCSD(T)/cc-pVTZ//UMP2/6-311G level, inconsistent with the results obtained at a higher level of theory (see Table S1, Supporting Information). To rationalize this difference, we turn to the PES features of the hydrogen

Chart 1



abstraction processes. The use of the low-level transition state geometry generally gives poor results, and in the five instances, this method actually gives *negative* barriers! “When the low level transition states occurs very late the higher level energy can already have dropped below the energy of the reactants”.<sup>37</sup> Furthermore, bearing in mind the limitation (the use of harmonic frequencies for the highly anharmonic intramolecular vibrations) inherent in our calculations and considering Yang and Zhang’s suggestion<sup>15</sup> (the 6-311G basis set is insufficient for the title reaction), the CCSD(T)/cc-pVTZ//UMP2/6-311G scheme is unreliable. Expectedly and interestingly, as shown in Figure 1, the distances of  $\text{H}\cdots\text{F}$  are 1.356, 1.373, 1.320, 1.293, and 1.280 Å in **TS8-P1**, **TS9-P1**, **TS10-P2**, **TS11-P3**, and **TS12-P4**, respectively, decreasing gradually, which is in line with the trend of their relative energies except for **TS8-P1**.

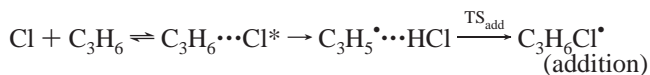
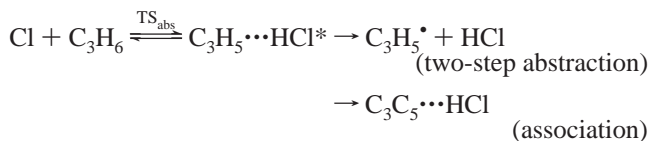
It must be stressed that an exhaustive search on the PES failed to locate a transition structure like that in Chart 1 (corresponding to a conventional addition–elimination mechanism. Indeed, in previous theoretical studies<sup>21,43</sup> on the reactions of the Cl atom with isoprene and propene, this ringlike transition structure wasn’t located on the PESs. Therefore, the addition–elimination mechanism can be safely ruled out from the **F + propene** reaction.

**3.2. Experimental Implication.** Recently, Ran et al. carried out the kinetic study for the title reaction.<sup>5</sup> In their experiment, three reaction channels were observed:  $\text{CH}_3 + \text{C}_2\text{H}_3\text{F}$ ,  $\text{H} + \text{C}_3\text{H}_3\text{F}$ , and  $\text{HF} + \text{C}_3\text{H}_5$ . They also suggested that the  $\text{CH}_3$  formation is the most important channel for this chemical reaction, while both H and HF formation channels are also significant, and a long-lived complex formation is the most important mechanism with some minor contributions from a direct reaction mechanism. However, they did not provide unambiguous information on the site-specific effect in the H-atom elimination channel and the HF formation channel because there are three chemically different hydrogen sites in the propene molecule, although it is important to better understand the roles of different reaction mechanisms in these reaction channels. On the other hand, they did not give any information about the formation of  $\text{CH}_2\text{F} + \text{CH}_2=\text{CH}_2$ . In our calculations, it is shown that, among the possible pathways, the most feasible one is that where the atomic radical F and propene approach each other to form the initial weakly bound complex **1** before F addition to the C=C double bond. The elimination of methyl from isomer **3** leads to the main products **P5**,  $\text{CH}_3 + \text{C}_2\text{H}_3\text{F}$ . And, one can find that the atomic radical F directly abstracting the allylic hydrogen of  $\text{CH}_3\text{CH}=\text{CH}_2$  is the most favorable

pathway, while the abstraction of the vinyl terminal hydrogen should be less competitive in the HF formation channels. Furthermore, the most favorable eliminated H atom is the vinyl inner hydrogen ( $H_c$ ) after the **F** addition to the vinyl inner carbon in the H formation channels.

Finally, it should be pointed out that, in the title reaction, all of the intermediates and transition states involved in the possible addition–isomerization–elimination channels lie below the reactants. Once the addition reaction proceeds, the reaction system immediately enters a deep potential well. As shown in Figure 2a, the exit barriers are moderately high for all of the addition–isomerization–elimination channels. Undoubtedly, these features will make the formation of long-lived reaction complexes stand a good chance. This is qualitatively consistent with the experimental results.<sup>5</sup>

**3.3. Comparison with the Cl + Propene Reaction.** To give a deeper understanding of the reaction mechanism of **F** + propene, it is worthwhile to compare the title reaction with the analogous reaction **Cl** + propene, which has been extensively studied both experimentally<sup>44–46</sup> and theoretically.<sup>21</sup> One can find that there exists both similarity and discrepancy. In a recent experiment<sup>46</sup> for the **Cl** + **propene** reaction, Kaiser and Wallington observed experimentally that the addition of a Cl atom to the double bond is the major reaction channel at high pressure (>100 Torr), while at low pressure (<10 Torr), the addition reaction rate slows and the removal of an H atom from the  $CH_3$  group in  $C_3H_6$  is the dominant channel. They suggest that, at lower pressure, the allylic radical may be formed by two processes: (1) a pressure-independent direct abstraction and (2) an addition–elimination process which depends inversely on the total pressure. However, in a recent detailed theoretical investigation, based on the structures on the PESs, Braña and Sordo<sup>21</sup> suggest that the previously established mechanism<sup>44</sup> consisting of direct abstraction and addition–elimination steps is instead made up of addition through an intermediate and two-step abstraction step processes; that is, no direct abstraction mechanism exists on the potential energy surface. They proposed a formally similar general mechanism that can be written as follows:



In general, the potential energy surface features of the two analogous reactions **X** + propene (**X** = **F** and **Cl**) are very similar. Both reactions involve the same initial association pathway; that is, the halogen atoms and unsaturated hydrocarbons approach each other, forming the initial weakly bound complex before halogen-atom addition to the  $C=C$  double bond with no barrier. Furthermore, the types of the five different chemical environments of the H atom picked up by the halogen atoms are identical in the abstraction processes. Theoretically, the allylic hydrogen abstractions

by both **F** and **Cl** are exergonic processes and involve small barriers. Therefore, both kinetic and thermodynamic considerations support the viabilities of this type of H atom picked up. However, because the **F** atom with higher electronegativity strongly attracts a lone pair of electrons located at the H atom, the electron density on the H atom is reduced, which leads to an increasing of the reactivity of bond scission between adjacent atoms. Thus, it is obvious that, for the **Cl** + **propene** reaction, the H-abstraction channel is a two-step abstraction process, while for the **F** + **propene** reaction, the H-abstraction channel belongs to a direct abstraction mechanism typically. On the other hand, the increasing reactivity of atomic radical **F** also leads to the mechanism discrepancies as well as the different product distribution between the two reactions. First, for the **Cl** + **propene** reaction, the product  $CH_3 + C_2H_3Cl$  cannot be obtained experimentally. While, for the **F** + **propene** reaction, the  $C_2-C_3$  bond cleavage of **I3** to yield the main product  $CH_3 + C_2H_3F$  plays an absolute role.<sup>5</sup> Second, for the **Cl** + **propene** reaction, no literature reported the H-atom formation channel either experimentally or theoretically. However, for the **F** + **propene** reaction, the H-atom formation channel cannot be neglected because the branching ratio of the  $H + C_3H_5$  channel is about 15%. Third, we obtained a new feasible minor product **P6** ( $C_2H_4 + CH_2F$ ), which may play a role in high-temperature ranges for the **F** + **propene** reaction in this work, while no information about the product ( $C_2H_4 + CH_2Cl$ ) was reported experimentally or theoretically.

Before ending, we would like to make a short comment on the expected behavior of the rate constants at various experimental pressures about the title reaction. As pointed out by Braña and Sordo<sup>21,43</sup> and Kaiser and Wallington<sup>46</sup> on the analogous reactions of **Cl** + **propene** and **Cl** + **isoprene**, we conjecture that, for the **F** + propene reaction, even at high pressure, the activated complexes **I1** and **In** ( $n = 8-12$ ) cannot be stabilized efficiently by third-body collisions as the stabilization energies of these complexes are very small. Taking into account that the **TS1-2**, **TS1-3**, and direct H-abstraction transition states' (**TS8-P1**, **TS9-P1**, **TS10-P2**, **TS11-P3**, and **TS12-P4**) barriers are negligible, there will be no accumulation of **I1** and **In** ( $n = 8-12$ ) ( $C_3H_6 \cdots F$ ). As the formation intermediates **I2** and **I3** are exergonic processes and the energy barriers associated with **TS1-2** and **TS1-3** are much lower than that of direct H-abstraction transition states, consequently, the  $C_3H_6F^{\bullet}$  radical should be the most favored product. Therefore, the major reaction pathways should be the addition–isomerization–elimination reaction channel at the high-pressure limit. At low pressure, the addition–isomerization–elimination reaction rate decreases and the direct hydrogen abstraction reaction becomes the domination channel. Indeed, as a consequence of very small stabilization of **In** ( $n=8-12$ ) by third-body collisions,  $C_3H_5^{\bullet}$  production should increase as compared with that obtained under high-pressure conditions. At sufficiently low pressure, the stabilization channel to form the  $C_3H_6F^{\bullet}$  radical becomes insignificant. In this pressure regime, the rate constant for allyl formation from the direct abstraction channel might be a constant value.

#### 4. Conclusions

A detailed theoretical survey on the complicated doublet PESs of the reaction of atomic radical F with propene,  $\text{CH}_3\text{-CH=CH}_2$ , has been performed at the UMP2 and CCSD(T) levels of theory. Two different mechanisms and four reaction channels are revealed in the present study. The most competitive reaction channel is the  $\text{CH}_3$  radical formation channel. The  $\text{CH}_2\text{F}$  radical formation channel may be more favorable than the H-atom formation channel in high-temperature ranges on the doublet PESs. The pathway Path1 RP5 is the most competitive pathway in the addition–isomerization–elimination reaction mechanism. While in the direct hydrogen abstraction reaction mechanism, the main pathway is Path11 RP1. No addition–elimination mechanism exists on the potential energy surface, which is consistent with previous theoretical studies for the reactions of the Cl atom with propene and isoprene. On the basis of the analysis of the energetics of all channels through which the addition and abstraction reactions proceed, we expect that the actual feasibility of the reaction channels may depend on the reaction conditions in the experiment. Consequently, future experimental studies on the title reaction are highly desirable under various pressures and temperatures. The present work will provide useful information for understanding the processes of atomic radical F reaction with other unsaturated hydrocarbons. Despite the analogous chemical reactivity with the well-known atomic radical Cl, the barrier-free atomic radical F reaction with propene is expected to be of unique importance.

**Acknowledgment.** This work is supported by the National Natural Science Foundation of China (Nos. 20073014 and 20103003), Excellent Young Teacher Foundation of the Ministry of Education of China, Excellent Young Foundation of Jilin Province, and Technology Development Project of Jilin Province (No. 20050906-6). The authors thank Prof. Jose A. Sordo (Universidad de Oviedo) for helpful discussions. The authors are also thankful for the reviewers' invaluable comments.

**Supporting Information Available:** Table S1 gives the energies, entropies, enthalpies, and Gibbs free energies. Table S2 shows the vibrational frequencies of the reactant, some important intermediates, transition states, and products. The CCSD(T) single-point energies along the IRC are given in Table S3. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### References

- (1) Parson, J. M.; Lee, Y. T. Crossed Molecular Beam Study of  $\text{F} + \text{C}_2\text{H}_4, \text{C}_2\text{D}_4$ . *J. Chem. Phys.* **1972**, *56*, 4658–4666.
- (2) Parson, J. M.; Shobatake, K.; Lee, Y. T.; Rice, S. A. Unimolecular Decomposition of the Long-Lived Complex Formed in the Reaction  $\text{F} + \text{C}_4\text{H}_8$ . *J. Chem. Phys.* **1973**, *59*, 1402–1415.
- (3) Shobatake, K.; Parson, J. M.; Lee, Y. T.; Rice, S. A. Unimolecular Decomposition of Long-Lived Complexes of Fluorine and Substituted Mono-Olefins, Cyclic Olefins, and Dienes. *J. Chem. Phys.* **1973**, *59*, 1416–1426.
- (4) Farrar, J. M.; Lee, Y. T. The Question of Energy Randomization in the Decomposition of Chemically Activated  $\text{C}_2\text{H}_4\text{F}$ . *J. Chem. Phys.* **1976**, *65*, 1414–1426.
- (5) Ran, Q.; Yang, C.-H.; Lee, Y. T.; Shen, G.; Yang, X. Dynamics of the F Atom Reaction with Propene. *J. Chem. Phys.* **2004**, *121*, 6302–6308.
- (6) Ran, Q.; Yang, C. H.; Lee, Y. T.; Shen, G.; Wang, L.; Yang, X. Molecular Beam Studies of the F Atom Reaction with Propyne: Site Specific Reactivity. *J. Chem. Phys.* **2005**, *122*, 044307(1–8).
- (7) Manion, J. A.; Tsang, W. Hydrogen Atom Attack on Fluorotoluenes: Rates of Fluorine Displacement. *Isr. J. Chem.* **1996**, *36*, 263–273.
- (8) Moehlmann, J. G.; Gleaves, J. T.; Hudgens, J. W.; McDonald, J. D. Infrared Chemiluminescence Studies of the Reaction of Fluorine Atoms with Monosubstituted Ethylene Compounds. *J. Chem. Phys.* **1974**, *60*, 4790–4799.
- (9) Zvijac, D. J.; Mukamel, S.; Ross, J. Polarization of the Strongest Rydberg Transitions of 1, 3-Butadiene. *J. Chem. Phys.* **1977**, *67*, 2007–2008.
- (10) Hase, W. L.; Bhalla, K. C. A Classical Trajectory Study of the  $\text{F} + \text{C}_2\text{H}_4 \rightarrow \text{C}_2\text{H}_4\text{F} \rightarrow \text{H} + \text{C}_2\text{H}_3\text{F}$  Reaction Dynamics. *J. Chem. Phys.* **1981**, *75*, 2807–2819.
- (11) Clark, D. T.; Scanlan, I. W. A Note on the Transition State of Radical Addition Reactions. *Chem. Phys. Lett.* **1978**, *55*, 102–106.
- (12) Kato, S.; Morokuma, K. Potential Energy Characteristics and Energy Partitioning in Chemical Reactions: *Ab Initio* MO Study of  $\text{H}_2\text{CCH}_2\text{F} \rightarrow \text{H}_2\text{CCHF} + \text{H}$  Reaction. *J. Chem. Phys.* **1980**, *72*, 206–217.
- (13) Schlegel, H. B. *Ab Initio* Molecular Orbital Studies of Atomic Hydrogen + Ethylene and Atomic Fluorine + Ethylene. 1. Comparison of the Equilibrium Geometries, Transition Structures, and Vibrational Frequencies. *J. Phys. Chem.* **1982**, *86*, 4878–4882.
- (14) Schlegel, H. B.; Bhalla, K. C.; Hase, W. L. *Ab Initio* Molecular Orbital Studies of Atomic Hydrogen + Ethylene and Atomic Fluorine + Ethylene. 2. Comparison of the Energetics. *J. Phys. Chem.* **1982**, *86*, 4883–4888.
- (15) Zhang, M. B.; Yang, Z. Z. Computational Study on the Reaction  $\text{CH}_2\text{CH}_2 + \text{F} \rightarrow \text{CH}_2\text{CHF} + \text{H}$ . *J. Phys. Chem. A* **2005**, *109*, 4816–4823.
- (16) Robinson, G. N.; Continetti, R. E.; Lee, Y. T. The Translational Energy Dependence of the  $\text{F} + \text{C}_2\text{H}_4 \rightarrow \text{H} + \text{C}_2\text{H}_3\text{F}$  Reaction Cross Section near Threshold. *J. Chem. Phys.* **1990**, *92*, 275–284.
- (17) Harper, W. W.; Nizkorodov, S. A.; Nesbitt, D. J. Quantum State-Resolved Reactive Scattering of  $\text{F} + \text{CH}_4 \rightarrow \text{HF}(v,J) + \text{CH}_3$ : Nascent  $\text{HF}(v,J)$  Product State Distributions. *J. Chem. Phys.* **2000**, *113*, 3670–3680.
- (18) Harper, W. W.; Nizkorodov, S. A.; Nesbitt, D. Differential Scattering Dynamics of  $\text{F} + \text{CH}_4 \rightarrow \text{HF}(v,J) + \text{CH}_3$  via High-Resolution IR Laser Dopplerimetry. *Chem. Phys. Lett.* **2001**, *335*, 381–387.
- (19) Shiu, W.; Lin, J. J.; Liu, K. Reactive Resonance in a Polyatomic Reaction. *Phys. Rev. Lett.* **2004**, *92*, 103201(1–4).
- (20) Lin, J. J.; Zhou, J.; Shiu, W.; Liu, K. State-Specific Correlation of Coincident Product Pairs in the  $\text{F} + \text{CD}_4$  Reaction. *Science* **2003**, *300*, 966–969.

- (21) Braña, P.; Sordo, J. A. Theoretical Approach to the Mechanism of Reactions between Halogen Atoms and Unsaturated Hydrocarbons: The Cl + Propene Reaction. *J. Comput. Chem.* **2003**, *24*, 2044–2062.
- (22) (a) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.11; Gaussian, Inc.: Pittsburgh, PA, 1998. (b) 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 C02; Gaussian, Inc.: Wallingford, CT, 2004.
- (23) Schlegel, H. B. Potential Energy Curves Using Unrestricted Møller–Plesset Perturbation Theory with Spin Annihilation. *J. Chem. Phys.* **1986**, *84*, 4530–4534.
- (24) Sponer, J.; Hobza, P. J. Interaction Energies of Hydrogen-Bonded Formamide Dimer, Formamidinium Dimer, and Selected DNA Base Pairs Obtained with Large Basis Sets of Atomic Orbitals. *J. Phys. Chem. A* **2000**, *104*, 4592–4597.
- (25) Nguyen, M. T.; Creve, S.; Vanquickenborne, L. G. Difficulties of Density Functional Theory in Investigating Addition Reactions of the Hydrogen Atom. *J. Phys. Chem.* **1996**, *100*, 18422–18425.
- (26) Gonzalez, C.; Schlegel, H. B. Reaction Path following in Mass-Weighted Internal Coordinates. *J. Phys. Chem.* **1990**, *94*, 5523–5527.
- (27) Purvis, G. D.; Bartlett, R. J. A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982**, *76*, 1910–1918.
- (28) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (29) Ignatyev, I. S.; Xie, Y.; Allen, W. D.; Schaefer, H. F. Mechanism of the C<sub>2</sub>H<sub>5</sub> + O<sub>2</sub> Reaction. *J. Chem. Phys.* **1997**, *107*, 141–155.
- (30) Schlegel, H. B.; Sosa, C. Ab Initio Molecular Orbital Calculations on F + H<sub>2</sub> → HF + H and OH + H<sub>2</sub> → H<sub>2</sub>O + H Using Unrestricted Møller–Plesset Perturbation Theory with Spin Projection. *Chem. Phys. Lett.* **1988**, *145*, 329–333.
- (31) McDouall, J. J. W.; Schlegel, H. B. Analytical Gradients for Unrestricted Hartree–Fock and Second-Order Møller–Plesset Perturbation Theory with Single Spin Annihilation. *J. Chem. Phys.* **1989**, *90*, 2363–2369.
- (32) Farnell, L.; Pople, J. A. Radom, L. Structural Predictions for Open-Shell Systems: A Comparative Assessment of ab Initio Procedures. *J. Phys. Chem.* **1983**, *87*, 79–82.
- (33) Sekusak, S.; Liedl, K. R.; Sabljic, A. Reactivity and Regioselectivity of Hydroxyl Radical Addition to Halogenated Ethenes. *J. Phys. Chem. A* **1998**, *102*, 1583–1594.
- (34) Schlegel, H. B. Potential Energy Curves Using Unrestricted Møller–Plesset Perturbation Theory with Spin Annihilation. *J. Chem. Phys.* **1986**, *84*, 4530–4534.
- (35) Stanton, J. F. On the Extent of Spin Contamination in Open-Shell Coupled-Cluster Wave Functions. *J. Chem. Phys.* **1994**, *101*, 371–374.
- (36) Sosa, C.; Schlegel, H. B. Calculated Barrier Heights for OH + C<sub>2</sub>H<sub>2</sub> and OH + C<sub>2</sub>H<sub>4</sub> Using Unrestricted Møller–Plesset Perturbation Theory with Spin Annihilation. *J. Am. Chem. Soc.* **1987**, *109*, 4193–4198.
- (37) Malick, D. K.; Petersson, G. A.; Montgomery, J. A., Jr. Transition States for Chemical Reactions I. Geometry and Classical Barrier Height. *J. Chem. Phys.* **1998**, *108*, 5704–5713.
- (38) Huber, K. P.; Herzberg, G. Constants of Diatomic Molecules. In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, 2005; p 20899. <http://webbook.nist.gov> (accessed Jun 2006). Data prepared by J. W. Gallagher and R. D. Johnson, III.
- (39) Lide, D. R.; Christensen, D. An Improved Structure Determination for Vinyl Fluoride. *Spectrochim. Acta* **1961**, *17*, 665–668.
- (40) Ochsenfeld, C.; Kaiser, R. I.; Lee, Y. T. A Coupled-Cluster *ab Initio* Study of Triplet C<sub>3</sub>H<sub>2</sub> and the Neutral–Neutral Reaction to Interstellar C<sub>3</sub>H. *J. Chem. Phys.* **1997**, *106*, 4141–4151 and references therein.
- (41) Geng, C. Y.; Li, J. L.; Huang, X. R.; Sun, C. C. A Barrier-Free Atom–Molecule Reaction: F + HONO. *Chem. Phys.* **2006**, *324*, 474–482. (b) Li, J. L.; Geng, C. Y.; Huang, X. R.; Sun, C. C. Atomic Radical-Molecule Reactions F + CH<sub>3</sub>C≡CH: Mechanistic Study. *Theor. Chem. Acc.* (accepted).
- (42) Coxon, J. A.; Hajigeorgiou, P. G. Isotopic Dependence of Born–Oppenheimer Breakdown Effects in Diatomic Hydrides: The B<sup>1</sup>Σ<sup>+</sup> and X<sup>1</sup>Σ<sup>+</sup> States of HF and DF. *J. Mol. Spectrosc.* **1990**, *142*, 254–278.
- (43) Braña, P.; Sordo, J. A. Mechanistic Aspects of the Abstraction of an Allylic Hydrogen in the Chlorine Atom Reaction with 2-Methyl-1,3-Butadiene (Isoprene). *J. Am. Chem. Soc.* **2001**, *123*, 10348–10353.
- (44) Stutz, J.; Ezell, M. J.; Ezell, A. A.; Finlayson-Pitts, B. J. Rate Constants and Kinetic Isotope Effects in the Reactions of Atomic Chlorine with *n*-Butane and Simple Alkenes at Room Temperature. *J. Phys. Chem. A* **1998**, *102*, 8510–8519.

(45) Pilgrim, J. S.; Taatjes, C. A. Infrared Absorption Probing of the Cl + C<sub>3</sub>H<sub>6</sub> Reaction: Rate Coefficients for HCl Production between 290 and 800 K. *J. Phys. Chem. A* **1997**, *101*, 5776–5782.

(46) Kaiser, E. W.; Wallington, T. J. Pressure Dependence of the Reaction Cl + C<sub>3</sub>H<sub>6</sub>. *J. Phys. Chem.* **1996**, *100*, 9788–9793.

CT050233M



## Analysis of the Primary Photocycle Reactions Occurring in the Light, Oxygen, and Voltage Blue-Light Receptor by Multiconfigurational Quantum-Chemical Methods

Tatiana Domratcheva,\* Roman Fedorov,<sup>†</sup> and Ilme Schlichting

Max-Planck Institute for Medical Research, Department of Biomolecular Mechanisms,  
Jahnstrasse 29, 69120 Heidelberg, Germany

Received January 8, 2006

**Abstract:** The photocycle reactions occurring between the flavin mononucleotide cofactor and the reactive cysteine residue in the blue-light photoreceptor domain light, oxygen, and voltage (LOV) were modeled for a system consisting of lumiflavin and thiomethanol. The electronic structure and energies of the reactive species were estimated using the CASSCF and MCQDPT2 quantum-chemical methods. The reaction pathway for the S–C4a covalent adduct formation in the triplet state was determined. Concerted electron and proton transfer from the thiol to the flavin in the triplet electronic state results in a biradical complex that is, however, unstable because its structure corresponds to a triplet–singlet crossing. The covalent adduct dissociation in the ground electronic state is a reverse of the photoreaction proceeding via a single energy barrier for hydrogen transfer. Thus, both photo- and dark reactions were found to be single-step chemical transformations occurring without stable intermediates. The photoreaction yielding the S–C4a covalent adduct is an intrinsic property of the isoalloxazine–thiol complex in the specific geometry arranged by the protein in LOV. The S–C4a covalent adduct between lumiflavin and thiomethanol is rather stable implying that in LOV its dissociation is facilitated by the protein.

### Introduction

To increase the efficiency of photosynthesis, plants use blue-light receptor proteins called phototropins<sup>1</sup> that are involved in phototropism, light-induced stomatal opening, and chloroplast movement to places with appropriate light intensity. Phototropins are light-driven autophosphorylating protein kinases that contain two similar, but not identical, light-receptor domains LOV1 and LOV2.<sup>2–4</sup> Each light, oxygen, and voltage (LOV) domain binds noncovalently a flavin mononucleotide (FMN) chromophore that undergoes a photoreaction upon blue-light absorption. Photoexcited FMN and a conserved cysteine (Cys) residue form a S–C4a flavin–cysteinyll covalent adduct (Scheme 1) on a microsecond

scale.<sup>5,6</sup> In the dark, the photoproduct dissociates into the initial components within minutes. The LOV photoproduct triggers autophosphorylation of the kinase, which initiates the phototropism-signaling reaction cascade.

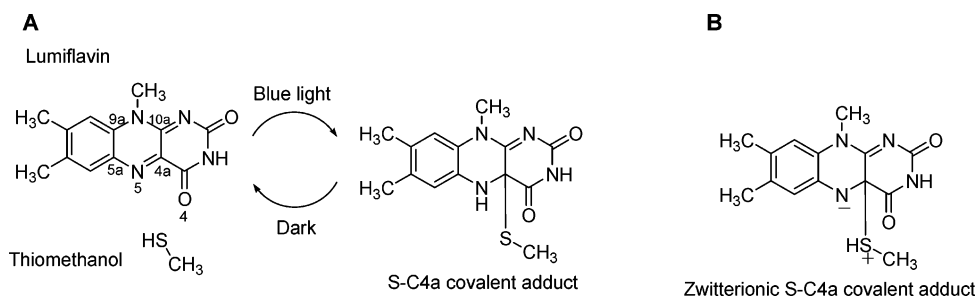
Light-induced changes in absorbance have been detected for the LOV domains.<sup>7,8</sup> Upon illumination by blue light, the FMN absorption band at 447 nm bleaches. A transient absorption band at about 700 nm of a triplet flavin species is formed within nanoseconds. Then, a broad band of the covalent adduct appears with a maximum at 390 nm. In the dark, the FMN absorption band at 447 nm recovers because of thermal dissociation of the adduct into the initial components.

Several hypotheses of the mechanism of the covalent adduct formation in LOV have been put forward. The increase of the basic properties of flavins upon photoexcitation is well-known<sup>9–12</sup> and expected to play a role in the LOV photochemistry.<sup>13,14</sup> In the triplet state, a protonated cation of FMN (HFMN<sup>+</sup>) can be formed. An ionic mecha-

\* Corresponding author phone: +49-6221-486-504; fax: 49 6221 486 585; e-mail: Tatjana.Domratcheva@mpimf-heidelberg-mpg.de.

<sup>†</sup> Current address: Department of Biophysical Chemistry, OE4350 Hannover Medical School, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany.

## Scheme 1



nism of the S–C4a covalent adduct formation<sup>8,14</sup> implies an interaction of the thiolate anion with the C4a atom of HFMN<sup>+</sup>. However, there is no experimental evidence for stable or transient ionic species occurring in the LOV photocycle. FTIR spectra showed that the cysteine was protonated in the LOV dark-adopted state,<sup>15</sup> while electron paramagnetic resonance (EPR) measurements excluded a protonation of the triplet FMN intermediate.<sup>16,17</sup> Flavins are potent oxidants for redox-active amino acids such as cysteine.<sup>18</sup> Hence, a radical mechanism has been proposed.<sup>16,17,19,20</sup> An ionic radical pair was predicted to form as a result of a one-electron reduction of FMN by the cysteine. At cryogenic temperature, the radical pair recombines resulting in the S–C4a zwitterionic covalent adduct that at room temperature converts to the S–C4a covalent adduct by proton transfer. However, the FMN radicals were detected only in light-excited LOV mutants<sup>17,19–21</sup> but not in the wild-type LOV.<sup>16</sup> In our previous work, we discussed a concerted mechanism based on quantum-chemical analysis of the crystal structures of the LOV1 domain.<sup>13</sup> Proton transfer from the cysteine to the N5 atom of the triplet FMN is concerted with the S–C4a covalent bond formation.

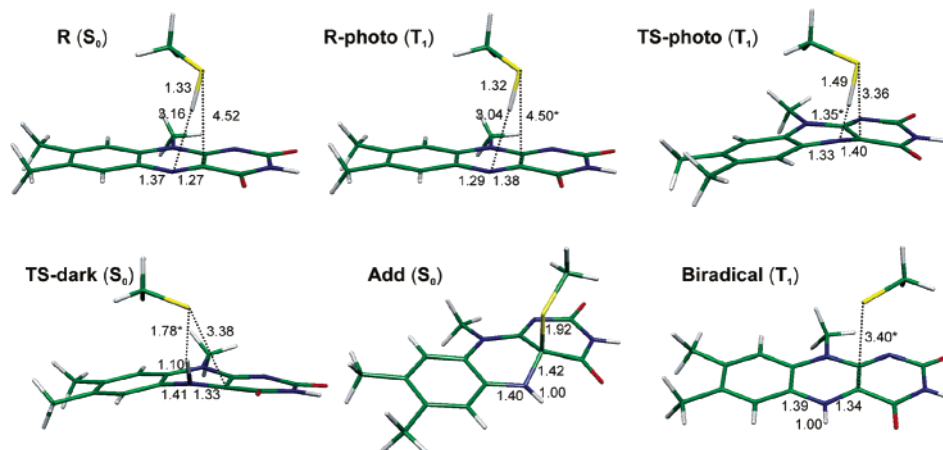
After the X-ray crystal structures of both dark-adopted and light-illuminated states of the LOV domains were determined,<sup>13,22,23</sup> the photoreaction mechanism was addressed by means of quantum-chemical calculations.<sup>13,24,25</sup> The analysis of the electronic structure of the isoalloxazine chromophore in the ground singlet and excited triplet states by complete active space self-consistent field (CASSCF) calculations<sup>13</sup> favored a concerted mechanism. Energy estimates by density functional theory (DFT) showed that a radical pair between FMN-related molecules and a thiol is more favorable than the corresponding ion pair.<sup>24</sup> A stable radical pair intermediate along the photoreaction pathway was predicted by Hartree–Fock- (HF-)based quantum-mechanical/molecular-mechanical (QM/MM) modeling supporting the radical mechanism hypothesis.<sup>25</sup> Nevertheless, the LOV quantum-chemical analysis is far from being complete. To describe the electronic excitations, the excited-state reaction pathway, and the S–C4a covalent bond dissociation in the ground state, it is necessary to go beyond single-configurational methods such as HF and DFT. Experimental data,<sup>26</sup> the QM/MM model,<sup>25</sup> and our previous calculations<sup>13</sup> indicate strongly that the primary photoreaction in LOV is localized within the FMN-reactive Cys subsystem. Therefore, a model system consisting of an isoalloxazine chromophore and a thiol can be used for the high-level quantum-chemical analysis.

In the present study, we analyze the LOV photocycle reactions using a lumiflavin-thiomethanol model complex (Scheme 1A) and CASSCF calculations supplemented with second-order multiconfigurational perturbation theory (MC-QDPT2). The electronic structure of the photospecies is described, and accurate energy estimates are obtained. The derived energy profile of both photo- and dark reactions is consistent with a concerted mechanism.

### Computational Details

A model system consisting of lumiflavin and thiomethanol is depicted in Scheme 1A; the atom numbering used throughout the text is indicated. Lumiflavin contains the dimethylisoalloxazine ring of FMN; thiomethanol represents the side chain of the reactive cysteine. The starting geometries were taken from the crystal structures of LOV1 from *Chlamydomonas reinhardtii*<sup>13</sup> (reagent and covalent adduct complex PDB codes 1n9l and 1n9o, respectively). Geometry optimization of the model complexes was carried out in the internal coordinates.<sup>27</sup> Restrained geometry optimization was used to ensure the arrangement of the reagents as in the LOV domain, where the side chain of the cysteine is located above the isoalloxazine ring. The optimized structures are in good agreement with the ones obtained in the QM/MM study of LOV<sup>25</sup> where the geometry of a quantum-chemically described lumiflavin–thiomethanol complex was restrained by the molecular-mechanical model of the LOV protein.

CASSCF<sup>28</sup> and MCQDPT<sup>29</sup> calculations with a standard Gaussian basis set 6-31G(d) containing polarization d functions for the heavy atoms were carried out to derive the energy profile of the photocycle reactions and to evaluate the electronic and absorption properties of the model complexes. The lumiflavin–thiomethanol complex has 160 electrons, which are described by 351 single-electron wave functions—molecular orbitals (MOs)—when the 6-31G(d) basis set is used. The active space is indicated by the numbers in parentheses CASSCF(*n,m*), where *n* is the number of electrons and *m* is the number of MOs. The multireference configurational interaction (MRCI) method was used to validate the choice of the active space. The configuration interaction (CI) calculations accounting for single and double electronic excitations (CI–SD) were done with the molecular orbitals obtained from the CASSCF calculations (hereafter referred to as MRCI–SD calculations). A total of 65 MOs were considered to be core orbitals and 256 MOs to be virtual, yielding a MRCI–SD solution accounting for electronic excitations within a window of 30 MOs with 30 electrons sufficiently large for our purposes because it



**Figure 1.** Structures of the model complexes optimized by the CASSCF(2,2)/6-31G(d) method. The electronic state is indicated in parentheses. Selected distances are specified in angstroms. The distances marked by an asterisk were fixed during geometry optimization.

contained all  $\pi$  MOs ( $7\pi$  and  $7\pi^*$ ). The dynamic electron correlation not accounted for by the CASSCF(2,2) or (4,4) method was taken into consideration by single-point MC-QDPT2 calculations.

To optimize geometries, state-specific CASSCF(2,2) calculations were used. To describe the charge distribution, Löwdin<sup>30</sup> and Mulliken<sup>31</sup> atomic electrostatic charges were calculated. Both methods provided consistent results, but the Mulliken charges were somewhat overestimated. Therefore, Löwdin atomic charges are presented in the text. To evaluate the electronic transition energies of the reactive species, MCQDPT2/CASSCF calculations with the state-averaged energy functional were performed. The CASSCF equations were solved with the averaged density from the two electronic states involved in the transition with weight coefficients 0.5 and 0.5. The calculations were done using the PC GAMESS version<sup>32</sup> of the GAMESS (U.S.) quantum-chemical program package.<sup>33</sup> The program Molekel<sup>34</sup> was used to create the figures.

We extensively compare our results on the “free complex” (i.e., not embedded in the protein) with the QM/MM calculations<sup>25</sup> to analyze the effect of the protein matrix, which is not included in our model, as well as the effects of electron correlation, which are not accounted for in the QM/MM model.

## Results and Discussion

In the photoreaction, the excited triplet flavin interacts with the cysteine to form the S–C4a-covalent adduct (Scheme 1A). In the dark-reaction, occurring in the ground electronic state, the photoproduct dissociates. To obtain the energy profile of the photocycle, the following structures were optimized by the CASSCF(2,2) method: (1) the lumiflavin–thiomethanol complex **R** optimized in the ground state (dark-adapted state of LOV absorbing at 447 nm), (2) the lumiflavin–thiomethanol complex **R-photo** optimized in the triplet state (the spectroscopically characterized triplet intermediate absorbing at 660 and 715 nm), (3) the highest-energy structure for the H-transfer reaction between S–H and N5 in the triplet state **TS-photo**—a model for the photoreaction transition state, (4) the triplet **Biradical**

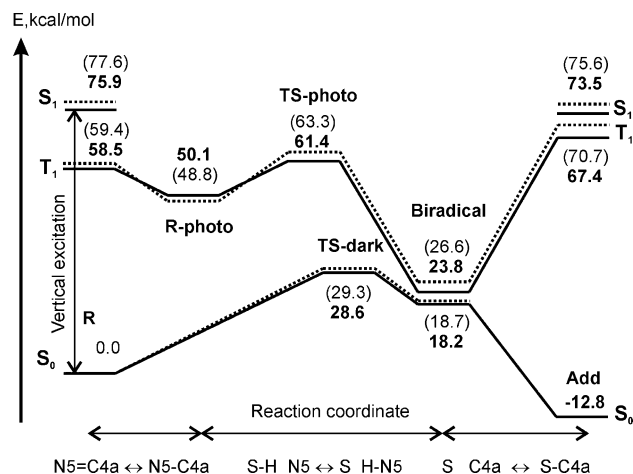
complex between two radicals SCH<sub>3</sub> and N5H-semiquinone, which is a product of the H-transfer reaction, (5) the S–C4a covalent adduct **Add** optimized in the ground state (the photoproduct state of LOV absorbing at 390 nm), and (6) the structure **TS-dark** in the ground state similar to **TS-photo** and representing a transition state for the **Add** conversion into the initial complex **R**. All structures are presented in Figure 1.

The calculated energies and CI coefficients of the wave functions obtained by different multiconfigurational methods are presented in Table 1. The state-specific CASSCF(2,2) solutions were obtained with the active space containing the highest occupied and lowest uncoupled molecular orbitals (HOMO and LUMO) of the ground-state Hartree–Fock electronic configuration. For complexes **R** in S<sub>0</sub> and S<sub>1</sub> and **Add** in S<sub>0</sub>, the CASSCF(2,2) wave functions mainly consist of a single determinant. For geometries **Biradical** and **TS-dark** in the S<sub>0</sub> state, a two-configurational wave function is required to account for the dissociated S–C4a covalent bond. The CASSCF(2,2) solutions for the triplet state are identical to the ROHF solutions. To determine if the active space needs to be extended, the MRCI–SD calculations with CASSCF(2,2) MOs were performed. The MRCI–SD energies are lower, compared to the CASSCF(2,2) energies, but the CI coefficients obtained by the two methods are in agreement. Only minor contributing configuration state functions (CSFs) whose MRCI coefficients do not exceed 0.1 (the largest value, 0.12, was obtained for the S<sub>1</sub> solution at the **R** geometry) corresponding to the excitations outside the (2,2) active space were found. These excitations occur within a pair of the  $\pi, \pi^*$  MOs localized on the dimethylbenzene ring. The orbitals were included into the active space to obtain CASSCF(4,4) solutions. For the **Add** geometry in S<sub>0</sub>, only double excitations from HOMO to LUMO were found to be significant. Accordingly, the energies of **Add** calculated by the CASSCF(2,2) and MRCI–SD methods differ only by 0.3 kcal/mol. Our attempts to obtain the CASSCF(4,4) solution for **Add** that would be similar to the solutions obtained for the other model complexes were unsuccessful because the CASSCF iterations did not converge. For the other model complexes, the extension of the

**Table 1.** Total Energies (Hartree) and Compositions of the Wave Functions Given by the CI Coefficients of the Dominant Configuration State Functions (CSFs)

structure <sup>a</sup>	R		R	R-photo	TS-photo	Biradical	Biradical	Add	TS-dark
electronic state	S <sub>0</sub>	S <sub>1</sub>	T <sub>1</sub>	T <sub>1</sub>	T <sub>1</sub>	T <sub>1</sub>	S <sub>0</sub>	S <sub>0</sub>	S <sub>0</sub>
CAS(2,2)	-1304.576423 (1) <sup>b</sup> 0.981 (3) -0.196	-1304.398784 (2) 0.996	-1304.443019 (2) 1.000	-1304.467340 (2) 1.000	-1304.424206 (2) 1.000	-1304.515333 (2) 1.000	-1304.520331 (1) 0.820 (3) -0.573	-1304.565059 (1) 0.992 (3) -0.124	-1304.479388 (1) 0.916 (3) -0.399
MCQDPT2	-1307.440091	-1307.319140	-1307.346847	-1307.360405	-1307.342207	-1307.402101	-1307.411409	-1307.460488	-1307.394552
MRCI-SD <sup>c</sup>	-1304.627336 (1) 0.952 (3) -0.174 (10) -0.088	-1304.424324 (2) 0.950 (5) 0.120 (6) 0.113 (7) 0.070	-1304.503981 (2) 0.948 (5) 0.065	-1304.525280 (2) 0.954 (8) -0.059 (6) 0.051	-1304.479042 (2) 0.957 (8) -0.055	-1304.575831 (2) 0.961 (8) 0.055	-1304.557026 (1) 0.851 (3) -0.573 (10) -0.070	-1304.565597 (1) 0.992 (3) -0.124	-1304.525402 (1) 0.914 (3) -0.322 (10) -0.073
CAS(4,4)	-1304.591664 (1) 0.969 (3) -0.193 (10) -0.146	-1304.421838 (2) 0.957 (6) -0.183 (7) -0.121	-1304.463504 (2) 0.977 (8) -0.154 (5) 0.070	-1304.489505 (2) 0.976 (8) -0.158 (4) 0.065	-1304.432498 (2) 0.995 (8) -0.078	-1304.532676 (2) 0.986 (8) -0.118 (4) -0.091	-1304.534528 (1) 0.810 (3) -0.566 (10) -0.116 (11) 0.081		-1304.493832 (1) 0.907 (3) -0.392 (10) -0.134 (11) 0.058
MCQDPT2	-1307.444489	-1307.320818	-1307.349785	-1307.366718	-1307.343690	-1307.402157	-1307.414647		-1307.397714
MRCI-SD <sup>d</sup>	-1304.631149 (1) 0.950 (3) -0.172 (10) -0.126	-1304.428096 (2) 0.943 (6) -0.164 (7) -0.108 (5) -0.107	-1304.509025 (2) 0.945 (8) -0.135 (5) 0.077	-1304.531495 (2) 0.950 (8) -0.141 (9) 0.063	-1304.486881 (2) 0.953 (8) -0.071 (5) 0.059	-1304.588301 (2) 0.954 (8) -0.106	-1304.562081 (1) 0.853 (3) -0.464 (10) -0.101		-1304.531630 (1) 0.913 (3) -0.312 (10) -0.116

<sup>a</sup> The geometries were optimized by the CASSCF(2,2) method. <sup>b</sup> The CSFs specified by the occupancy numbers of the MOs within the (4,4) active space: (1) 2200, (2) 2110, (3) 2020, (4) 2101, (5) 1210, (6) 1111, (7) 1111, (8) 0112, (9) 0211, (10) 0202, and (11) 0022. <sup>c</sup> MOs obtained from the CASSCF(2,2) calculations were used. <sup>d</sup> MOs obtained from the CASSCF(4,4) calculations were used.



**Figure 2.** Energy diagram of the photocycle reactions between lumiflavin and thiomethanol derived from the single-point MCQDPT2/CASSCF(2,2) (solid line, bold numbers) and MCQDPT2/CASSCF(4,4) (dashed line, numbers in parentheses) calculations. The geometries of the model complexes were optimized by the CASSCF(2,2) method.

active space to (4,4) results in lowering the absolute energies of the electronic states without any significant changes of the relative energies (see also Figure 2) and the CI coefficients. Thus, for all considered geometries, the minimum active space (2,2) is sufficient for the description of the dynamic correlation.

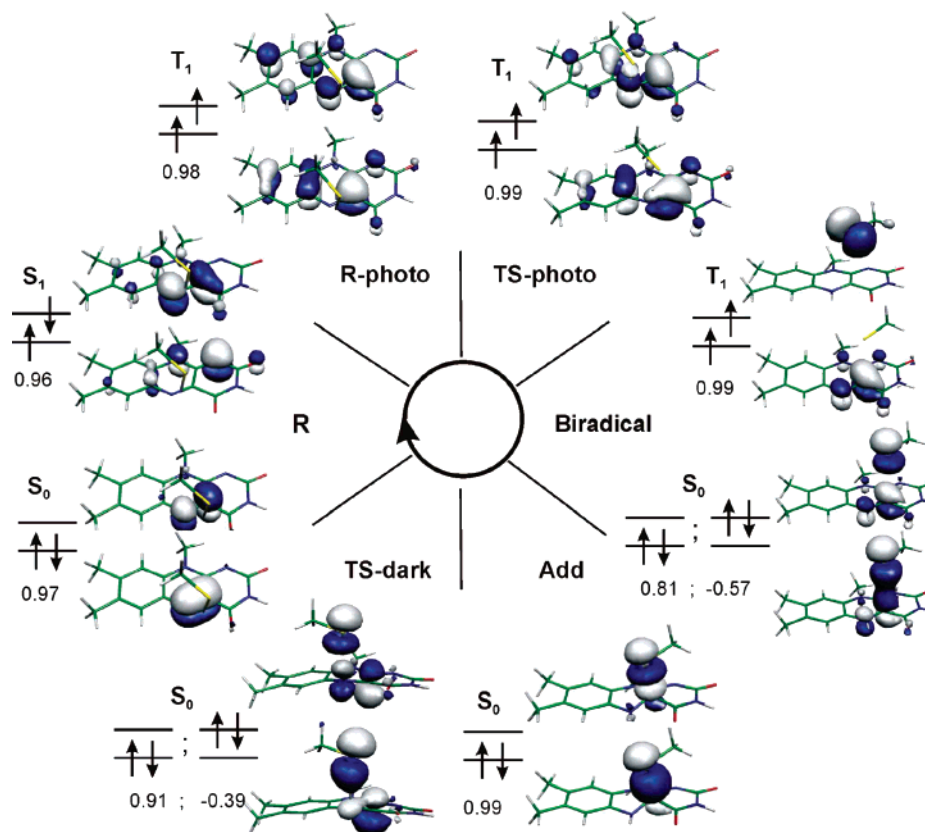
The data from Table 1 is illustrated in Figure S1 in the Supporting Information. The diagrams show how the energies of the electronic states change when different electron correlation methods are employed. Accounting for dynamic correlation decreases the energies of the transition states **TS-photo** and **TS-dark** more than those of the reactant and product structures (Figure S1-B, Supporting Information). Therefore, the MCQDPT2 correction is significant for estimating the energy barrier heights for the photo- and dark reactions. Electrostatic charge distributions in the model complexes calculated by the CASSCF methods are presented in Table 2. As it is indicated for the **R** structure, virtually the same results were obtained with the (2,2) and (4,4) active space.

The unrestrained geometry optimization of the lumiflavin–thiomethanol complex results in an equilibrium structure where thiomethanol is located aside the isoalloxazine ring forming a hydrogen bond with the O4 atom (Figure S2 in the Supporting Information). There is no local minimum corresponding to a complex where thiomethanol occupies a position above the isoalloxazine ring. To obtain structure **R**, the geometry optimization in internal coordinates was carried out until the thiomethanol molecule started to move to the side of the isoalloxazine ring. In the resulting structure, the geometries of the thiomethanol and lumiflavin molecules are identical to those in the equilibrium structure. The S–C4a distance of 4.52 Å is similar to that of one of the Cys isomers in LOV1 (4.4 Å)<sup>13</sup> and a little longer than that in LOV2 (4.2 Å).<sup>23</sup>

In Figure 2, the energy profiles of the photocycle reactions derived from the MCQDPT2/CASSCF(2,2) and (4,4) calculations are superimposed. The relative energies of the model complexes (with respect to the energy of **R** in  $S_0$ ) evaluated by these two methods agree within 0.5–3.3 kcal/mol. Unless otherwise stated, we will refer to the MCQDPT2/CASSCF(2,2) energy estimates (bold-type numbers in Figure 2). Vertical excitation to the  $S_1$  state increases the energy of **R** to 75.9 kcal/mol. The vertical energy of the first triplet state  $T_1$  is 58.5 kcal/mol. To obtain the structure of the triplet species, the geometry of complex **R** was further optimized in the  $T_1$  state with the S–C4a distance fixed at 4.5 Å. The resulting structure **R-photo** is a complex between triplet lumiflavin and thiomethanol with a relative energy of 50.1 kcal/mol. Structural relaxation of the triplet lumiflavin involves an elongation of the N5–C4a bond from 1.27 to 1.38 Å and a shortening of the N5–C5 bond from 1.37 to 1.29 Å. In the triplet state, the H transfer between S–H and N5 has a 11.3 kcal/mol energy barrier [14.5 kcal/mol according to MCQDPT2/CASSCF(4,4)]. The barrier corresponds to the geometry **TS-photo** with S–H and H–N5 distances of 1.49 and 1.35 Å, respectively. Proton transfer coupled with electron transfer yields a biradical intermediate. Structure **Biradical** in Figure 1 was optimized with the S–C4a distance fixed at 3.40 Å, which was similar to the distance in **TS-photo**. Unrestrained geometry optimization resulted in a structure with the SCH<sub>3</sub> fragment aside of the

**Table 2.** Charge Distribution and Dipole Moments

method	CAS(2,2)		CAS(4,4)			CAS(2,2)	
structure	<b>R</b>	<b>R</b>	<b>R-photo</b>	<b>TS-photo</b>	<b>Biradical</b>	<b>Add</b>	<b>TS-dark</b>
electronic state	$S_0$	$S_0$	$T_1$	$T_1$	$T_1$	$S_0$	$S_0$
Löwdin Atomic Charges (au)							
C5a	-0.023	-0.022	+0.132	+0.082	+0.048	+0.055	+0.006
N5	-0.050	-0.051	-0.233	-0.275	-0.214	-0.327	-0.204
C4a	-0.010	-0.009	+0.027	+0.042	-0.026	-0.052	-0.005
S	-0.067	-0.067	-0.071	-0.232	+0.026	+0.118	-0.053
H	+0.114	+0.114	+0.121	+0.232	+0.327	+0.310	+0.282
Net Charges (au)							
HSCH <sub>3</sub>	-0.003	-0.003	-0.005	-0.093			
SCH <sub>3</sub>					+0.007	+0.095	-0.089
dipole moments (Debye)	10.0	10.0	10.8	12.2	6.6	5.5	9.6



**Figure 3.** Electronic structure of the model complexes. Shown frontier molecular orbitals and the CI coefficients were optimized for a selected electronic state by the CASSCF(4,4) method. For the **Add** structure, the CASSCF(2,2) results are presented. The geometries of the model complexes were optimized by the CASSCF(2,2) method.

isalloxazine ring (Figure S2 in the Supporting Information). **Biradical** contains two almost neutral radical fragments: HN5-semiquinone and thiomethyl radical  $\text{SCH}_3$ . The relative energies of **Biradical** are 23.8 and 18.2 kcal/mol in the  $T_1$  and  $S_0$  electronic states, respectively. In  $S_0$ , the energy of the complex decreases upon shortening of the S–C4a distance yielding the covalent adduct. A small energy difference between the two electronic states in the **Biradical** geometry indicates a possible crossing between the  $T_1$  and  $S_0$  states via a covalent bond formation between S and C4a. In the fully optimized structure **Add**, the equilibrium S–C4a bond distance is 1.92 Å. In the ground electronic state, the **Add** energy is 12.8 kcal/mol lower compared to that of **R**. A reverse of the photoreaction results in the dissociation of the covalent adduct via **TS-dark** with an energy barrier of 41.4 kcal/mol. The **TS-dark** structure is similar to **TS-photo** and has N–H and S–H distances of 1.10 and 1.78 Å, respectively, and a S–C4a distance of 3.38 Å.

The electronic structure of the model complexes occurring along the reaction coordinate is illustrated in Figure 3. In **R**, the electron excitation corresponds to a loss of the N5–C4a double bond. The  $S_1$  and  $T_1$  states of lumiflavin are of  $\pi$ – $\pi^*$  character. In  $T_1$ , the negative charge on N5 increases compared to the ground electronic state (Table 2). Interestingly, the positive charge on C5a occurs significantly higher than on C4a because of stabilization by the dimethylbenzene ring. In **R-photo**, unpaired electrons are rather delocalized. The product of the photoreaction **Biradical** has unpaired electrons localized on S, C4a, and N5. At the **Biradical**

**Table 3.** Energies of the Electronic Transitions (eV), Estimated by the State-Averaged MCQDPT2/CASSCF(2,2) and (4,4) Methods for the Singlet and Triplet Species, Respectively

molecular system	CASSCF ( $f^a$ )	MCQDPT2	exptl. LOV17, solution <sup>35</sup>
<b>R</b> ( $S_0$ – $S_1$ )	4.53 (0.655)	2.86	2.77
<b>R-photo</b> ( $T_1$ – $T_2$ )	1.99 (0.010)	1.64	1.73, 1.88
<b>Add</b> ( $S_0$ – $S_1$ )	5.13 (0.476)	3.37	3.18
lumiflavin <sup>b</sup> ( $S_0$ – $S_1$ )	4.56 (0.670)	2.88	2.80
lumiflavin <sup>c</sup> ( $T_1$ – $T_2$ )	2.20 (0.011)	1.49	1.95

<sup>a</sup> Oscillator strength. <sup>b</sup>  $S_0$  RHF/6-31G(d) equilibrium geometry. <sup>c</sup>  $T_1$  ROHF/6-31G(d) equilibrium geometry.

geometry, intersystem crossing takes place and the S–C4a  $\sigma$  bond forms in the ground state. Upon the **Add** dissociation, the N5=C4a double bond is partially restored in **TS-dark**.

To further characterize the reactive complexes and to confirm that they can represent the LOV photospecies in terms of the electronic structure, the energies of the first electronic transitions  $S_0$ – $S_1$  and  $T_1$ – $T_2$  were evaluated by the state-averaged MCQDPT2/CASSCF calculations. The results are presented in Table 3. For comparison, the electronic transition energies in a free lumiflavin molecule in the ground and triplet electronic states were calculated in the RHF and ROHF/6-31G(d) equilibrium geometries, respectively. The  $S_0$ ,  $S_1$ ,  $T_1$ , and  $T_2$  electronic states do not change within the reactant complexes with respect to a free lumiflavin molecule; therefore, the nature of the  $S_0$ – $S_1$  and  $T_1$ – $T_2$  transitions is the same, and the energies can be

compared. The CASSCF(2,2) method alone overestimates the transition energies, which can be improved by the MCQDPT2 correction.

All energy differences presented in Table 3 correspond to the  $\pi$ - $\pi^*$  electron excitations within the isoalloxazine fragment. The  $S_0$ - $S_1$  energy is virtually identical in the reactant species **R** and in the isolated lumiflavin molecule and somewhat larger compared to the experimental numbers. The calculated energy of the  $T_1$ - $T_2$  transition in both **R-photo** and free lumiflavin is significantly red-shifted when compared to the experimental results. Interestingly, despite that only  $\pi$  and  $\pi^*$  MOs of lumiflavin are involved in the transition, its energy is sensitive to the presence of the thiomethanol fragment, showing a blue shift in **R-photo** compared to lumiflavin. The  $T_1$ - $T_2$  transition has a small oscillator strength and, probably, is not visible in the experimental spectrum. The triplet species in LOV is characterized by a broad absorption peak with two maxima at 715 and 650–660 nm,<sup>5,7</sup> very similar to what is observed for the triplet flavins in a water solution.<sup>36</sup> In recent time-dependent DFT calculations,<sup>37,38</sup> the absorption of the triplet flavins around 700 nm was associated with the  $T_1$ - $T_4$  and  $T_1$ - $T_5$  transitions whose calculated energies were in satisfactory agreement with experimental results. Clearly, this subject requires more extensive analysis. Similarly to complex **R**, the predicted  $S_0$ - $S_1$  transition energy for the **Add** structure is overestimated with respect to the experimentally observed one. For the adduct complex, one should keep in mind that the transition energy calculated for the free compound is compared to the observations for the covalent adduct in the LOV protein. A larger difference between the calculated and observed transition energy of 0.2 eV obtained for **Add** compared to 0.1 eV obtained for **R** suggests that the absorption of the free covalent adduct may be slightly blue-shifted with respect to that of LOV. We could not confirm this suggestion by comparing experimental data because no data on a free flavin S-C4a adduct was available. However, it is known as a general trend that the low-energy absorption band of flavin C4a adducts around 380 nm is slightly shifted to the longer wavelength in the enzymes and also at a low temperature (77 K).<sup>39</sup>

The calculations of the electronic transitions in flavins are of great practical interest. Recently, absorption properties of different isoalloxazine molecules have been analyzed using DFT.<sup>37,38,40</sup> The  $S_0$ - $S_1$  transition in lumiflavin and riboflavin was predicted to be 3.05<sup>37</sup> and 3.07 eV,<sup>38</sup> respectively. Apparently, our MCQDPT2/CASSCF calculations with the minimal active space and a modest basis set are in better agreement with experimental results. Thus, multiconfigurational methods can be a very useful tool for spectroscopic applications in flavins.

The energy barrier structures **TS-photo** and **TS-dark** were determined using constrained optimization along the approximate reaction coordinate for hydrogen transfer between N5 and S. The reaction coordinate was specified by a decreasing N5-H or S-H distance for the photo- or dark reaction, respectively. Two dihedral angles determining the angle between S-N5 and the C5a-N5-C4a plane were fixed, ensuring the position of thiomethanol above the

isoalloxazine ring. The other internal coordinates were optimized. In the obtained partly optimized geometry, the N5-H or S-H distance was further decreased and a new optimization step was performed to locate the following point of the energy scan. Such a stepwise procedure allows maximum structural relaxation of the molecular complex necessary to approach the intrinsic reaction coordinate. The calculated energy profiles are presented in Figures S3 and S4 in the Supporting Information. The drop of the energy of the complex is coupled with the major structural rearrangements such as breaking and formation of the N-H or S-H bonds when the system arrives at the product valley on the potential energy surface. In the photoreaction, the relaxation of the geometry after the hydrogen transfer to N5 includes the movement of the CH<sub>3</sub> group of thiomethanol from above the benzene ring to above the pyrimidine ring. The reverse movement was observed in the ground electronic state along the dark-reaction energy scan before the system reached the **TS-dark** geometry. A similar change in the conformation of the Cys side chain upon the covalent adduct formation is present in the X-ray crystal structures of the dark-adopted and light-illuminated LOV domains<sup>13</sup>. In the dark-reaction, the hydrogen transfer from N5-H to S follows the dissociation of the S-C4a covalent bond.

The highest-energy structures **TS-photo** and **TS-dark** are approximations for the saddle points. For both geometries, the calculated Hessians reveal two negative curvatures: one along the hydrogen-transfer intrinsic coordinate and the other for the rotation of the thiomethanol fragment. The corresponding imaginary frequencies (in cm<sup>-1</sup>) are 835*i* (N5-H stretching) and 62*i* (SCH<sub>3</sub> rotation) in **TS-photo** and 1572*i* (S-H stretching) and 41*i* (SCH<sub>3</sub> rotation) in **TS-dark**. A saddle-point search starting from the **TS-photo** and **TS-dark** geometries was complicated by the movement of the thiol fragment away from its position above the isoalloxazine ring, yielding either the reactant or biradical equilibrium structure (Figure S2, Supporting Information). Nevertheless, we expect **TS-photo** and **TS-dark** to lie in the vicinity of the corresponding saddle points. Indeed, the **TS-photo** structure is similar to the saddle point located and confirmed by the vibrational analysis in the ROHF/6-31(2d,2p) QM/MM study<sup>25</sup> whose energy is 30.2 kcal/mol above the triplet state reactant. The lower activation energy in our model results from accounting for dynamic electron correlation. **TS-dark** is an analogue of **TS-photo** in the ground electronic state corresponding to the same chemical reaction—an addition of a thiol group to the N5=C4a double bond of an isoalloxazine. The energy barrier in the triplet state is significantly smaller than in the ground electronic state. The activation energies obtained for the photo- and dark reactions might be overestimated. Accounting for the vibrational energy typically decreases the energy barriers by several kilocalories per mole. In photoreactions, the population of the higher vibrational levels is common, which also decreases energy barriers and facilitates conversions. Both the photo- and dark reactions are single-step chemical transformations proceeding without intermediates.

The product of the photoreaction **Biradical** is unstable because of intersystem crossing. To characterize the crossing

geometry, we performed state-averaged CASSCF(4,3) calculations for the  $S_0$  and  $T_1$  electronic states (Figure S5 in the Supporting Information). The active space consisting of four electrons in three MOs was found (on the basis of the MRCI analysis for the state-averaged CASSCF wave function) to be sufficient to describe these electronic states simultaneously. In the **Biradical** geometry with a S–C4a distance of 3.40 Å, the  $S_0$ – $T_1$  energy difference calculated by the MCQDPT2/CASSCF(4,3) is 4.6 kcal/mol, which is in good agreement with the single-state estimates indicated in Figure 2. If the S–C4a distance increases, the  $S_0$ – $T_1$  energy difference becomes smaller. In a **Biradical**-like structure with a S–C4a distance of 4.5 Å, the MCQDPT2/CASSCF(4,3)  $S_0$ – $T_1$  energy difference is 1.4 kcal/mol. This structure lies in the vicinity of the  $S_0$ / $T_1$  crossing. A triplet intermediate similar to **Biradical** with an 18 kcal/mol energy barrier for the intersystem crossing was identified in LOV in the QM/MM model.<sup>25</sup> The predicted high stability of the biradical state is an artifact of the Hartree–Fock method used to compare the energies of the  $S_0$  and  $T_1$  electronic states. Our analysis shows (Table 1 and Figure 3) that the  $T_1$  electronic state can be described by a single-determinant wave function, while the wave function of the  $S_0$  electronic state should contain at least two determinants with occupied bonding  $\pi$  and antibonding  $\pi^*$  MOs. The  $S_0$  energy of **Biradical** calculated with the two-configurational self-consistent field wave function is 41.3 kcal/mol lower than the Hartree–Fock energy. The decay of the triplet state biradical occurs via S–C4a covalent bond formation. It is very likely that, in different LOV domains, the  $S_0$ – $T_1$  energy difference for the crossing geometry can vary, resulting in different kinetics of the triplet state decay.

Structure **Add** aligns with the corresponding crystal structure (PDB: 1N9O)<sup>13</sup> with a root-mean-square deviation of 0.078 Å. According to the single-point MCQDPT2 calculations, the reaction energy of the addition of thiomethanol to the N5=C4a double bond of lumiflavin is –12.8 kcal/mol. The CASSCF method favors the reactant structure over the adduct by 7 kcal/mol, revealing a considerable contribution of the dynamic electron correlation to the adduct stability. To check the effect of the dynamic electron correlation, the geometries of the reactant and product complexes were optimized by the MP2/6-31G(d) method, which gave the reaction energy of –11.8 kcal/mol consistent with the single-point MCQDPT2 calculations. Interestingly, the S–C4a MP2 equilibrium distance is 1.87 Å, somewhat shorter than the CASSCF(2,2) equilibrium distance of 1.92 Å. The latter probably reflexes the tendency of the CASSCF(2,2) to underestimate the product stability. The HF/6-31G(d) equilibrium distance is 1.87 Å, in agreement with the MP2 optimization as well as the QM/MM optimization in LOV.<sup>25</sup> Despite the same equilibrium S–C4a distance, the HF method predicts a reaction energy of 1.3 and 8.7 kcal/mol<sup>25</sup> for the free compound and the LOV domain, respectively. Apparently, the stability of the covalent adduct is a property of a free compound. Therefore, we suggest that the protein matrix destabilizes the covalent adduct, which dissociates, however, with a high energy barrier.<sup>41,42</sup> To estimate the activation energy of the LOV dark-reaction using

our model, the energy of **TS-dark** should be compared to the energy of **R** rather than the energy of **Add**. The relative energy of **TS-dark** gives an upper limit for the activation energy of the dark reaction. The decrease of the energy barrier depends on how much the S–C4a covalent adduct is destabilized in the protein and can be different in LOV1 and LOV2 domains.

Blue-light illumination of the LOV at  $T = 77$  K yields a photoadduct absorbing at around 395 and 405 nm for LOV1 and LOV2, respectively.<sup>16</sup> This species, which is red-shifted compared to the room-temperature photoproduct absorbing at 390 nm, has been suggested to be a zwitterionic covalent adduct (Scheme 1B). The proposed mechanism of the photoreaction implied electron transfer from the thiol to FMN followed by recombination of the electron pair and a bond formation without proton transfer. We tested this hypothesis using our molecular model. Clearly, electron transfer from the thiol to the flavin does not take place upon excitation because complex **R-photo** consists of two uncharged fragments: the triplet lumiflavin and thiomethanol (Table 2). To form the zwitterion, electron transfer from the thiol to lumiflavin must be coupled with a shortening of the S–C4a distance. We calculated the energies of a series of lumiflavin–thiomethanol complexes with gradually decreasing S–C4a distances and with all other internal coordinates optimized in the triplet electronic state (Figure S6 in the Supporting Information). When the S–C4a distance was shortened to 2.4 Å, we observed a break of the C–S bond in thiomethanol, resulting in a release of the CH<sub>3</sub> radical. The lumiflavin–thiomethanol complex with a S–C4a distance of 2.6 Å is 12.7 kcal/mol above **R-photo** in the triplet state. The calculated positive net charge on SHCH<sub>3</sub> of +0.034 au is inconsistent with the suggestion that thiomethanol is an electron donor. We were not able to locate a minimum energy structure corresponding to the zwitterionic species in the ground electronic state. Thus, our results do not support the hypothesis that there is a pathway resulting in a zwitterionic S–C4a complex in the  $T_1$  electronic state. Most likely, the cysteine does not reduce the triplet flavin unless it is deprotonated and electron-transfer coupled with proton transfer takes place in LOV. The specific location of the proton accepted by FMN within the S–H group situated above the FMN ring explains why the rate of the photoreaction is not pH-dependent, whereas it decreases in D<sub>2</sub>O compared to that in a H<sub>2</sub>O solution.<sup>8,41,42</sup>

Although our model does not account for the protein matrix, it allows suggestions concerning the role of the protein in the primary photoreaction of LOV. The positioning of the thiol above the isoalloxazine ring, which is arranged by the protein, is crucial for the chemistry because it ensures the interaction of the p atomic orbital centered on the sulfur atom with the  $\pi^*$  MO of the flavin, resulting in the S–C4a bond. The photoreaction in the free complex is unlikely to occur because the position of the thiol above the isoalloxazine ring is not stabilized by any intermolecular interaction. A **Biradical**-like geometry necessary for the efficient intersystem crossing is ensured by the protein. We suggest that the S–C4a covalent adduct spontaneously decays because the reactant state of the cofactor is favored by specific cofactor–



protein interactions including the change of the Cys conformation upon covalent binding as well as hydrogen bonds with the protein. One can notice that the hydrogen bonds between the FMN cofactor and the protein have shorter distances in the dark-adopted crystal structure model (PDB 1n9l) with respect to the distances in the covalent adduct model (PDB 1n9o) assuming that the LOV protein may work as a spring to facilitate the dissociation of the S–C4a adduct. Interestingly, the above prediction excludes acid–base catalysis of the dark reaction by the protein. This agrees with the crystal structures of LOV, where no potentially catalytic residues are present in the vicinity of the isoalloxazine chromophore.<sup>13,22,23</sup> Movement of the sulfur atom away from C4a to pick up the proton is possibly coupled with the transition between the two conformations of the reactive cysteine observed in the LOV1 crystal structure.<sup>13</sup>

To conclude, the mechanism of the primary photocycle reactions in LOV, which is rooted in the chemical properties of FMN and Cys and a specific geometry of the reagents arranged by the protein, can be accurately analyzed on a model system using MCQDPT2/CASSCF quantum-chemical calculations. In particular, the obtained electronic transition energies are in good agreement with the experimental data of LOV. The electronic structure of the reactants, adduct, and the triplet state species is adequately described by single-determinant wave functions. One should account for the dynamic electron correlation to estimate the energy barrier of the H transfer and the reaction energy. A multiconfigurational representation of the wave function is important to describe the formation and dissociation of the C–S4a bond and to estimate the stability of the biradical intermediate. Both the light-induced and dark reactions proceed with a concerted mechanism with a single energy barrier corresponding to a hydrogen transfer. The protein plays a minor role in stimulating the chemical reactivity of the FMN cofactor in LOV, being, however, crucial for the specific positioning of the cysteine above the isoalloxazine ring as well as the destabilization of the photoproduct and determination of the conformational space of the dark reaction.

**Acknowledgment.** We are very grateful to Roger Goody for his generous hospitality and support and acknowledge the DFG (FOR526) for funding.

**Supporting Information Available:** Cartesian coordinates of the model complexes, the diagrams illustrating the data presented in Table 1, absolute energies calculated with the state-averaged energy functional, results of the energy scanning for the photo- and dark reactions, energy profile for the **Biradical** recombination. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Christie, J. M.; Raymond, P.; Powell, G. K.; Bernasconi, P.; Raibekas, A. A.; Liscum, E.; Briggs, W. R. *Science* **1998**, *282*, 1698–1701.
- (2) Christie, J. M.; Salomon, M.; Nozue, K.; Wada, M.; Briggs, W. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8779–8783.
- (3) Christie, J. M.; Briggs, W. R. *J. Biol. Chem.* **2001**, *276*, 11457–11460.

- (4) Taylor, B. L.; Zhulin, I. B. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 479–506.
- (5) Salomon, M.; Christie, J. M.; Knieb, E.; Lempert, U.; Briggs, W. R. *Biochemistry* **2000**, *39*, 9401–9410.
- (6) Salomon, M.; Eisenreich, W.; Durr, H.; Schleicher, E.; Knieb, E.; Massey, V.; Rudiger, W.; Muller, F.; Bacher, A.; Richter, G. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12357–12361.
- (7) Kottke, T.; Heberle, J.; Hehn, D.; Dick, B.; Hegemann, P. *Biophys. J.* **2003**, *84*, 1192–1201.
- (8) Swartz, T. E.; Corchnoy, S. B.; Christie, J. M.; Lewis, J. W.; Szundi, I.; Briggs, W. R.; Bogomolni, R. A. *J. Biol. Chem.* **2001**, *276*, 36493–36500.
- (9) Song, P. S. *Photochem. Photobiol.* **1968**, *7*, 311.
- (10) Miura, R. *Chem. Rec.* **2001**, *1*, 183–194.
- (11) Drossler, P.; Holzer, W.; Penzkofer, A.; Hegemann, P. *Chem. Phys.* **2002**, *282*, 429–439.
- (12) Kowalczyk, R. M.; Schleicher, E.; Bittl, R.; Weber, S. *J. Am. Chem. Soc.* **2004**, *126*, 11393–11399.
- (13) Fedorov, R.; Schlichting, I.; Hartmann, E.; Domratcheva, T.; Fuhrmann, M.; Hegemann, P. *Biophys. J.* **2003**, *84*, 2474–2482.
- (14) Kennis, J. T.; Crosson, S.; Gauden, M.; van Stokkum, I. H.; Moffat, K.; van Grondelle, R. *Biochemistry* **2003**, *42*, 3385–3392.
- (15) Ataka, K.; Hegemann, P.; Heberle, J. *Biophys. J.* **2003**, *84*, 466–474.
- (16) Schleicher, E.; Kowalczyk, R. M.; Kay, C. W. M.; Hegemann, P.; Bacher, A.; Fischer, M.; Bittl, R.; Richter, G.; Weber, S. *J. Am. Chem. Soc.* **2004**, *126*, 11067–11076.
- (17) Kay, C. W.; Schleicher, E.; Kuppig, A.; Hofner, H.; Rudiger, W.; Schleicher, M.; Fischer, M.; Bacher, A.; Weber, S.; Richter, G. *J. Biol. Chem.* **2003**, *278*, 10973–10982.
- (18) Heelis, P. F.; Parsons, B. J.; Phillips, G. O.; McKellar, J. F. *Photochem. Photobiol.* **1979**, *30*, 343–347.
- (19) Bittl, R.; Kay, C. W.; Weber, S.; Hegemann, P. *Biochemistry* **2003**, *42*, 8506–8512.
- (20) Kottke, T.; Dick, B.; Fedorov, R.; Schlichting, I.; Deutzmann, R.; Hegemann, P. *Biochemistry* **2003**, *42*, 9854–9862.
- (21) Richter, G.; Weber, S.; Romisch, W.; Bacher, A.; Fischer, M.; Eisenreich, W. *J. Am. Chem. Soc.* **2005**, *127*, 17245–17252.
- (22) Crosson, S.; Moffat, K. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2995–3000.
- (23) Crosson, S.; Moffat, K. *Plant Cell* **2002**, *14*, 1067–1075.
- (24) Neiss, C.; Saalfrank, P. *Photochem. Photobiol.* **2003**, *77*, 101–109.
- (25) Dittrich, M.; Freddolino, P. L.; Schulten, K. *J. Phys. Chem. B* **2005**, *109*, 13006–13013.
- (26) Losi, A.; Kottke, T.; Hegemann, P. *Biophys. J.* **2004**, *86*, 1051–1060.
- (27) Pulay, P.; Fogarasi, G. *J. Chem. Phys.* **1992**, *96*, 2856–2860.
- (28) Schmidt, M. W.; Gordon, M. S. *Annu. Rev. Phys. Chem.* **1998**, *49*, 233–266.
- (29) Nakano, H. *J. Chem. Phys.* **1993**, *99*, 7983–7992.
- (30) Lowdin, P. O. *Phys. Rev.* **1955**, *97*, 1474–1489.
- (31) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833–1840.

- (32) Granovsky, A. A. PC Gamess. <http://classic.chem.msu.su/gran/gamess/index.html> (accessed Mar 2004).
- (33) Schmidt, M. W.; Baldrige, 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*, 1347–1363.
- (34) Flükiger, H. P.; Lüthi, S.; Portmann, J. *MOLEKEL 4.2*; Weber and Swiss Center for Scientific Computing: Manno, Switzerland, 2003.
- (35) Sikorska, E.; Khmelinskii, I. V.; Koput, J.; Sikorski, M. *THEOCHEM* **2004**, *676*, 155–160.
- (36) Sakai, M.; Takahashi, H. *J. Mol. Struct.* **1996**, *379*, 9–18.
- (37) Neiss, C.; Saalfrank, P.; Parac, M.; Grimme, S. *J. Phys. Chem. A* **2003**, *107*, 140–147.
- (38) Sikorska, E.; Khmelinskii, I.; Komasa, A.; Koput, J.; Ferreira, L. F. V.; Herance, J. R.; Bourdelande, J. L.; Williams, S. L.; Worrall, D. R.; Insinska-Rak, M.; Sikorski, M. *Chem. Phys.* **2005**, *314*, 239–247.
- (39) Ghisla, S.; Massey, V.; Lhoste, J. M.; Mayhew, S. G. *Biochemistry* **1974**, *13*, 589–597.
- (40) Insinska-Rak, M.; Sikorska, E.; Herance, J. R.; Bourdelande, J. L.; Khmelinskii, I. V.; Kubicki, M.; Prukala, W.; Machado, I. F.; Komasa, A.; Ferreira, L. F. V.; Sikorski, M. *Photochem. Photobiol. Sci.* **2005**, *4*, 463–468.
- (41) Corchnoy, S. B.; Swartz, T. E.; Lewis, J. W.; Szundi, I.; Briggs, W. R.; Bogomolni, R. A. *J. Biol. Chem.* **2003**, *278*, 724–731.
- (42) Guo, H.; Kottke, T.; Hegemann, P.; Dick, B. *Biophys. J.* **2005**, *89*, 402–412.

CT0600114

## Allylic H-Abstraction Mechanism: The Potential Energy Surface of the Reaction of Propene with OH Radical

Milan Szori,<sup>†</sup> Christa Fittschen,<sup>‡</sup> Imre G. Csizmadia,<sup>†,§</sup> and Bela Viskolcz<sup>\*,†</sup>

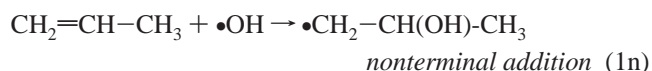
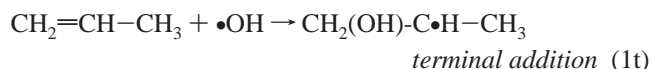
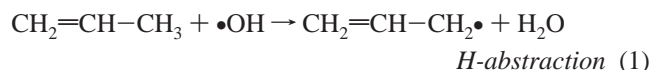
Department of Chemistry and Chemical Informatics, Faculty of Education, University of Szeged, Szeged, Boldogasszony sgt. 6, Hungary 6725, PhysicoChimie des Processus de Combustion et de l'Atmosphère - UMR 8522, Centre d'Etudes et de Recherches Lasers et Applications Université de Lille 1, F-59655 Villeneuve d'Ascq, Cedex, France, and Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 3H6

Received April 13, 2006

**Abstract:** The allylic H-atom abstraction reaction plays a more dominant role, especially at lower temperature, than addition reactions in the case of the  $\text{CH}_2=\text{CH}-\text{CH}_3 + \bullet\text{OH}$  system. Different computational methods including ab initio as well as density functional methods have been used to examine allylic H-abstraction. Both the energetically less favorable direct H-abstraction and the more favorable indirect H-abstractions have been investigated. Using first principles computations, for the indirect abstraction, a stable  $\pi$ - or reactantlike as well as a late productlike complex were found on the potential energy surface. Based on higher level single point calculations (QCISD(T)/6-311+G(3df,2p)), a new activation enthalpy value,  $\Delta^\ddagger H^\circ = 0.3 \pm 2$  kJ/mol, is suggested for the title reaction. The computed reaction enthalpy  $\Delta_r H^\circ = -124.7 \pm 2$  kJ/mol is in good agreement with the experimental value. The stability of the initial  $\pi$ -complex was found to be  $\Delta H^\circ_{\pi\text{-complex}} = -7.1$  kJ/mol. The product complex between the transition state and the product was found with the stability of  $-127.2$  kJ/mol.

### 1. Introduction

Many radical-molecule reactions exhibit negative activation energies, such as the addition of OH radical<sup>1</sup> or Cl atom to olefins.<sup>2</sup> The situation becomes more complicated in the case of substituted olefins, like propene. H-atom abstraction (1) may play an important role in the overall kinetics besides reaction channels of the terminal (1t) and nonterminal (1n) additions, both of which lead to the formation of hydroxy-alkyl radicals:<sup>3</sup>



Despite what is mentioned above, the dominance of quasi barrierless additional channels was reported in several cases.<sup>4</sup> However, H-abstraction reactions are critical in complex reaction systems such as the combustion of hydrocarbon fuels,<sup>5</sup> atmospheric chemistry,<sup>6</sup> and autoignition<sup>7</sup> as well as various processes of biological systems.<sup>8</sup> During the past few decades, OH radical reactions involving different organic compounds were of special interest in gas-phase kinetic studies.<sup>9–11</sup> The most recent experimental study focused upon reactions used in the detection of OH radical decay at low temperatures by laser induced fluorescence (LIF) spectroscopy.<sup>12</sup> The branching ratios of elementary steps in most

\* Corresponding author e-mail: viskolcz@jgytf.u-szeged.hu.

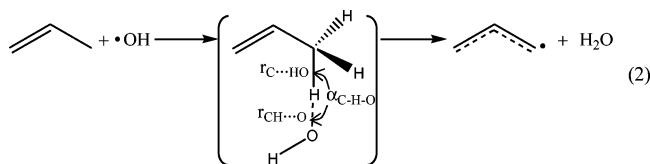
<sup>†</sup> University of Szeged.

<sup>‡</sup> Centre d'Etudes et de Recherches Lasers et Applications Université de Lille 1.

<sup>§</sup> University of Toronto.

cases were not distinguishable by a direct technique where only one species is detected. If a chemical system is small enough, then ab initio calculations may be utilized to produce an accurate reactive potential energy surface (PES),<sup>13</sup> thus permitting current chemical reaction rate theories to be tested. Besides extensive experimental studies,<sup>14,15</sup> a number of theoretical calculations were carried out on the elementary steps involving the reaction for the OH radical with different unsaturated hydrocarbons.<sup>16</sup> The reaction of the simplest system ( $\text{H}_2\text{C}=\text{CH}_2 + \text{OH}$ ) has been studied extensively.<sup>17–19</sup> However, propene is the next step toward the understanding of OH reaction with the unsaturated hydrocarbon homologue, and only a few theoretical studies have been reported for the propene + OH system.<sup>20,21</sup> Surprisingly, H-atom abstraction has not yet been studied using high level quantum chemistry calculations, and neither have theoretical comparisons with addition channels been well established.

Allylic hydrogen may be extracted via two possible H-atom abstraction channels which can play a significant role in the overall kinetics of the OH and propene system. One of them, indirect abstraction, can start from a so-called  $\pi$ -complex, the other is the direct H-abstraction. Higher unsaturated hydrocarbons (including higher degrees of unsaturation) react with OH radicals, via additions, as well as hydrogen atom abstractions.<sup>22</sup> A prototype of the allylic H-abstraction is shown in eq 2:



For an indirect abstraction the central structure in square brackets in eq 2 would be a minimum energy structure labeled as a reactant (or  $\pi$ ) complex. For a direct abstraction it would be a transition state.

Both QCISD and CCSD theoretical models yield results that have been shown to be relatively stable with regard to a modest spin contamination in the unrestricted case.<sup>23</sup> Radom et al. suggested that UMP2 should be used with caution, since this method is too erratic for general use for radicals.<sup>24</sup> The density functional theory (DFT), which uses gradient-corrected exchange and correlation potentials, has been shown to be an efficient and accurate tool to calculate molecular properties such as structural geometry and vibrational frequencies.<sup>25–28</sup> The unrestricted DFT wave function is typically considerably less spin contaminated than the corresponding UHF wave function for a given open shell system.<sup>29,30</sup> A number of DFT transition state property studies have been carried out<sup>31,32</sup> and assessed by comparing the computed results with experimental and other theoretical results.

In the present paper, different computational methods were applied and analyzed for similarities and differences in their respectively computed PESs. One of the aims of this study is to test systematically the different levels of theory for their efficiency and accuracy with respect to the given system and, furthermore, to demonstrate the importance of the allylic

H-atom abstraction. One has to note that some comparisons were published previously,<sup>33,34</sup> for a similar type of reaction. The experience gained will help to choose a suitable method for describing hydrogen abstraction reactions from unsaturated hydrocarbons and their derivatives such as polyunsaturated fatty acids (PUFA).

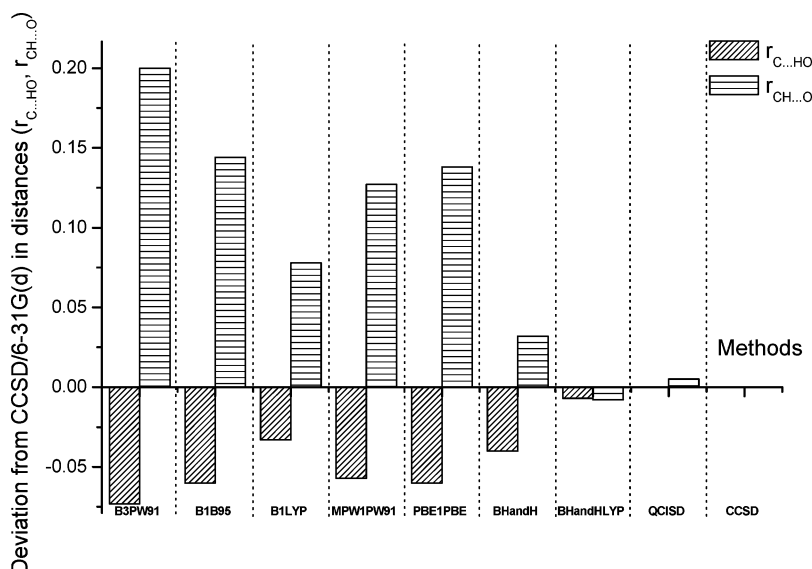
## 2. Computational Methods

To generate potential energy surfaces (PESs), in the vicinity of the H-abstraction transition state, relaxed scans were carried out by systematically altering two variables, namely  $r_{\text{C}\cdots\text{HO}}$  (i.e. the C–H bond to be broken) and  $r_{\text{CH}\cdots\text{O}}$  (i.e. the H–O bond to be formed), as shown in eq 2. Values of the  $r_{\text{C}\cdots\text{HO}}$  parameter were changed between 1.1 Å and 1.3 Å, while the  $r_{\text{CH}\cdots\text{O}}$  parameter was varied from 1.2 Å to 1.6 Å. The step sizes were 0.04 Å in both directions. The quadratic convergence SCF procedure<sup>35</sup> was used in each scan. The potential energy surface with all tested methods combined utilized the standard 6-31G(d) split valence basis set.<sup>36</sup>

For geometry optimizations, a number of standard methods were used, namely ab initio HF,<sup>37–39</sup> MP2,<sup>40–43</sup> QCISD,<sup>44</sup> and CCSD.<sup>45–48</sup> A set of density functional theory (DFT)<sup>49</sup> methods such as B3PW91, B1B95, B1LYP, MPW1PW91, PBE1PBE, BHandH, and BHandHLYP was also applied to characterize the transition state. Two DFT methods are presented in the present paper as two differently performing examples. These were the B3LYP<sup>50,51</sup> and the BH&HLYP<sup>52</sup> methods, exemplifying in geometry points of view a not very accurate and a relatively accurate DFT methods, respectively. The remaining DFT results, listed in Table S1 of the Supporting Information, show that they produce geometries scattered around the most accurate QCISD and CCSD methods (Figure 1) which also produced the best activation enthalpies for the reaction using a large basis set. The minima and saddle points of the PESs (i.e. reactants, products, complexes, and transition states) were checked by frequency analysis at each level of theory. The zero point vibrational energy (ZPVE) was scaled by standard factors.<sup>53</sup>

To increase the accuracy of the computed PES, a non-iterative addition of triple excitations was taken into consideration using the QCISD(T)<sup>44</sup> and CCSD(T)<sup>44</sup> methods. The QCISD(T)-PES and the CCSD(T)-PES were mapped out by QCISD(T)/6-31G(d) and CCSD(T)/6-31G(d) single point calculations using CCSD/6-31G(d) geometries, denoted as QCISD(T)/6-31G(d)//CCSD/6-31G(d) and CCSD(T)/6-31G(d)//CCSD/6-31G(d), respectively.

Basis set effects on the electronic energy were also studied with single point calculations for the BH&HLYP method using BH&HLYP/6-31G(d) geometries and also at the QCISD(T) and CCSD(T) level of theories using CCSD/6-31G(d) and BH&HLYP geometries. Two types of split valence basis sets<sup>36</sup> were used which were also augmented with diffuse and polarization functions {6-31G(d), 6-311G-(d,p), 6-311+G(d,p), 6-311+G(3df,2p)} as well as Dunning's correlation consistent basis sets<sup>54–58</sup> {cc-pVxZ, aug-cc-pVxZ (where x is D, T, and Q for cc-pVxZ and D, T for aug-cc-pVxZ)}. The effects of using these different basis sets were explored.



**Figure 1.** A comparison of different DFT methods with high level ab initio (QCISD and CCSD) calculations in terms of geometrical parameters ( $r_{C...OH}$  and  $r_{CH...O}$ ) of the H-abstraction transition state associated with the  $CH_2=CHCH_3 + \bullet OH$  system.

The basis set superposition error (BSSE) is considered for the complexes by using the simple counterpoise (CP) method<sup>59</sup>

$$\Delta E^{CP} = E^{AB} - [E^{A(AB)} + E^{B(AB)}] \quad (3)$$

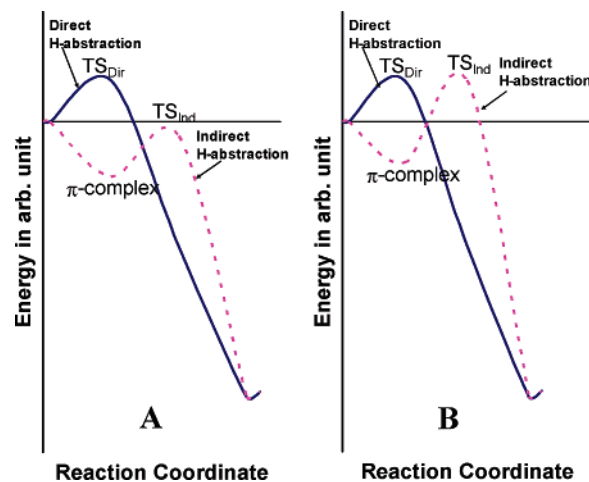
where  $E^{AB}$  is the energy of the supermolecule (e.g.  $\pi$ -complex),  $E^{A(AB)}$  is the energy of molecule A (e.g. propene) with the basis set of supermolecule AB, and  $E^{B(AB)}$  is the energy of molecule B (e.g. hydroxyl radical) with the basis set of the AB supermolecule. This method was applied to the QCISD(T) and CCSD(T) single point calculations for the final analysis of the reaction profile (using "Counterpoise=2" keyword in Gaussian03).

For testing the IRCMax<sup>60</sup> method, calculations are carried out on the QCISD(T)/6-311+G(3df,2p)//BH&HLYP/6-31G(d) level of theory. All of the calculations were carried out using the GAUSSIAN03 program package.<sup>61</sup>

### 3. Scope

The allylic H-abstraction can be envisaged with the following possibilities: (i) a direct abstraction mechanism, (ii) an indirect abstraction mechanism, and (iii) a mixture of a direct and indirect abstraction mechanism.

To decide which one of the above possibilities is operative it is necessary to search for transition states. If only one TS exists, then it can be analyzed whether (i) or (ii) is the actual mechanism. If a pair of TSs are located, then (iii) is operative, and the relative barrier heights for the two TSs will predetermine which mechanism is dominantly operative as it is illustrated in Figure 2. If either (ii) or (iii) is operative, then one must search for one or two reactants or  $\pi$ -complexes. If only one  $\pi$ -complex exists, then it should be common for both the H-abstraction and the OH addition (Figure 3A). If there are two  $\pi$ -complexes formed, then one is for the H-abstraction and one is for the OH addition (Figure 3B): (i) one  $\pi$ -complex is formed which is common for both H-abstraction and OH addition (Figure 3A) and (ii) two



**Figure 2.** Two possible reaction energy profiles involving two alternative H-abstraction mechanisms showing that direct and indirect H-abstractions compete with each other. Note that  $TS_{Dir}$  is always positive, even though the barriers are usually low, while  $TS_{Ind}$  may be either positive or negative. A: indirect H-abstraction is favored because  $TS_{Ind} < TS_{Dir}$  and B: direct H-abstraction is favored because  $TS_{Ind} > TS_{Dir}$ .

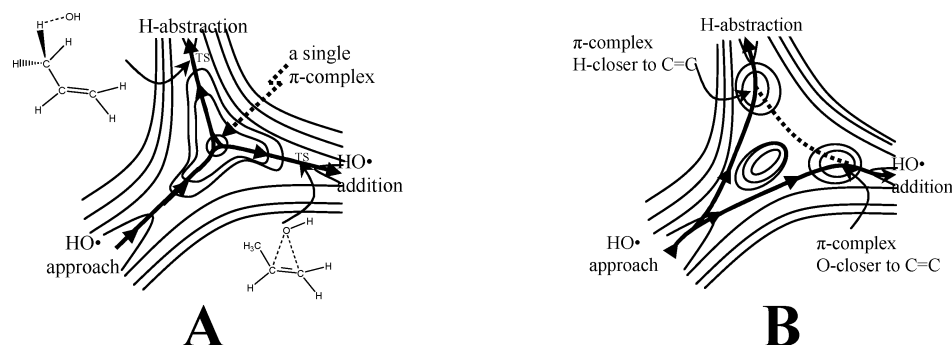
$\pi$ -complexes are formed, one is for the H-abstraction and one is for the OH addition (Figure 3B).

In this paper we shall seek to answer these questions but not necessarily in the order listed above. Before all of these questions could be answered, a full exploration of the theoretical methods and basis sets applied must be carried out for the computational accuracy which will make the conclusions reliable.

## 4. Results and Discussion

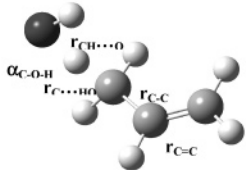
### 4.1. Geometries of the Critical Points and Along the RC.

Along each path there may be several minimum energy points corresponding to either reactants, reaction intermediates, or products. This study explores the geometries of the maxima and minima along the hydrogen abstraction reaction



**Figure 3.** Model potential energy surfaces for the  $\text{CH}_2=\text{CHCH}_3 + \bullet\text{OH}$  system involving addition as well as allylic H-abstraction. A: two reaction paths involving a single  $\pi$ -complex and B: two reaction paths involving two separate  $\pi$ -complexes

**Table 1.** Geometric Parameters of Reactants, Products, and First-Order Saddle Point (TS) Computed at the BH&HLYP/6-31G(d), QCISD/6-31G(d), and CCSD/6-31G(d) Levels of Theory and Available Experimental Results

		propene + OH				TS			allyl radical + H <sub>2</sub> O			
		BH&HLYP	CCSD	QCISD	Exp.	BH&HLYP	CCSD	QCISD	BH&HLYP	CCSD	QCISD	Exp.
$r_{\text{C}=\text{C}}$	in Å	1.322	1.338	1.339	1.341 <sup>a</sup>	1.331	1.343	1.344	1.379	1.390	1.390	1.3869 <sup>c</sup>
$r_{\text{C}-\text{C}}$	in Å	1.495	1.504	1.504	1.506 <sup>a</sup>	1.467	1.480	1.480	1.379	1.391	1.391	1.3869 <sup>c</sup>
$r_{\text{C}\cdots\text{HO}}$	in Å	1.089	1.098	1.099	1.117 <sup>a</sup>	1.224	1.231	1.232	-	-	-	-
$\alpha_{\text{C}=\text{C}-\text{C}}$	in degree	125.1	124.7	124.7	124.3 <sup>a</sup>	124.5	124.1	124.1	124.8	124.4	124.4	123.96 <sup>c</sup>
$\alpha_{\text{C}-\text{C}-\text{H}}$	in Å	111.1	111.1	111.1	110.7 <sup>a</sup>	109.5	109.1	109.1	-	-	-	-
$\alpha_{\text{C}-\text{H}-\text{O}}$	in degree	-	-	-	-	173.7	171.7	171.9	-	-	-	-
$\alpha_{\text{C}=\text{C}-\text{C}-\text{H}}$	in degree	120.7	120.4	120.4	n.a.	-105.6	-104.9	-104.7	-	-	-	-
$\pi_{\text{C}-\text{C}-\text{H}-\text{O}}$	in degree	-	-	-	-	76.7	52.0	56.0	-	-	-	-
$\pi_{\text{C}-\text{H}(\text{O}-\text{H})}$	in degree	-	-	-	-	-27.9	-9.3	-12.7	-	-	-	-
$r_{\text{O}-\text{H}}$	in Å	0.969	0.984	0.984	0.971 <sup>b</sup>	0.966	0.981	0.981	0.957	0.970	0.970	0.9575 <sup>a</sup>
$\alpha_{\text{H}(\text{O}-\text{H})}$	in degree	-	-	-	-	97.9	97.2	97.1	104.6	104.0	104.0	104.51 <sup>a</sup>
$r_{\text{CH}\cdots\text{O}}$	in Å	-	-	-	-	1.309	1.317	1.318	0.957	0.970	0.970	0.9575 <sup>a</sup>

<sup>a</sup> – cExperimental geometries from refs 62–64, respectively.

of the  $\text{CH}_3\text{CH}=\text{CH}_2 + \text{OH}$  system. As pointed out in the previous section, such a path of hydrogen abstraction may be at least partially, if not completely, different from the path of the addition of the OH radical to the double bond.

**4.1.1. Reactants and Products.** Table 1 shows, besides the data of TS obtained initially, the calculated geometrical parameters of reactants and products at the BH&HLYP/6-31G(d), QCISD/6-31G(d), and CCSD/6-31G(d) levels of theory. Both reactants and product are in very good agreement with the available experimental data:<sup>62–64</sup> the largest deviation from experimental data in the geometrical parameters of the reactants was found in the allylic C–H bond length (the bond to be broken during the reaction), but these deviations were smaller than 0.02 Å at the BH&HLYP/6-31G(d) and smaller than 0.03 Å at the CCSD/6-31G(d) level of theory. In terms of bond angle deviation, the angular deviations of the carbon chain (denoted  $\alpha_{\text{C}=\text{C}-\text{C}}$ ) were less

than 0.8 degrees using DFT and half of that value at the CCSD level of theory. On the side of the products, the deflection between experimental and theoretical methods was about the same in the angles, and in the distance deviation it was found to be around 0.01 Å. However, it is clear that the quality of the geometry can be improved somewhat by increasing the number of basis functions, but our aim was to obtain reasonable geometries not only for this system but also for larger alkenes.

**4.1.2. Potential Energy Surfaces and Location of the TS in H-Abstraction.** The selected geometry parameters (namely the  $r_{\text{CH}\cdots\text{O}}$ ,  $r_{\text{C}\cdots\text{HO}}$ , and  $\alpha_{\text{C}-\text{H}-\text{O}}$ ) and enthalpies are listed and compared in Table 2 specifically for TS of the reaction 2. This demonstrates that at the HF level of theory, the TS structure is almost halfway between the reactant and product. However, high level ab initio calculations such as CCSD show that the TS became more and more reactantlike.

**Table 2.** Standard Reaction and Activation Enthalpies in kJ/mol and Relevant Geometrical Parameters of the Indirect H-Abstraction TS Structure for the  $\text{CH}_2=\text{CHCH}_3 + \bullet\text{OH} \rightarrow \text{CH}_2=\text{CHCH}_2\bullet + \text{H}_2\text{O}$  Reaction Computed at Different Levels of Theory Using a 6-31G(d) Split Valence Basis Set

method	$\Delta_r H^\circ$	$\Delta^\ddagger H^\circ$	$r_{\text{CH}\cdots\text{O}}$ (Å)	$r_{\text{C}\cdots\text{HO}}$ (Å)	$\alpha_{\text{C-H-O}}$
B3LYP	-102.2				
HF	-70.2 <sup>a</sup>	88.3 <sup>a</sup>	1.270	1.276	176.3
MP2	-81.6 <sup>a</sup>	47.7 <sup>a</sup>	1.293	1.216	163.4
BH&HLYP	-87.9	18.8	1.309	1.224	173.7
QCISD	-89.8 <sup>a</sup>	30.7 <sup>a</sup>	1.318	1.232	171.9
CCSD	-89.2	32.0	1.317	1.231	171.6
exp.	-128.53 <sup>b</sup>	-3.7 <sup>c</sup>			

<sup>a</sup> Values of ZPVE are scaled. Scale factors used are from ref 53.

<sup>b</sup> The experimental reaction enthalpy is calculated from data given in refs 65–67. <sup>c</sup> The experimental reaction rate is published in ref 72.

Generally, the transition state structure obtained by the CCSD/6-31G(d) methods can be considered to be the most accurate geometry.

Within the coupled cluster method, the  $\text{C}\cdots\text{HO}$  bond length of the TS stretched to 1.231 Å from the original 1.098 Å, amounting to an increase of 0.133 Å. However, the coupled cluster method revealed that the  $\text{CH}\cdots\text{O}$  bond distance was 1.317 Å, which is about 0.347 Å away from the final H–O bond length of the product  $\text{H}_2\text{O}$  (0.970 Å). These two geometrical parameters indicate that the allylic H-atom abstraction has an early transition state. In addition, the TS corresponded to a nearly linear approach between the species, since the optimized  $\alpha_{\text{C-H-O}}$  was 171.7°. The TS structures listed in Tables 1 and 2 obtained by BH&HLYP/6-31G(d) and QCISD/6-31G(d) are very similar to those found by CCSD/6-31G(d). It is important to note that this TS could not be found at the B3LYP/6-31(d) optimization level.

The differences in reaction enthalpy ( $\Delta_r H^\circ$ ) and activation enthalpy ( $\Delta^\ddagger H^\circ$ ) are rather large and show a marked disagreement with the experimental values as listed in Table 2. It is rather difficult to suggest that there is a convergence in energies of the different methods used up until now. All theoretical methods underestimate the reaction exothermicity by a large margin of 25–60 kJ/mol. The calculated activation enthalpy was also overestimated by about the same amount. Thus, the various methods, used so far, produced values which were too high, with respect to both in terms of thermodynamics and kinetics. To conclude, higher post-HF methods coupled with relatively small basis sets do not permit accurate calculations of the enthalpy profile of this reaction. Clearly, larger basis sets are required to achieve a desirable level of accuracy. The nature of the potential energy surface particularly in the vicinity of the transition state can help to estimate the error of our calculations as well as to select methods using even a modest basis set in order to determine how to compute a more accurate potential energy surface.

The relative energy can be calculated according to the following equation using any (x) level of theory (e.g. x =

BH&HLYP) for the generation of a potential energy surface (4)

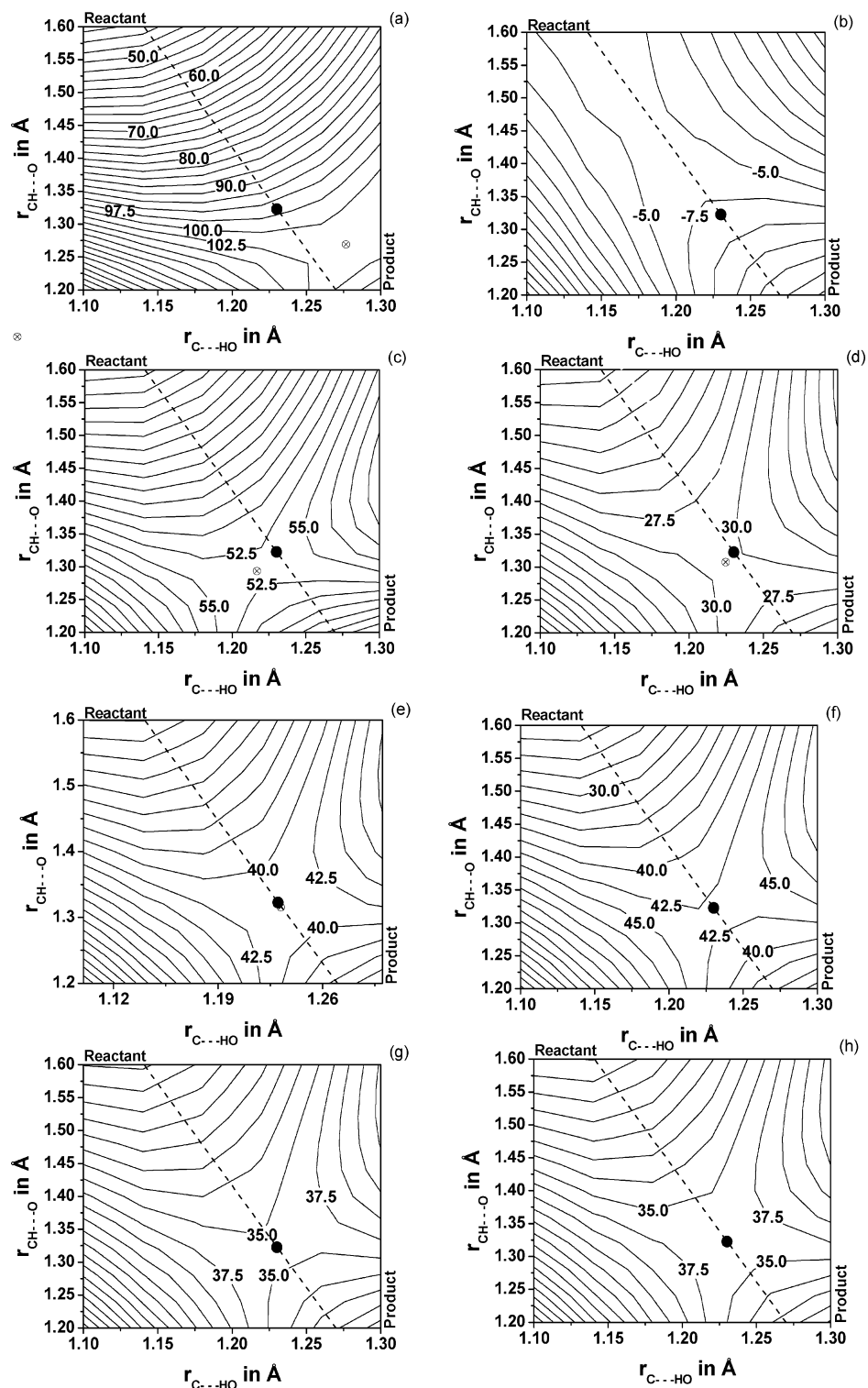
$$\Delta E_x = E_x(r_{\text{C}\cdots\text{HO}}, r_{\text{CH}\cdots\text{O}}) - [E_x(\text{propene}) + E_x(\text{OH})] \quad (4)$$

where  $E_x(r_{\text{C}\cdots\text{HO}}, r_{\text{CH}\cdots\text{O}})$  represents the electronic energy of a given structure (determined by a pair of fixed variables:  $r_{\text{C}\cdots\text{HO}}$  and  $r_{\text{CH}\cdots\text{O}}$ ). The quantity of  $[E_x(\text{propene}) + E_x(\text{OH})]$  stands for the total electronic energy of the structures of propene and OH radical, each being optimized at the x level of theory.

Figure 4 shows contour map presentations of potential energy surfaces (PESs), which illustrate the change of energy in the vicinity of the transition state of the reaction computed at various levels of theory. The minimum energy pathway (MEP) at the CCSD/6-31G(d) level of theory is denoted by the dotted line. Each diagram has contour lines as calibration for relative energy. Selected values of relative energies are shown to demonstrate the characteristics of a given PES. The progress of the reaction is made evident by clearly marking the reactant and product side of the PES at the top of the left-hand side and at the bottom of the right-hand side, respectively, in each of the figures.

Figure 4h illustrates the PES computed at the CCSD(T)/6-31G(d)//CCSD/6-31G(d) level of theory. This is the highest level of theory applied for the scan and is therefore assumed to be the closest to the “exact” surface. For this reason, it was used in all subsequent discussions as the reference surface. This figure shows that the electronic energy of the well-defined transition state is located between 35.0 and 37.5 kJ/mol.

It is clear from Figure 4 that there are some significant differences between the lower level PESs and the reference PES. In comparing the PESs generated at various levels of theory in each diagram of Figure 4, the large dot (●) and the dotted line symbolize the position of TS and MEP, respectively, on the reference PES {namely, CCSD(T)/6-31G(d)//CCSD/6-31G(d)}. The location of the TS (symbolized by ⊗) in HF-PES (denoted by a in Figure 4) is at 1.270 Å and 1.276 Å,  $r_{\text{CH}\cdots\text{O}}$  and  $r_{\text{C}\cdots\text{HO}}$ , respectively. These values are also summarized in Table 2. We should note that the shape of the B3LYP surface (b) is significantly different from the remainder of PESs. Surprisingly, there is no transition state (TS) on this surface, as only a path or a channel was found instead of a well-localized TS. The TS (symbolized by ⊗) which belongs to the higher level MP2 ab initio calculation (denoted by c) occurs at 1.293 Å and 1.216 Å for  $r_{\text{CH}\cdots\text{O}}$  and  $r_{\text{C}\cdots\text{HO}}$ , respectively. This suggests that at this level of theory, the correlation energy treatment shifts the position of the TS structure. The BH&HLYP surface (denoted by d) is an additional step toward the CCSD surface. Both  $r_{\text{CH}\cdots\text{O}}$  (1.309 Å) and  $r_{\text{C}\cdots\text{HO}}$  (1.224 Å) are larger than those associated with the MP2 surface. QCISD (denoted by e) and CCSD (denoted by f) are noticeably similar to each other. The shapes of these surfaces seem analogous: 1.327 Å and 1.231 Å for  $r_{\text{CH}\cdots\text{O}}$  and  $r_{\text{C}\cdots\text{HO}}$ , respectively, in both cases. The QCISD(T) and CCSD(T) surfaces (denoted as g and h, respectively, in Figure 4) were generated on CCSD/6-31G(d) geometries with single point calculations and had no deviation in the geometric aspect.

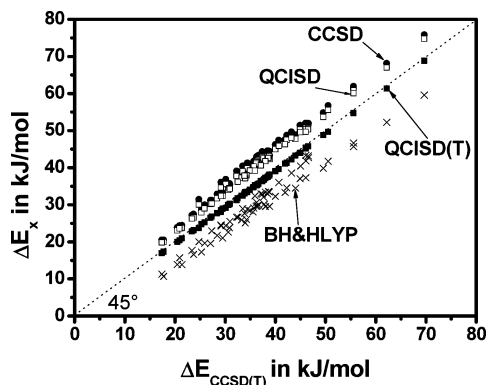


**Figure 4.** Potential energy surfaces (PES) around the TS of  $\text{CH}_2=\text{CHCH}_3 + \bullet\text{OH} \rightarrow \text{CH}_2=\text{CHCH}_2\bullet + \text{H}_2\text{O}$  reaction computed at different levels of theory: (a) HF, (b) B3LYP, (c) MP2, (d) BH&HLYP, (e) QCISD, (f) CCSD, (g) QCISD(T), and (h) CCSD(T). The symbol  $\otimes$  indicates the place of the TS structure in the given method. The dotted line symbolizes the minimum energy pathway (MEP) and  $\bullet$  symbolizes the position of the TS obtained at the CCSD/6-31G(d) level of theory which is the “reference PES” (shown also in Figure 1).

The relative energies show remarkable deviations from the reference surface. The highest energy (101.5 kJ/mol) belongs to the TS structure found using the HF/6-31G(d) level of theory, as shown in the first diagram (a) in Figure 4. However, the relative energy difference was decreased by increasing the electron correlation: MP2, QCISD, and CCSD

(methods are denoted by c, e, and f of Figure 4, respectively) and the barrier heights were 52.8, 41.1, and 42.5 kJ/mol, respectively. Clearly, the deviation between the latter pair of values is not significant. The BH&HLYP surface (denoted by d of Figure 4) predicts the lowest barrier height of 29.6 kJ/mol among the calculated surfaces. It is interesting to note





**Figure 5.** A comparison of relative energies obtained at the QCISD(T)/6-31G(d) (■), CCSD/6-31G(d) (●), QCISD/6-31G(d) (□), and BH&HLYP/6-31G(d) (×) levels of theory with respect to CCSD(T)/6-31G(d)//CCSD/6-31G(d) energies.

**Table 3.** Linear ( $y = mx + b$ ) Fitted Parameters for the Correction of Relative Energy Values, Obtained at Various Levels of Theory against Relative Energies Computed at the CCSD(T)/6-31G(d) Level of Theory

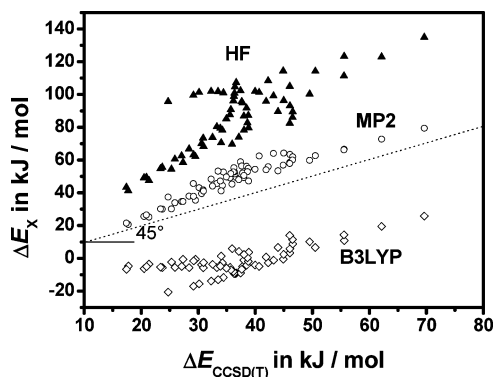
model	$m$	$b$	$R^2$
HF	1.714	23.2	0.816
B3LYP	0.617	-102.8	0.741
MP2	1.178	5.7	0.942
BH&HLYP	0.946	-4.6	0.984
QCISD	1.052	2.3	0.997
CCSD	1.062	3.1	0.996
QCISD(T)	0.991	-0.5	0.999

that although the TS was not found on the B3LYP surface (b of Figure 4), it gave relative energy differences in the  $-5$  and  $-7.5$  kJ/mol range virtually at every grid point which are the closest to the experimental TS energy of  $-3.7$  kJ mol $^{-1}$  (see Table 2).

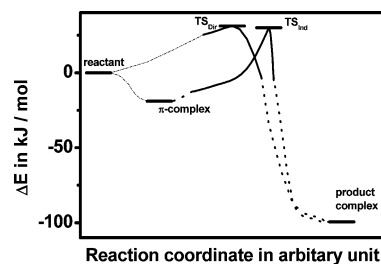
The correlation of the different level of theories with relative CCSD(T)/6-31G(d) energies were examined. As Figure 5 shows, there is a noticeable correlation between the CCSD(T)/6-31G(d) and CCSD/6-31G(d) relative energies. Not surprisingly, the situation is the same with QCISD(T)/6-31G(d) and QCISD/6-31G(d). The fitting of BH&HLYP/6-31G(d) energy values to the CCSD(T)/6-31G(d) is also a good one. Linear fitting was carried out, as shown in (5)

$$\Delta E_x(r_{C\cdots HO}, r_{CH\cdots O}) = m\Delta E_{CCSD(T)}(r_{C\cdots HO}, r_{CH\cdots O}) + b \quad (5)$$

where  $x$  is QCISD(T), QCISD, CCSD, MP2, HF, BH&HLYP, and B3LYP.  $\Delta E_x$  represents the relative energy calculated by the given method ( $x$ ). The parameters  $m$  and  $b$  are fitted according to eq 5 and are summarized in Table 3. The correlation coefficient is up to 0.98 in the QCISD(T), QCISD, CCSD, and BH&HLYP cases indicating a relatively close approximation to the primary standard PES computed at the CCSD(T). At QCISD(T) the slope is the closest to unity: it is 0.991. In addition, the  $y$ -intercept is almost zero ( $-0.5$  kJ/mol). However, QCISD and CCSD gave similar results for slope, but the deflection of the intercept increases to 2.3 and 3.1 kJ/mol, respectively. At BH&HLYP, the fitted parameter  $m$  is 0.946 which is slightly smaller, and parameter



**Figure 6.** A comparison of relative energies obtained at the HF/6-31G(d) (▲), MP2/6-31G(d) (○), and B3LYP/6-31G(d) (◇) levels of theory with respect to CCSD(T)/6-31G(d)//CCSD/6-31G(d) energies.

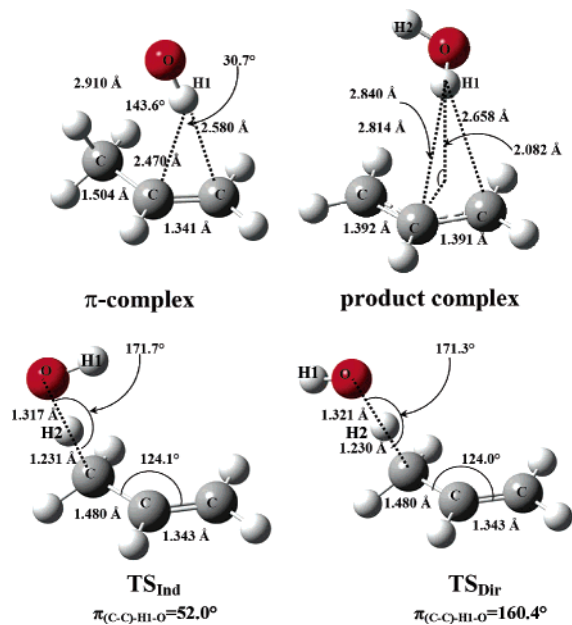


**Figure 7.** Relative BH&HLYP energy profile of indirect and direct H-abstractions: IRC calculation (solid curvature), normal optimization step (dotted curvature), and supplementary curvature (dash-dotted, thin line). Optimized critical points are represented by solid horizontal lines.

$b$  is also somewhat lower at  $-4.6$  kJ/mol. However, the remainder of the methods when plotted against CCSD(T) (Figure 6) do not show any obvious correlation, especially in the region between 20 and 50 kJ/mol of relative energies which corresponds to the value around the TS. In regards to these data the MP2 points show the best tendency for correlation in Figure 6, but the actual correlation coefficient of this linear fit is under 0.95.

#### 4.1.3. Location of Reactant ( $\pi$ ) and Product Complexes.

The fact that the experimental enthalpy of activation is negative ( $-3.7$  kJ/mol) with respect to the reactant state (Table 2) means that H-abstraction should be relevant. The negative value of this activation enthalpy might predict that this reaction channel goes through a reactant or a  $\pi$ -complex of the OH plus propene and not via direct linear collision. However, one needs to address the question if it is possible to go from the  $\pi$ -complex to the TS of H-abstraction reaction as well as back to the separate reactants. Both direct and indirect H-abstraction channels were explored using IRC calculation at the BH&HLYP/6-31G(d) level of theory. Due to the numerical problem, these calculations were crashed after certain steps. The latest steps were used to optimize further (with the same level of theory) to reach the complexes. The indirect reaction pathway was followed along a long range using IRC (Figure 7). The latest point has energy under the reactant energy level by 12.6 kJ/mol. Consequently, there should be a prereaction complex for this channel. Then normal optimization was used to find it, and the  $\pi$ -complex

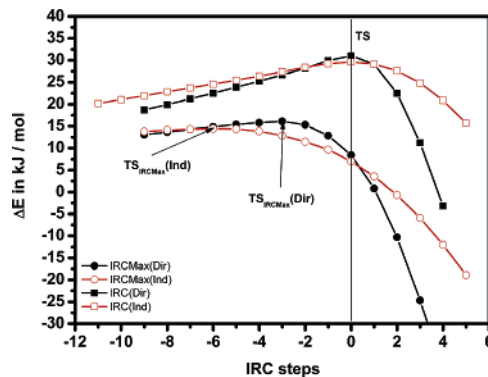


**Figure 8.** Top: Geometric parameters of reactant (so-called  $\pi$ ), product complexes. Bottom: transition states (TS) for indirect and direct H-abstractions showing geometrical distinctions. All structure were computed at the CCSD/6-31G(d) levels of theory.

was characterized (Figure 8). The direct pathway was broken when OH was about 1.46 Å away from the hydrogen of the methyl group. Relative energy of this structure was 25.3 kJ/mol. The structures of these complexes were also determined by full geometry optimization at the CCSD/6-31G(d) level of theory. For both mechanisms, the same product complex was found (Figure 8).

In the  $\pi$ -complex, the hydroxyl radical with the hydrogen end is situated at the double bond 2.470 Å away from the central carbon. It means that the hydrogen is 0.11 Å closer than the terminal carbon of the ethylene group. The oxygen is orientated toward the hydrogen of the methyl group with a distance of 2.910 Å and an (O-H...C) angle of 143.6° as indicated in Figure 8. This arrangement shows the relation to the H-abstraction. The single carbon-carbon bond has not changed at all when the  $\pi$ -complex is formed from the reactants (1.504 Å). The double bond has changed slightly (from 1.338 Å to 1.341 Å). A similar complex structure was obtained using the MP2/6-311G(d,p) level of theory by Vivier-Bunge et al.<sup>20</sup> This result can also confirm our BH&HLYP geometry. The product complex is very weakly bound and does not have a symmetric structure, which is depicted on the right-hand side of Figure 8. The water molecule is not in the symmetry plane of the allyl radical; however, carbon-carbon bond lengths are almost the same (1.392 Å and 1.391 Å). As is indicated in Figure 8, one of two hydrogens (H1) of the water molecule is at a 2.082 Å distance from the plane of the allyl radical. The other hydrogen (H2) and oxygen in the water molecule are closer to one end of the allyl radical than to the other. The distance between the central carbon atom and the hydrogen (H1) projected to the plane of allyl radical is 1.931 Å.

**4.1.4. Details of Reaction Profiles around the Transition States.** The IRCMax calculations are carried out on the



**Figure 9.** Comparisons of the IRC and IRCMax calculations at both direct (Dir) and indirect (Ind) channels. IRC and IRCMax were carried out at the BH&HLYP/6-31G(d) and QCISD(T)/6-311+G(3df,2p)//BH&HLYP/6-31G(d) levels of theory, respectively.

**Table 4.** Critical Geometric Parameters at the TS of Direct (TS<sub>Dir</sub>) and Indirect (TS<sub>Ind</sub>) Abstractions Computed at BH&HLYP/6-31G(d) and Obtained with the IRCMax Method

methods	$r_{C\cdots HO}$ (Å)	$r_{CH\cdots O}$ (Å)	$E_{rel}$ (kJ/mol)
	TS <sub>Dir</sub>		
TS (BH&HLYP/6-31G(d))	1.224	1.314	31.1
TS (IRCMax)	1.119	1.443	16.1
	TS <sub>Indir</sub>		
TS (BH&HLYP/6-31G(d))	1.224	1.309	29.6
TS (IRCMax)	1.118	1.439	14.5

QCISD(T)/6-311+G(3df,2p)//BH&HLYP/6-31G(d) level of theory for both direct and indirect H-abstractions. The relative energies are related to the levels of propene and OH without zero-point correction, and they are calculated without counterpoise correction. The relative energies are plotted against the IRC steps as shown in Figure 9. The zeroth step is the transition state (TS) itself at the BH&HLYP/6-31G(d) level of theory in both cases of direct and indirect H-abstractions. In IRCMax calculations, the TS structures become considerably more reactantlike as Table 4 shows. If one compares the values of  $r_{C\cdots HO}$  at the TS in IRCMax calculations (1.118 Å and 1.119 Å) with the experimental value of C-H bond in propene (1.117 Å), then these IRCMax calculations are not reasonable. The similarity of transition states at CCSD/6-31G(d) and at BH&HLYP/6-31G(d) also confirms that BH&HLYP geometry can be far more realistic. However, CCSD calculation at the larger basis set would give a more accurate structure of the TS, but these are very expensive to calculate. We also believe that the TS structures do not change so much.

**4.1.5. Molecular Geometries.** The deviation of BH&HLYP/6-31G(d) geometric parameters from the results obtained at the CCSD/6-31G(d) level of theory was studied in every grid point of the potential energy surface. For example, at the  $r_{CH\cdots O} = 1.1$  Å and  $r_{C\cdots HO} = 1.2$  Å point of the PES, the optimized parameter  $r_{C=C}$  was 1.330 Å using the BH&HLYP method and 1.342 Å obtained by the CCSD level of theory, so the difference in  $r_{C=C}$  was only 0.012 Å. Absolute

**Table 5.** Basis Set Effect on the Standard Activation and Reaction Enthalpies (in kJ/mol) for the  $\text{CH}_2=\text{CHCH}_3 + \bullet\text{OH} \rightarrow \text{CH}_2\bullet\text{CHCH}_2 + \text{H}_2\text{O}$  Reaction Computed at the BH&HLYP/6-31G(d) Geometry

basis set	no. of basis functions	$\Delta_r H^\circ$	$\Delta^\ddagger H^\circ$
6-31G(d)	74	-87.9	18.8
cc-pVDZ	91	-101.4	8.8
6-311G(d,p)	114	-105.8	11.6
6-311+G(d)	130	-114.5	10.7
aug-cc-pVDZ	155	-118.8	6.7
6-311+G(3df,2p)	219	-118.4	10.2
cc-pVTZ	218	-113.1	10.6
aug-cc-pVTZ	345	-117.6	10.7
cc-pVQZ	430	-116.5	10.7

deviations were calculated using this process at each grid point of the BH&HLYP surface.

The above example illustrates that, in general, the agreement is good; the largest deviation (LD) was 0.015 Å in the bond lengths, 7.9° in the bond angles, and 6.3° in the dihedral angles. Note that these LD values belong to those extreme grid points of the PES which are at places far from the minimum energy pathway (MEP), and these points are actually very few in number. Consequently, BH&HLYP in combination with the 6-31G(d) basis set (Table 1) yields fairly accurate geometrical parameters with respect to those obtained by the CCSD/6-31G(d) computations.

#### 4.2. Energetic Exploration of Reactants, Products, Complexes, and Transition States. 4.2.1. Basis Set Studies.

The effect of basis sets was also studied (see Tables 5 and 6). At the BH&HLYP/6-31G(d) geometry, single point calculations were carried out using various basis set sizes. As Table 5 shows, both reaction enthalpy ( $\Delta_r H^\circ$ ) and activation enthalpy ( $\Delta^\ddagger H^\circ$ ) tend to converge toward the experimental values, thus the results are qualitatively or semiquantitatively indicative. However, convergence is still limited for both  $\Delta^\ddagger H^\circ$  (10.7 kJ/mol) and  $\Delta_r H^\circ$  (-116.5 kJ/mol). The experimental result of reaction enthalpy is -128.53 kJ/mol<sup>65-67</sup> which is 12.03 kJ/mol lower than the limit (-116.5 kJ/mol) of the BH&HLYP energy result. Unfortunately, there is relevant deviation between the BH&HLYP  $\Delta^\ddagger H^\circ$  result (10.7 kJ/mol) and the experimentally estimated value (from -3.7 to +6.2 kJ/mol range depending on the extrapolation);<sup>72</sup> the error is relatively large.

**Table 6.** Basis Set Effect on the Stability of the  $\pi$ -Complex and the Product Complex as Well as Reaction and Activation Enthalpies Calculated (in kJ/mol) on CCSD/6-31G(d) Geometries

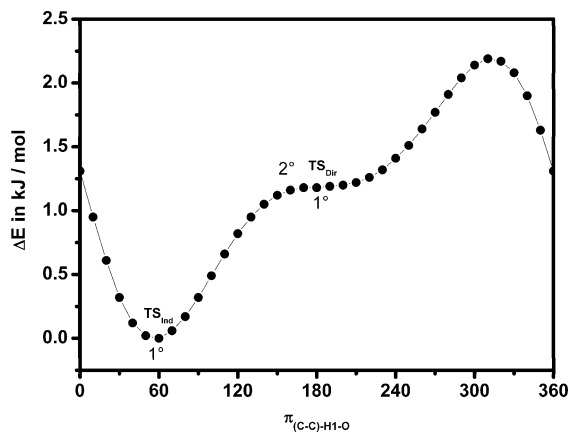
model	no. of basis functions	$\Delta H^\circ_{\pi\text{-complex}}$	$\Delta^\ddagger H^\circ$ direct	$\Delta^\ddagger H^\circ$ indirect	$\Delta H^\circ_{\text{product-complex}}$	$\Delta_r H^\circ$
CCSD(T)/6-31G(d)	74	-11.0	33.6	32.0	-96.3	-89.2
CCSD(T)/6-311G(d,p) <sup>a</sup>	114	-1.9	32.4	20.7	-106.1	-108.1
QCISD(T)/6-311G(d,p) <sup>a</sup>	114	-1.9	31.2	19.4	-107.0	-109.0
CCSD(T)/6-311+G(3df,2p) <sup>a</sup>	219	-7.1	7.3	1.9	-126.2	-123.8
QCISD(T)/6-311+G(3df,2p) <sup>a</sup>	219	-7.1	5.7	0.3	-127.2	-124.7
G3MP2/BHandHLYP <sup>b</sup>		-8.2	0.0	-1.6	-129.0	-126.6
exp. (from refs 63-66)				-3.7		-128.53

<sup>a</sup> With the inclusion of BSSE correction (see eq 3) using the counterpoise method of Bernadi and Boys.<sup>59</sup> <sup>b</sup> G3MP2 enthalpy are calculated on the BH&HLYP/6-31G(d) geometry.

These results lead to higher level single point calculations such as QCISD(T) and CCSD(T) at medium 6-311G(d,p) and large 6-311+G(3df,2p) split valence basis sets on the CCSD/6-31G(d) geometries and BH&HLYP. Note that there is not only remarkable similarity between CCSD/6-31G(d) and BH&HLYP/6-31G(d) geometries, but energetic deviations at high level single point calculations are also negligible (about 1 kJ/mol). At these single point calculations of allyl radical and the transition state, the spin-momentum operator expectation values ( $S^2$ ) are between 0.97 and 0.93. The QCISD(T) method can generally better handle the failure related to the spin contamination. Klippenstein et al.<sup>19</sup> reported a similar value for the transition state of OH addition to ethene due to significant multireference character and/or dynamic correlation to the wave function. Their MRCI calculation compared to their QCISD(T) results shows agreement with each other. In the case of the hydroxyl radical, the expectation value of the  $S^2$  operator is relatively close to 0.75, so the spin contamination may be considered acceptable in all cases. Table 5 presents the calculated reaction ( $\Delta_r H^\circ$ ) and activation ( $\Delta^\ddagger H^\circ$ ) enthalpies at different levels of theory.

BSSE was considered by the counterpoise method (see eq 3) both at the QCISD(T) and CCSD(T) levels of theory using 6-311G(d,p) and 6-311G+(3df,2p) basis sets: they are 23.2 and 10.3 kJ/mol for the activation enthalpy, for the two basis sets, respectively. In the case of the  $\pi$ -complex, BSSE was found to be moderate (6.4 and 3.9 kJ/mol) with those two basis sets. BSSE for the product complex were only 4.1 and 2.9 kJ/mol, respectively.

Both activation and reaction enthalpy change dramatically with the enhancement of the basis set (at the CCSD(T) and QCISD(T) level of theory). Using the 6-311G(d,p) basis set the stability of the  $\pi$ -complex decreases significantly. The activation enthalpy did not change for the direct abstraction, while extension of the basis set has a relevant effect on the indirect transition state: the barrier is decreased by 12 kJ/mol. Further increasing of the number of the basis set made the  $\pi$ -complex almost as stable as it was at 6-31G(d). The reaction enthalpy changes from the average of -108.5 kJ/mol to -124.7 kJ/mol. This effect probably is due to the formation of the allyl radical. This is a conjugated open shell system with a relevant multireference character. For comparison, the reaction G3MP2<sup>68</sup> enthalpy (-126.6 kJ/mol) and activation enthalpy for indirect abstraction (-1.6 kJ/mol)



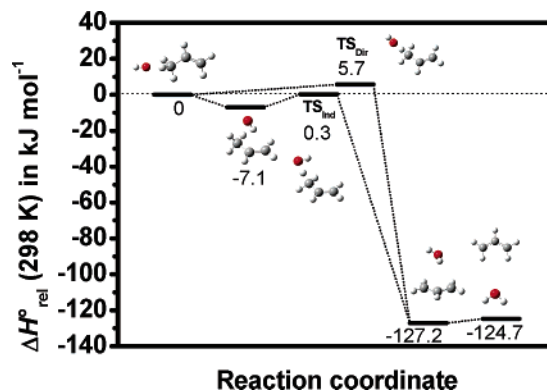
**Figure 10.** BH&HLYP/6-31G(d) potential energy curve along the  $\pi_{C-C-H1-O}$  dihedral angle. The approximate position of the transition states and second-order saddle points are marked by  $1^\circ$  and  $2^\circ$ .

were calculated. They were found to be in good agreement with both the experimental and our high level data.

**4.2.2. Search for Direct H-Abstraction TS and for a Second  $\pi$ -Complex.** Since the first transition state optimized is associated with the path through the first  $\pi$ -complex, therefore it has to be classified as the TS for indirect H-abstraction ( $TS_{Ind}$ ). Subsequently, the question of the existence of a second transition state corresponding to the direct H-abstraction ( $TS_{Dir}$ ) must be examined. As our BH&HLYP/6-31G(d) geometry optimization shows, the relevant geometrical parameters for  $TS_{Dir}$  are similar to those of the indirect transition state ( $TS_{Ind}$ ). Only the  $r_{CH\cdots O}$  is by  $0.004 \text{ \AA}$  larger at  $TS_{Ind}$ , but it is only an insignificant difference. The H of the OH radical is orientated far away from the  $\pi$ -bond as the torsion angles ( $\pi_{(C-C)-H1-O}$ ) were shown to be:  $160.4^\circ$  for the direct TS. That is  $52.0^\circ$  for the indirect TS. These two TS geometries are shown in the lower part of Figure 8; they appear to be rotamers. A semirigid rotational BH&HLYP/6-31G(d) potential along the  $\pi_{(C-C)-H1-O}$  dihedral angle is shown in Figure 10. Clearly, at appropriately low temperature the  $TS_{Ind}$  will dominate the rate of the reaction. At sufficiently high temperature both transition states are involved. Note that the barrier that separates the  $TS_{Dir}$  and  $TS_{Ind}$  is in fact a second-order saddle point. However, the direct TS did not exhibit a linear structure, yet it did not have a prereaction complex in this channel as discussed before. In the possession of activation enthalpies and activation entropies one may estimate the relative rate constant of the indirect (IND) and direct (DIR) hydrogen abstraction:

$$\frac{k_{Ind}}{k_{Dir}} = \exp \frac{\Delta^\ddagger S_{Ind} - \Delta^\ddagger S_{Dir}}{R} \cdot \exp \frac{-(\Delta^\ddagger H_{Ind} - \Delta^\ddagger H_{Dir})}{RT} = \exp \frac{\Delta\Delta^\ddagger S}{R} \cdot \exp \frac{-(\Delta\Delta^\ddagger H)}{RT} \quad (6)$$

The calculation of the activation entropy can be made in two ways: (a) via the harmonic oscillator approximation for all of the normal modes and (b) via the hindered rotor treatment of the low frequencies and harmonic oscillator approximation for the remaining modes.



**Figure 11.** Enthalpy profile of the  $CH_2=CHCH_3 + \bullet OH \leftrightarrow CH_2=CHCH_2\bullet + H_2O$  reaction. Enthalpies were obtained at the QCISD(T)/6-311+G(3df,2p)//CCSD/6-31G(d) level of theory. The sum of the enthalpies of formation of propene and the OH radical was chosen as the zero point for the enthalpy profile.

The difference in the activation entropy ( $\Delta\Delta^\ddagger S$ ) is  $6.1 \text{ J/(mol K)}$  with harmonic oscillators approximation, while (b) method produces  $10.9 \text{ J/(mol K)}$ . They can increase the rate constant of the indirect H-abstraction related to the direct H-abstraction by a factor of 2.1 and 3.9, respectively. The activation enthalpy of the direct H-abstraction channel proved to be larger by  $5.4 \text{ kJ/mol}$  compared to those of the indirect TS calculated by the QCISD(T)/6-311+G(3df,2p)//CCSD/6-31G(d) level of theory. This difference makes the indirect abstraction 8.8 times faster than the direct one at room temperature. To summarize, the indirect H-abstraction reaction can be 18.3 (using harmonic oscillator approximation) and 32.2 (with hindered rotor treatment) times faster than the direct reaction at room temperature.

A systematic effort has been made to locate a second  $\pi$ -complex which could be associated with the OH addition. Since such a second  $\pi$ -complex has not been found, it seemed reasonable to assume that the one and only  $\pi$ -complex is a common intermediate for both the indirect H-abstraction and the OH addition.

**4.2.3. Final Reaction Profile.** Finally, we suggest, based on our extended study, that the activation enthalpy ( $\Delta^\ddagger H^\circ$ ) of the title reaction should be  $0.3 \text{ kJ/mol}$ . The calculated reaction enthalpy ( $\Delta_r H^\circ$ ) is  $-124.7 \text{ kJ/mol}$ , which is in good agreement with the experimental value ( $-128.53 \text{ kJ/mol}$ ). Figure 11 shows the enthalpy profile of the reaction studied.

Alvarez-Idaboy et al. suggested<sup>20,21</sup> an activation energy value of  $-8.78 \text{ kJ/mol}$ , which is the result of their accurate PMP4/6-311+G(d,p)//MP2/6-311G(d,p) calculations,<sup>69</sup> as relative energy of addition channels compared to the entrance channel. This value is lower than the QCISD(T)/6-311G+(3df,2p) activation energy of the H-abstraction channel  $3.1 \text{ kJ/mol}$  which corresponds to the  $0.3 \text{ kJ/mol}$  activation enthalpy. Thus, addition reactions are probably faster than the H-abstraction in this case. However, both adducts have a fast unimolecular back reaction to the  $\pi$ -complex, and no subsequent reaction of the adducts exist<sup>20</sup> under the energy level of the propene+OH system. This is in contrast with the products of the H-abstraction.

## 5. Conclusions

The H-abstraction from propene with a hydroxyl radical was studied with different first principle computations, including B3LYP and BH&HLYP density functional methods (DFT), and HF, MP2, QCISD, QCISD(T), CCSD, and CCSD(T) ab initio methods combined with the 6-31G(d) split valence basis set, to explore the potential energy surface (PES) of the hydrogen abstraction reaction with special attention to the transition state and other stationary points. Relatively small deviations from the CCSD(T)/6-31G(d)//CCSD/6-31G(d) reference surface were found in the cases of the BH&HLYP, QCISD, QCISD(T), and CCSD level of theory. Linear relationships between the relative energies of these four and the reference methods have been found. The BH&HLYP method closely reproduces the geometrical parameters of the grid points of CCSD-PES, especially at the maxima and minima. It seems that the BH&HLYP method is very useful as it is a rapid way to compute reliable geometries for allylic H-abstraction. The BH&HLYP geometries of stationary points are not only similar to CCSD geometries but are also very close to available experimental data.

The energetic test shows that the BH&HLYP method tends to converge for both reaction and activation enthalpy as the basis set increases in size:  $\Delta_r H^\circ = -117.0$  kJ/mol and  $\Delta^\ddagger H^\circ = 10.7$  kJ/mol, respectively. There are acceptable deviations between the reaction and activation enthalpies obtained at the level of theory as was found previously.<sup>70,71</sup>

We suggest a new value, based on our detailed ab initio analysis (QCISD(T)/6-311+G(3df,2p)//CCSD/6-31G(d), for the activation enthalpy of H-abstraction  $\Delta^\ddagger H^\circ = 0.3 \pm 2.0$  kJ/mol ( $-3.7$  and  $+6.2$  kJ/mol experimental range<sup>72</sup>). The calculated and the experimental reaction enthalpies are in good agreement:  $\Delta_r H^\circ = -124.7 \pm 4.0$  kJ/mol and  $-128.53$  kJ/mol, respectively.

Both the theoretical value of  $\Delta^\ddagger H^\circ$  and the IRC calculation suggested that hydrogen abstraction is quite fast and predicted that the reaction cannot occur exclusively via direct collision but may eventuate also via a  $\pi$ -reactant complex intermediary. Although, the structures of the two H-abstraction transition states are quite similar, nevertheless their mechanisms show relevant differences. In addition, there is a competition between indirect H-abstraction as well as central and terminal additions. The structures of reactant ( $\pi$ ) and product complexes were optimized, and their stabilities were computed.

**Acknowledgment.** The authors thank M. Labadi for technical support. One of the authors (I.G.C.) thanks the Ministry of Education for a Szent-Györgyi Visiting Professorship. The authors are grateful to the Hungarian Scientific Research Fund (OTKA T046861 and F037648). We wish to thank the University of Szeged (MU-00094/2002) for computational time.

**Supporting Information Available:** Important geometry parameters of H-abstraction obtained using different DFT and high level ab initio methods (Table S1). This

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

## References

- (1) Taatjes, C. A.; Hansen, N.; McIlroy, A.; Miller, J. A.; Senosiain, J. P.; Klippenstein, S. J.; Qi, F.; Sheng, L.; Zhang, Y.; Cool, T. A.; Wang, J.; Westmoreland, P. R.; Law, M. E.; Kasper, T.; Kohse-Höinghaus, K. *Science* **2005**, *308*, 1887–1889.
- (2) Braña, P.; Sordo, J. A. *J. Comput. Chem.* **2003**, *24*, 2044–2062.
- (3) Atkinson, R. *Chem. Rev.* **1986**, *86*, 69–201.
- (4) Atkinson, R. *J. Phys. Chem. Ref. Data* Monograph No.1 1989.
- (5) Baulch, D. L.; Cobos, C. J.; Cox, R. A.; Frank, P.; Hayman, G.; Just, T.; Kerr, J. A.; Murrells, T.; Pilling, M. J.; Troe, J.; Walker, R. W.; Warnatz, J. *J. Phys. Chem. Ref. Data* **1994**, *23*, 847–1033.
- (6) Turpin, E.; Fittschen, C.; Tomas, A.; Devolder, P. *J. Atmos. Chem.* **2003**, *46*, 1–13.
- (7) Minetti, R.; Ribaucour, M.; Carlier, M.; Fittschen, C.; Sochet, L. R. *Combust. Flame* **1994**, *96*, 201–211.
- (8) Nagaoka, S.; Okauchi, Y.; Urano, S.; Nagashima, U.; Mukait, K. *J. Am. Chem. Soc.* **1990**, *112*, 8921–8924.
- (9) Chuong, B.; Stevens, P. S. *J. Phys. Chem. A* **2000**, *104*, 5230–5237.
- (10) Hippler, H.; Viskolcz, B. *Phys. Chem. Chem. Phys.* **2000**, *2*, 3591–3596.
- (11) Szilagy, I.; Dobe, S.; Berces, T.; Marta, F.; Viskolcz, B. *Z. Phys. Chem.* **2004**, *218*, 479–492.
- (12) Vakhin, A. B.; Lee, S.; Heard, D. E.; Smith, I. W. M.; Leone, S. R. *J. Phys. Chem. A* **2001**, *105*, 7889–7895; **2003**, *107*, 10055–10062.
- (13) Walch, S. P. *J. Chem. Phys.* **1993**, *98*, 3163–3178.
- (14) Zellner, R.; Lorenz, K. *J. Phys. Chem.* **1984**, *88*, 984–989.
- (15) Spangenberg, T.; Köhler, S.; Hansmann, B.; Wachsmuth, U.; Abel, B. *J. Phys. Chem. A* **2004**, *108*, 7527–7534.
- (16) Vereecken, L.; Peeters, J. *Chem. Phys. Lett.* **2001**, *333*, 162–168.
- (17) Bradley, J. N.; Hack, W.; Hoyermann, K.; Wagner, H. G. *J. Chem. Soc., Faraday Trans. 1* **1973**, *69*, 1889–1898.
- (18) Diau, E. W. G.; Lee, Y. P. *J. Chem. Phys.* **1992**, *96*, 377–386.
- (19) Greenwald, E. E.; North, S. W.; Georgievskii, Y.; Klippenstein, S. J. *J. Phys. Chem. A* **2005**, *109*, 6031–6044.
- (20) Diaz-Acosta, I.; Alvarez-Idaboy, J. R.; Vivier-Bunge, A. *Int. J. Chem. Kinet.* **1999**, *31*, 29–36.
- (21) Alvarez-Idaboy, J. R.; Diaz-Acosta, I.; Vivier-Bunge, A. *J. Comput. Chem.* **1998**, *19*, 811–819.
- (22) Atkinson, R.; Arey, J. *Chem. Rev.* **2003**, *103*, 4605–4638.
- (23) Chen, W.; Schlegel, H. B. *J. Chem. Phys.* **1994**, *101*, 5957–5968.
- (24) Henry, D. J.; Parkinson, C. J.; Radom, L. *J. Phys. Chem. A* **2002**, *106*, 7927–7936.
- (25) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974–12980.

- (26) Ziegler, T. *Chem. Rev.* **1991**, *91*, 651–667.
- (27) Handy, N. C.; Tozer, D. J.; Laming, G. J.; Murray, C. W.; Amos, R. D. *Isr. J. Chem.* **1993**, *33*, 331–344.
- (28) Johnson, B. G.; Gill, P. M.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 5612–5626.
- (29) Baker, J.; Scheiner, A.; Andzelm, J. *Chem. Phys. Lett.* **1993**, *216*, 380–388.
- (30) Laming, G. J.; Handy, N. C.; Amos, R. D. *Mol. Phys.* **1993**, *80*, 1121–1134.
- (31) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063–1079.
- (32) Fan, L. Y.; Ziegler, T. *J. Am. Chem. Soc.* **1992**, *114*, 10890–10897.
- (33) Durant, J. L. *Chem. Phys. Lett.* **1996**, *256*, 595–602.
- (34) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 2936–2941.
- (35) Bacskay, G. B. *Chem. Phys.* **1981**, *61*, 385–404.
- (36) Hariharanm, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (37) Roothan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.
- (38) Pople, J. A.; Nesbet, R. K. *J. Chem. Phys.* **1954**, *22*, 571–572.
- (39) McWeeny, R.; Dierksen, G. *J. Chem. Phys.* **1968**, *49*, 4852–4856.
- (40) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
- (41) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503–506.
- (42) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 275–280.
- (43) Saebo, S.; Almlof, J. *Chem. Phys. Lett.* **1989**, *154*, 83–89.
- (44) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968–5975.
- (45) Cizek, J. *Adv. Chem. Phys.* **1969**, *14*, 35.
- (46) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910–1918.
- (47) Scuseria, G. E.; Janssen, C. L.; Schaefer, H. F., III *J. Chem. Phys.* **1988**, *89*, 7382–7387.
- (48) Scuseria, G. E.; Schaefer, H. F., III *J. Chem. Phys.* **1989**, *90*, 3700–3703.
- (49) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.
- (50) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (51) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (52) Note that BH&HLYP means  $0.5 \times E_X^{\text{HF}} + 0.5 \times E_X^{\text{LSDA}} + 0.5 \times \Delta E_X^{\text{Becke88}} + E_C^{\text{LYP}}$  functional.
- (53) Scaling factors for ZPVE: HF(0.9135), MP2(0.9670), QCISD(0.9776), and B3LYP(0.9806) from the following: Scott, A. P.; Random, L. *J. Phys. Chem.* **1996**, *100*, 16502–16513.
- (54) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (55) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (56) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (57) Peterson, K. A.; Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 7410–7415.
- (58) Wilson, A.; van Mourik, T.; Dunning, T. H., Jr. *J. Mol. Struct. (THEOCHEM)* **1996**, *388*, 339–349.
- (59) Boys, S. F.; Bernardi, F. *Mol. Phys.* **2002**, *100*, 65–73.
- (60) Malick, D. K.; Petersson, G. A.; Montgomery, J. A., Jr. *J. Chem. Phys.* **1998**, *108*, 5704.
- (61) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, revision C.02*; Gaussian, Inc.: Wallingford, CT, 2004.
- (62) *CRC Handbook of Chemistry and Physics*, 77th ed.; Lide, D. R., Frederikse, H. P. R., Eds.; CRC Press: Boca Raton, FL, 1996–1997.
- (63) *JANAF Thermochemical Tables*, 3rd ed.; Chase, M. W., Jr., Davies, C. A., Downey, J. R., Frurip, D. J., McDonald, R. A., Syverud, A. N., Eds.; National Bureau of Standards: Washington, DC, 1985; Vol. 14.
- (64) Hirota, E.; Yamada, C.; Okunishi, M. *J. Chem. Phys.* **1992**, *7*, 2963–2970.
- (65) Turányi, T.; Zalotai, L.; Dóbbé, S.; Bérces, T. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2568–2578.
- (66) Ruscic, B.; Boggs, J. E.; Burcat, A.; Császár, A. G.; Demaison, J.; Janoschek, R.; Martin, J. M. L.; Morton, M. L.; Rossi, M. J.; Stanton, J. F.; Szalay, P. G.; Westmoreland, P. R.; Zabel, F.; Bérces, T. *J. Phys. Chem. Ref. Data* **2005**, *34*, 573–656.
- (67) Frenkel, M.; Marsh, K. N.; Wilhoit, R. C.; Kabo, G. J.; Roganov, G. N. *Thermodynamics of Organic Compounds in the Gas State*; Thermodynamics Research Center: College Station, TX, 1994.
- (68) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650.
- (69) Krishnan, R.; Pople, J. A. *Int. J. Quantum Chem.* **1978**, *14*, 91–100.
- (70) Truong, T. N.; Duncan, W. *J. Chem. Phys.* **1994**, *101*, 7408–7414.
- (71) Durant, J. L. *Chem. Phys. Lett.* **1996**, *256*, 595–602.
- (72) Tsang, W. *J. Chem. Ref. Data* **1991** 221–273. They proposed for rate constant for the reaction studied as follows:  $k(T) = 2.8 \times 10^{11} (T/298)^2 \cdot \exp(1.25 \text{ kJ}/RT) \text{ cm}^3/(\text{molecule} \cdot \text{s})$ , where  $T = 300\text{--}2500 \text{ K}$  and uncertainty = 1.20.

# JCTC

Journal of Chemical Theory and Computation

## Atomic Level Anisotropy in the Electrostatic Modeling of Lone Pairs for a Polarizable Force Field Based on the Classical Drude Oscillator

Edward Harder,<sup>†</sup> Victor M. Anisimov,<sup>‡</sup> Igor V. Vorobyov,<sup>‡</sup> Pedro E. M. Lopes,<sup>‡</sup>  
Sergei Y. Noskov,<sup>†</sup> Alexander D. MacKerell Jr.,<sup>\*,‡</sup> and Benoît Roux<sup>\*,†</sup>

*Institute for Molecular Pediatric Sciences, Gordon Center for Integrative Science,  
University of Chicago, Chicago, Illinois 60637,  
and Department of Pharmaceutical Sciences,  
School of Pharmacy, University of Maryland, Baltimore, Maryland 21201*

Received May 23, 2006

**Abstract:** Electron pairs in the valence shell of an atom that do not participate in the bonding of a molecule (“lone pairs”) give rise to a concentrated electron density away from the atom center. To account for the asymmetry in the electron charge density that arises from lone pairs, an electrostatic model is developed that is parametrically anisotropic at the atomic level. The model uses virtual interaction sites with partial charges that are associated but not coincident with the nuclei. In addition, the model incorporates anisotropic atomic polarizabilities. The protocol previously outlined in Anisimov et al. [*J. Chem. Theory Comput.* **2005**, *1*, 153] for parametrizing the electrostatic potential energy of a polarizable force field using classical Drude oscillators is extended to incorporate additional lone pair parameters. To probe the electrostatic environment around the lone pairs, the static (molecule alone) and perturbed (molecule in the presence of a test charge) electrostatic potential (ESP) are evaluated and compared to high level quantum mechanical (QM) electronic structure calculations. The parametrization of the virtual sites relies on data from the QM static ESP. The contribution to the perturbed ESP from the electronic polarization of the molecule is used to resolve the components of the atomic polarizability tensor. The model is tested in the case of four molecules: methanol, acetone, methylamine, and pyridine. Interaction energies with water and sodium are used to assess the accuracy of the model. The results are compared with simpler models placing all the charge on the nuclei as well as using only isotropic atomic polarizabilities. Analysis shows that the addition of virtual sites reduces the average error relative to the QM calculations. In contrast to models with atom centered charges, the virtual site models correctly predict the minimum energy conformation for acetone and methanol, with water, to be closely coordinated with the lone pair direction. Furthermore, addition of anisotropic atomic polarizabilities to the virtual site model allows for precise fitting to the local perturbed QM ESP.

### 1. Introduction

Computer simulations of atomistic models used to investigate biological phenomena often employ simple potential functions that balance computational efficiency with a suitable

level of accuracy for the microscopic interactions. Commonly used models such as CHARMM,<sup>1</sup> OPLS/AA,<sup>2</sup> AMBER,<sup>3</sup> and GROMOS<sup>4</sup> approximate the electrostatic potential that surrounds a molecule by using point charges of fixed magnitude placed on (or coordinated to) atomic positions. The underlying assumption in such models is that the electron density of the molecule remains unaffected by external perturbations. However, it is well-understood that the electron density of real molecules is not static but responds to fluctuating electric

\* Corresponding author e-mail: roux@uchicago.edu alex@outerbanks.umaryland.edu.

<sup>†</sup> University of Chicago.

<sup>‡</sup> University of Maryland.

fields arising from neighboring entities. Even in neat liquid water such fluctuations are significant and comparable in magnitude to the size of the induced molecular dipole ( $\approx 1$  D).<sup>5</sup> The importance of these effects in the case of biological macromolecular systems, with variations in microscopic environments ranging from hydrophobic to highly polar, is expected to be considerable. The need for explicit inclusion of induced polarization effects has been explicitly documented in studies concerned with the structural rearrangements in liquid water,<sup>6–8</sup> the hydration of small ions,<sup>9–13</sup> the structure of water–alcohol liquid mixtures,<sup>14</sup> the dielectric constant of pure nonpolar solvents,<sup>15</sup> the solvation of water around a protein surface with varying polarity,<sup>16</sup> and the conduction of ions through a membrane channel.<sup>17,18</sup>

To make progress toward the parametrization of a polarizable biomolecular force field we have chosen to represent electronic induction using classical Drude oscillators.<sup>15,19–21,48</sup> Similar to commonly used inducible point dipole models,<sup>22–27</sup> Drude oscillators represent electronic induction using a pair of charges of equal magnitude and opposite sign connected by a harmonic spring.<sup>28</sup> In the present model the Drude spring connects the negative Drude charge to its associated heavy atom. The familiar self-consistent field (SCF) condition of induced polarization is reproduced if the massless particles are allowed to relax instantaneously to their local energy minima for any given fixed configuration of the atoms in the system. The model reduces to the standard model with induced point dipoles in the limit of a large spring constant. For simplicity, the initial implementation of the polarizable force field with classical Drude oscillators assumed that the fixed charges were centered on the nuclei and that the local atomic polarizabilities were isotropic.<sup>21</sup> Models parameterized in accordance with this framework show good agreement with experimental properties, including the diffusivity and dielectric constant of ethanol/water mixtures<sup>14</sup> and crystal structures of DNA.<sup>21</sup> Nevertheless, it may be necessary to go beyond this basic framework to achieve the needed level of accuracy for detailed microscopic interactions. In particular, the presence of nonbonding electronic density situated away from the nucleus, so-called “lone pairs”, breaks the spherical symmetry assumed by the atom-centered electrostatic models. It is intuitively obvious that the local electrostatic field arising from lone pairs could be poorly described by an atom-centered centrosymmetric model. Furthermore, it is likely that induced electronic polarization in the neighborhood of the lone pairs is anisotropic. In a rational effort to develop a computationally useful and accurate force field, a quantitative characterization is essential.

The goal of this article is to address these issues. The strategy employed in the current effort focuses primarily on the accuracy of microscopic interactions rather than the properties of the bulk phase. The target interaction data, used to fit the electrostatic model, are supplied by high level electronic structure calculations of the surrounding electrostatic potential (ESP). The performance of three electrostatic models at different levels of complexity is examined: an atom-centered charge model with isotropic Drude oscillators, a model with virtual sites representing the lone pairs and

isotropic Drude oscillators, and a model with virtual sites representing the lone pairs together with anisotropic Drude oscillators. The ability of the different electrostatic models to accurately represent microscopic interactions is examined in the case of four molecules: methanol, acetone, methylamine, and pyridine. The details of the model are presented in section 2.1. The parameter fitting protocol, generalized to include anisotropic atomic polarizabilities and virtual sites, is given in section 2.2, and all remaining computational details are described in section 3. Section 4.1 presents a general discussion of the electrostatic properties associated with lone pairs. An examination of the validity of a linear polarization response is presented in section 4.3. Dimer energies with water and sodium and a comparison between QM and the MM models is presented in section 4.2. The paper is concluded with a summary of the main results and an outlook for future force field development.

## 2. Methods

**2.1. Model.** To model the electronic polarization of a given atom, a mobile auxiliary particle carrying a charge  $q^D$  is introduced and attached to the atom by a harmonic spring. A charge of opposite sign is added to the atom and the electroneutral pair forms a classical Drude oscillator, which polarizes in response to an external field. In the present implementation, the charge of the mobile Drude particle is chosen to be negative,<sup>15,21,48</sup> because this yields a more realistic electrostatic response in molecular systems (see also the discussion in section 4.3 below). The model comprises the “core” charges  $q_i$ , associated either with atomic or virtual lone pair sites, as well as the Drude charges  $q_i^D$  and  $-q_i^D$ , associated with the electroneutral oscillators. For a fixed atomic geometry, the electrostatic response of a molecule is determined by the potential energy

$$U_{\text{Drude}} = U_{\text{self}} + U_{\text{elec}} \quad (1)$$

The first term in eq 1 is the harmonic self-energy of the Drude oscillators. For a given oscillator, it can be written as

$$\begin{aligned} U_{\text{self}} &= \frac{1}{2} \mathbf{d} \cdot \mathbf{K}^{(D)} \cdot \mathbf{d} \\ &= \frac{1}{2} ([\mathbf{K}_{11}^{(D)}] d_1^2 + [\mathbf{K}_{22}^{(D)}] d_2^2 + [\mathbf{K}_{33}^{(D)}] d_3^2) \end{aligned} \quad (2)$$

where  $d_1$ ,  $d_2$ , and  $d_3$  are the projection of the Drude displacement vector  $\mathbf{d}$  on orthogonal axis defined using a local intramolecular reference frame. For example  $d_1 = \mathbf{d} \cdot \hat{n}_{A,B}$  where  $\hat{n}_{A,B}$  is a unit vector directed between atoms A and B in the molecule of interest. One may note that, due to the intramolecular electrostatic interactions between induced oscillators, even a model with isotropic spring constants can give rise to a molecular polarizability that is anisotropic. However, further analysis shows that this level of anisotropy is inadequate to properly model the local anisotropy around lone pairs. This generalization to anisotropic atomic polarizabilities has been implemented in CHARMM.<sup>31</sup>

The second term,  $U_{\text{elec}}$ , corresponds to the sum over all Coulombic interactions between the core charges  $q_i$  located at  $\mathbf{r}_i$  and the Drude charges  $q_i^D$  and  $-q_i^D$ , located at  $\mathbf{r}_i$ , and



$\mathbf{r}_i^D = \mathbf{r}_i + \mathbf{d}_i$ , respectively. The interactions of the various pairs of charges are treated according to the topological bonding order determined from the atoms in the molecule. As in standard force fields, the interactions between core charges corresponding to 1–2 and 1–3 pairs are excluded. Similarly, the interactions of the Drude oscillators with core charges are excluded for 1–2 and 1–3 pairs. The interactions involving all core charges and all Drude oscillators are included without screening for all 1–4 pairs and beyond. The interactions of the Drude oscillators corresponding to 1–2 and 1–3 pairs are screened by the function  $S_{ij}$

$$S_{ij}(r_{ij}) = 1 - \left( 1 + \frac{ar_{ij}}{2(\alpha_i\alpha_j)^{1/6}} \right) e^{-ar_{ij}/(\alpha_i\alpha_j)^{1/6}} \quad (3)$$

It should be noted that the screening is applied to the interaction of the electroneutral pair, including both the mobile charge  $q^D$  and its countercharge  $-q^D$  located on the atom. The empirical dimensionless coefficient  $a$  is chosen to be 2.6, consistent with previous work.<sup>19,29,30</sup>

In accord with the Born–Oppenheimer approximation in quantum mechanics, the electronic degrees of freedom in the model are relaxed to their energy minimum for any given nuclear configuration. The result is an equilibrium between the force of the Drude spring and the electrostatic force from the total external electric field,  $\mathbf{K}_i^D \mathbf{d}_i = -q_i^D \mathbf{E}_i$ , where  $\mathbf{E}_i$  is the total electric field at the position of the Drude particle,  $\mathbf{r}^D$ . This condition can be written in a form analogous to the self-consistent field (SCF) equation for atomic point dipoles,  $\mu_i = \alpha_i \mathbf{E}_i$ . This yields the definition of the atomic polarizability tensor as  $\alpha_i = (q_i^D)^2 [\mathbf{K}^D]^{-1}$ .

Three models are constructed with these fundamental elements: an atom-centered charge model with isotropic Drude oscillators denoted NOLP+ISO, a model with virtual lone pair sites and isotropic Drude oscillators denoted LP+ISO, and a model with virtual lone pair sites and anisotropic Drude oscillators denoted LP+ANISO.

**2.2. Parameter Fitting Protocol.** The parametrization protocol for the core charges and polarizabilities used in the present work is an extension of the method documented by Anisimov et al.<sup>21</sup> The parametrization is achieved by comparison to analogous QM electronic structure computations. The ESP map is computed on a set of grid points surrounding the molecule. To measure the electronic response, a series of perturbed ESP maps is computed by placing a single  $+0.5e$  test charge at chemically relevant positions around the molecule. The same calculation is repeated using our MM model, restricting the atomic polarizabilities to be isotropic. The approach is similar in spirit to work by Friesner and co-workers used in the parametrization of polarizable point dipole and fluctuating charge models.<sup>27,32–35</sup> The set of electrostatic parameters (charges and polarizabilities) is obtained by optimizing the function

$$\chi^2[\alpha, q] = \sum_{p, \text{grid}} [\phi_{p, \text{grid}}^{\text{QM}} - \phi_{p, \text{grid}}^{\text{MM}}]^2 + \chi_r^2 \quad (4)$$

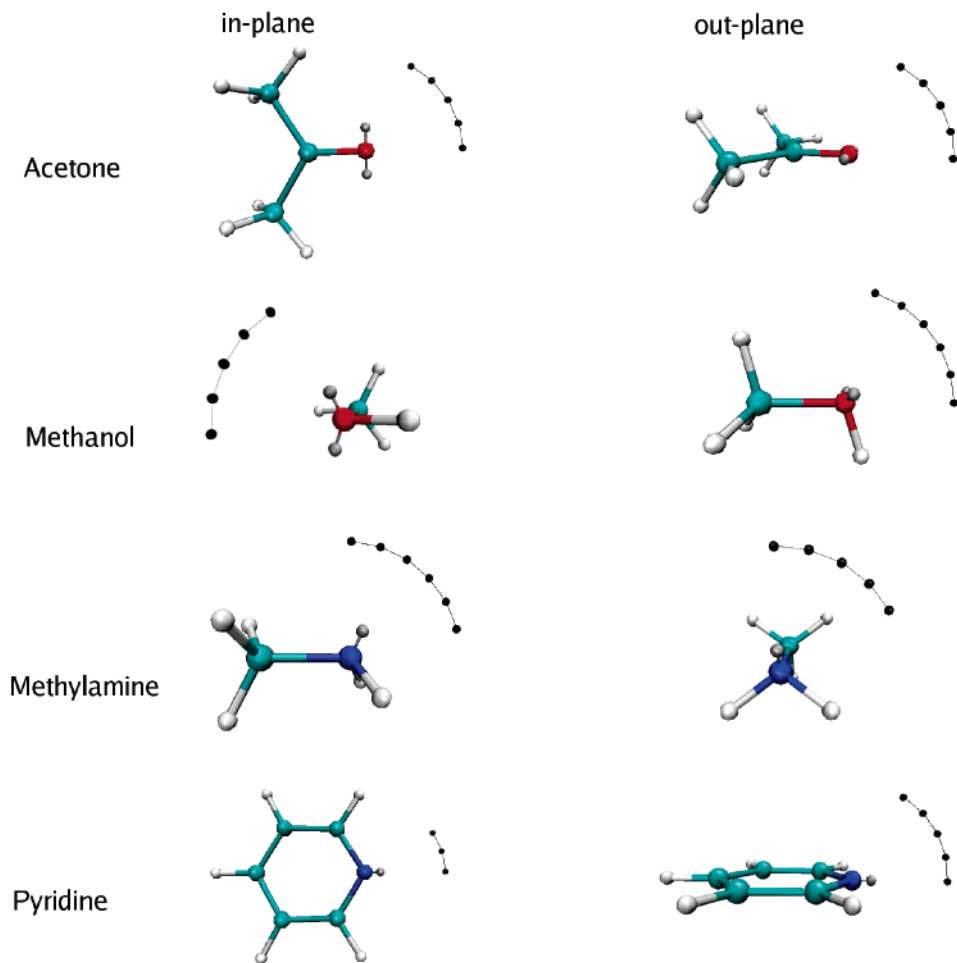
where the index  $p$  sums over different positions of the test charge around the model compound. Eq 4 measures the

deviations of the ESP between the MM and QM results. The optimal parameters are determined with an additional restraint,  $\chi_r^2$ . Such a restraint is necessary because the charge fitting problem is underdetermined,<sup>36,37</sup> and an unrestrained fit (though it may fit the ESP well at the grid points) can often lead to a set of charges with poor chemical significance and limited usefulness for a force field. This problem is particularly evident with buried atoms whose partial charges contribute little to the ESP.<sup>36</sup> The restrained fitting scheme employed here is similar to the RESP scheme of Bayly et al.<sup>37</sup> The  $\chi_r^2$  term in eq 4 provides a penalty for deviations of the fitted parameters relative to chosen reference values. The functional form of the restraint potential is a flat well potential and deviations within a defined interval  $|q_{\text{ref}} - q_{\text{flat}}, q_{\text{ref}} + q_{\text{flat}}|$  receive no penalty, while larger deviations feel a parabolic restraint given by

$$\chi_r^2 = \sum_i w_i [q_i - q_{\text{ref},i}]^2 \quad (5)$$

where  $w$  is a weighting factor for the restraint. The reference values are adopted from two sources. The PARAM22 force field of CHARMM<sup>1,38,39</sup> is used to provide a set of reference values for the atomic charges.<sup>21</sup> The atomic polarizabilities of Miller,<sup>40</sup> derived from experimental gas-phase molecular polarizabilities, are used as a reference for the Drude oscillator parameters. The Miller parameters assign additive atomic polarizabilities to atom types based on the hybridization state of the atom including hydrogen atoms. The reference polarizabilities were constructed by adding the Miller polarizabilities of hydrogen atoms to their covalently bonded heavy atom.<sup>21</sup>

The MM models containing virtual sites and anisotropic polarizabilities have additional parameters (geometry of the virtual sites and the components of the force constant tensor) that must be fitted to QM data. The virtual site geometry is determined iteratively using the aforementioned charge/polarizability fitting procedure. An initial guess is obtained from an atoms in molecules (AIM)<sup>41</sup> analysis of the electron density, by which the positions of lone pairs can be mapped to local maxima in the negative of the Laplacian of the density. The charges are then fit using the above protocol. The reference charge values for the lone pair containing atoms are shifted to the virtual sites, and the charge on the corresponding atom site is restrained to zero during the fitting procedure. Further refinement of the lone pair positions is accomplished by comparing the static (unperturbed) ESP,  $\phi_{\text{stat}}$ , along an arc spanning a plane containing the lone pair(s) at a distance of 2 Å from the reference atom (i.e. oxygen or nitrogen for the molecules studied) site. Using this geometry, all core charges and isotropic atomic polarizabilities are resolved from the RESP electrostatic potential fitting procedure outlined above. The generalization to anisotropic atomic polarizabilities in the model requires an additional step in the parametrization process. A  $+0.5e$  test charge is placed on the arc positions, and the perturbed ESP ( $\phi_{\text{pert}}$ ) is evaluated at the position of the test charge. Defining the contribution arising from the polarization of the molecule in response to external perturbation as  $\phi_{\text{pol}}$ , equal to the difference between the perturbed and unperturbed ESP, the



**Figure 1.** Location of arc point positions used to probe the local ESP. The arc point positions are represented by black spheres.

components of the force constant matrix  $\mathbf{K}$  are determined, through a trial and error procedure, to match the QM result for  $\phi_{\text{pol}}$ .

The atom LJ parameters, with the exception of oxygen or nitrogen, are taken from the PARAM22 force field of CHARMM.<sup>1,38,39</sup> For interaction energies with water, the LJ parameters on oxygen or nitrogen are varied to best reproduce the QM interaction energies. For interaction energies with the sodium cation, the LJ on the oxygen or nitrogen and if necessary the LJ parameters on the virtual sites are varied to best reproduce the QM interaction energies. The model parameters are given in the Supporting Information.

### 3. Computational Details

QM calculations were carried out using the Gaussian 03 suite of programs.<sup>42</sup> The first step in the parametrization is to obtain optimized geometries. The MP2(fc)/6-31G(d) level of theory and basis set found to give molecular geometries consistent with experiment<sup>43</sup> was used in the present work.

The QM ESP maps were evaluated using density functional theory with the B3LYP functional<sup>44,45</sup> and the aug-cc-pVDZ basis set. This combination has been shown to give good agreement with molecular polarizabilities and gas-phase dipole moments.<sup>21</sup> Minimization of eq 4 to arrive at the electrostatic parameters of the force field was carried out using the FITCHARGE module in CHARMM.<sup>31</sup> The QM ESP arcs used to parametrize the virtual site location, and

polarizability components of the anisotropic Drude oscillator were also evaluated using the B3LYP/aug-cc-pVDZ combination.

Optimized interaction energies and geometries for the model compounds with individual water molecules or sodium ions were obtained by minimizing along selected intermolecular distance coordinates while keeping the geometries of the model compounds and water fixed. The interaction energy is evaluated as the difference between the resultant minimum energy conformer and the respective monomers. MP2 calculations with the 6-311+G\* basis set were done to locate the position of the minima for each dimer (both water and sodium). This was followed by MP2 calculations with the 6-311+G(3df,2p) basis set and the removal of basis set superposition error.<sup>46</sup> Polarization and diffuse functions were included to give a faithful representation of the polarization response. The LJ minimum for the MM models was varied until good agreement for the lowest energy conformation was achieved.

Computations designed to probe the breakdown of linear response used electrostatic perturbations with a test charge as large as  $2.0e$  within  $2 \text{ \AA}$  of the carbonyl oxygen of acetone. To achieve an accurate representation of the electronic response in the presence of such a large perturbation, MP2 calculations with the 6-311+G(3df,2p) basis set were again used in the evaluation of  $\phi_{\text{stat}}$ ,  $\phi_{\text{pol}}$ , and the interaction energy.

The size of the force constant in eq 2 is not a variable in

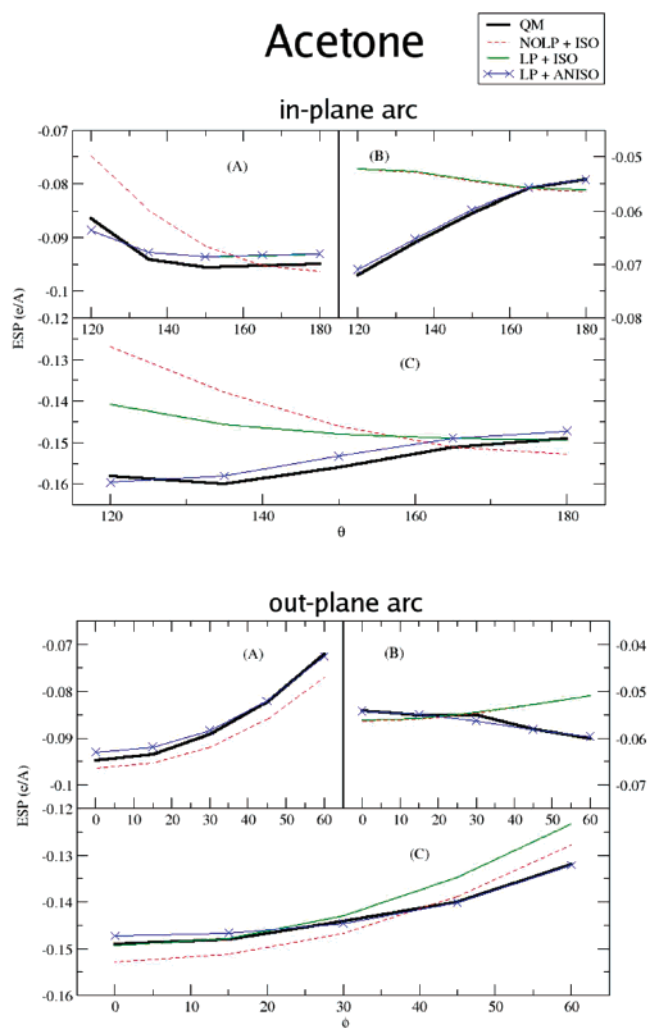
the parameter optimization. The value is chosen to be sufficiently large such that Drude oscillator displacements, for typical field strengths found in biomolecular systems, remain small relative to interatomic distances. This is done to ensure that the model does not deviate strongly from the point dipole limit. Previous work with an isotropic polarizable Drude model of water used a value of  $k_D = 1000$  kcal/mol/Å<sup>2</sup>.<sup>19</sup> Consistent with this choice the components of the anisotropic Drude oscillators were chosen so that the trace of the tensor has a similar value (i.e.  $\text{tr}\{\mathbf{K}_D\} \approx 1000$  kcal/mol/Å<sup>2</sup>).

## 4. Results and Discussion

### 4.1. Parametrization and the Electrostatic Potential Arcs.

The ESP calculated along “in-plane” and “out-plane” arcs, used to probe the electrostatic environment of the lone pairs, are defined by two angles and the distance from the reference heavy atom. For acetone and methanol, the in-plane arc spans a plane containing both virtual sites and the reference heavy atom. For methylamine and pyridine the in-plane arc spans a plane containing the virtual site and the heavy atoms of the molecule. The in-plane arc, which varies with the angle  $\theta$ , is illustrated in Figure 1 for each molecule. The out-plane arc spans a plane orthogonal to the first. The out-plane arc, which varies with the angle  $\Phi$ , is shown in Figure 1 for each molecule. In the figure, the arc point positions are represented by black spheres. For acetone, methylamine, and pyridine  $\theta$  is the C–O–X (or C–N–X) angle, where C is the carbonyl carbon of acetone or the carbon 4 atom of pyridine and X denotes positions along the arc. The angle  $\Phi$  is defined by the  $X_{\text{in-plane}}\text{--O--X}$  (or  $X_{\text{in-plane}}\text{--N--X}$ ) positions where  $X_{\text{in-plane}}$  is the in-plane arc position at  $\theta = 180$  for acetone and pyridine and  $\theta = 105$  for methylamine. For methanol  $\Phi$  is the C–O–X angle and  $\theta$  is the  $X_{\text{out-plane}}\text{--O--X}$  (or  $X_{\text{out-plane}}\text{--N--X}$ ) angle where  $X_{\text{out-plane}}$  is at  $\Phi = 105$ . The total ESP at the position of the  $0.5e$  test charge is presented in panel C of Figures 2–5 for the four molecules. The static and polarization contributions to the ESP are presented in panels A and B, respectively. In each plot the black solid line is the QM result, the red dashed line is the model optimized using the original protocol (NOLP+ISO), the green solid line is the MM model with virtual sites and isotropic Drude oscillators (LP+ISO), and the blue-cross line is the MM model with virtual sites and anisotropic Drude oscillators on the reference heavy atoms (LP+ANISO).

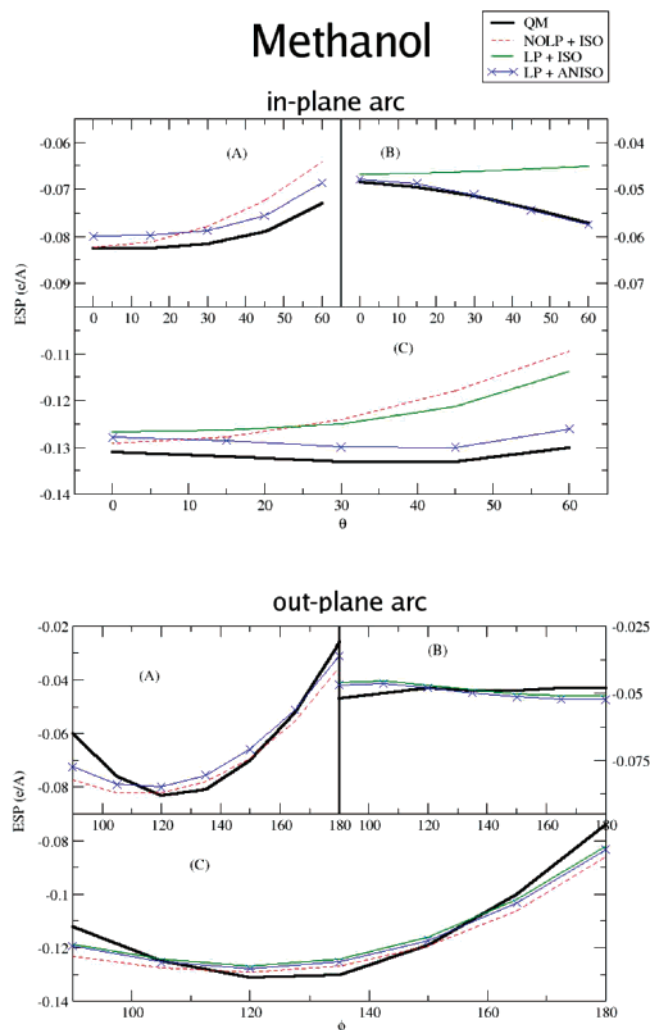
The position of the virtual sites is not part of the function minimized in the charge fitting protocol (see eq 4). As stated in section 2.2, an initial guess for the virtual site geometry follows from an AIM analysis of the electron density. In an iterative process the geometry is modified followed by the minimization of eq 4 until qualitative agreement between the QM  $\phi_{\text{stat}}$  and the MM  $\phi_{\text{stat}}$  (see panel A) is obtained. This is the LP+ISO model. The final virtual site geometries of each model are reported in the Supporting Information. The improved agreement between QM and the MM models using virtual sites with respect to the in-plane arcs is clear for all four molecules. The NOLP MM models do not give a strong enough  $\phi_{\text{stat}}$  at arc positions that are approximately coordinated with the lone pairs ( $\theta = 120$  for acetone,  $\theta =$



**Figure 2.** The ESP along an arc 2 Å from the oxygen of an acetone monomer is plotted in panel (A). The polarization contribution and total ESP from acetone in the presence of a  $0.5e$  test charge is plotted in panels (B) and (C), respectively. The perturbed ESP is calculated at the position of the  $0.5e$  test charge.

60 for methanol,  $\theta = 110$  for methylamine, and  $\theta = 180$  for pyridine). This is corrected by the addition of virtual sites, which yields an unperturbed ESP in better agreement with the QM result. It may be noted that the LP+ANISO and LP+ISO models are nearly equivalent with respect to the unperturbed ESP, making only the plot for LP+ANISO visible in the plots. Similar improvement is also seen for the out-plane arc in methylamine and pyridine where addition of a virtual site increases the strength of  $\phi_{\text{stat}}$  along the lone pair direction leading to improved agreement with the QM static ESP. For acetone and methanol it appears that the orthogonal out-plane arc is already well represented by the NOLP models, and addition of the virtual site provides little improvement.

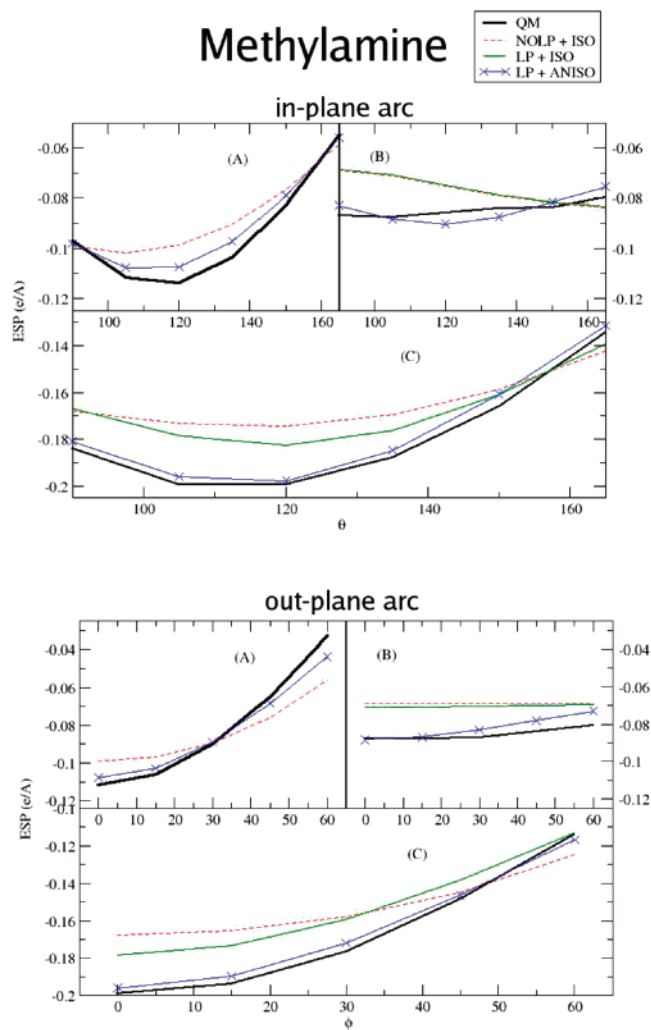
The plots of the polarization ESP (panel B) show the effect lone pairs have on the electronic response of the molecule. The standard MM model, which uses isotropic atomic polarizabilities (ISO), gives a response along the arc that is approximately uniform. Some anisotropy is present in the response of the ISO models due to the screened Coulomb



**Figure 3.** The ESP along an arc 2 Å from the oxygen of a methanol monomer is plotted in panel (A). The polarization contribution and total ESP from methanol in the presence of a 0.5e test charge is plotted in panels (B) and (C), respectively. The perturbed ESP is calculated at the position of the 0.5e test charge.

interactions between induced Drude dipoles. However, over the local coordinates probed by the arcs, this effect is small and qualitatively opposes the trend seen in the anisotropy of the QM response. The local QM response reflects a stronger polarization around the lone pair regions, which seems intuitive given the greater electron density in that region. The magnitude in the discrepancy between the standard MM model and QM is not small for the 0.5e test charge. In fact for the in-plane arc of acetone the difference in the ESP is approximately 0.02 e/Å comparable in magnitude to the improvement found in the static field by the addition of virtual sites. For lone pair containing heavy atoms, the components of the Drude force constant tensor ( $\mathbf{K}_D$ ) are varied to model the local asymmetry in the electronic response (see eq 2).

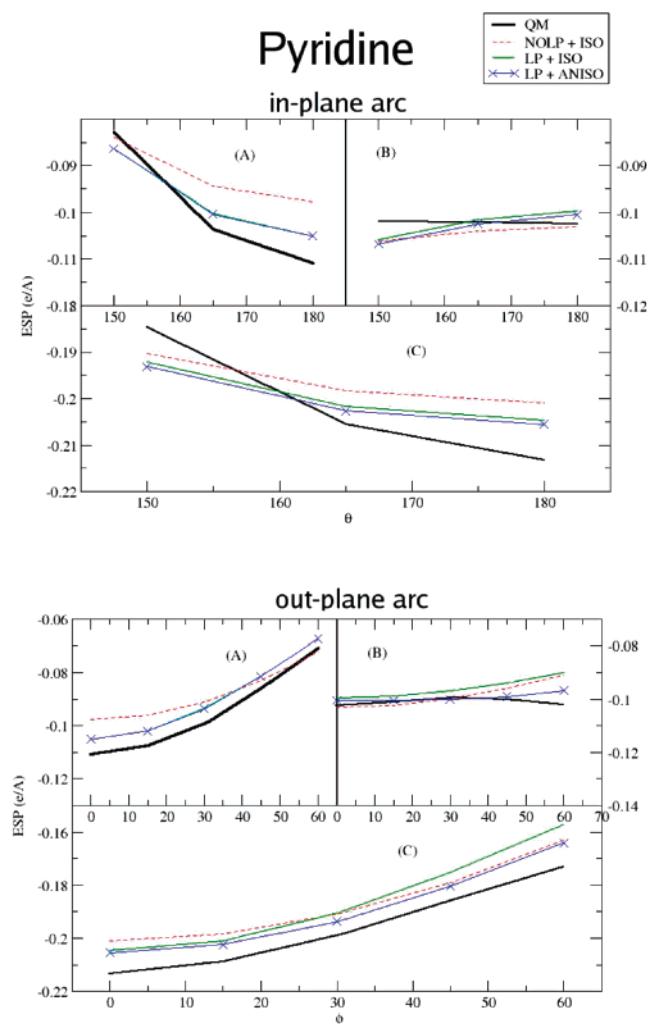
Like the virtual site geometry the relative magnitude of the components of  $\mathbf{K}_D$  is not part of the parameter fitting function, eq 4. The polarization ESP computed on the aforementioned QM arcs is used to parametrize the components of the force constant tensor. The value of the Drude



**Figure 4.** The ESP along an arc 2 Å from the nitrogen of a methylamine monomer is plotted in panel (A). The polarization contribution and total ESP from methylamine in the presence of a 0.5e test charge is plotted in panels (B) and (C), respectively. The perturbed ESP is calculated at the position of the 0.5e test charge.

oscillator charge, initially fit against eq 4 for the LP+ISO model, is scaled as the components of the force constant tensor are varied (keeping the trace approximately constant) until satisfactory agreement with the response curves from QM is found. This yields the LP+ANISO model. The polarization ESP for the resulting model is plotted as the blue-cross curve. By strengthening the polarizability along the LP direction, excellent agreement with the QM response is achieved for the models of acetone, methanol, and methylamine. The response curve for pyridine, which is approximately isotropic, already shows good agreement with the ISO models.

**4.2. Interaction Energies.** Interaction energies with a single water molecule are computed to probe the hydrogen bonding environment of the model compounds and serve to test the new MM model. In addition, we investigate interactions between the model compounds and a monovalent cation (sodium) to test the model under strong electric field conditions. Comparison to QM interaction energies are made for dimer conformations that correspond to an optimization



**Figure 5.** The ESP along an arc 2 Å from the nitrogen of a pyridine monomer is plotted in panel (A). The polarization contribution and total ESP from pyridine in the presence of a 0.5e test charge is plotted in panels (B) and (C), respectively. The perturbed ESP is calculated at the position of the 0.5e test charge.

along an intermolecular distance coordinate between the oxygen or nitrogen atom of the model compound and the hydrogen bond forming hydrogen of water or sodium ion. The conformations resulting from this partial minimization are presented in Figure 6. The conformation labels indicate whether the water/ion is coincident with the previously defined in-plane or out-plane and the angle  $\theta$ . The model/sodium conformations (not illustrated) probe the same coordinates.

For water, the results for the optimized energies and the difference between QM and MM are presented in Table 1. The average error over the suite of conformations is presented in Table 2. The results for acetone and methanol illustrate the failure of the NOLP+ISO MM model to predict the location of the lowest energy orientation (coordinated with the lone pair in QM). The LP models (both ISO and ANISO) lead to the correct prediction for the ordering of interaction energies and significantly improve the average error over all conformations. For methylamine a significant improvement is seen in the interaction energies with water. The

**Table 1.** Interaction Energy and the Energy Difference between QM and MM from a Partial Radial Optimization of a Model Compound/Water Dimer<sup>a</sup>

orientation	QM	NOLP + ISO		LP + ISO		LP + ANISO	
	$E_{\min}$	$E_{\min}$	$E_{\text{diff}}$	$E_{\min}$	$E_{\text{diff}}$	$E_{\min}$	$E_{\text{diff}}$
Acetone							
in-120	-5.1	-3.8	1.3	-4.7	0.4	-4.9	0.2
in-180	-4.1	-4.0	0.1	-3.6	0.5	-3.6	0.5
out-180	-3.8	-4.0	0.2	-3.7	0.1	-3.7	0.1
Methanol							
out-105	-4.6	-4.6	0.0	-4.6	0.0	-4.7	0.1
in-105	-4.3	-4.7	0.4	-4.1	0.2	-4.1	0.2
in-180	-2.1	-2.5	0.4	-2.0	0.1	-2.0	0.1
Methylamine							
in-105	-6.4	-6.5	0.1	-6.2	0.2	-6.4	0.0
out-105	-4.8	-5.4	0.6	-4.5	0.3	-4.6	0.2
in-150	-2.1	-3.5	1.4	-2.3	0.2	-2.3	0.2
Pyridine							
in-180	-6.3	-6.3	0.0	-6.3	0.0	-6.3	0.0
in-150	-4.9	-5.2	0.3	-5.0	0.1	-5.0	0.1
out-180	-4.4	-5.3	0.9	-4.1	0.3	-4.2	0.1

<sup>a</sup> All energies are in kcal/mol. The orientations correspond to conformations illustrated in Figure 6.

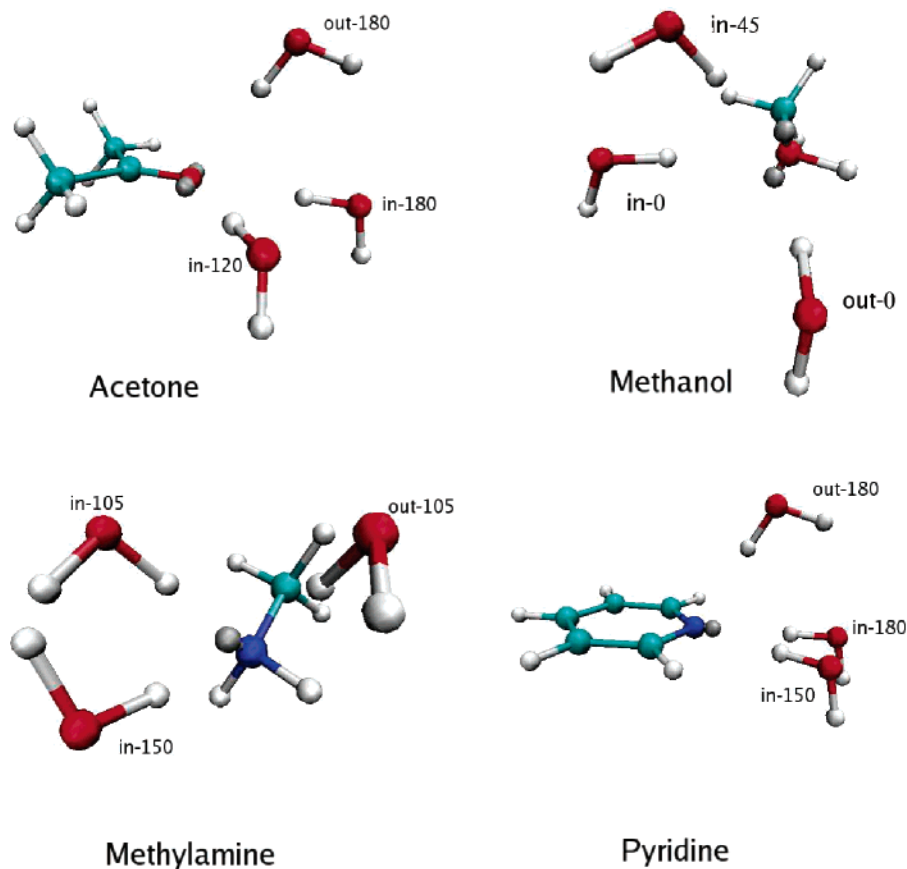
**Table 2.** Average Error (AE) of MM with Respect to QM for Interaction Energies (kcal/mol) of Model Compound/Water Dimers from Partial Radial Optimization

molecule	NOLP + ISO AE ( $E_{\min}$ )	LP + ISO AE ( $E_{\min}$ )	LP + ANISO AE ( $E_{\min}$ )
acetone	0.6	0.3	0.3
methanol	0.3	0.1	0.1
methylamine	0.7	0.2	0.1
pyridine	0.4	0.1	0.1

energy difference between QM and MM is reduced by greater than 1 kcal/mol for the conformation labeled “in-150”, and the average error falls from 0.7 kcal/mol in the NOLP model to 0.2–0.1 kcal/mol for the LP models. A similar improvement is seen with pyridine where the energy difference for the out-of-plane conformation “out-180” is reduced by 0.6–0.8 kcal/mol for the LP models. The addition of anisotropic Drude oscillators gives a modest additional improvement to the observed results leading to smaller energy differences with the QM results relative to the LP+ISO model.

For sodium, the results for the optimized energies and the difference between QM and MM are presented in Table 3. To achieve a satisfactory level of accuracy, it was necessary to place additional LJ interactions centered on the positions of the virtual lone pair sites. The need for such an empirical correction possibly reflects the breakdown of linear response caused by the large distortion in the electron density in the presence of a large electrostatic perturbation. The average errors presented in Table 4 show a marked improvement for the models that use virtual sites relative to the NOLP model. Including anisotropy leads to a further reduction in the average error.

**4.3. Linear Response Regime.** For polarizable MM models the response of the polarizable degrees of freedom



**Figure 6.** Location of water after partial radial optimization of model compound/water dimer.

**Table 3.** Interaction Energy and the Energy Difference between QM and MM from a Partial Radial Optimization of a Model Compound/Sodium Dimer<sup>a</sup>

orientation	QM	NOLP + ISO		LP + ISO		LP + ANISO	
	$E_{\min}$	$E_{\min}$	$E_{\text{diff}}$	$E_{\min}$	$E_{\text{diff}}$	$E_{\min}$	$E_{\text{diff}}$
Acetone							
in-120	-24.5	-24.7	0.2	-24.6	0.1	-24.5	0.0
in-180	-29.9	-31.9	2.0	-30.9	1.0	-29.7	0.2
out-180	-23.2	-25.1	1.9	-23.9	0.7	-24.7	1.5
Methanol							
in-45	-20.2	-19.6	0.6	-19.8	0.4	-20.9	0.7
in-0	-21.7	-22.1	0.4	-21.2	0.5	-21.5	0.2
out-0	-13.6	-14.2	0.6	-13.5	0.1	-13.7	0.1
Methylamine							
in-105	-26.2	-26.0	0.2	-26.5	0.3	-27.2	1.0
out-105	-11.9	-15.0	3.1	-13.4	1.5	-13.1	1.2
in-150	-20.8	-21.8	1.0	-22.1	1.3	-21.0	0.2
Pyridine							
in-180	-29.8	-29.9	0.1	-29.9	0.1	-30.0	0.2
in-150	-25.7	-26.5	0.8	-27.2	1.5	-27.3	1.6
out-180	-23.8	-21.1	2.7	-22.8	1.0	-23.5	0.3

<sup>a</sup> All energies are in kcal/mol. The orientations correspond to conformations similar to those illustrated in Figure 6 for water.

is typically assumed to be linearly proportional to an electrostatic perturbation. This approximation is exact for a model potential energy that is quadratic in the polarizable degrees of freedom such as polarizable point dipole models<sup>22</sup> and fluctuating charge models<sup>47</sup> where the polarizable degrees

**Table 4.** Average Error (AE) of MM with Respect to QM for Interaction Energies (kcal/mol) of Model Compound/Sodium Dimers from Partial Radial Optimization

molecule	NOLP + ISO AE ( $E_{\min}$ )	LP + ISO AE ( $E_{\min}$ )	LP + ANISO AE ( $E_{\min}$ )
acetone	1.4	0.6	0.6
methanol	0.5	0.3	0.3
methylamine	1.4	1.0	0.8
pyridine	1.2	0.9	0.7

of freedom are the magnitude of the dipoles and atomic centered charges, respectively. This condition is not strictly held in the Drude oscillator model, where the electric field is evaluated at the position of the mobile auxiliary particle. For weak perturbations and/or a very stiff spring constant the Drude displacements are much smaller than the interatomic distance ( $r_{ij}$ ), and the effective response is linear. For large electrostatic perturbations the Drude model, like the real QM molecule, will deviate from linear response. In this nonlinear regime we do not expect the Drude model to successfully model the QM data. Such limitations notwithstanding, the present implementation of the Drude model with a negative (rather than the positive) Drude particle<sup>21,48</sup> does deviate from the linear response regime in the same direction as the real QM case. In this section we investigate the level of accuracy possible with these MM models (both LP models are investigated) by measuring the degree to which the QM molecule response deviates from the linear regime for electrostatic charge perturbations of varying magnitude (0.5e, 1.0e, 2.0e) and compare with the Drude models.

**Table 5.** Breakdown of Interaction Energy between Acetone and a 0.5e Test Charge<sup>a</sup>

model	$q\phi_{\text{stat}}$	$0.5q\phi_{\text{pol}}$	$E^{\text{lr}}$	$E$	$E_{\text{diff}}$
$\theta = 120, d = 2.286 \text{ \AA}$					
QM	-10.7	-3.8	-14.5	-14.4	0.1
LP+ISO	-11.2	-2.7	-13.9	-13.9	0.0
LP+ANISO	-11.2	-3.4	-14.6	-14.5	0.1
$\theta = 180, d = 2.1883 \text{ \AA}$					
QM	-13.1	-3.4	-16.5	-16.3	0.2
LP+ISO	-13.4	-3.4	-16.8	-16.7	0.1
LP+ANISO	-13.4	-3.3	-16.7	-16.6	0.1

<sup>a</sup> All values are in kcal/mol. Calculations for conformations at  $\theta = 120$  and  $\theta = 180$  are presented. Shown are the static ( $q\phi_{\text{stat}}$ ) and polarization ( $0.5q\phi_{\text{pol}}$ ) contribution to the energy predicted from linear response ( $E^{\text{lr}}$ ), the actual interaction energy ( $E$ ), and the difference between them ( $E_{\text{diff}}$ ). The models presented include the QM molecule and the LP+ISO and LP+ANISO Drude oscillator models.

The total interaction energy between a polarizable molecule and a test charge ( $q_{\text{test}}$ ) can be written as

$$E = \int_0^{q_{\text{test}}} dq \phi(q) \quad (6)$$

where  $\phi(q)$  is the total electrostatic potential from the molecule at the position of the perturbing charge. In the linear response approximation, it is assumed that  $\phi(q) = A + Bq$ , and the interaction energy becomes

$$E^{\text{lr}} = Aq_{\text{test}} + \frac{1}{2}Bq_{\text{test}}^2 \quad (7)$$

The constant  $A$  corresponds to the electrostatic potential in the absence of a perturbing charge,  $\phi_{\text{stat}} = \phi(q = 0)$ . The constant  $B$ , a measure of the change in the electrostatic potential to the presence of a test charge, is thus  $(\phi(q_{\text{test}}) - \phi_{\text{stat}})/q_{\text{test}}$  which is equal to  $\phi_{\text{pol}}/q_{\text{test}}$ . The computation of  $E^{\text{lr}}$  follows from a similar breakdown of the ESP into static ( $\phi_{\text{stat}}$ ) and polarization ( $\phi_{\text{pol}}$ ) contributions performed in the electrostatic model parametrization (see section 2.2). A series of single point calculations of the ESP are carried out at positions corresponding to the minimum energy for the model compound/sodium dimer. The static field contribution to  $E^{\text{lr}}$ , corresponding to the first term in eq 7, is taken from the ESP in the absence of a test charge. The contribution to  $E^{\text{lr}}$  from polarization, the second term in eq 7, is taken from the difference in the ESP in the presence and absence of a test charge. To examine the validity of linear response, the computed value for  $E^{\text{lr}}$  is compared to the actual QM interaction energy, taken as the difference between the energy of the complex and the energy of the monomer. The results were particularly sensitive to the basis set chosen. Therefore the results for a large basis set including diffuse and polarization functions (6-311+G(3df,2p)) is presented in this article. Along with the interaction energy ( $E$  and  $E^{\text{lr}}$ ) the aforementioned static and polarization components of  $E^{\text{lr}}$  are recorded for acetone in Tables 5–7 for the 0.5e, 1.0e, and 2.0e charge, respectively. The conformations corresponding to the in-120 and in-180 labels are similar to those plotted for the model compound/water dimer (see Figure 6). The same calculations were then carried out for the LP+ISO and LP+ANISO MM models.

**Table 6.** Breakdown of Interaction Energy between Acetone and a 1.0e Test Charge<sup>a</sup>

model	$q\phi_{\text{stat}}$	$0.5q\phi_{\text{pol}}$	$E^{\text{lr}}$	$E$	$E_{\text{diff}}$
$\theta = 120, d = 2.286 \text{ \AA}$					
QM	-21.3	-19.3	-40.6	-38.0	2.6
LP+ISO	-22.3	-11.8	-34.1	-33.6	0.5
LP+ANISO	-22.5	-15.8	-38.3	-36.9	1.4
$\theta = 180, d = 2.1883 \text{ \AA}$					
QM	-26.2	-15.3	-41.5	-40.2	1.3
LP+ISO	-26.8	-14.9	-41.7	-40.7	1.0
LP+ANISO	-26.8	-14.2	-41.0	-40.2	0.8

<sup>a</sup> All values are in kcal/mol. Calculations for conformations at  $\theta = 120$  and  $\theta = 180$  are presented. Shown are the static ( $q\phi_{\text{stat}}$ ) and polarization ( $0.5q\phi_{\text{pol}}$ ) contribution to the energy predicted from linear response ( $E^{\text{lr}}$ ), the actual interaction energy ( $E$ ), and the difference between them ( $E_{\text{diff}}$ ). The models presented include the QM molecule and the LP+ISO and LP+ANISO Drude oscillator models.

**Table 7.** Breakdown of Interaction Energy between Acetone and a 2.0e Test Charge<sup>a</sup>

model	$q\phi_{\text{stat}}$	$0.5q\phi_{\text{pol}}$	$E^{\text{lr}}$	$E$	$E_{\text{diff}}$
$\theta = 120, d = 2.286 \text{ \AA}$					
QM	-42.9	-151.0	-193.9	-146.7	47.2
LP+ISO	-44.7	-56.3	-101.0	-94.6	6.4
LP+ANISO	-44.7	-130.1	-174.8	-124.9	49.9
$\theta = 180, d = 2.1883 \text{ \AA}$					
QM	-52.4	-115.0	-167.4	-128.8	38.6
LP+ISO	-53.7	-78.4	-132.1	-119.1	13.0
LP+ANISO	-53.6	-71.1	-124.7	-115.0	9.7

<sup>a</sup> All values are in kcal/mol. Calculations for conformations at  $\theta = 120$  and  $\theta = 180$  are presented. Shown are the static ( $q\phi_{\text{stat}}$ ) and polarization ( $0.5q\phi_{\text{pol}}$ ) contribution to the energy predicted from linear response ( $E^{\text{lr}}$ ), the actual interaction energy ( $E$ ), and the difference between them ( $E_{\text{diff}}$ ). The models presented include the QM molecule and the LP+ISO and LP+ANISO Drude oscillator models.

For the 0.5e charge, linear response holds almost exactly in both the QM and MM calculations. As expected, from fitting to the ESP, the LP+ANISO model improves the accuracy of the polarization contribution to the interaction energy and gives the best overall agreement with the QM result. For the 1.0e test charge, the energies begin to deviate by approximately 0.5–2.5 kcal/mol from the linear response regime. Somewhat encouraging is that the deviation in the interaction energy caused by nonlinear effects arising from finite Drude displacements in the MM model are in the same direction as the QM calculation. Only the LP+ANISO model gives interaction energies that lie within the linear response resolution.

For the 2.0e charge, the deviations from linear response become quite severe ( $\approx 50$  kcal/mol). The large deviations would appear to reflect poorly on the effectiveness of linear response models (point dipole and fluctuating charge) and approximate linear response model (Drude oscillators) to accurately model interactions with divalent ions such as calcium and magnesium. However one should keep in mind that this experiment is performed in the gas phase. As noted in previous work,<sup>20,35,49</sup> contributions to the electronic response in the gas phase from diffuse portions of the electron wave function are reduced in a condensed phase

environment due to repulsive interactions with neighboring molecules. It may be the case that this would decrease not only the polarizability but also the degree to which the response of these molecules deviates from the linear response regime. If these deviations are indicative of the response in aqueous ionic systems, it may be necessary to incorporate additional features (e.g. an anharmonic restoring force to the Drude particles) to properly model the polarization effects when linear response breaks down.

**4.4. Conclusions.** In the present manuscript we investigate the components of the electrostatic potential, in the vicinity of functional groups containing lone pairs, for a set of small molecules. Included are the nitrogen of pyridine and methylamine, the carbonyl group of acetone, and the hydroxyl group of methanol. Three models, incorporating virtual sites to represent the lone pairs and the local polarization anisotropy, were parametrized and compared with ab initio electronic structure calculations for a system containing the molecule and a perturbing charge of magnitude  $+0.5e$ . Use of virtual sites representing lone pairs and anisotropic atom-based polarizabilities allows for the model to better reproduce the anisotropy of interactions with the environment, thereby improving the ability of the model to reproduce atomic detail interactions. Only the model using both virtual sites and anisotropic Drude oscillators is able to capture all the features of the local QM electrostatic potential. Dimer energies with water and sodium show a marked improvement for models that incorporate virtual sites in the representation of lone pairs relative to QM calculations. The MM model that incorporates virtual sites and anisotropic atomic polarizabilities for lone pairs will be implemented into an empirical force field that explicitly includes electronic polarizability via Drude oscillators.

Notably absent from the preceding analysis is a water model. Although water does indeed have lone pairs, the use of a TIP4P-like<sup>50</sup> virtual site (charge on oxygen is shifted off the atom along the HOH bisector) by the Drude water model (SWM4-NDP<sup>20</sup>) that will accompany the rest of the biomolecular force field appears to provide the level of detail necessary to accurately reproduce the water dimer geometry. Furthermore the nearly isotropic experimental molecular polarizability<sup>51</sup> of water appears to make the addition of an anisotropic Drude oscillator unnecessary for this molecule.

**Acknowledgment.** Financial support to A.D.M. is acknowledged from the NIH (GM 51501) and to B.R. and A.D.M. from the NIH (GM 072558). Computer time allocations were received from DOD ACS Major Shared Resource Computing and PSC Pittsburgh Supercomputing Center. E.H. would like to thank Troy Whitfield and Sergey Krupin for helpful discussions.

**Supporting Information Available:** The model parameters, minimized interacting dimer distances, and energies. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fisher, 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.; Prodhorn, B.; Reiher, W. E., III; Roux, B.; Schlenkerich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wirkiewicz-Kuczera, D.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (2) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (3) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R., Jr.; K. M. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (4) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Huenenberger, P. H.; Krueger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GRO-MOS96 Manual and User Guide*; Hochschulverlag AG/ETH: Zurich, 1996.
- (5) Silvestrelli, P. L.; Parrinello, M. *J. Chem. Phys.* **1999**, *111*, 3572–3580.
- (6) Harder, E.; Eaves, J. D.; Tokmakoff, A.; Berne, B. J. *Proc. Natl. Acad. Sci.* **2005**, *102*, 11611–11616.
- (7) Asbury, J. B.; Steinel, T.; Kwak, K.; Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L.; Fayer, M. D. *J. Chem. Phys.* **2004**, *121*, 12431–12446.
- (8) Xu, H.; Stern, H. A.; Berne, B. J. *J. Phys. Chem. B* **2002**, *106*, 2054–2060.
- (9) Stuart, S. J.; Berne, B. J. *J. Phys. Chem.* **1996**, *100*, 11934–11943.
- (10) Jungwirth, P.; Tobias, D. J. *J. Phys. Chem. B* **2002**, *106*, 6361–6373.
- (11) Herce, D. H.; Perera, L.; Darden, T. A.; Sagui, C. *J. Chem. Phys.* **2005**, *122*, 024513.
- (12) Archontis, G.; Leontidis, E.; Andreou, G. *J. Phys. Chem. B* **2005**, *109*, 17957–17966.
- (13) Lamoureux, G.; Roux, B. *J. Phys. Chem. B* **2006**, *110*, 3308–3322.
- (14) Noskov, S. Y.; Lamoureux, G.; Roux, B. *J. Phys. Chem. B* **2005**, *109*, 6705–6713.
- (15) Vorobyov, I. V.; Anisimov, V. M.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2005**, *109*, 18988–18999.
- (16) Kim, B.; Young, T.; Harder, E.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* **2005**, *109*, 16529–16538.
- (17) Allen, T. W.; Andersen, O. S.; Roux, B. *Biophys. J.* **2006**, *90*, 3447–3468.
- (18) Allen, T. W.; Andersen, O. S.; Roux, B. *Proc. Natl. Acad. Sci.* **2004**, *101*, 117–122.
- (19) Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025–3039.
- (20) Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185–5197.
- (21) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I.; Huang, N.; Roux, B.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2005**, *1*, 153–168.
- (22) van Belle, D.; Froeyen, M.; Lippens, G.; Wodak, S. J. *Mol. Phys.* **1992**, *77*, 239–255.
- (23) Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1994**, *98*, 4180–4187.
- (24) Mountain, R. D. *J. Chem. Phys.* **1995**, *103*, 3084–3090.



- (25) Gao, J.; Habibollahzadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460–16467.
- (26) Dang, L. X.; Chang, T. *J. Chem. Phys.* **1997**, *106*, 8149–8159.
- (27) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515–1531.
- (28) Drude, P. *The Theory of Optics*; Longmans: Green, New York, 1902.
- (29) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341–350.
- (30) van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399–2407.
- (31) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. B., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. *CHARMM: The Energy Function and Its Parametrization with an Overview of the Program. In Encyclopedia of Computational Chemistry*; John Wiley & Sons: Chichester, 1998; Vol. 1.
- (32) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741–754.
- (33) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, *103*, 4730–4737.
- (34) Liu, Y.-P.; Kim, K.; Berne, B. J.; Friesner, R. A.; Rick, S. W. *J. Chem. Phys.* **1998**, *108*, 4739–4755.
- (35) Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, *115*, 2237–2251.
- (36) Singh, U. C.; Kollman, P. A. *J. Comput. Chem* **1984**, *5*, 129–145.
- (37) Singh, U. C.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (38) Yin, D. Ph.D. Thesis, University of Maryland, 1997.
- (39) Chen, I.-J.; Yin, D.; MacKerell, A. D., Jr. *J. Comput. Chem* **2002**, *23*, 199–213.
- (40) Miller, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 8533–8542.
- (41) Bader, R. F. W. *Atoms in Molecules – A Quantum Theory*; Oxford University Press Inc.: New York, 1990.
- (42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*; Gaussian Inc.: Pittsburgh, PA.
- (43) Huang, N.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2002**, *106*, 7820–7827.
- (44) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (45) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (46) Boys, S.-F.; Bernardi, F. *Mol. Phys.* **2002**, *100*, 65–73.
- (47) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141–6156.
- (48) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. *Chem. Phys. Lett.* **2006**, *418*, 245–249.
- (49) Morita, A.; Kato, S. *J. Chem. Phys.* **1999**, *110*, 11987–11998.
- (50) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (51) Murphy, W. F. *J. Chem. Phys.* **1977**, *67*, 5877–5882.

CT600180X

## Ab Initio Finite-Temperature Electronic Absorption Spectrum of Formamide

Nicholas A. Besley<sup>†</sup> and Nikos L. Doltsinis<sup>\*‡</sup>

*School of Chemistry, University of Nottingham, Nottingham, NG7 2RD, United Kingdom, and Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, Bochum, D-44780 Germany*

Received July 27, 2006

**Abstract:** A combination of Car–Parrinello molecular dynamics (CP–MD) and high-level ab initio quantum chemical calculations has been used to calculate the electronic absorption spectrum of formamide at finite temperatures. Thermally broadened spectra have been obtained by averaging over a large number of single-point multireference configuration interaction excitation energies calculated for geometries sampled from a CP–MD simulation. Electronic excitation spectra of possible contaminants ammonia and formamidic acid have also been computed. Ammonia exhibits a strong peak in the shoulder region of the experimental formamide spectrum at 6.5 eV, and formamidic acid has a strong absorption above 7.5 eV. The calculations reproduce the shape of the experimental absorption spectrum, in particular, the low-energy shoulder of the main peak, and demonstrate how finite-temperature electronic absorption spectra can be computed from first principles.

### 1. Introduction

A major goal of current research in molecular biology is the determination and understanding of protein structure. X-ray crystallography can determine protein structure at the atomic level. However, many proteins cannot be crystallized, and there may be differences between the crystal and solvated structures. These disadvantages coupled with time-resolved spectroscopy, which can monitor the evolution of protein structure, have led to continued interest in spectroscopic probes of protein structure.<sup>1,2</sup>

Electronic circular dichroism (CD) spectroscopy is used extensively as a measure of the helical content of proteins. Theoretical calculations of protein CD hold the key to establishing and quantifying the link between the measured spectra and the underlying structural information. To date, the most accurate theoretical calculations of protein circular dichroism spectroscopy use small amides as models of the backbone chromophore.<sup>3</sup> This role as a model for the repeating unit in the protein backbone has motivated many studies

into the electronic structure of amides. In particular, many theoretical and experimental investigations of the excited states of formamide in the gas phase<sup>4–15</sup> and in solution<sup>4,15–22</sup> have been reported. In addition to characterizing the spectroscopy of an important system, these studies provide a basis for improving our understanding and simulation of the spectroscopy of proteins.

In the gas phase,<sup>13</sup> the electronic spectrum of formamide is dominated by an intense band at 7.4 eV, often labeled the  $V_1$  band, arising from a  $\pi_{nb}\pi^*$  (nonbonding  $\pi$  orbital to antibonding  $\pi^*$  orbital) transition. At about 5.5 eV is a much weaker band arising from the  $n\pi^*$  (lone pair on an oxygen to an antibonding  $\pi^*$  orbital) transition. The spectrum is also characterized by a number of sharp peaks that are attributed to Rydberg excitations. At a higher energy is the Q band that was originally assigned to a  $\pi_b\pi^*$  excitation, although later work has shown it to arise from a superposition of Rydberg excitations.<sup>14</sup>

In recent years, there has been a number of theoretical studies of the excited states of formamide employing a range of methodologies. Multireference configuration interaction (MRCI) calculations were reported by Hirst et al.<sup>10</sup> These provided an accurate description of the Rydberg states, but

\* Corresponding author e-mail: nikos.doltsinis@theochem.rub.de.

<sup>†</sup> University of Nottingham.

<sup>‡</sup> Ruhr-Universität Bochum.

the predicted  $\pi_{\text{nb}}\pi^*$  excitation energy was too high. This was attributed to Rydberg-valence mixing and was not improved significantly with the use of a larger active space. Rydberg-valence mixing occurs because the average state MRCI calculation includes a number of Rydberg states in addition to the valence  $n\pi^*$  and  $\pi_{\text{nb}}\pi^*$  states. The Rydberg states are optimized at the “cost” of the valence states, resulting in the valence states being too diffuse with corresponding energies that are too high. The complete active space self-consistent field with multiconfigurational perturbation theory (CASSCF/CASPT2) study of Serrano-Andrés and Fülcher overcame this problem with a two-step procedure comprising a calculation of the Rydberg state energies followed by a subsequent calculation with the Rydberg states “deleted” to determine the valence state properties.<sup>11</sup> These calculations predicted a value of 7.41 eV for the  $\pi_{\text{nb}}\pi^*$  excitation energy, in agreement with experimental results.

These calculations represent each electronic transition with a single energy and oscillator strength. The direct calculation of electronic spectra represents an important challenge for theory. Broadening of the spectral lines can occur through thermal and electronic effects, resulting in the spectral bands observed in experimental results. The electronic spectrum of formamide has been simulated using a combination of molecular dynamics and quantum chemical calculations to model the thermal broadening of the spectral lines. Doltsinis and Sprik<sup>14</sup> used a combination of Car–Parrinello molecular dynamics (CP–MD) and time-dependent density functional theory (TDDFT) to calculate the electronic spectrum of formamide at room temperature. However, the sampling of configuration space in this study was poor, and the TDDFT Rydberg excitation energies were lacking the important Rydberg corrections. Later work<sup>15</sup> used classical molecular dynamics simulation with TDDFT to model the electronic spectra of a number of amides, including formamide. Both of these studies reproduced important spectral features observed in experimental results but failed to explain certain characteristics of the experimental spectra such as the low-energy shoulder of the main peak around 7.4 eV.

In this work, we reinvestigate the excited states of formamide with MRCI. An improved description of thermal broadening is achieved by averaging over a large number of finite temperature configurations sampled from CP–MD simulations. The thermally averaged results using CP–MD will be compared in detail to those obtained using a classical force field. To partially mimic nuclear quantum effects which smear out the distribution function of the atomic nuclei, we have also carried out CP–MD simulations at increased temperatures<sup>23,24</sup> and studied the effect on the absorption spectrum. Furthermore, we discuss for the first time the absorption spectra of the possible contaminants ammonia and formamidic acid.

## 2. Computational Details

CP–MD<sup>25,26</sup> simulations of formamide have been carried out at three different temperatures, 300, 400, and 500 K. The calculations were performed in a periodically repeated orthorhombic unit cell of size  $10 \times 8 \times 10 \text{ \AA}$  using the BLYP exchange-correlation functional<sup>27,28</sup> and a plane-wave

basis truncated at 70 Ry in conjunction with Troullier–Martins pseudopotentials.<sup>29</sup> The system was first brought to thermal equilibrium in a thermostated run over more than 2 ps using a Nosé–Hoover chain<sup>30,31</sup> for each degree of freedom. In the production run of about 15 ps in length, a single Nosé–Hoover chain was used for the whole system to reproduce the canonical ensemble. The propagation of both the fictitious electronic and the nuclear degrees of freedom was carried out with a time step of 4 au; a fictitious mass of 400 au was ascribed to the former. Analogous CP–MD simulations were performed for gas-phase ammonia using the same settings as those for formamide. Merely the unit cell was changed to a simple cubic box of length 9 Å. From each production run, a set of 100 molecular structures was extracted at intervals of 1000 MD steps (roughly 100 fs).

Classical molecular dynamics simulations were performed using the CHARMM program<sup>32</sup> with the CHARMM22 all-hydrogen parameters.<sup>33</sup> All simulations had a time step of 1 fs and consisted of a heating time of 6 ps in which the temperature was raised from 0 to 300 K followed by an equilibrium period of 12 ps. In the first simulation, denoted class-short, 100 structures were drawn at 20 fs intervals from a 200 ps simulation. The second set of 100 structures, denoted class-long, were taken from 10 independent 200 ps simulations at 2 ps intervals. These classical simulations are described in more detail elsewhere.<sup>15</sup>

We would like to emphasize at this point that we obtain thermally broadened electronic excitation spectra by averaging over 100 vertical excitation spectra calculated at the different geometries sampled from a ground-state trajectory. A full quantum-mechanical treatment would also require calculation of the overlap between the nuclear wave functions of the ground state and all excited states. However, this is currently unfeasible for the large number of electronic states and structures considered here.

Excited-state calculations were performed using the MOLPRO suite of programs.<sup>34</sup> Reference orbitals for the MRCI calculations were obtained from state averaged multiconfigurational self-consistent field calculations (MCSCF). The MP2/6-31+G\*\* equilibrium geometry with  $C_s$  symmetry from an earlier study<sup>10</sup> was used. The (9,0;16,6) active space comprising nine  $a'$  and zero  $a''$  closed orbitals with seven  $a'$  and six  $a''$  active orbitals was chosen. The active  $a'$  orbitals include the  $n$ , 3s, 3p<sub>x</sub>, 3p<sub>y</sub>, 3d<sub>xy</sub>, 3d<sub>x<sup>2</sup>-y<sup>2</sup></sub>, and 3d<sub>z<sup>2</sup></sub> orbitals, while the active  $a''$  orbitals are the  $\pi_b$ ,  $\pi_{\text{nb}}$ ,  $\pi^*$ , 3p<sub>z</sub>, 3d<sub>xz</sub>, and 3d<sub>yz</sub> orbitals (with formamide in the  $xy$  plane). In the MRCI calculations, 10  $A'$  states and nine  $A''$  states were calculated using the projection procedure introduced by Knowles and Werner.<sup>35</sup> However, these calculations give poor excitation energies and properties for the  $n\pi^*$  and  $\pi_{\text{nb}}-3s$  states. The source of this problem lies with the MCSCF reference calculation, which predicts the  $n\pi^*$  state to lie above the  $\pi_{\text{nb}}3s$  state. To obtain accurate values for these states, a second state averaged MRCI calculation over the two lowest  $A''$  states was performed. The d-aug-cc-pVDZ and d-aug-cc-pVTZ basis sets<sup>36–38</sup> were used for these calculations. Structures drawn from molecular dynamics simulations have  $C_1$  symmetry. For these calculations, the projection procedure was employed to determine the lowest

**Table 1.** MRCI/d-aug-cc-pVDZ Excitation Energies,  $\Delta E$  (in eV); Permanent and Transition Dipole Moments,  $\mu$  and  $\mu^T$  (in debye); and Oscillator Strengths,  $\sigma$ , for Formamide<sup>a</sup>

state	$\Delta E$	$\mu_x$	$\mu_y$	$\mu_z$	$ \mu $	$\mu_x^T$	$\mu_y^T$	$\mu_z^T$	$\sigma$	assignment
1A'		4.31	-1.13	0.00	4.46					G. S.
2A'	6.41	-3.13	-0.20	0.00	3.13	-0.03	0.02	0.00	0.000	n3s
3A'	7.09	1.43	3.17	0.00	3.48	1.17	-1.10	0.00	0.069	n3p
4A'	7.33	3.14	-3.07	0.00	4.39	-2.17	-1.02	0.00	0.160	$\pi_{nb}\pi^*$
5A'	7.44	-0.16	-0.18	0.00	0.24	-1.33	0.61	0.00	0.060	n3p
6A'	7.76	-0.10	3.49	0.00	3.49	-1.61	-0.30	0.00	0.079	$\pi_{nb}3p$
7A'	8.21	-0.53	-1.45	0.00	1.54	0.02	-0.18	0.00	0.001	n3d
8A'	8.33	-0.70	2.40	0.00	2.50	0.37	0.08	0.00	0.004	$\pi_{nb}3d$
9A'	8.40	-7.92	-0.33	0.00	7.93	0.96	0.16	0.00	0.030	n3d
10A'	8.53	6.35	0.81	0.00	6.40	-0.46	0.00	0.00	0.007	n3d
1A''	5.76	1.83	-0.24	0.00	1.84	0.00	0.00	0.09	0.000	n $\pi^*$
2A''	6.11	-3.96	-0.15	0.00	3.96	0.00	0.00	-0.98	0.022	$\pi_{nb}3s$
3A''	7.11	-3.52	0.37	0.00	3.54	0.00	0.00	-0.15	0.006	$\pi_{nb}3p$
4A''	7.46	4.75	-0.04	0.00	4.75	0.00	0.00	-0.41	0.005	$\pi_{nb}3p$
5A''	7.50	0.17	1.53	0.00	1.54	0.00	0.00	-0.47	0.006	n3p
6A''	8.15	-4.93	0.12	0.00	4.93	0.00	0.00	0.20	0.001	$\pi_{nb}3d$
7A''	8.18	0.32	0.28	0.00	0.42	0.00	0.00	-0.28	0.002	$\pi_{nb}3d$
8A''	8.42	3.55	-0.32	0.00	3.56	0.00	0.00	-0.17	0.001	n3d
9A''	8.47	2.24	4.56	0.00	5.08	0.00	0.00	-0.29	0.003	$\pi_{nb}3d$

<sup>a</sup> The coordinates (in bohr) of the C, N, and O atoms are (-0.16, -0.78, 0), (2.17, 0.32, 0), and (-2.16, 0.39, 0), respectively.

20 states. Again, the poor MCSCF reference led to problems for the  $n\pi^*$  and  $\pi_{nb}3s$  states. The excitation energies and transition dipole moments for the  $n\pi^*$ ,  $\pi_{nb}3s$ , and  $n3s$  states were taken from separate state averaged MRCI calculations over the four lowest states. All calculations used the d-aug-cc-pVDZ basis set.

In addition to formamide, calculations have also been performed to assess the effect of likely contaminants on the absorption spectrum. The excited states of formamidic acid have been computed at the ground-state equilibrium MP2/6-31+G\*\* structure with  $C_s$  symmetry. A similar protocol to formamide was used, with the (9,0;16,6) active space and d-aug-cc-pVDZ basis set. For formamidic acid, averaging over structural snapshots drawn from MD simulations was not performed and a spectrum was generated by representing transitions with a Gaussian function with a bandwidth of 0.2 eV to mimic the thermal broadening. Ammonia was also studied at the MRCI/d-aug-cc-pVDZ level. An active space of (2;10) was chosen and the lowest five states computed using the projection procedure. The electronic absorption spectrum was generated on the basis of 100 snapshots drawn from the CP-MD simulation.

### 3. Results and Discussion

Table 1 shows the computed excitation energies and permanent and transition dipole moments for the equilibrium ( $C_s$ ) structure computed with the d-aug-cc-pVDZ basis set. The main focus of studies of amide excited states is the  $\pi_{nb}\pi^*$  state because this transition dominates the electronic absorption and CD spectra. The calculations predict an excitation energy of 7.33 eV. This lies close to the value from the experiment of 7.4 eV. Previous studies have shown accurate calculation of this excited state to be problematic because of Rydberg-valence mixing.<sup>10,11</sup> Our calculations differ from this previous work because a much larger basis set is used.

In particular, the d-aug-cc-pVDZ contains a number of diffuse basis functions that are appropriate for describing Rydberg states. This indicates that the Rydberg-valence mixing and associated poor description of the  $\pi_{nb}\pi^*$  excited state is an artifact of the small basis set. This is reasonable because a large basis set should be sufficiently flexible to describe both Rydberg and valence states and, in conjunction with MRCI with a large active space, should be accurate. The calculations predict  $\pi_{nb}\pi^*$  to be the most intense, although the computed oscillator strength is less than the estimated experimental value in solution.<sup>11</sup>

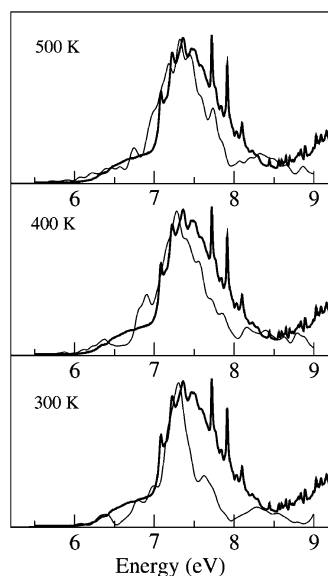
The other important valence state is the  $n\pi^*$  state. This state is weak and difficult to characterize in experiments. The predicted excitation energy of 5.76 eV is in agreement with experimental and previous calculations.<sup>11,19</sup> Most of the intensities for the remaining Rydberg states are low. The exceptions are the  $\pi_{nb}3s$  and  $n3p$  states on the low-energy side of the  $\pi_{nb}\pi^*$  state and  $n3p$ ,  $\pi_{nb}3p$ , and  $n3d$  excitations on the high-energy side. Table 2 shows results for the larger d-aug-cc-pVTZ basis set. There is little change in the calculated excitation energies. However, for some of the higher-lying Rydberg states, there are some significant changes in the computed dipole moments. This indicates that there are not sufficient basis functions in the smaller basis set to describe these states. Overall, there is good agreement between the absorption spectra predicted with the two basis sets, and the smaller d-aug-cc-pVDZ basis set is used in subsequent calculations.

Figure 1 shows the computed spectra based on CP-MD simulations at 300, 400, and 500 K with the experimental spectrum reported by Gingell et al.<sup>13</sup> shown in bold. At 300 K, the spectrum generally has the correct shape with an intense band at the correct energy arising predominantly from the  $\pi_{nb}\pi^*$  transition. The calculated spectrum also shows the formation of a shoulder between 6 and 7 eV comprising  $\pi3s$

**Table 2.** MRCI/d-aug-cc-pVTZ Excitation Energies,  $\Delta E$  (in eV); Permanent and Transition Dipole Moments,  $\mu$  and  $\mu^T$  (in debye); and Oscillator Strengths,  $\sigma$ , for Formamide<sup>a</sup>

state	$\Delta E$	$\mu_x$	$\mu_y$	$\mu_z$	$ \mu $	$\mu_x^T$	$\mu_y^T$	$\mu_z^T$	$\sigma$	assignment
1A'		4.31	-1.15	0.00	4.46					G. S.
2A'	6.45	-3.10	-0.14	0.00	3.10	-0.01	0.02	0.00	0.000	n3s
3A'	7.13	1.21	3.13	0.00	3.35	1.10	-1.11	0.00	0.066	n3p
4A'	7.40	3.37	-2.88	0.00	4.43	-2.13	-0.97	0.00	0.154	$\pi_{nb}\pi^*$
5A'	7.49	-0.56	-0.58	0.00	0.81	1.48	-0.54	0.00	0.071	n3p
6A'	7.86	0.16	3.96	0.00	3.96	-1.41	-0.35	0.00	0.063	$\pi_{nb}3p$
7A'	8.25	-0.65	-1.15	0.00	1.32	0.01	-0.17	0.00	0.001	n3d
8A'	8.34	1.77	1.93	0.00	2.61	0.02	0.04	0.00	0.000	$\pi_{nb}3d$
9A'	8.49	-2.47	-0.24	0.00	2.48	0.81	0.25	0.00	0.023	n3d
10A'	8.58	1.56	0.72	0.00	1.72	0.97	0.10	0.00	0.031	n3d
1A''	5.76	1.82	-0.20	0.00	1.84	0.00	0.00	0.08	0.000	n $\pi^*$
2A''	6.17	-3.97	-0.92	0.00	3.97	0.00	0.00	-0.98	0.023	$\pi_{nb}3s$
3A''	7.07	-0.32	-0.66	0.00	0.73	0.00	0.00	-0.14	0.001	$\pi_{nb}3p$
4A''	7.52	4.62	0.07	0.00	4.63	0.00	0.00	-0.37	0.004	$\pi_{nb}3p$
5A''	7.54	0.15	1.53	0.00	1.54	0.00	0.00	0.50	0.007	n3p
6A''	8.22	2.36	0.19	0.00	2.36	0.00	0.00	-0.46	0.006	$\pi_{nb}3d$
7A''	8.26	0.58	1.27	0.00	1.39	0.00	0.00	-0.12	0.000	$\pi_{nb}3d$
8A''	8.45	2.88	-0.52	0.00	2.92	0.00	0.00	-0.12	0.000	n3d
9A''	8.54	1.08	3.89	0.00	4.03	0.00	0.00	-0.15	0.001	$\pi_{nb}3d$

<sup>a</sup> The coordinates (in bohr) of the C, N, and O atoms are (-0.16, -0.78, 0), (2.17, 0.32, 0), and (-2.16, 0.39, 0), respectively.

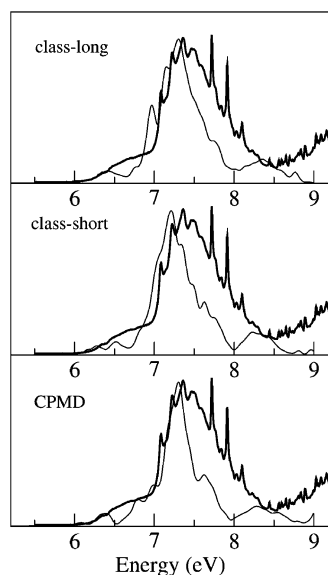


**Figure 1.** Theoretical simulations of the electronic absorption spectrum of formamide based on CP-MD simulations at different temperatures. The experimental spectrum<sup>13</sup> is shown in bold.

and  $n3p$  transitions. Although, the bands arising from these excitations are distinct and have not merged. On the high-energy side of the  $\pi_{nb}\pi^*$  band, there is little evidence of the sharp Rydberg bands observed in experimental results. The calculations only include the lowest 20 states, so there is no attempt to describe the spectrum above 8.5 eV. However, the main deficiency in the computed spectrum is that the  $\pi_{nb}\pi^*$  band is too narrow. At the higher temperatures, there is significant further broadening on the  $\pi_{nb}\pi^*$  band. At 500 K, there is reasonable agreement between the calculated and experimental spectra. Furthermore, the  $\pi3s$  and  $n3p$  bands merge, producing a broad shoulder on the low-energy side of the  $\pi_{nb}\pi^*$  band. On the high-energy side of the  $\pi_{nb}\pi^*$

band, the sharp bands corresponding to Rydberg excitations can be distinguished. We should bear in mind, at this point, that in the CP-MD simulations the atomic nuclei are treated as classical particles. Compared to a full quantum-mechanical description, the thermal distribution of classical nuclei is too narrow. However, it has been demonstrated previously<sup>23,24</sup> that nuclear quantum fluctuations can be emulated very well by CP-MD simulations at a higher temperature. It is not clear a priori, though, how to map the temperature of a classical simulation onto the real (quantum) temperature. Our observation that the classical simulation at 500 K yields closer agreement with experimental results than the lower temperature simulations seems to indicate that the 500 K run gives the best description of experimental conditions at 300 K.

Figure 2 shows spectra computed from classical and CP-MD simulations at 300 K. The classical-short spectrum can be compared directly to the CP-MD spectrum because the structural sampling is the same. The general shape of the spectra are similar with the features evident in the CP-MD spectrum also present in those from the classical simulation. The  $\pi_{nb}\pi^*$  band is slightly lower in energy and is also broader. There is also some further broadening in the spectrum from the longer simulation. Analysis of the MD trajectories shows that there are some significant differences in the structural parameters between the CP-MD and classical simulations. In particular, the C-N bond length is significantly shorter in the classical MD simulation. However, the prediction of the gross features of the spectrum appear not very sensitive to the quality of the simulation. Table 3 shows the ensemble average predictions of the  $n\pi^*$  and  $\pi_{nb}\pi^*$  excitation energies. These predicted excitation energies are sensitive to the simulation, with some significant differences arising between classical and CP-MD simulations. For all simulations, there is a decrease in the excitation energy compared to the values at the planar minimum energy



**Figure 2.** Theoretical simulations of the electronic absorption spectrum of formamide based on different MD simulations at 300 K. The experimental spectrum<sup>13</sup> is shown in bold.

**Table 3.** Ensemble Average  $n\pi^*$  and  $\pi_{nb}\pi^*$  Excitation Energies (in eV) and Oscillator Strengths in Parentheses

simulation	$n\pi^*$	$\pi_{nb}\pi^*$
CP-MD 300 K	5.61 (0.002)	7.27 (0.128)
CP-MD 400 K	5.63 (0.002)	7.28 (0.131)
CP-MD 500 K	5.57 (0.004)	7.28 (0.131)
classical (short) 300 K	5.74 (0.001)	7.23 (0.124)
classical (long) 300 K	5.72 (0.002)	7.23 (0.117)

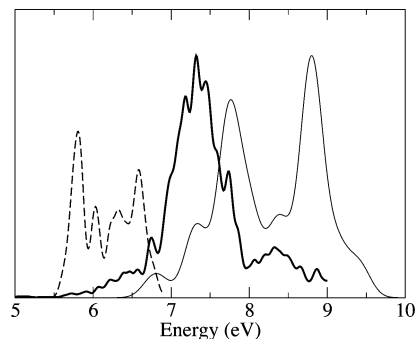
structure. This has been attributed to the structure of the excited-state being nonplanar. Consequently, the nonplanar structures drawn from the simulation will tend to stabilize the excited state relative to the ground-state, resulting in a red shift.<sup>15</sup> However, this red shift is smaller than the typical values reported in earlier work.

Formamic acid is formed by a proton transfer from the  $\text{NH}_2$  group to the  $\text{C}=\text{O}$  group. In the gas phase, the barrier for this process is very large, but it is lowered significantly by hydrogen bonding.<sup>39</sup> Because formamic acid is a possible contaminant, we have also computed its electronic spectrum to determine whether it accounts for remaining deficiencies in the computed spectrum. Table 4 shows the properties of the computed excited states. The calculations show that formamic acid has a number of intense transitions, in particular, for the  $\pi_{nb}\pi^*$  and  $\pi_{nb}3p$  states. The  $\pi_{nb}\pi^*$  transition is blue-shifted by 0.41 eV. A much larger change is observed for the  $1A''$  state, which is calculated to lie at 6.77 eV compared to 5.76 eV for formamide. This state is also labeled  $n\pi^*$ ; however, the lone pair orbital is now associated with the nitrogen atom. Like formamide, there are also a number of additional Rydberg states with significant oscillator strengths. Figure 3 shows the electronic spectrum for formamic acid. The spectrum indicates that formamic acid would make little contribution to the spectrum on the low-energy side of the formamide  $\pi_{nb}\pi^*$  band. However, formamic acid does have intense transitions on the high-energy side of the formamide  $\pi_{nb}\pi^*$  band.

**Table 4.** MRCI/d-aug-cc-pVDZ Excitation Energies,  $\Delta E$  (in eV); Permanent and Transition Dipole Moments,  $\mu$  and  $\mu^T$  (in debye); and Oscillator Strengths,  $\sigma$ , for Formamic Acid<sup>a</sup>

state	$\Delta E$	$\mu_x$	$\mu_y$	$\mu_z$	$ \mu $	$\mu_x^T$	$\mu_y^T$	$\mu_z^T$	$\sigma$	assignment
$1A'$		0.39	-1.38	0.00	1.43					G. S.
$2A'$	7.30	-2.62	0.80	0.00	2.73	0.05	1.17	0.00	0.038	$n3s$
$3A'$	7.74	-1.28	-0.15	0.00	1.29	-1.87	-0.88	0.00	0.125	$\pi_{nb}\pi^*$
$4A'$	8.02	0.41	-1.24	0.00	1.30	-0.96	-0.72	0.00	0.044	$n3p$
$5A'$	8.37	-0.48	4.22	0.00	4.25	0.15	1.14	0.00	0.042	$n3p$
$6A'$	8.80	5.35	-0.96	0.00	5.44	-2.07	-0.48	0.00	0.151	$\pi_{nb}3p$
$7A'$	9.10	0.45	-0.77	0.00	0.89	0.81	-0.45	0.00	0.032	$n3d$
$8A'$	9.20	0.38	-3.76	0.00	3.78	-0.05	0.09	0.00	0.000	$n3d$
$9A'$	9.26	6.66	-2.05	0.00	6.97	-0.28	-0.36	0.00	0.007	$n3d$
$10A'$	9.42	0.97	-1.50	0.00	1.78	0.22	-0.07	0.00	0.020	$\pi_{nb}3d$
$1A''$	6.77	1.61	0.05	0.00	1.61	0.00	0.00	0.70	0.012	$n\pi^*$
$2A''$	6.89	-0.56	-0.20	0.00	0.60	0.00	0.00	-0.49	0.006	$\pi_{nb}3s$
$3A''$	7.38	-0.69	0.20	0.00	0.72	0.00	0.00	0.69	0.013	$\pi_{nb}3p$
$4A''$	7.97	0.16	4.05	0.00	4.05	0.00	0.00	0.69	0.014	$\pi_{nb}3p$
$5A''$	8.41	-0.67	1.04	0.00	1.23	0.00	0.00	-0.58	0.011	$n3p$
$6A''$	8.63	-1.74	-1.16	0.00	2.09	0.00	0.00	0.36	0.004	$\pi_{nb}3d$
$7A''$	8.76	3.50	-4.01	0.00	5.32	0.00	0.00	0.58	0.011	$\pi_{nb}3d$
$8A''$	8.79	-0.60	-5.52	0.00	5.56	0.00	0.00	-0.09	0.000	$\pi_{nb}3d$

<sup>a</sup> The coordinates (in bohr) of the C, N, and O atoms are (0.17, -0.78, 0), (2.09, 0.68, 0), and (-2.21, 0.14, 0), respectively.



**Figure 3.** Theoretical simulations of the electronic absorption spectrum of formamide (bold line), formamic acid (solid line), and ammonia (broken line).

It is interesting to note that the electronic spectrum of formamide in solution does show an intense band above 7.8 eV.<sup>5</sup> This band is not accounted for by condensed-phase ab initio calculations of formamide.<sup>19,20</sup> Formamic acid is more likely to be present in solution and may provide a possible explanation for this band.

Ammonia is another contaminant that is likely to be present. Figure 3 shows a simulated 300 K spectrum for ammonia at low energies. This spectrum is based on MRCI/d-aug-cc-pVDZ on the 100 snapshots drawn from a CP-MD simulation. The spectrum shows that ammonia has an intense transition in the 5.5–7 eV region of the spectrum. This coincides with the low-energy shoulder of the  $\pi_{nb}\pi^*$  band of formamide. In the experiment,<sup>13</sup> the effects of ammonia contamination were removed; however, it is likely that some residual effects remain. Another possible contaminant is dimers of formamide. Calculation of the excited states of a formamide dimer at the MRCI level (even with  $C_{2h}$  symmetry) with a suitable basis set and active space is currently beyond our computational resources.

#### 4. Conclusions

The simulation of the electronic absorption spectrum of formamide at finite temperatures has been investigated using a combination of CP–MD and high-level ab initio quantum chemistry. The excited states of formamide have been computed with MRCI. It is shown that, in conjunction with a sufficiently large basis set, the experimental excitation energies are reproduced accurately. To obtain a thermally broadened spectrum, we have averaged over a large number of single-point MRCI excitation spectra calculated for geometries sampled from a CP–MD simulation. Comparison of the computed spectra with the experimental spectrum at 300 K shows that spectra computed at 500 K yield a much better agreement with experimental results. One possible reason for this may be the absence of nuclear quantum effects in the simulation. Previous studies have shown that nuclear quantum fluctuations may be emulated by CP–MD simulations at higher temperatures.

The accuracy of the computed absorption spectrum has allowed other aspects of the experimental spectrum to be explored. The effects on the spectrum of possible contamination by ammonia and formamidic acid have been studied. Formamidic acid can contribute at high energies and may be important for electronic spectra of formamide measured in solution. Ammonia can also make significant contributions, in particular, to the shoulder region of the experimental formamide spectrum at 6.5 eV. Overall, the calculations demonstrate that electronic absorption spectra at finite temperatures can be simulated accurately from first principles using a combination of CP–MD and quantum chemistry. However, this remains a demanding task because accurate quantum chemistry in conjunction with extensive averaging over molecular structure is required.

**Acknowledgment.** N.A.B. is grateful to the Engineering and Physical Sciences Research Council for funding, in particular for the award of an Advanced Research Fellowship (GR/R77636). N.L.D. gratefully acknowledges funding by Deutsche Forschungsgemeinschaft.

#### References

- (1) Chen, E. F.; Wittung-Stafshede, P.; Kliger, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 3811.
- (2) Rubstov, I. V.; Wang, J.; Hochstrasser, R. M. *J. Chem. Phys.* **2003**, *118*, 7733.
- (3) Besley, N. A.; Hirst, J. D. *J. Am. Chem. Soc.* **1999**, *121*, 9636.
- (4) Basch, H.; Robin, M. B.; Kuebler, N. A. *J. Chem. Phys.* **1967**, *47*, 1201.
- (5) Basch, H.; Robin, M. B.; Kuebler, N. A. *J. Chem. Phys.* **1968**, *49*, 5007.
- (6) Harding, L. B.; Goddard, W. A. *J. Am. Chem. Soc.* **1975**, *97*, 6300.
- (7) Stenkamp, L. Z.; Davidson, E. R. *Theor. Chim. Acta* **1977**, *44*, 405.
- (8) Nitzsche, L. E.; Davidson, E. R. *J. Chem. Phys.* **1978**, *68*, 3103.
- (9) Oliveros, E.; Riviere, M.; Teichtel, C.; Malrieu, P. *Chem. Phys. Lett.* **1978**, *57*, 220.
- (10) Hirst, J. D.; Hirst, D. M.; Brooks, C. L. *J. Phys. Chem.* **1996**, *100*, 13487.
- (11) Serrano-Andrés, L.; Fülischer, M. *J. Am. Chem. Soc.* **1996**, *118*, 12190.
- (12) Szalay, P. G.; Fogarasi, G. *Chem. Phys. Lett.* **1997**, *270*, 406.
- (13) Gingell, J. M.; Mason, N. J.; Zhao, H.; Walker, I. C.; Siggel, M. R. F. *Chem. Phys.* **1997**, *220*, 191.
- (14) Doltsinis, N. L.; Sprik, M. *Chem. Phys. Lett.* **2000**, *330*, 563.
- (15) Besley, N. A.; Oakley, M. T.; Cowan, A. J.; Hirst, J. D. *J. Am. Chem. Soc.* **2004**, *126*, 13502.
- (16) Del Bene, J. E. *J. Chem. Phys.* **1975**, *62*, 1961.
- (17) Sobolewski, A. *Photochem. Photobiol.* **1995**, *89*, 89.
- (18) Krauss, M.; Webb, S. P. *J. Chem. Phys.* **1997**, *107*, 5771.
- (19) Besley, N. A.; Hirst, J. D. *J. Phys. Chem. A* **1998**, *102*, 10791.
- (20) Besley, N. A.; Hirst, J. D. *J. Am. Chem. Soc.* **1999**, *121*, 8559.
- (21) Rocha, R. R.; Martins, V. M.; Coutinho, K.; Canuto, S. *Theor. Chem. Acc.* **2002**, *108*, 31.
- (22) Besley, N. A. *Chem. Phys. Lett.* **2004**, *390*, 124.
- (23) Kumar P.; Marx, D. *Chem. Phys. Phys. Chem.* **2006**, *8*, 573.
- (24) Marx, D.; Parrinello, M. *Z. Phys. D: At., Mol. Clusters* **1997**, *41*, 253.
- (25) CPMD 3.9; MPI für Festkörperforschung: Stuttgart, Germany; IBM Zurich Research Laboratory: Zurich, Switzerland.
- (26) Marx, D.; Hutter, J. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; NIC: Jülich, Germany, 2000. [www.theochem.ruhr-uni-bochum.de/go/cprev.html](http://www.theochem.ruhr-uni-bochum.de/go/cprev.html) (accessed Sept 2006).
- (27) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098.
- (28) Lee, C.; Yang, W.; Parr, R. C. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.
- (29) Troullier, N.; Martins, J. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 1993.
- (30) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511.
- (31) Hoover, W. G. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695.
- (32) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
- (33) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Wiorcikiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
- (34) MOLPRO is a package of ab initio programs written by H.-J. Werner and P. J. Knowles, with contributions from J. Almlöf, R. D. Amos, A. Berning, M. J. O. Deegan, F. Eckert, S. T. Elbert, C. Hampel, R. Lindh, W. Meyer, A. Nicklass, K. Peterson, R. Pitzer, A. J. Stone, P. R. Taylor, M. E. Mura, P. Pulay, M. Schuetz, H. Stoll, T. Thorsteinsson, and D. L. Cooper.
- (35) Knowles, P. J.; Werner, H.-J. *Theor. Chim. Acta* **1992**, *84*, 95.

- (36) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.
- (37) Kendall, R.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
- (38) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 2975.
- (39) Wang, X.-C.; Nichols, J.; Feyereisen, M.; Gutowski, M.; Boatz, J.; Haymet, A. D. J.; Simons, J. *J. Phys. Chem.* **1991**, *95*, 10419.

CT600244Z



## Classical Calculation of Transient Absorption Spectra Monitoring Ultrafast Electron Transfer Processes

Igor Uspenskiy, Birgit Strodel, and Gerhard Stock\*

*Institute of Physical and Theoretical Chemistry, J. W. Goethe University,  
Max-von-Laue-Str. 7, D-60438 Frankfurt, Germany*

Received June 28, 2006

**Abstract:** Classical formulations are considered that allow for the calculation of time- and frequency-resolved pump–probe spectra of nonadiabatically coupled molecular systems. When the semiclassical Franck–Condon approximation in the theoretical framework of the doorway-window formalism is employed, various first- and second-order expressions for the classical doorway and window functions are derived. Moreover, a classical analogue of the electronic dipole transition operator is employed. When established models describing ultrafast photoinduced electron transfer are adopted, it is found that the first-order approximations give rise to spurious structures of the time-resolved signal, which indicate that these approximations fail to correctly account for the averaging effect caused by finite pulses. The higher-order approximations, on the other hand, are shown to give a fairly accurate description of the transient absorption spectrum. By comparing to exact quantum-mechanical calculations, the merits and shortcomings of the various approaches as well as the generally achievable accuracy of a classical modeling of optical spectra is discussed.

### I. Introduction

Significant progress in femtosecond time-resolved spectroscopy has made it possible to observe elementary photochemical reactions in real time.<sup>1</sup> Because the interpretation of these experiments often represents a nontrivial problem, a theoretical modeling of photoinduced processes and their spectroscopy is needed. In principle, the theory of nonlinear spectroscopy is well-developed.<sup>2</sup> In the majority of theoretical studies, however, the description of the chemical dynamics is based on rather simple models, usually a few-level system or a set of harmonic oscillators, which—in the case of condensed-phase spectroscopy—may be coupled to a thermal bath. To facilitate the description of photophysical and photochemical reactions, several groups have constructed ab initio based multidimensional model Hamiltonians and performed time-dependent wave packet calculations.<sup>3</sup> The main bottleneck of this approach is the cumbersome and often impossible task of precomputing the relevant multidimensional potential energy surfaces of the system. To

circumvent this problem, the theoretical description of photochemical reactions in terms of direct ab initio molecular-dynamics simulations has recently become popular.<sup>4–6</sup> Because the nuclear dynamics is treated classically in this approach, it is necessary to develop formulations that allow us to calculate nonlinear time- and frequency-resolved spectra from classical trajectories.<sup>7–17</sup>

In many cases, however, the theoretical description of photoinduced molecular dynamics is complicated by the fact that the underlying Born–Oppenheimer assumption of non-interacting adiabatic potential-energy surfaces may break down.<sup>18–20</sup> Nonadiabatic photoprocesses such as internal conversion, cis–trans photoisomerization, and electron transfer are prime examples. To account for this situation, multidimensional time-dependent wave packet calculations on coupled potential-energy surfaces as well as simulations of the corresponding femtosecond time- and frequency-resolved spectra have been reported.<sup>3,8,20–22</sup> To achieve a classical description of nonadiabatic quantum dynamics, various semiclassical theories as well as mixed quantum-classical schemes have been proposed,<sup>23–26</sup> most notably the “mean-field trajectory” approach and the “surface-hopping”

\* Corresponding author fax: +49-69-798-29709; e-mail: stock@theochem.uni-frankfurt.de.

approach. These methods share the concept that the electronic degrees of freedom are represented by a wave function description, while the nuclear degrees of freedom are represented by a trajectory description. Only a few papers,<sup>27–29</sup> though, have employed these methods to calculate optical spectra of nonadiabatically coupled systems.

In this work, we present several classical approximations to calculate time- and frequency-resolved pump–probe spectra. To this end, we generalize various versions of the so-called semiclassical Franck–Condon approximation<sup>7,8,30</sup> to the case of nonadiabatically coupled potential-energy surfaces. Alternatively, we employ the mapping approach<sup>31</sup> or classical electron analogue model<sup>32</sup> to define the classical analogon of the electronic dipole transition operator.<sup>33</sup> When these formulations are employed, explicit expressions for various cases of electron transfer are derived and compared. As representative examples, we adopt two established models describing ultrafast photoinduced electron transfer,<sup>34,35</sup> for which exact quantum-mechanical reference calculations are possible. Performing a detailed numerical study, the merits and shortcomings of the various approaches as well as the generally achievable accuracy of a classical modeling of optical spectra is discussed.

## II. Quantum-Mechanical Formulation

**A. Model.** As a well-studied model of photoinduced electron transfer,<sup>36</sup> we consider an electronic three-state system, comprising an energetically well-separated electronic ground state  $|\psi_0\rangle$  and two nonadiabatically coupled excited states  $|\psi_1\rangle$  and  $|\psi_2\rangle$  (see Figure 1a). It is assumed that the electron-transfer process takes place between the two excited electronic states after excitation of the system at time  $t = 0$  by an ultrashort laser pulse from  $|\psi_0\rangle$  to the optically bright state  $|\psi_2\rangle$ . When a diabatic electronic representation is adopted, the molecular Hamiltonian of an electron-transfer system can be written as<sup>37</sup>

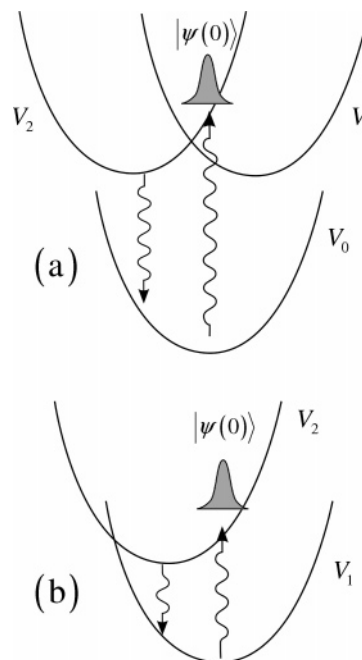
$$H = \sum_{n,m} |\psi_n\rangle h_{nm} \langle \psi_m| \\ = \sum_n |\psi_n\rangle h_n \langle \psi_n| + \{ |\psi_1\rangle V_{12} \langle \psi_2| + \text{h.c.} \} \quad (2.1)$$

comprising the vibrational Hamiltonian  $h_n = T + V_n$  in the diabatic electronic state  $|\psi_n\rangle$  and the off-diagonal diabatic coupling elements  $V_{12}$ . To be specific, we adopt a spin-boson type of model,<sup>38</sup> that is, a constant diabatic coupling  $V_{12} = V_{21} = g$  and harmonic diabatic potentials

$$V_n(x) = E_n + \sum_j \frac{1}{2} \omega_j x_j^2 + \kappa_j^{(n)} x_j \quad (2.2)$$

where  $E_n$  is the vertical excitation energy,  $\omega_j$  denotes the frequency of the  $j$ th vibrational mode, and  $\kappa_j^{(n)}$  represents the gradient of the excited-state potential-energy surface  $V_n$  along this mode. Throughout this paper, we set  $\hbar \equiv 1$  and use dimensionless vibrational coordinates and momenta.

**B. Electron-Transfer Dynamics.** To study the dynamics of the electron-transfer model (eq 2.1), we solve the time-dependent Schrödinger equation to obtain the time-dependent wave function  $|\Psi(t)\rangle$  as well as the density operator  $\rho(t) =$



**Figure 1.** Schematic view of the diabatic potential-energy curves  $V_n(x)$  pertaining to the two electron-transfer models under consideration. At time  $t = 0$ , the system is prepared by an impulsive pump pulse into a nonstationary state  $|\Psi(0)\rangle$ , whose time evolution is interrogated by a time-delayed probe pulse which may induce internal or external electronic transitions. (a) External case, i.e., the radiation field induces transitions from the two coupled excited-state potentials  $V_1$  and  $V_2$  to the electronic ground-state potential  $V_0$ . (b) Internal case, i.e., the radiation field induces transitions between the two coupled electronic states.

$|\Psi(t)\rangle\langle\Psi(t)|$  of the system. As a suitable initial condition, we assume that at time  $t = 0$  the system is impulsively prepared in the excited electronic  $|\psi_2\rangle$ . A key quantity in the discussion of photoinduced electron-transfer processes is the time-dependent population probability of the optically excited diabatic electronic state<sup>20</sup>

$$P_2(t) = \text{Tr}\{|\psi_2\rangle\langle\psi_2|\rho(t)\} \quad (2.3)$$

As  $P_2(t)$  is constant in the absence of nonadiabatic coupling, the diabatic population directly monitors the electron-transfer process of interest. Furthermore, we consider the time-dependent probability distribution along a specific vibrational mode  $x$

$$P_n(x,t) = \text{Tr}\{|\psi_n\rangle\langle x|\langle\psi_n|\rho(t)\} \quad (2.4)$$

in order to illustrate the vibrational motion of the laser-induced wave packet on the coupled potential-energy curves  $V_1$  and  $V_2$ .

So far, the theory has been elaborated using *diabatic* electronic states  $|\psi_n\rangle$ . A diabatic electronic representation is advantageous for the interpretation of spectroscopic data, because in the vicinity of surface-crossings the electronic transition dipole operator is only smooth in the diabatic representation.<sup>3</sup> On the other hand, we may employ a unitary transformation that diagonalizes the diabatic potential matrix, which yields *adiabatic* electronic states  $|\psi_n^{\text{ad}}\rangle$  and the

adiabatic potential-energy surfaces

$$W_{1/2} = \frac{1}{2}(V_1 + V_2) \mp \frac{1}{2}\sqrt{(V_2 - V_1)^2 + 4V_{12}} \quad (2.5)$$

For example, to discuss nonadiabatic relaxation processes such as electron transfer, it appears natural to consider the adiabatic population probability

$$P_2^{\text{ad}}(t) = \text{Tr}\{|\psi_2^{\text{ad}}\rangle\langle\psi_2^{\text{ad}}|\rho(t)\} \quad (2.6)$$

which monitors the decay of the “upper” adiabatic state  $|\psi_2^{\text{ad}}\rangle$  into the lower-lying electronic states.

**C. Transient Absorption.** In typical pump–probe experiments, the molecular system is prepared at time  $t = 0$  by a first laser pulse (the “pump”  $\epsilon_I$ ) into a nonstationary state, whose time evolution is interrogated by a second laser pulse (the “probe”  $\epsilon_{II}$ ) at the delay time  $t_d$ . Let us furthermore assume that the excitation is resonant, thus resulting in a population of the excited state, and that the pump and probe pulses do not overlap, thus facilitating the interpretation of signals. Within the electric dipole approximation, the interaction between the molecular system and the external electric field is then given by

$$H_{\text{int}}(t) = -\hat{\mu}[\epsilon_I(t) + \epsilon_{II}(t)]$$

$$\hat{\mu} = |\psi_0\rangle\langle\mu_{02}| + |\psi_2\rangle\langle\mu_{20}| \quad (2.7)$$

$$\epsilon_i(t) = 1/\sqrt{4\pi\alpha\tau_i^2} e^{-(t-t_i)^2/(4\alpha\tau_i^2)} e^{-i\omega_i(t-t_i)} + \text{c.c.} \quad (2.8)$$

with  $\alpha = 1/(16 \ln 2)$ . The laser pulses are characterized by their carrier frequencies  $\omega_I$  and  $\omega_{II}$  and their durations  $\tau_I$  and  $\tau_{II}$ , respectively. They are centered at times  $t = t_i$ , where  $t_I = 0$  and  $t_{II} = t_d$ , representing the delay time of the probe pulse.

Adopting a doorway–window-type representation,<sup>2,39</sup> the quantum-mechanical pump–probe signal can be written as (see Appendix A)

$$I(t_d) = \text{Tr}\{\mathcal{W}e^{-i\hat{L}t_d}\mathcal{D}\rho_0\} \quad (2.9)$$

where Tr denotes the trace over the electronic and vibrational degrees of freedom and  $\rho_0$  represents the initial density operator of the molecular system prior to the interaction with the laser field. The Liouville operator  $\mathcal{D}$  denotes the doorway operator describing the preparation of the system at time  $t = 0$  by the pump pulse,  $\hat{\mathcal{L}} \cdots = [H, \cdots]$  is the molecular Liouvillian ( $\hbar \equiv 1$ ), accounting for the time evolution of the molecular system during  $0 \leq t \leq t_d$ , and  $\mathcal{W}$  represents the window operator, describing the interaction of the system with the probe laser at  $t = t_d$ .

Depending on the employed laser frequencies and the molecular system under consideration, several spectroscopic processes may contribute to the transient absorption signal (eq 2.9): the photoinduced excited-state population gives rise to stimulated emission and excited-state absorption, while the reduced population in the electronic ground state causes a bleach of the absorption band as well as impulsive stimulated Raman scattering.<sup>2,40</sup> In this work, we are particularly interested in the excited-state contribution to the transient absorption. In the limit of ultrashort probe pulses,

the excited-state pump–probe signal measures the diabatic population probability  $P_2(t)$  defined in eq 2.3<sup>20</sup> and therefore directly monitors the electron-transfer process of interest.

### III. Classical Description of Nonadiabatic Dynamics

The classical description of the nonadiabatic dynamics has been the subject of numerous articles<sup>23,24,26</sup> and is only briefly reviewed here for further reference. The starting point of most mixed quantum-classical schemes such as the mean-field trajectory and the surface-hopping methods is to expand the total wave function in electronic states  $|\psi_n\rangle$

$$|\Psi(t)\rangle = \sum_n d_n(\mathbf{x}, t) |\psi_n\rangle \quad (3.1)$$

where  $d_n(\mathbf{x}, t)$  denotes the vibrational wave function pertaining to the electronic state  $|\psi_n\rangle$ . Employing Hamiltonian eq 2.1, we obtain for the time-dependent Schrödinger equation

$$i\dot{d}_n = \sum_m h_{nm}(\mathbf{x}) d_m \quad (3.2)$$

To introduce the classical-path approximation, we assume that the nuclear dynamics of the system can be described by classical trajectories; that is, the position operator  $\hat{\mathbf{x}}$  is approximated by its mean value, that is, the trajectory  $\mathbf{x}(t)$ . As a consequence, the quantum-mechanical operators of the nuclear dynamics [e.g.,  $h_{nm}(\mathbf{x})$ ] become classical functions which depend parametrically on  $\mathbf{x}(t)$ . In the same way, the nuclear wave functions  $d_n(\mathbf{x}, t)$  become complex-valued coefficients  $d_n[\mathbf{x}(t), t]$ . As the electronic dynamics are evaluated along the classical path of the nuclei, the approximation thus accounts for the reaction of the quantum degrees of freedom to the dynamics of the classical degrees of freedom.

The back-reaction of the classical degrees of freedom to the dynamics of the quantum degrees of freedom may be described either self-consistently (in the mean-field trajectory method) or via a hopping algorithm (in the surface-hopping method). In the former, the classical force  $F = \dot{p}_j$  acting on the nuclear degrees of freedom  $x_j$  is given as an average over the quantum degrees of freedom

$$\dot{p}_j = -\left\langle \Psi(t) \left| \frac{\partial H}{\partial x_j} \right| \Psi(t) \right\rangle \quad (3.3)$$

In the surface-hopping approach, on the other hand, the coupling of quantum and classical degrees of freedom is realized via instantaneous hops between coupled potential-energy surfaces, while the trajectories always propagate on a single adiabatic surface.<sup>23</sup>

Because in mixed quantum-classical methods electronic and nuclear degrees of freedom are treated on a different dynamical footing, these theories do not necessarily provide a satisfying classical picture of nonadiabatic dynamics. As an alternative approach to incorporate quantum degrees of freedom in a classical formulation, it has recently been proposed to utilize quantum-mechanical bosonization techniques, that is, to represent discrete electronic states by continuous harmonic oscillators, which possess a well-defined classical limit.<sup>31,41</sup> This is achieved by the mapping

relations

$$|\psi_n\rangle\langle\psi_m| \rightarrow \frac{1}{2}(X_n - iP_n)(X_m + iP_m) \quad (3.4)$$

$$|\psi_n\rangle \rightarrow |0_1, \dots, 1_n, \dots, 0_N\rangle \quad (3.5)$$

where  $X_n$  and  $P_n$  are position and momentum operators of the  $n$ th oscillator with commutation relations  $[X_n, P_m] = i\delta_{nm}$ , and  $|0_1, \dots, 1_n, \dots, 0_N\rangle$  denotes a harmonic-oscillator eigenstate with a single quantum excitation in the mode  $n$ .<sup>42</sup> Inserting eq 3.4 into eq 2.1, we obtain the boson representation of the electron-transfer system

$$H = \frac{1}{2} \sum_{n,m} h_{nm} (X_n X_m + P_n P_m - \delta_{nm}) \quad (3.6)$$

As the mapping Hamiltonian (eq 3.6) contains only continuous operators, the quantum-mechanical system has a well-defined classical analogue. The transition to classical mechanics is performed by changing from the Heisenberg operators  $y_k(t)$  ( $y_k = X_n, P_n, x_j, p_j$ ) obeying Heisenberg's equations of motion ( $i\dot{y}_k = [y_k, H]$ ) to the corresponding classical functions obeying Hamilton's equations (e.g.,  $\dot{X}_k = \partial H / \partial P_k$ ). In this classical limit, the formalism can be shown to recover the classical electron analogue model of Meyer and Miller.<sup>32</sup> To make contact with the mean-field trajectory method, the real-valued electronic variables  $X_n$  and  $P_n$  may be replaced by the complex-valued variables  $d_n = (X_n + iP_n)/\sqrt{2}$ , which have a similar meaning to the electronic coefficients in eq 3.1.<sup>43</sup> Although the equations of motion of both formulations are similar, the mapping formulation treats electronic and nuclear degrees of freedom on the same dynamical footing and is therefore superior to the mean-field trajectory method. The mapping formalism has been employed to facilitate the treatment of nonadiabatic dynamics in various theoretical approaches.<sup>41,44–49</sup> Furthermore, the formulation has been applied to a variety of systems with nonadiabatic dynamics,<sup>43,50–53</sup> including the description of photoinduced electron-transfer and internal-conversion processes.

To obtain the classical expectation value of an observable, say, the time-dependent diabatic population probability  $P_2(t)$  defined in eq 2.3, the quantum-mechanical trace is replaced by a phase-space average

$$P_2(t) = \int d\Gamma_0 \rho(\Gamma_0) |d_2(t)|^2 \quad (3.7)$$

where  $\Gamma_0$  denotes the trajectory  $\Gamma_t = \{x_j(t), p_j(t), d_n(t)\}$  at  $t = 0$  and  $\rho(\Gamma_0)$  represents a phase-space distribution function describing the quantum-mechanical initial state of the system (see section 5.1 for details). To classically calculate adiabatic quantities, for example, the adiabatic population probability defined in eq 2.6, we use

$$P_2^{\text{ad}}(t) = \int d\Gamma_0 \rho(\Gamma_0) |a_2(t)|^2 \quad (3.8)$$

where  $a_n(t)$  are the electronic coefficients that are obtained when the total wave function is expanded in adiabatic basis states, that is,  $|\Psi(t)\rangle = \sum_n a_n(t) |\psi_n^{\text{ad}}\rangle$ . By exploiting the unitary transformation between the diabatic and the adiabatic

representation, the adiabatic coefficients  $a_n(t)$  are readily obtained from the diabatic coefficients  $d_n(t)$  and vice versa.<sup>32,54</sup>

#### IV. Classical Calculation of Electronic Spectra

In principle, we may use any of the above three methods (mean-field, surface hopping, and mapping) for a classical calculation of electronic spectra. In all formulations, the nonadiabatic dynamics of the molecular system are described by complex-valued variables  $d_n(t)$  [or  $a_n(t)$ ] for the electronic states and by the trajectory  $\mathbf{x}(t)$  for the nuclear degrees of freedom. As derived in Appendix B, the classical limit of the pump–probe signal (eq 2.9) can be written in the form

$$I(t_d) = \int d\Gamma_0 W(t_d) D(0) \rho(\Gamma_0) \quad (4.1)$$

where—in direct analogy to the quantum-mechanical expression— $D(0)$  and  $W(t_d)$  are phase-space functions describing the pump and the probe processes. To calculate these functions, we introduce four classical approximations. The details of the derivations are given in Appendix B. The first three are based on the semiclassical Franck–Condon approximation,<sup>7,8,30</sup> which assumes that the Condon approximation holds (i.e., that  $\mu_{nm}$  is coordinate-independent) and that we may employ a short-time approximation of the type  $e^{-i\hbar n t} e^{i\hbar n t} \approx e^{-i(V_m - V_n)t}$ . The last formulation exploits the mapping transformation to directly calculate the dipole correlation function.<sup>33</sup>

**A. First-Order Franck–Condon Approximation.** The simplest way to generalize the standard Franck–Condon approximation to the case of vibronically coupled systems is to neglect the nonadiabatic coupling  $V_{12}$  during the duration of the short laser pulses. Doing so, we obtain for the classical doorway and window functions

$$D(0) = \frac{\mu^2}{4} e^{-\alpha \tau_1^2 \{\omega_1 - [V_2(0) - V_0(0)]\}^2} \quad (4.2)$$

$$W(t_d) = \frac{\mu^2}{4} e^{-\alpha \tau_1^2 \{\omega_{11} - [V_2(t_d) - V_0(t_d)]\}^2} |d_2(t_d)|^2 \quad (4.3)$$

Equations 4.1–4.3 offer a simple classical interpretation of time-resolved pump–probe spectroscopy, which becomes clear by outlining their numerical implementation: (i) Generate an ensemble of trajectories that represents the initial ground-state phase-space distribution  $\rho(\Gamma_0)$  in eq 4.1. (ii) Because of the excitation of the system by a finite pump pulse at time  $t = 0$ , these trajectories are lifted to the optically allowed excited electronic state and are weighted in eq 4.2 by the resonance condition  $\omega_1 = V_2(0) - V_0(0)$ . (iii) Subsequently, the trajectories are propagated on the coupled excited electronic states, using some nonadiabatic classical method. (iv) At time  $t = t_d$ , the contribution of each trajectory to the stimulated emission signal is given in eq 4.3 by the Gaussian factor describing time-dependent resonance condition  $\omega_{11} = V_2[x(t_d)] - V_1[x(t_d)]$  multiplied with electronic population probability  $|d_2(t_d)|^2$  of the bright electronic state.

Equations 4.1–4.3 represent the simplest way to classically calculate pump–probe spectra. Obviously, it hardly requires

additional effort in the trajectory calculations. The nonadiabatic electron-transfer dynamics of the system enter through the time-dependent electronic population probability  $|d_2(t_d)|^2$ . In the absence of nonadiabatic coupling, the latter factor is constant and the above result becomes equivalent to previous results.<sup>7–16</sup> Because the underlying Franck–Condon approximation employs a short-time expansion of the propagator to the first order, the result will be referred to as the FC1 approximation.

### B. Second-Order Franck–Condon Approximation.

Using a short-time expansion that is valid in the second order,<sup>55</sup> the classical doorway and window functions can be derived as

$$D(0) = \frac{\mu^2}{4} \int_{-\infty}^{\infty} dt e^{-t^2/(2\alpha\tau_1^2)} e^{-\alpha\tau_1^2\{\omega_1-[V_2(t)-V_0(t)]\}^2} \quad (4.4)$$

$W(t_d) =$

$$\frac{\mu^2}{4} \int_{-\infty}^{\infty} dt e^{-(t-t_d)^2/(2\alpha\tau_1^2)} e^{-\alpha\tau_1^2\{\omega_1-[V_2(t)-V_0(t)]\}^2} |d_2(t)|^2 \quad (4.5)$$

Similar to the first-order result, the second-order Franck–Condon (FC2) approximation accounts for the electron-transfer dynamics through the electronic population probability  $|d_2(t_d)|^2$ . In the absence of nonadiabatic coupling, eq 4.5 recovers the standard second-order results.<sup>55</sup> Interestingly, it is also equivalent to a semiclassical approximation of the Liouville von Neumann equation,<sup>13,17</sup> which has recently been generalized to the nonadiabatic case.<sup>29</sup> Besides the resonance conditions  $\omega_i = V_2 - V_0$ , the FC2 approximation also accounts for the time window  $|\epsilon_i(t)|^2$  created by the laser pulses. Hence, the uncertainty relation between the time resolution (with respect to the delay time  $t_d$ ) and the frequency resolution (with respect to the carrier frequencies  $\omega_i$ ) is correctly reproduced in the second order (but not by the first-order approximations 4.2 and 4.3).

The price to pay for this improvement is that we need to propagate classical trajectories during the interaction with the laser pulses. In the numerical implementation of the FC2 approximation, this raises the question whether the trajectories are propagated on the ground-state or excited-state potentials (in the quantum-mechanical formulation, the system propagates during interaction with the electric field in both electronic states according to the dipole operator  $e^{-iH_m}\mu_{mm}e^{iH_m}$ ). The usual solution to this well-known ambiguity of the classical approximation is to choose the ground-state potential for the pump process and the excited-state potential for the probe process.<sup>2</sup>

While the latter works well for the calculation of the window function, the calculation of the doorway function requires some additional effort. This is because we also need to generate an appropriate excited-state phase-space distribution from the initial ground-state distribution  $\rho(\Gamma_0)$ . (Note that the FC1 approximation assumes that these two distributions are the same.) Converting the time integral in eq 4.4 into a sum

$$D(0) \propto \sum_j e^{-t_j^2/(2\alpha\tau_1^2)} e^{-\alpha\tau_1^2\{\omega_1-[V_2(t_j)-V_0(t_j)]\}^2} \quad (4.6)$$

we may consider the  $t_j$ 's as “hopping times” at which the trajectories are lifted from the ground to the excited electronic state. This suggests the following scheme for the calculation of the doorway function: (i) For every trajectory of the initial ground-state phase-space distribution  $\rho(\Gamma_0)$ , we randomly select a hopping time  $t_j$  from their Gaussian distribution. (ii) For  $t < t_j$ , the trajectory is propagated on the ground-state potential. (iii) At  $t = t_j$ , the trajectory is lifted to the excited electronic state and weighted by the Gaussian resonance factor. For  $t > t_j$ , the trajectory is propagated on the excited-state potential.

**C. Nonadiabatic Franck–Condon Approximation.** Taking the vibronic coupling during the interaction with the laser pulses explicitly into account, Diltthey et al.<sup>56</sup> derived a generalization of the first-order Franck–Condon approximation to nonadiabatic systems. It leads to the classical doorway and window functions

$$D(0) = \frac{\mu^2}{4} \sum_{n=1,2} S_{2n}^2(0) e^{-2\alpha\tau_1^2\{\omega_1-[W_n(0)-V_0(0)]\}^2}$$

$$W(t_d) = \frac{\mu^2}{4} \sum_{n=1,2} S_{2n}^2(t_d) e^{-2\alpha\tau_1^2\{\omega_1-[W_n(t_d)-V_0(t_d)]\}^2} |a_n(t_d)|^2 \quad (4.7)$$

where the matrix  $\mathbf{S}$  with elements

$$S_{2n}(t) = \sqrt{\frac{V_2(t) - V_1(t)}{2\sqrt{[V_2(t) - V_1(t)]^2 + V_{12}^2(t)} + \frac{(-1)^n}{2}} \quad (4.8)$$

represents the unitary transformation between the diabatic and adiabatic representations.<sup>37</sup> In contrast to the approximations FC1 and FC2, which describe transitions between diabatic potentials, the Franck–Condon approximation including vibronic coupling (hereafter referred to as FCC) exhibits the adiabatic resonance conditions  $\omega = W_n - V_0$ . They account for transitions from both coupled adiabatic electronic potentials  $W_n$  to the electronic ground state  $V_0$ , which are weighted by its population probability  $|a_n(t_d)|^2$ . A further coordinate dependency arises from the prefactor  $S_{2n}^2$ , which represents the relation between the diabatic and the adiabatic energy gap. With respect to the short-time approximation of the propagator, eq 4.7 is a first-order expression and therefore suffers from the same shortcomings as eq 4.3.

**D. Classical Dipole Function.** Apart from the classical evaluation of the various Franck–Condon approximations, the mapping formulation provides an alternative approach to calculate the electronic dipole correlation function. The basic idea is to apply the mapping relation 3.4 directly to the electronic transition dipole operator, that is,

$$\begin{aligned} \hat{\mu}_{02}(t) &= e^{iHt} |\psi_0\rangle \mu_{02} \langle \psi_2 | e^{-iHt} \\ &\rightarrow \mu_{02} \frac{1}{2} [X_0(t) - iP_0(t)][X_2(t) + iP_2(t)] \equiv \mu_{02}(t) \end{aligned} \quad (4.9)$$

This leads to the classical doorway and window functions

$$\begin{aligned} D(0) &= \left| \int_{-\infty}^{\infty} dt \epsilon_1(t) \mu_{20}(t) \right|^2 \\ W(t_d) &= \left| \int_{-\infty}^{\infty} dt \epsilon_2(t) \mu_{02}(t) \right|^2 \end{aligned} \quad (4.10)$$

As this expression was first introduced within the theoretical framework of the classical electron analogue model,<sup>27,33</sup> eq 4.10 will be referred to as the CEA approximation. Compared to approximations FC1, FC2, and FCC, which employ a classical method (e.g., mean-field, surface hopping, and mapping) to describe nonadiabatic dynamics *and* some type of Franck–Condon approximation, the CEA result (eq 4.10) only requires a single approximation, that is, the classical limit of the mapping relations (eq 3.4). A closer analysis shows that the structures of the CEA result (eq 4.10) and the FC2 result (eq 4.5) are quite similar in that both expressions are weighted by the diabatic population probability  $|d_2|^2$  and account for the time window created by the laser pulse. In detail, however, the two expressions differ significantly.

**E. Internal Transitions.** In the optical spectroscopy of nonadiabatically coupled electronic states, two qualitatively different cases may occur: intramolecular and radiative couplings pertain to (a) *different* electronic transitions and (b) to the *same* electronic transitions. The two different situations are illustrated in Figure 1, which schematically shows two electron-transfer models with coupled electronic states  $|\psi_1\rangle$  and  $|\psi_2\rangle$ . So far, we have been concerned with situation a, also referred to as the “external transition” case. In situation b, referred to as the “internal transition” case, we have the transition dipole operator

$$\hat{\mu} = |\psi_1\rangle\mu_{12}\langle\psi_2| + |\psi_2\rangle\mu_{21}\langle\psi_1| \quad (4.11)$$

which connects the coupled diabatic electronic states  $|\psi_1\rangle$  and  $|\psi_2\rangle$  (instead of  $|\psi_0\rangle$  and  $|\psi_2\rangle$  as in the external case).

Performing the derivations of the various classical approximations of the pump–probe signal for the internal case, there are only small changes for the Franck–Condon approximations FC1 and FC2. Because of the dipole operator in eq 4.11, the resonance factors in approximations 4.3 and 4.5 change to  $\omega - (V_2 - V_1)$ . Moreover, because both electronic states  $|\psi_1\rangle$  and  $|\psi_2\rangle$  are populated because of the vibronic coupling, in the internal case, the classical doorway and window functions describes both emission and absorption between these states. In the FC2 approximation, for example, we obtain

$$D(0) = \frac{\mu^2}{4} \int_{-\infty}^{\infty} dt e^{-t^2/(2\alpha\tau_p^2)} e^{-\alpha\tau_p^2\{\omega_1-[V_2(t)-V_1(t)]\}^2} (|d_2(t)|^2 - |d_1(t)|^2)$$

$$W(t_d) = \frac{\mu^2}{4} \int_{-\infty}^{\infty} dt e^{-(t-t_d)^2/(2\alpha\tau_p^2)} e^{-\alpha\tau_p^2\{\omega_1-[V_2(t)-V_1(t)]\}^2} (|d_2(t)|^2 - |d_1(t)|^2) \quad (4.12)$$

and a similar result for the FC1 approximation.

Because the FCC approximation (eq 4.7) takes the nonadiabatic interaction during the radiative transitions explicitly into account, the results for the classical doorway and window functions differ in the external and internal cases.<sup>56</sup> In the latter, we obtain

$$D(0) = \frac{\mu^2}{4} \left[ S_{22}^2(0) - \frac{1}{2} \right] e^{-2\alpha\tau_p^2\{\omega_1-[W_2(0)-W_1(0)]\}^2}$$

$$W(t_d) = \frac{\mu^2}{4} \left[ S_{22}^2(t_d) - \frac{1}{2} \right] e^{-2\alpha\tau_p^2\{\omega_1-[W_2(t_d)-W_1(t_d)]\}^2} (|a_2(t_d)|^2 - |a_1(t_d)|^2) \quad (4.13)$$

In the CEA approximation for the internal case, we obtain the classical dipole function  $\mu_{12}(t)$  instead of  $\mu_{02}(t)$  in eq 4.10. As shown in the Appendix, this gives rise to a population term  $|d_2(t)|^2|d_1(t)|^2$  instead of the population difference obtained in eq 4.12. In practice, however, it has proven advantageous to consider emission and absorption separately, see eq B7.

**F. Ground-State Contribution.** Besides the excited-state contribution discussed so far, the transient absorption spectrum also contains contributions from the electronic ground state. On one hand, the reduced population in the electronic ground state causes a bleach of the absorption band, which is obtained by calculating the linear absorption of the probe pulse (e.g., at the FC1 level, the bleach signal is  $\propto e^{-\alpha\tau_p^2\{\omega_1-[V_2(0)-V_0(0)]\}^2}$ ). On the other hand, finite pump pulses give rise to stimulated impulsive Raman scattering.<sup>2,40</sup> As shown in Appendix A, this ground-state contribution can also be cast in doorway–window expression 2.9 and contains the Liouville operators given in eq A6. A simple classical approximation of the latter can be derived as follows.

Let us first consider the first-order short-time approximation to the ground-state doorway operator

$$\mathcal{D}_g\rho_0 = \mu^2 \int_{-\infty}^{\infty} dt_2 \epsilon_1(t_2) e^{-i(V_2-V_0)t_2} \times \int_{-\infty}^{t_2} dt_1 \epsilon_1(t_1) e^{i(V_2-V_0)t_1} \langle\psi_0|\rho_0|\psi_0\rangle |\psi_0\rangle\langle\psi_0| \quad (4.14)$$

where the electronic projector states that the molecular system evolves in the electronic ground state after the interaction with the pump field. Although the expression formally looks quite similar to the corresponding doorway operator for the excited-state contribution, its further evaluation is more cumbersome because of the time ordering of the integrals. To obtain a simple and computationally convenient classical expression for the ground-state contribution, we choose to simply ignore the time ordering (i.e., we shift the upper integral limit from  $t_2$  to  $\infty$ ). Then, the classical evaluation is straightforward and yields the FC1 approximation  $D_g(0) = \mu^2/4 e^{-\alpha\tau_p^2\{\omega_1-[V_2(0)-V_0(0)]\}^2}$ . Apart from the fact that the trajectories evolve in the electronic ground state after the interaction with the pump field, the latter is identical to the excited-state expression (eq 4.2). Ignoring the integral time ordering in the second-order expansion (eq B2), we obtain the FC2 approximation of the ground-state contribution

$$D_g(0) = \frac{\mu^2}{4} \int_{-\infty}^{\infty} dt e^{-t^2/(2\alpha\tau_p^2)} e^{-\alpha\tau_p^2\{\omega_1-[V_2(t)-V_0(t)]\}^2} \quad (4.15)$$

$$W_g(t_d) = \frac{\mu^2}{4} \int_{-\infty}^{\infty} dt e^{-(t-t_d)^2/(2\alpha\tau_p^2)} e^{-\alpha\tau_p^2\{\omega_1-[V_2(t)-V_0(t)]\}^2} |d_0(t)|^2 \quad (4.16)$$

which again looks quite similar to the corresponding excited-state expressions 4.4 and 4.5. Note that the molecular system

evolves in the electronic ground state between the interaction with two laser pulses, while it evolves in the excited electronic state during the pulses.

The ground-state doorway function (eq 4.15) can be evaluated using the following scheme. (i) For every trajectory of the initial ground-state phase-space distribution  $\rho(\Gamma_0)$ , we randomly select two hopping times  $t_a$  and  $t_b$  from the Gaussian distribution  $e^{-t^2/(2\alpha\tau_p^2)}$ , for which we require that  $t_a < t_b$ . (ii) For  $t < t_a$ , the trajectory is propagated on the ground-state potential. (iii) When the trajectory to the excited electronic state at  $t = t_a$  is lifted, the trajectory is weighted by the Gaussian resonance factor and is propagated on the excited-state potential for  $t_a < t < t_b$ . (iv) At  $t = t_b$ , the trajectory is moved back to the ground-state potential and propagates in the electronic ground state until the arrival of the probe laser field. Then, the ground-state window function (eq 4.16) is evaluated in a similar manner.

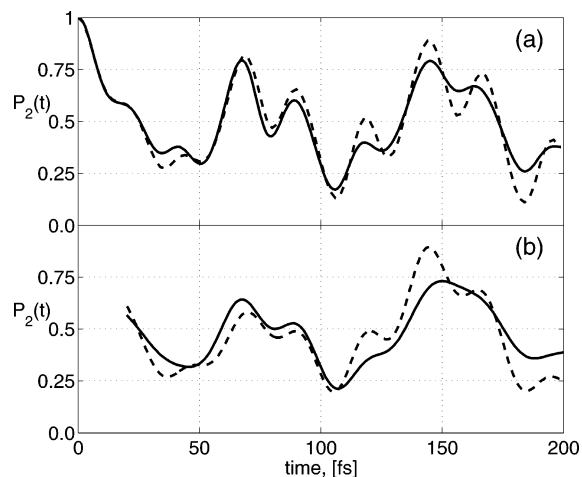
## V. Computational Results

**A. Computational Details.** The computational methods employed for the quantum-mechanical wave packet propagations have been described in detail in ref 20. In short, the time-dependent Schrödinger equation is converted into a numerically tractable problem by expanding the state vector in a direct-product basis constructed from diabatic electronic states and harmonic-oscillator states for the vibrational degrees of freedom. This results in a system of coupled first-order differential equations, which are solved using a Runge–Kutta–Merson scheme with an adaptive step size. The quantum-mechanical transient absorption spectrum (eq 2.9) has been calculated (numerically) exactly by using a nonperturbative evaluation of the electronic polarization.<sup>20,57</sup>

The classical mapping method and its numerical implementation have been described in refs 26 and 41. Basically, Hamilton's equations pertaining to the classical Hamiltonian (eq 3.6) are solved for the electronic ( $X_n, P_n$ ) and nuclear ( $x_j, p_j$ ) degrees of freedom, using again a Runge–Kutta–Merson integrator. To calculate quasiclassical averages  $\int d\Gamma \rho(\Gamma) \dots$  (as, e.g., in eq 3.7), the quantum-mechanical initial state  $\rho_0 = \rho_{el}\rho_{vib}$  is to be represented by a suitable phase-space distribution  $\rho(\Gamma)$ . In this work, the vibrational system was assumed to be in its harmonic ground state,  $\rho_{vib} = |0\rangle\langle 0|$ , which was represented by a Wigner distribution.<sup>8</sup> Within the mapping formalism, the electronic initial state  $\rho_{el} = |\psi_0\rangle\langle\psi_0|$  is mapped onto the harmonic-oscillator eigenstate  $|1_0 0_1 0_2\rangle$  containing a single quantum excitation in the first mode [see eq 3.5]. As is well-known, the Wigner distribution of the latter state is not positive-definite and therefore is problematic in a quasiclassical implementation. We therefore change to classical action-angle variables  $\{N_n, Q_n\}$  using the transformations  $X_n = \sqrt{2N_n+1} \sin Q_n$  and  $P_n = \sqrt{2N_n+1} \cos Q_n$ , sample the angles  $Q_n$  from  $[0, 2\pi]$ , and fix the actions  $N_n$  to represent the quantum number of the oscillator.<sup>26</sup> This yields

$$\rho_{el}(N_1, N_2) = \delta\left(N_1 - \frac{1}{2}\right) \delta\left(N_2 - \frac{3}{2}\right) \quad (5.1)$$

where the factors  $1/2$  reflects the zero-point energy of the harmonic oscillators. Converged quasiclassical averages for



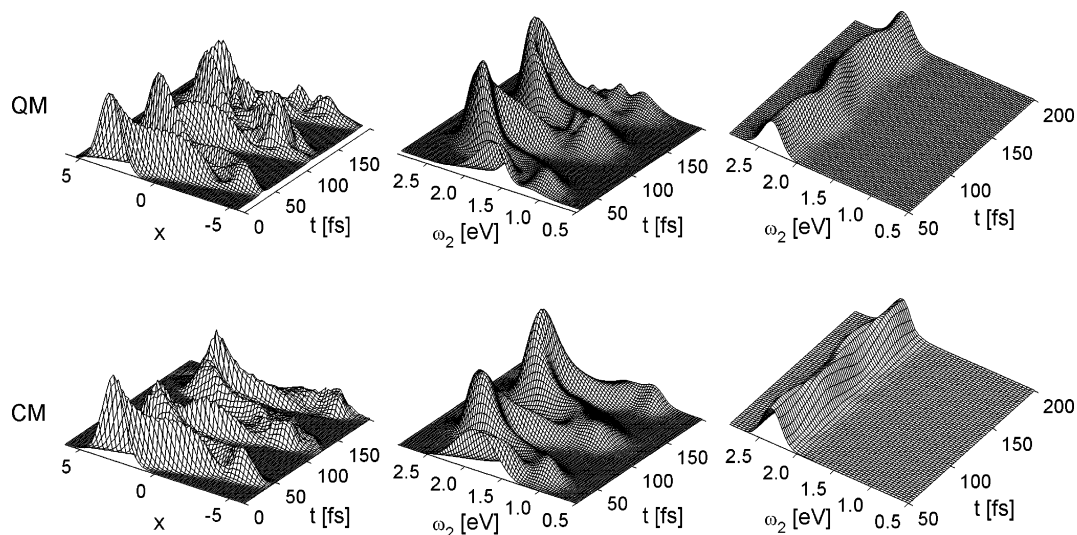
**Figure 2.** Time-dependent population  $P_2(t)$  of the optically excited diabatic electronic state, as obtained for the one-mode electron-transfer model. Compared are classical mapping results (solid lines, eq 3.7) and quantum-mechanical reference calculations (dashed line, eq 2.3) for the cases of (a) impulsive excitation and (b) excitation by a Gaussian pump pulse of 20 fs duration (shown only for  $t > 25$  fs, i.e., after the pump pulse).

the time-resolved signals (e.g., as in Figure 2) and for time- and frequency-resolved spectra (e.g., as in Figure 3) are obtained by running  $10^4$  trajectories.

**B. One-Mode Three-State Model.** To consider the case of external radiative transitions, we adopt the one-mode three-state model proposed in ref 45, which comprises an energetically well-separated electronic ground state  $|\psi_0\rangle$  and two nonadiabatically coupled excited states  $|\psi_1\rangle$  and  $|\psi_2\rangle$ . The parameters of the model are  $E_0 = -1.775$  eV,  $E_1 = E_2 = 0$ ,  $g = 0.05$  eV,  $\omega = g$ ,  $\kappa^{(0)} = -3g$ , and  $\kappa^{(2)} = -\kappa^{(1)} = g/2$ . The diabatic potential-energy curves  $V_n(x)$  of the model are shown in Figure 1a.

In a first step, it is instructive to study the quantum-mechanical electron-transfer dynamics of the system. To this end, Figure 2a shows the population probability  $P_2(t)$  (see eq 2.3) of the diabatic electronic state  $|\psi_2\rangle$ . As schematically indicated in Figure 1a, it is assumed that at time  $t = 0$  this state is impulsively excited by an ultrashort laser pulse. Following an ultrafast initial decay, the quantum-mechanical result for  $P_2(t)$  is seen to exhibit quasiperiodic oscillations with periods of  $\approx 24$  and 70 fs. A closer analysis shows that these dynamics are caused by the two shortest vibronic periodic orbits of the system.<sup>34,45</sup>

To further illustrate the motion of the laser-induced wave packet on the coupled potential-energy curves  $V_1$  and  $V_2$ , Figure 3 shows the time-dependent probability distribution  $P_2(x,t)$  defined in eq 2.4. As a consequence of the impulsive  $|\psi_0\rangle \rightarrow |\psi_2\rangle$  excitation, the wave function at time  $t = 0$  is a Gaussian centered at  $x_0 = 3$ . With increasing time, the wave packet is seen to undergo oscillations along  $x$  with a period of  $\approx 70$  fs, which roughly corresponds to the vibrational frequency  $\omega$  of the model. Note that the nuclear motion is directly linked to an oscillation of the electronic population; that is, the vibrational dynamics trigger electronic transitions between the two coupled states  $|\psi_1\rangle$  and  $|\psi_2\rangle$ .<sup>58</sup>

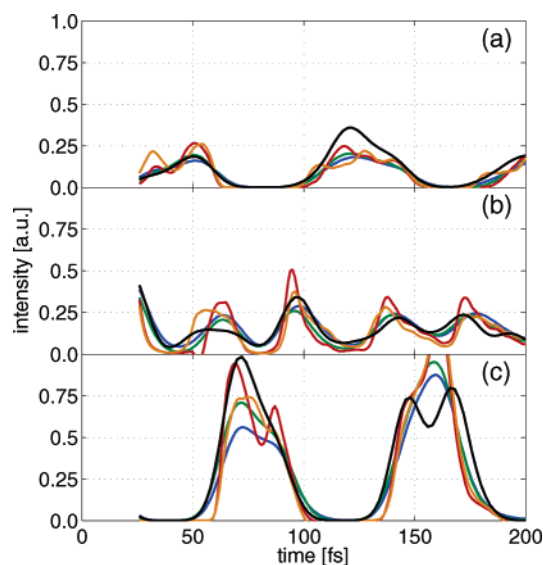


**Figure 3.** Comparison of the quantum-mechanical (QM, upper panels) and classical (CM, lower panels) electron-transfer dynamics and spectroscopy of the one-mode model. Shown are (left) the time-dependent nonadiabatic wave packet motion as described by the probability density  $P_2(x,t)$  in eq 2.4 as well as the excited-state (stimulated emission, middle) and the ground-state (impulsive Raman scattering, right) contributions to the transient absorption spectrum. The classical spectra are calculated using the FC2 approximation in eqs 4.5 and 4.16.

We now turn to the classical mapping calculation for the one-mode three-state model. As discussed before,<sup>58</sup> the nonadiabatic quantum dynamics of the system are nicely reproduced by the classical formulation. While we find almost quantitative agreement of the classical and quantum-mechanical diabatic populations in Figure 2a, the comparison of the corresponding time-dependent probability distributions  $P_2(x,t)$  in Figure 3 reveals the limits of the classical approximation. Although the mapping calculation catches the main features of the nonadiabatic wave packet motion, it is seen to miss the finer details and substructures at longer times.

It is interesting to study to what extent the electron-transfer dynamics are reflected in the excited-state contribution to the transient absorption spectrum. Assuming impulsive excitation and probe pulses of 20 fs duration, Figure 3 shows the quantum-mechanical pump–probe signal as a function of the delay time  $t_d$  and the probe carrier frequency  $\omega_{\text{II}}$ . As discussed by several authors,<sup>9,40,59</sup> the excited-state pump–probe signal nicely maps the time evolution of the quantum-mechanical wave function. To obtain a first impression on the overall performance of a classical description, Figure 3 also shows the corresponding transient absorption signal as obtained from the second-order Franck–Condon approximation (eq 4.5). As in the case of the wave packet dynamics discussed above, the classical transient absorption spectrum is in excellent agreement with the quantum reference calculation for times up to 150 fs. Only at longer times, the classical description deteriorates and fails to resolve the finer details of the spectrum.

We are now in a position to study the accuracy of the various classical approximations to the transient absorption introduced above, that is, the Franck–Condon approximation to first-order (FC1), second-order (FC2), and vibronic coupling (FCC), as well as the classical electron analogue model (CEA) approximation. To this end, Figure 4 shows cuts of the above-discussed transient absorption spectrum



**Figure 4.** Cuts of the excited-state contribution to the transient absorption spectrum at probe carrier frequencies (a) 1.25 eV, (b) 1.75 eV, and (c) 2.25 eV, obtained for the one-mode electron-transfer model. The classical approximations FC1 (red lines, eq 4.3), FC2 (green lines, eq 4.5), FCC (orange lines, eq 4.7), and CEA (blue lines, eq 4.10) are compared to exact quantum calculations (black lines). at probe carrier frequencies  $\omega_{\text{II}} =$  (a) 1.25 eV, (b) 1.75 eV, and (c) 2.25 eV, corresponding to the low-, middle- and high-frequency regions, respectively, of the spectrum shown in Figure 3. While all classical approximations account for the time evolution of the spectrum at least qualitatively, clear differences are observed. First, we note that the two first-order approximations FC1 and FCC may give rise to spurious structures of the time-resolved signal, which indicate that these approximations fail to correctly account for the averaging effect caused by finite pulses. Although the FCC description clearly represents an improvement over the simple FC1 approximation, its higher computational effort



compared to the FC1 and FC2 calculations renders this method less attractive.

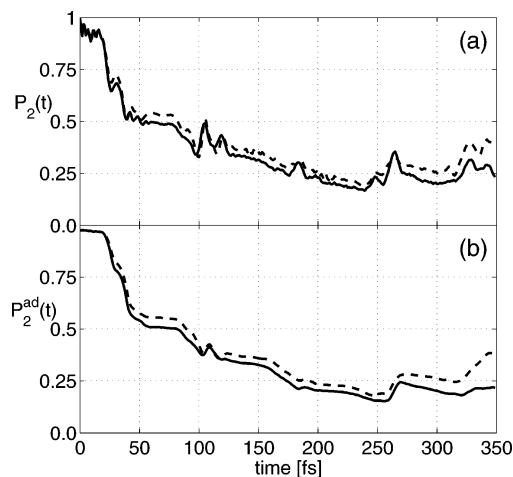
The FC2 (eq 4.5) and the CEA (eq 4.10) classical descriptions, on the other hand, are clearly superior to the first-order approximations. Considering that their derivations are based on quite different assumptions, the two formulations give surprisingly similar results, at least for the simple model under consideration. Their main deviation from the quantum calculation—the failure to reproduce the dip of the pump–probe signal at 2.25 eV—is caused by the limits of the classical mapping calculations (it already shows up in the classical wave packet calculation in Figure 3) and not by the approximate calculation of the spectrum.

So far, we have restricted the discussion to the case of impulsive excitation ( $\tau_1 = 0$ ). To study the effect of finite pump pulses on the electron-transfer dynamics, Figure 2b shows the diabatic population  $P_2(t)$  in the case of a resonant ( $\omega_1 = E_2 - E_0$ ) pump pulse of 20 fs duration. As is expected, the finite pulse tends to smear out the time evolution of electronic population dynamics. When the FC2 doorway function (eq 4.4) is used, the corresponding classical calculation is seen to somewhat exaggerate this broadening effect, particularly at longer times. Because the latter also occurs in the case of impulsive excitation, it again seems to be caused less by the approximate calculation of the doorway function rather than by the approximate calculation of the nonadiabatic dynamics through the classical mapping calculations.

Finally, we wish to study the performance of the classical approximation to calculate the ground-state contribution to the transient absorption spectrum. Employing the FC2 ground-state doorway and windows functions 4.15 and 4.16, Figure 3 compares quantum and classical results of the time-resolved ground-state pump–probe signal, sometimes referred to as “impulsive stimulated Raman” emission.<sup>40</sup> As the name indicates, the ground-state contribution mainly reflects vibrational dynamics in the electronic ground state and therefore yields only little information on the excited-state electron-transfer dynamics. The coherent vibrational motion seen in the quantum calculation is nicely reproduced by the classical results.

**C. Three-Mode Two-State Model.** To describe the situation of internal radiative transitions, we consider a two-state three-mode system describing the photoinduced electron-transfer process of the mixed-valence system  $(\text{NH}_3)_5\text{Ru}^{\text{III}}\text{NCRu}^{\text{II}}(\text{CN})_5^-$  investigated by Barbara and co-workers.<sup>60</sup> On the basis of the model of Wang and Thoss,<sup>35</sup> we choose the parameters (in eV)  $E_1 = 0$ ,  $E_2 = 1.177$ ,  $g = 0.186$ ,  $\omega_1 = 0.2615$ ,  $\omega_2 = 0.0565$ ,  $\omega_3 = 0.0198$ ,  $\kappa_n^{(1)} = 0$ ,  $\kappa_1^{(2)} = 0.2003$ ,  $\kappa_2^{(2)} = 0.1474$ , and  $\kappa_3^{(2)} = 0.0496$ . The diabatic potential-energy curves  $V_n(x_2)$  of the model along the most important vibrational mode  $x_2$  are shown in Figure 1b.

Assuming again that the diabatic electronic state  $|\psi_2\rangle$  is impulsively excited at time  $t = 0$ , Figure 5a shows the quantum-mechanical diabatic population probability  $P_2(t)$  of the two-state three-mode model. The diabatic population is seen to exhibit a complex stepwise decay within the first  $\approx 200$  fs. Also shown in Figure 5b is the adiabatic population

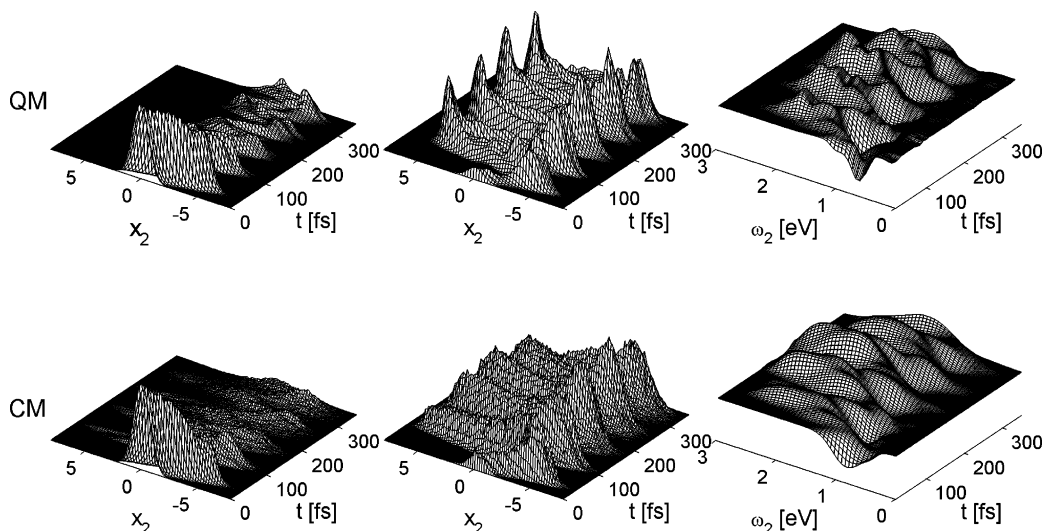


**Figure 5.** Time evolution of the electronic population probability of the (a) diabatic and (b) adiabatic optically excited electronic state. Compared are classical mapping results (solid lines, eqs 3.7 and 3.8) and quantum-mechanical reference calculations (dashed lines, eqs 2.3 and 2.6), as obtained for the three-mode electron-transfer model.

probability  $P_2^{\text{ad}}(t)$  (see eq 2.6), which is surprisingly similar to the diabatic population  $P_2(t)$ . Unlike the simple one-mode model discussed above, however, the coupling between the electronic and the vibrational dynamics is not as straightforward for the three-mode model. As an example, Figure 6 shows the time-dependent probability densities  $P_2(x_2, t)$  and  $P_1(x_2, t)$ , reflecting the time evolution of the nonadiabatic wave packet motion on the diabatic electronic states  $|\psi_2\rangle$  and  $|\psi_1\rangle$ , respectively. (For brevity, we restrict the discussion to the coordinate  $x_2$ , which represents the most important vibrational mode of the model.) Starting at time  $t = 0$  in the upper electronic state at position  $x_2 = 0$ , the wave packet reaches the crossing region within  $\approx 20$  fs and bifurcates into two components evolving on the two coupled electronic surfaces. The excited-state component of the wave packet is seen to decay within a few hundred femtoseconds, while in the electronic ground-state potential, the vibrational motion is only weakly damped.

Let us turn to the transient absorption spectrum of the three-mode model. As the ground-state contribution essentially reflects simple harmonic vibrational motion similar to the one-mode case, we restrict the discussion to the excited-state contribution to the spectrum shown in Figure 6. The three-mode model is seen to give rise to pronounced wave packet motion, exhibiting stimulated emission ( $I > 0$ ) as well as transient absorption ( $I < 0$ ) from the hot ground state. We note that the latter reflects parts of the wave packet that decayed via the curve crossing to the electronic ground state. The wave packet dynamics mostly reflects the motion along the  $x_2$  coordinate, which is strongly shifted ( $\kappa_2^{(2)}/\omega_2 \approx 1$ ) and shows a period of  $T_{\omega_2} \approx 70$  fs. This is because the high-frequency vibration  $\omega_1$  ( $T_{\omega_1} \approx 16$  fs) is too fast to be resolved by 20 fs pulses, and the low-frequency vibration  $\omega_3$  ( $T_{\omega_3} \approx 200$  fs) is quite slow on the time scale considered and only affects a minor overall widening of the spectrum around  $\approx 200$  fs.

Finally, we wish to consider the performance of various classical models. To this end, Figures 5 and 6 compare the



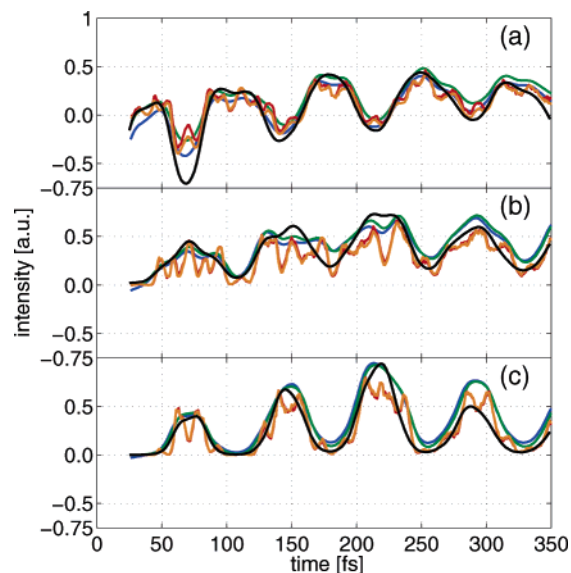
**Figure 6.** Comparison of the quantum-mechanical (QM, upper panels) and classical (CM, lower panels) electron-transfer dynamics and spectroscopy of the three-mode model. Shown are the time-dependent nonadiabatic wave packet motion as described by the probability densities (left)  $P_2(x_2, t)$  and (middle)  $P_1(x_2, t)$  (eq 2.4), as well as (right) the excited-state contribution to the transient absorption spectrum. The classical spectrum is calculated using the FC2 approximation in eq 4.5.

quantum and classical electronic population probabilities and probability densities, respectively. The simple classical mapping formulation model is seen to reproduce the complex structures of the electronic populations and the nonadiabatic wave packet motion with surprising accuracy. The classical calculation only misses fine details of the vibrational structures, such as the complete rephasing of the wave packet at the turning point of the potential at  $x_2 \approx 5$ . Similarly, the classical transient absorption spectrum calculated at the FC2 level is found to be in excellent agreement with the quantum result. The only difference to be seen in Figure 6 is that the classical spectrum reaches somewhat further toward the high-frequency side.

To study the accuracy of the various classical approximations to the transient absorption for the three-mode model, Figure 7 shows cuts of the transient absorption spectrum at probe carrier frequencies (a) 1.0 eV, (b) 1.5 eV, and (c) 2.0 eV. Again, it is found that the two first-order approximations FC1 and FCC give rise to spurious structures of the time-resolved signal, which indicate that these approximations fail to correctly account for the averaging effect caused by finite pulses. The FC2 and the CEA approximations are found to be quite similar and give a fairly accurate description of the transient absorption spectrum of the three-mode model.

## VI. Conclusions

We have outlined a classical approach to the calculation of time- and frequency-resolved pump–probe spectra of nonadiabatically coupled molecular systems. We have generalized the first- and second-order semiclassical Franck–Condon approximations to the case of vibronically coupled potential-energy surfaces (FC1 and FC2), proposed a classical version of a nonadiabatic version of the Franck–Condon approximation (FCC), and employed a classical analogue of the electronic dipole function (CEA). When established models describing ultrafast photoinduced electron transfer are adopted,<sup>34,35</sup> it has been found that the two first-order



**Figure 7.** Cuts of the excited-state contribution to the transient absorption spectrum at probe carrier frequencies (a) 1.0 eV, (b) 1.5 eV, and (c) 2.0 eV, obtained for the three-mode electron-transfer model. The classical approximations FC1 (red lines, eq 4.3), FC2 (green lines, eq 4.5), FCC (orange lines, eq 4.13), and CEA (blue lines, eqs 4.10 and B7) are compared to exact quantum calculations (black lines).

approximations FC1 and FCC give rise to spurious structures of the time-resolved signal, which indicate that these approximations fail to correctly account for the averaging effect caused by finite pulses. The FC2 and CEA approximations, on the other hand, were found to be quite similar and gave a fairly accurate description of the transient absorption spectrum of the electron-transfer models under consideration. As the numerical implementation of the derived doorway and window functions at the FC2 level (eqs 4.4 and 4.5) is rather straightforward, the generalized second-order Franck–Condon approximation appears to be the method of choice.

Apart from the excited-state contribution for which a

number of classical formulations has been proposed, we have also considered the classical evaluation of the ground-state contribution to the transient absorption spectrum. Although additional assumptions need to be invoked in order to implement the FC2 approximation for this case, the resulting classical calculations compare well to the quantum reference results. Furthermore, we have devised a practical scheme (eq 4.6) to account for effects of pump pulses of finite duration.

Throughout this paper, we have employed the mapping approach in order to achieve a classical description of nonadiabatic quantum dynamics. The classical limit of the mapping formulation has been shown to work surprisingly well for the electron-transfer models under consideration. Nevertheless, the derivations of the various classical approximations are by no means restricted to the mapping formulation but can be used as well by employing other classical models of nonadiabatic quantum dynamics, such as the mean-field trajectory or the surface-hopping method.

**Acknowledgment.** We thank Michael Thoss for numerous inspiring and helpful discussions. This work has been supported by the Frankfurt Center for Scientific Computing, the Fonds der Chemischen Industrie, and the Deutsche Forschungsgemeinschaft.

## Appendix A: Quantum Description

Considering the external case defined by eqs 2.1 and 2.7, we wish to derive the quantum-mechanical result for the pump–probe signal (eq 2.9). Prior to the interaction with the laser field, we assume that the system is in its electronic and vibrational ground state  $|\Psi_0\rangle = |\psi_0\rangle|\phi_0\rangle$ . When time-dependent perturbation theory is employed with respect to the field–matter interaction, the wave function after the interaction with the pump field can be written as<sup>2</sup>

$$|\Psi_A\rangle = i \int_{-\infty}^{\infty} dt' \epsilon_1(t') \hat{\mu}(t') |\Psi_0\rangle \equiv A |\Psi_0\rangle \quad (\text{A1})$$

where the operator  $A$  accounts for the  $|\psi_0\rangle \rightarrow |\psi_2\rangle$  absorption process and  $\hat{\mu}(t) = e^{iHt} \hat{\mu} e^{-iHt}$  represents the transition dipole operator in the Heisenberg representation. Similarly, the wave function after the interaction with the second pulse reads

$$|\Psi_{EA}\rangle = i \int_{-\infty}^{\infty} dt' \epsilon_2(t') \hat{\mu}(t') |\Psi_A\rangle \equiv E |\Psi_A\rangle \quad (\text{A2})$$

Assuming that the measured pump–probe signal is directly proportional to the field-induced population in the final electronic state, the excited-state spectroscopic signal is given by

$$\begin{aligned} I &= \langle \Psi_A | E^\dagger E | \Psi_A \rangle \\ &= \int_{-\infty}^{\infty} dt_2 \int_{-\infty}^{\infty} dt_1 \epsilon_2(t_2) \epsilon_2(t_1) \langle \Psi_A | \hat{\mu}(t_1) \hat{\mu}(t_2) | \Psi_A \rangle \end{aligned} \quad (\text{A3})$$

When the initial density operator  $\rho_0 = |\Psi_0\rangle\langle\Psi_0|$  is introduced, this can be rewritten as

$$I = \text{Tr}\{A e^{-iHt} E \rho_0 E^\dagger e^{iHt} A^\dagger\} \quad (\text{A4})$$

$$= \text{Tr}\{\mathcal{W} e^{-i\mathcal{L}t} \mathcal{D} \rho_0\} \quad (\text{A5})$$

where we defined the Liouville operators  $\mathcal{D}\rho = A\rho A^\dagger$ ,  $\mathcal{L}\rho = [H, \rho]$ , and  $\mathcal{W}\rho = E\rho E^\dagger$  to recover the doorway–window expression in eq 2.9.

Similarly, the ground-state signal (because of impulsive stimulated Raman scattering) discussed in section IV.F is obtained by projecting the fourth-order wave function  $|\Psi_{EAEA}\rangle$  on the initial state  $|\Psi_0\rangle$ .<sup>2</sup> The resulting expression can be cast in doorway–window form (eq A5) with the Liouville operators

$$\mathcal{D}_g = E^\dagger A = \int_{-\infty}^{\infty} dt_2 \int_{-\infty}^{t_2} dt_1 \epsilon_1(t_2) \epsilon_1(t_1) \hat{\mu}(t_2) \hat{\mu}(t_1)$$

$$\mathcal{W}'_g = E^\dagger A = \int_{-\infty}^{\infty} dt_2 \int_{-\infty}^{t_2} dt_1 \epsilon_{II}(t_2) \epsilon_{II}(t_1) \hat{\mu}(t_2) \hat{\mu}(t_1) \quad (\text{A6})$$

Note that because of eq A6 the molecular system evolves in the electronic ground state between the interaction with two laser pulses, while in eq A4, the system evolves in the excited electronic state between pulses.

## Appendix B: Classical Approximations

To derive the FC1 result (eq 4.2), we employ the first-order short-time approximation  $e^{-iH_m t} e^{iH_n t} \approx e^{-i(V_m - V_n)t}$  to the Heisenberg dipole operator  $\hat{\mu}(t)$ .<sup>7,8,30</sup> This allows us to perform the time integrations in eq A3 and yields for the doorway and window operators

$$\begin{aligned} \mathcal{D}\rho_0 &= \frac{\mu^2}{4} e^{-\alpha\tau_1^2[\omega_1 - (V_2 - V_0)]^2} \langle \psi_0 | \rho_0 | \psi_0 \rangle |\psi_2\rangle\langle\psi_2| \\ \mathcal{W}' &= \frac{\mu^2}{4} e^{-\alpha\tau_1^2[\omega_{II} - (V_2 - V_0)]^2} |\psi_2\rangle\langle\psi_2| \end{aligned} \quad (\text{B1})$$

By replacing the quantum-mechanical trace in eq A5 by the classical phase-space average  $\int d\Gamma_0 \dots$  defined in eq 3.7, the classical limit of the doorway–window expression (eq A5) is given by eq 4.1 with the classical doorway and window functions of eq 4.2.

When the second-order short-time approximation<sup>55</sup> is employed,

$$\begin{aligned} e^{iH_2 t_1} \mu_{20} e^{iH_0(t_2 - t_1)} \mu_{02} e^{iH_2 t_2} \approx \\ e^{iH_2(t_2 - t_1)/2} \mu_{20} e^{i(V_2 - V_0)(t_2 - t_1)} \mu_{02} e^{iH_2(t_2 - t_1)/2} \end{aligned} \quad (\text{B2})$$

only one of the two time integrations in eq A3 can be performed analytically. This yields the window operators<sup>55</sup>

$$\mathcal{W}' = \frac{\mu^2}{4} \int_{-\infty}^{\infty} dt e^{-(t-t_0)^2/(2\alpha\tau_1^2)} e^{-\alpha\tau_1^2[\omega_{II} - (V_2 - V_0)]^2} |\psi_2\rangle\langle\psi_2| \quad (\text{B3})$$

and a similar result for the doorway operator. In the classical limit, we obtain the FC2 result (eq 4.5).

Assuming that the matrix elements of the diabatic Hamiltonian commute (i.e.,  $[h_{nm}, h_{kl}] = 0$ ), Diltthey et al.<sup>56</sup> performed a resummation of the Heisenberg dipole operator expansion, which leads to closed expressions for the spectroscopic projection operators  $A$  (eq A1) and  $E$  (eq A2). In the external case, one obtains for the window operator

$$\mathcal{W} = \frac{\mu^2}{4} \sum_{n=1,2} S_{2n}^2 e^{-2\alpha\tau_{\text{II}}^2[\omega_{\text{II}}-(W_n-V_0)]^2} |\psi_n^{\text{ad}}\rangle\langle\psi_n^{\text{ad}}| - \frac{\mu^2}{4} e^{-\alpha\tau_{\text{II}}^2\{[\omega_{\text{II}}-(W_2-V_0)]^2+[\omega_{\text{II}}-(W_1-V_0)]^2\}} \left(S_{22}^2 - \frac{1}{2}\right) (|\psi_1^{\text{ad}}\rangle\langle\psi_2^{\text{ad}}| + \text{h.c.}) \quad (\text{B4})$$

where  $|\psi_n^{\text{ad}}\rangle$  and  $W_n$  denote adiabatic states and potentials, respectively, defined in eq 2.5, and the matrix  $\{S_{mn}\}$  is given in eq 4.8. Containing two resonance conditions, the last term in eq B4 is usually strongly suppressed and can be omitted. In the classical limit, the FCC result (eq 4.7) is recovered. In the internal case, we obtain<sup>56</sup>

$$\mathcal{W} = \frac{\mu^2}{4} \left(S_{22}^2 - \frac{1}{2}\right)^2 e^{-2\alpha\tau_{\text{II}}^2[\omega_{\text{II}}-(W_2-W_1)]^2} (|\psi_2^{\text{ad}}\rangle\langle\psi_2^{\text{ad}}| - |\psi_1^{\text{ad}}\rangle\langle\psi_1^{\text{ad}}|) \quad (\text{B5})$$

where again off-resonant terms have been neglected. In the classical limit, this yields eq 4.13.

The CEA result (eq 4.10) is directly obtained from eq A3 by employing the classical mapping of the electronic transition dipole operator in eq 4.9. The calculation is straightforward in the external case.<sup>27</sup> In the internal case, we recall that the quantum-mechanical pump–probe signal consists of two contributions, reflecting absorption and emission between the coupled electronic states  $|\psi_1\rangle$  and  $|\psi_2\rangle$ . In the CEA formulation, we also obtain two terms,  $\mu_{12}(t) + \mu_{21}(t)$ , which, however, collapse to the term  $2\text{Re}\mu_{21}(t)$  within approximation 4.9. Hence, in the internal case, both absorption and emission are described by a single dipole function and therefore in an averaged manner.

To motivate a way to improve upon this simple approximation, we change to classical action angle variables via  $\sqrt{N_n} e^{iQ_n} = (X_n + iP_n)/\sqrt{2}$ ,<sup>32</sup> thus obtaining for the mapped transition dipole function<sup>33</sup>

$$\mu_{12}(t) = \mu \sqrt{\left(N_1 + \frac{1}{2}\right)\left(N_2 + \frac{1}{2}\right)} e^{i(Q_2-Q_1)} \quad (\text{B6})$$

A simple way to consider absorption and emission separately is to replace the CEA dipole functions in eq B6 by

$$\begin{aligned} \tilde{\mu}_{12}(t) &= \mu_{12} \sqrt{\left(N_2 + \frac{1}{2}\right)} e^{i(Q_2-Q_1)} \\ \tilde{\mu}_{21}(t) &= \mu_{21} \sqrt{\left(N_1 + \frac{1}{2}\right)} e^{i(Q_1-Q_2)} \end{aligned} \quad (\text{B7})$$

where now—as in the quantum case—the emission term depends on the excited-state population  $N_2$  and the absorption term on the ground-state population  $N_1$ .

## References

- (1) Zewail, A. H. *J. Phys. Chem. A* **2000**, *104*, 5660–5694.
- (2) Mukamel, S. *Principles of Nonlinear Optical Spectroscopy*; University Press: Oxford, U. K., 1995.
- (3) Domcke, W.; Yarkony, D. R.; Köppel, H. *Conical Intersections: Electronic Structure, Dynamics and Spectroscopy*; World Scientific: Singapore, 2004.

- (4) Ben-Nun, M.; Quenneville, J.; Martínez, T. J. *J. Phys. Chem. A* **2000**, *104*, 5161–5175.
- (5) Doltsinis, N. L.; Marx, D. *J. Theor. Comput. Chem.* **2002**, *1*, 319–349.
- (6) Carloni, P.; Rothlisberger, U.; Parrinello, M. *Acc. Chem. Res.* **2002**, *35*, 455–464.
- (7) Mukamel, S. *J. Chem. Phys.* **1982**, *77*, 173–181.
- (8) Schinke, R. *Photodissociation Dynamics*; University Press: Cambridge, U. K., 1993.
- (9) Braun, M.; Meier, C.; Engel, V. *J. Chem. Phys.* **1995**, *103*, 7907–7911.
- (10) Braun, M.; Meier, C.; Engel, V. *J. Chem. Phys.* **1996**, *105*, 530–534.
- (11) Meyer, S.; Meier, C.; Engel, V. *J. Chem. Phys.* **1998**, *108*, 7631–7636.
- (12) Zadoyan, R.; Li, Z.; Martens, C. C.; Apkarian, V. A. *J. Chem. Phys.* **1994**, *101*, 6648–6657.
- (13) Li, Z.; Fang, J.-Y.; Martens, C. C. *J. Chem. Phys.* **1996**, *104*, 6919–6929.
- (14) Sterling, M.; Zadoyan, R.; Apkarian, V. A. *J. Chem. Phys.* **1996**, *104*, 6497–6506.
- (15) Ungar, L. W.; Cina, J. A. *Adv. Chem. Phys.* **1997**, *100*, 171–228.
- (16) Shen, Y.-C.; Cina, J. A. *J. Chem. Phys.* **1999**, *110*, 9793–9806.
- (17) Hartmann, M.; Pittner, J.; Bonačić-Koutecký, V. *J. Chem. Phys.* **2001**, *114*, 2106–2122.
- (18) Michl, J.; Bonačić-Koutecký, V. *Electronic Aspects of Organic Photochemistry*; Wiley: New York, 1990.
- (19) Bernardi, F.; Olivucci, M.; Robb, M. A. *Chem. Soc. Rev.* **1996**, *25*, 321–329.
- (20) Domcke, W.; Stock, G. *Adv. Chem. Phys.* **1997**, *100*, 1–169.
- (21) Reischl, B.; de Vivie-Riedle, R.; Rutz, S.; Schreiber, E. *J. Chem. Phys.* **1996**, *104*, 8857–8864.
- (22) Erdmann, M.; Engel, V. *J. Chem. Phys.* **2004**, *120*, 158–164.
- (23) Tully, J. C. *Faraday Discuss.* **1998**, *110*, 407–421.
- (24) Berne, B. J.; Cicciotti, G.; Coker, D. F. *Quantum and Classical Dynamics in Condensed Phase Simulations*; World Scientific: Singapore, 1998.
- (25) Gindensperger, E.; Meier, C.; Beswick, J. A. *Adv. Quantum Chem.* **2004**, *47*, 331–346.
- (26) Stock, G.; Thoss, M. *Adv. Chem. Phys.* **2005**, *134*, 243–375.
- (27) Stock, G.; Miller, W. H. *Chem. Phys. Lett.* **1992**, *197*, 396–404.
- (28) Yu, N.; Margulis, C. J.; Coker, D. F. *J. Chem. Phys.* **2004**, *120*, 3657–3664.
- (29) Hartmann, M.; Pittner, J.; Bonačić-Koutecký, V. *J. Chem. Phys.* **2001**, *114*, 2123–2136.
- (30) Lax, M. *J. Chem. Phys.* **1952**, *20*, 1752–1760.
- (31) Stock, G.; Thoss, M. *Phys. Rev. Lett.* **1997**, *78*, 578–581.
- (32) Meyer, H.-D.; Miller, W. H. *J. Chem. Phys.* **1979**, *70*, 3214–3223.

- (33) Stock, G.; Miller, W. H. *J. Chem. Phys.* **1993**, *99*, 1545–1555.
- (34) Diltthey, S.; Mehlig, B.; Stock, G. *J. Chem. Phys.* **2002**, *116*, 69–78.
- (35) Wang, H.; Thoss, M. *J. Phys. Chem. A* **2003**, *107*, 2126–2136.
- (36) Jortner, J.; Bixon, M. *Electron Transfer: From Isolated Molecules to Biomolecules*; Wiley: New York, 1999; Adv. Chem. Phys. Vols. 106–107.
- (37) Köppel, H.; Domcke, W.; Cederbaum, L. S. *Adv. Chem. Phys.* **1984**, *57*, 59–246.
- (38) Marcus, R. A.; Sutin, N. *Biochim. Biophys. Acta* **1985**, *811*, 265–322.
- (39) Yan, Y. J.; Mukamel, S. *Phys. Rev. A: At., Mol., Opt. Phys.* **1990**, *41*, 6485–6504.
- (40) Pollard, W.; Lee, S.-Y.; Mathies, R. A. *J. Chem. Phys.* **1990**, *92*, 4012–4029.
- (41) Thoss, M.; Stock, G. *Phys. Rev. A: At., Mol., Opt. Phys.* **1999**, *59*, 64–79.
- (42) The mapping of the operators in eq 3.4 preserves the commutation relations and leads to an exact identity of the electronic matrix elements of the propagator.<sup>31</sup>
- (43) Müller, U.; Stock, G. *J. Chem. Phys.* **1998**, *108*, 7516–7526.
- (44) Rabani, E.; Egorov, S. A.; Berne, B. J. *J. Phys. Chem. A* **1999**, *103*, 9539–9544.
- (45) Diltthey, S.; Stock, G. *Phys. Rev. Lett.* **2001**, *87*, 140404.
- (46) Liao, J.-L.; Voth, G. A. *J. Phys. Chem. B* **2002**, *106*, 8449–8455.
- (47) Shalashilin, D. V.; Child, M. S. *J. Chem. Phys.* **2004**, *121*, 3563–3568.
- (48) Novikov, A.; Kleinekathöfer, U.; Schreiber, M. *Chem. Phys.* **2004**, *296*, 149–158.
- (49) Shi, Q.; Geva, E. *J. Phys. Chem. A* **2004**, *108*, 6109–6116.
- (50) Grossmann, F. *Phys. Rev. A: At., Mol., Opt. Phys.* **1999**, *60*, 1791–1796.
- (51) Thoss, M.; Miller, W. H.; Stock, G. *J. Chem. Phys.* **2000**, *112*, 10282–10292.
- (52) Bonella, S.; Coker, D. *Chem. Phys.* **2001**, *268*, 189–200.
- (53) Bonella, S.; Coker, D. F. *J. Chem. Phys.* **2005**, *122*, 194102.
- (54) Stock, G. *J. Chem. Phys.* **1995**, *103*, 2888–2902.
- (55) Dietz, H.; Engel, V. *J. Phys. Chem. A* **1998**, *102*, 7406–7413.
- (56) Diltthey, S.; Hahn, S.; Stock, G. *J. Chem. Phys.* **2000**, *112*, 4910–4922.
- (57) Seidner, L.; Stock, G.; Domcke, W. *J. Chem. Phys.* **1995**, *103*, 3998–4011.
- (58) Diltthey, S.; Stock, G. *J. Phys. Chem. A* **2002**, *106*, 8483–8487.
- (59) Stock, G.; Domcke, W. *Phys. Rev. A: At., Mol., Opt. Phys.* **1992**, *45*, 3032–3040.
- (60) Kambhampati, P.; Song, D. H.; Kee, T. W.; Barbara, P. F. *J. Phys. Chem. A* **2000**, *104*, 10637–10644.

CT6002127

## Ewald Summation for Uniformly Charged Surface

Wen Yang,<sup>†,‡</sup> Xigao Jin,<sup>†</sup> and Qi Liao<sup>\*,†</sup>

*Beijing National Laboratory for Molecular Sciences, State Key Laboratory of Polymer Physics and Chemistry, Joint Laboratory of Polymer Science and Materials, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100080, P. R. China, and Graduate University of Chinese Academy of Sciences, Beijing 100080, P. R. China*

Received March 2, 2006

**Abstract:** We have developed an algorithm to calculate the long-range Coulomb interactions for the slab system with a uniformly charged surface that is periodic in two dimensions and finite extent in the third dimension. This method, which is a modification of the three-dimensional Ewald summation with a correction term, is tested via molecular dynamics simulations of the adsorption of co-ions on the charged surface. The simulation results of counterion distribution show a good agreement with theoretical prediction especially at low surface charge density. Furthermore, we compute the force exerted on the particles for different systems. The consistency of results demonstrates that the proposed algorithm is applicable to the system with a smoothly charged surface in only two dimensions and is benefited from the widely used Ewald summation method in all three dimensions.

### 1. Introduction

With recent advances in computer simulations of aqueous biological systems and nanotechnology, the accurate computation of long-range Coulombic interactions for charged systems is becoming more and more important. The well developed approach to this problem for the three-dimension (3D) periodic system is the Ewald summation technique.<sup>1</sup> Several optimization methods such as the smooth particle mesh Ewald (SPME) method<sup>2</sup> have been proposed in order to perform effectively for large systems.

However, for many situations especially in the simulations of surface and interfacial systems, the conventional Ewald summation method cannot be used directly and needed modification because the system is finite in one direction and infinite in the other two dimensions. To treat the electrostatic interactions of such slab geometry the two-dimensional Ewald summation (EW2D) technique was first reported by Parry<sup>3</sup> and was subsequently extended by various authors.<sup>4,5</sup> Unfortunately, the applications of these techniques are still limited for complex systems such as membranes and surfactants since the direct use of the EW2D method is not

as simple and efficient as the EW3D method.<sup>6,7</sup> Recently, to solve this problem, an approximation method added a correction term to the three-dimensional Ewald summation (EW3D) technique for the simulation cell which is sufficiently elongated in the direction of nonperiodicity to create large empty spaces outside the system so as to minimize an artificial influence from the periodic images in that direction.<sup>8</sup> Particularly, Yeh et al.<sup>9</sup> introduced the shape-dependent correction term given by Smith<sup>10</sup> into the EW3D technique to calculate the electrostatic forces for systems with slab geometry. It has been shown that this method termed as the three-dimensional Ewald summation with the correction term (EW3DC) is a good approximation to the exact EW2D method with a significant reduction in computing time.

An alternative analytical method which needed to evaluate trigonometric functions and Bessel functions of imaginary arguments to deal with the long-range electrostatic interactions in the three-dimensional system was proposed by Lekner.<sup>11</sup> Taking advantage of the symmetry given by the periodicity of the lattice summation, the charge–charge interaction was calculated and the interaction energy was obtained by integrating the force expression. Based on Lekner's summation technique, several authors<sup>12–16</sup> derived expressions for the electrostatic force in systems periodic in

\* Corresponding author e-mail: qiliao@iccas.ac.cn.

<sup>†</sup> Institute of Chemistry, Chinese Academy of Sciences.

<sup>‡</sup> Graduate University of the Chinese Academy of Sciences.

one or two dimensions. Recently, the comparison between the Ewald quasi-2D and Lekner summation methods has been carried out by Mazars,<sup>17</sup> which shows that they are in close agreement when correctly implemented for the computation.

Although the charged surface with molecular details is modeled by the EW3DC method mentioned above, many charged surfaces could be simulated by a coarse-grained model of a uniformly charged surface to make the calculations more efficient. The coarse-grained model of the uniformly charged surface may be described by just one parameter, i.e. the charge density at the surface. For this kind of model, we would like to substitute the atomistic detail of the surface by a uniformly charged surface to speed up the simulation. The purpose of our work is to extend the method based on the EW3DC technique to the electrostatic interaction calculation of smooth surface systems with uniform charges in computer simulations. Section 2 describes the expression of Ewald summation we developed for a two-dimensional (2D) system with a uniformly charged surface. To test this method we performed simulations on the counterion distribution next to the uniformly charged surface. Results for counterion density profiles compared with the Poisson–Boltzmann (PB) theory are presented in section 3. For a more direct test, force calculations of particles for different systems have been carried out, and the results are also shown in section 3. Finally, we draw a conclusion of our work in section 4.

## 2. Algorithm

We first give a brief review of the EW2D technique<sup>9</sup> and then present our method for treating Coulomb interactions in 3D systems with a smoothly charged surface. It is known that the electrostatic energy for a system of  $N$  point charges  $q_i$  at position  $\mathbf{r}_i$  can be written as

$$E = \frac{1}{2} \sum_n \sum_{i,j=1}^N \frac{q_i q_j}{|\mathbf{r}_{ij} + nL|} \quad (1)$$

The sum over  $n$  takes into account all of the periodic images of the charges, and the prime indicates that in the case of  $i = j$  the term  $n = 0$  must be omitted. We are omitting all factors of  $1/(4\pi\epsilon_0)$  for clarification ( $\epsilon_0$  is the vacuum permittivity). For long-range potentials, this sum is conditionally convergent, which means the result depends on the order in which it adds up the terms. If we add up our infinite system by sphere layers,<sup>1</sup> the electrostatic energy could be obtained by the Ewald sum, and the final result is

$$E = E^r + E^k + E^s + E^d \quad (2)$$

where the contribution from real space  $E^r$ , the contribution from reciprocal space  $E^k$ , the self-energy  $E^s$ , and the shape-dependent dipole correction for the spherical geometry  $E^d$  are respectively given by<sup>1,10</sup>

$$E^r = \frac{1}{2} \sum_{i,j} \sum_m q_i q_j \frac{\text{erfc}(\kappa|\mathbf{r}_{ij} + mL|)}{|\mathbf{r}_{ij} + mL|} \quad (3)$$

$$E^k = \frac{1}{2\pi V} \sum_{i,j} \sum_{\mathbf{k} \neq 0} q_i q_j \left( \frac{4\pi^2}{k^2} \right) \exp\left(-\frac{k^2}{4\kappa^2}\right) \cos(\mathbf{k} \cdot \mathbf{r}_{ij}) \quad (4)$$

$$E^s = -\frac{\kappa}{\sqrt{\pi}} \sum_i q_i^2 \quad (5)$$

$$E^d = \frac{2\pi}{(1 + 2\epsilon_s)V} \left( \sum_i q_i \mathbf{r}_i \right)^2 \quad (6)$$

By modifying the shape-dependent dipole correction  $E^d$  to the geometry of a rectangle plate, i.e.

$$E^d = \frac{2\pi}{V} \left( \sum_i q_i \mathbf{r}_{i,z} \right)^2 \quad (7)$$

Yeh et al.<sup>9</sup> show that the 3D Ewald summation with the correction term can be used with sufficient accuracy in the calculation of electrostatic energy of a system with slab geometry, even better than the 2D Ewald sum. More importantly, the computing time by using modification of the 3D Ewald sum is about 10 times faster than the time for the 2D Ewald sum.

In MD simulation, the evaluation of force is much more important. The force  $F_i$  exerted on particle  $i$  is obtained by differentiating the electrostatic energy  $E$ , and the results are given by

$$F_i = F_i^r + F_i^k + F_i^d \quad (8)$$

where

$$F_i^r = q_i \sum_j q_j \sum_m \left( \frac{2\kappa}{\sqrt{\pi}} \exp(-\kappa^2|\mathbf{r}_{ij} + mL|^2) + \frac{\text{erfc}(\kappa|\mathbf{r}_{ij} + mL|)}{|\mathbf{r}_{ij} + mL|} \right) \frac{\mathbf{r}_{ij} + mL}{|\mathbf{r}_{ij} + mL|^2} \quad (9)$$

$$F_i^k = \frac{q_i}{V} \sum_j q_j \sum_{\mathbf{k} \neq 0} \frac{4\pi\mathbf{k}}{k^2} \exp\left(-\frac{k^2}{4\kappa^2}\right) \sin(\mathbf{k} \cdot \mathbf{r}_{ij}) \quad (10)$$

$$F_x^d = F_y^d = 0, \quad F_{i,z}^d = -\frac{4\pi q_i}{V} \sum_j q_j \mathbf{r}_{j,z} \quad (11)$$

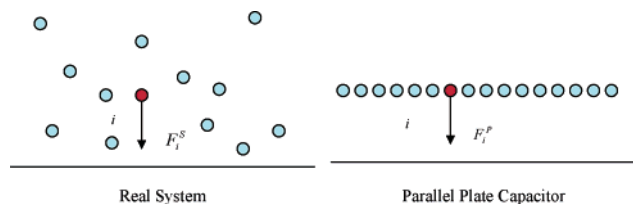
Consider a uniformly charged surface with charge density  $\rho = q_i^s/L_x L_y$ , the electric field acting on the particles  $i$  around the surface is given by<sup>18,19</sup>

$$E_i = \frac{\rho}{2\epsilon_0} \quad (12)$$

The  $x$ ,  $y$ , and  $z$  components of the force are given by

$$F_i^S = \left( F_{i,x}^S = 0, F_{i,y}^S = 0, F_{i,z}^S = \frac{\rho q_i}{2\epsilon_0} \right) \quad (13)$$

and the co-ions of the surface charges will distribute around the surface. When we put the co-ions in the simulation box, the simulation system is not electroneutral if the surface charge  $q_i^s$  is not included in the Ewald summation of eqs



**Figure 1.** The force contributed from surface charges  $i$  in real system and the one in the parallel plate capacitor.

9–11, although the real physical system is still electroneutral. Thus we could not directly obtain the electrostatic energy and Coulombic force by eqs 2 and 8. In the following parts, with the modification of the EW3DC technique, we put forward an approach to solve the Ewald summation for the system with a smoothly charged surface.

The Coulombic force on particle  $i$ , if we include the charges of surface in the Ewald sum, is given by

$$F_i = F_i^r + F_i^k + F_i^d = F_i^{r,S} + F_i^{r,NS} + F_i^{k,S} + F_i^{k,NS} + F_i^d \quad (14)$$

where  $F_i^{r,S}$  and  $F_i^{k,S}$  are contributions from surface charges  $q_i^S$ , and  $F_i^{r,NS}$  and  $F_i^{k,NS}$  are contributions from other charged particles  $q_i$ . Since there is no possible surface structure to obtain the positions of the surface charges, the EW3DC technique cannot be used directly to calculate the electrostatic interactions between surface charges and non-surface charges.

As shown in Figure 1, given that the surface charge density is the same, the force  $F_i^S$  (eq 13) contributed from surface charges in a real system is equal to the force  $F_i^P$  in the parallel plate capacitor. From Yeh et al.'s modification,<sup>9</sup> the force  $F_i^P$  in the parallel plate capacitor could be calculated with sufficient accuracy by

$$F_i^S = F_i^P = F_i^{r,P} + F_i^{k,P} + F_i^{d,P} \quad (15)$$

In the parallel plate capacitor system, the system is electroneutral if we include the surface charge in the calculations of the Ewald summation; we have  $F_i^P = F_i^{r,S} + F_i^{k,S} + F_i^{d,P}$  because the force contributed from nonsurface charges is zero in the parallel plate capacitor system. Thus the force  $F_i$  in the system with a smoothly charged surface at position  $r_{S,z}$  is given by

$$F_i = F_i^S - F_i^{d,S} + F_i^{r,NS} + F_i^{k,NS} + F_i^d \quad (16)$$

where

$$F_i^S = (0, 0, F_{i,z}^S) = \left(0, 0, \frac{\rho q_i}{2\epsilon_0}\right) \quad (17)$$

$$F_i^{d,S} = \left(0, 0, -\frac{4\pi q_i \rho L_x L_y (\mathbf{r}_{S,z} - \mathbf{r}_{i,z})}{V}\right) \quad (18)$$

$$F_i^d = \left(0, 0, -\frac{4\pi q_i}{V} \left[\sum_j q_j \mathbf{r}_{j,z} + \rho L_x L_y \mathbf{r}_{S,z}\right]\right) \quad (19)$$

$$F_i^{r,NS} = q_i \sum_{j \in S} q_j \sum_m \left( \frac{2\kappa}{\sqrt{\pi}} \exp(-\kappa^2 |\mathbf{r}_{ij} + mL|^2) + \frac{\text{erfc}(\kappa |\mathbf{r}_{ij} + mL|)}{|\mathbf{r}_{ij} + mL|} \right) \frac{\mathbf{r}_{ij} + mL}{|\mathbf{r}_{ij} + mL|^2} \quad (20)$$

$$F_i^{k,NS} = \frac{q_i}{V} \sum_{j \in S} q_j \sum_{\mathbf{k} \neq 0} \frac{4\pi \mathbf{k}}{k^2} \exp\left(-\frac{k^2}{4\kappa^2}\right) \sin(\mathbf{k} \cdot \mathbf{r}_{ij}) \quad (21)$$

Note that eqs 20 and 21 have the same forms of the Ewald summation but perform the force evaluation for the nonsurface particles. The Coulombic interaction between the surface charges and co-ions is merged into eqs 17–19 and has nothing to do with the position of the surface. In conclusion, the eqs 16–21 give the formula for the force in the system with a uniformly charged surface of charge density  $\rho$ .

### 3. Molecular Dynamics Simulations of Co-ions on a Uniformly Charged Surface

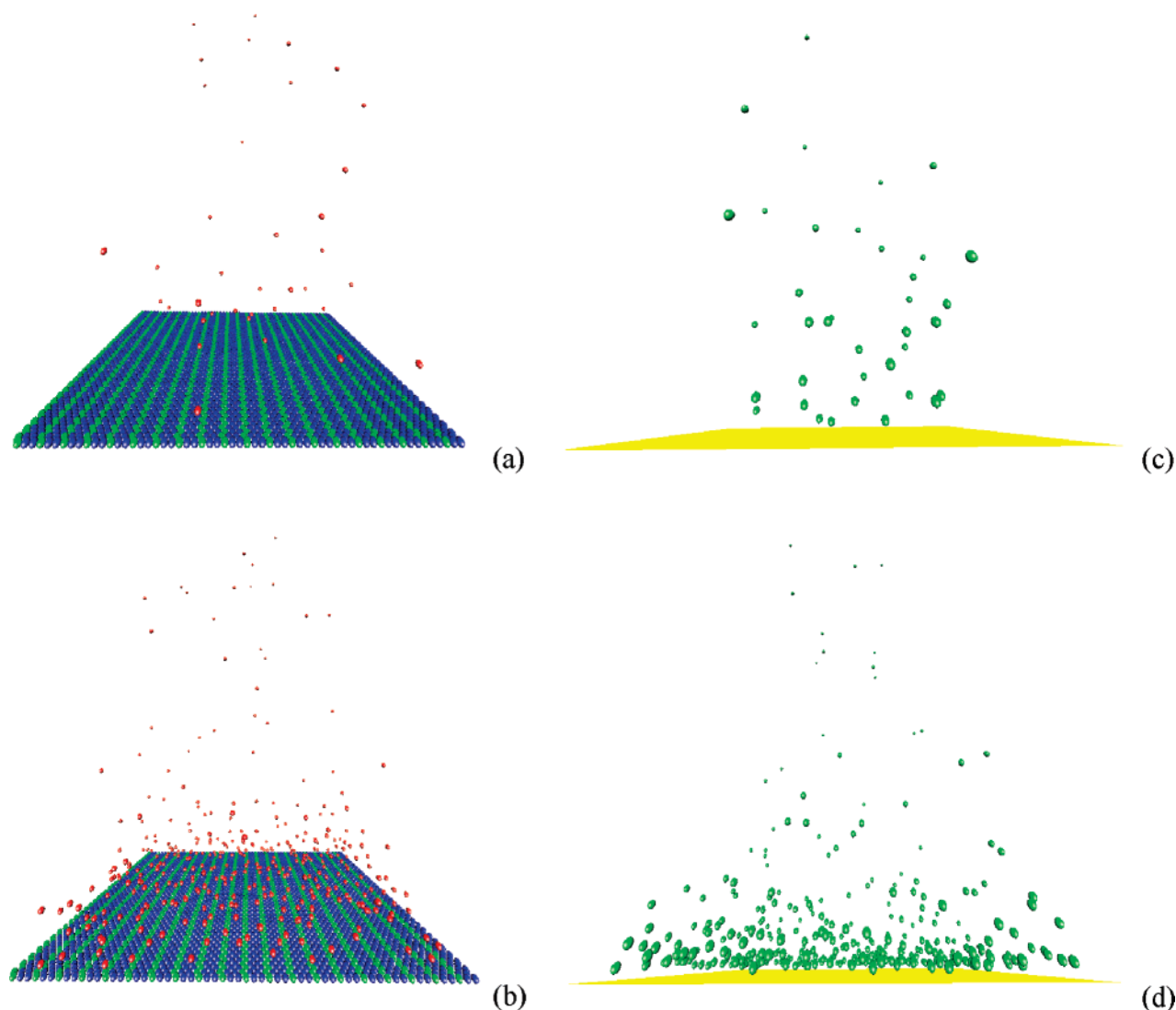
**3.1. Model.** To test the validity of the algorithm referred to above, we performed the simulations of co-ions adsorbed on the full-atom surface (simulation I) and smooth surface (simulation II), respectively, with the same charge density  $\rho$  of the surface. Figure 2 shows the snapshots of the adsorption of co-ions in simulation I and simulation II at different surface charge densities.

For simulation I, we chose a simulation box with the orthorhombic boundary condition having dimension  $L_z = 480.0 \sigma$  ( $\sigma$  is the length unit in the simulation) normal to the surfaces eight times larger than the lateral dimension  $L = 60.0 \sigma$ . Two impermeable rectangle slabs of thickness  $\sigma$  constructed by a number of 420 atoms are respectively located at  $z = 1/8 L_z$  and  $z = -1/8 L_z$  of the central simulation box perpendicular to the  $z$ -axis. Different numbers of atoms carrying charge are uniformly distributed on the slab at  $z = -1/8 L_z$  in order to obtain various surface charge density  $\rho$ . To maintain the electroneutrality of the system, monovalent co-ions are confined in the  $z$  direction between two slabs. Therefore, the closest approaches of the co-ions to the impenetrable surfaces are  $z = \pm(1/8)L_z \mp \sigma$ . We employed a shifted Lennard-Jones (LJ) potential to describe the pure repulsive excluded volume interactions between any pair of particles in the simulations.

$$U_{\text{LJ}}^S(r) = \begin{cases} 4\epsilon_{\text{LJ}} \left[ \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] + \epsilon_{\text{LJ}} & r \leq 2^{1/6} \sigma \\ 0 & r > 2^{1/6} \sigma \end{cases} \quad (22)$$

$\sigma$  and  $\epsilon_{\text{LJ}}$  are respectively the LJ units of length and energy. The strength of the short-range interactions is controlled by the parameter  $\epsilon_{\text{LJ}}$ . As far as simulation II is concerned, the charged smooth surface is substituted by the effect of the external field. We developed our modification code based on the DL\_POLY version 2.12 software package.<sup>20</sup> To avoid the co-ions moving out of the simulation box when the distance between ions and surface is less than  $\sigma$ , a potential provided by the software package named containing wall with the expression  $U = A(R_0 - r)^{-n}$  is employed to confine





**Figure 2.** Snapshots of the adsorption of co-ions on the full-atoms surface with the charge densities of 0.01(a) and 0.1(b) and the smooth surface with the charge densities of 0.01(c) and 0.1(d). For (a) and (b), co-ions are red spheres, charged atom on the surface – green spheres, and neutral ones – blue spheres. For (c) and (d), co-ions are green spheres, and the charged surface is yellow.

**Table 1.** Dependence of Co-ions Number  $N$  on Surface Charge Density  $\rho$  for Both Simulation I and Simulation II

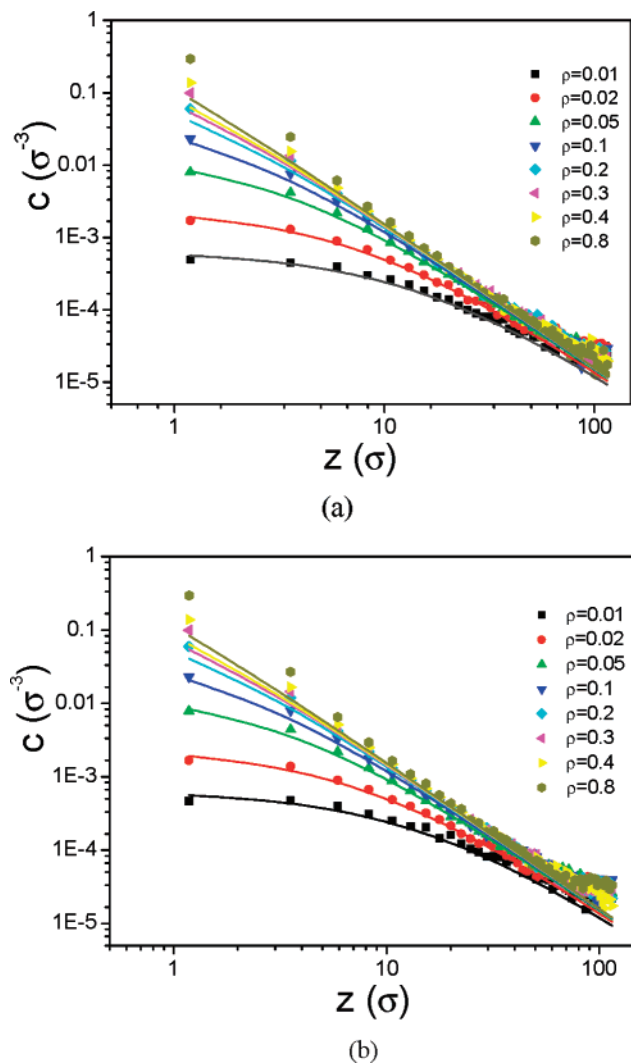
$\rho$	0.01	0.02	0.05	0.1	0.2	0.3	0.4	0.8
$N$	37	73	182	364	728	1092	1455	2910

the co-ions between  $z = (1/8)L_z - \sigma$  and  $z = -(1/8)L_z + \sigma$ . The interactions between any couple of co-ions are the same as the ones performed in simulation I. Solvent molecules, which are not included explicitly in simulation I and simulation II, are modeled by a continuum with the dielectric constant  $\epsilon$ . In such a medium, the electrostatic interactions between nonsurface charges in the simulation box and their periodic images are computed by the smoothed particle mesh Ewald (SPME) algorithm<sup>2</sup> implemented in the DL\_POLY version 2.12 software package.<sup>20</sup>

After the initial conformation was constructed as above (see Table 1 for details), we performed the molecular dynamics simulations (MD) for both simulation I and simulation II with different surface charge densities at constant temperature  $T = 1.0\epsilon_{LJ}/k_B$  using the Langevin thermostat,

where  $k_B$  is the Boltzmann factor. The propagation of the trajectories was done using the Verlet algorithm with a time step equal to  $\Delta t = 0.005\tau_{LJ}$ , where  $\tau_{LJ} = \sqrt{m\sigma^2/k_B T}$  is the standard time unit for Lennard-Jones fluid. In all simulations, a period of 450 000 MD steps was recorded in every 2500 MD step runs for further data analysis after the equilibration period of 50 000 MD steps, which is long enough to relax the whole system.

**3.2. Results and Discussion. A. Adsorption of Co-Ions on the Surface.** The results of the co-ions density profile with different surface charge densities for simulation I (full-atoms surface) and simulation II (smooth surface) are presented in Figure 3 (parts a and b, respectively). From these two figure parts, we easily observe that the density of co-ions decays with distance  $z$  from the plane and at a large distance becomes independent of the surface charge density because of the screening of the surface charge by the cloud of co-ions. Meanwhile, the data from simulation II are almost identical to those obtained by simulation I when the parameters used in all the simulations are the same.



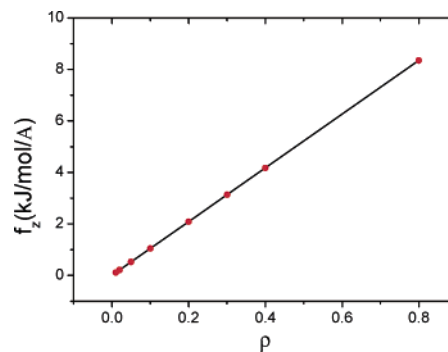
**Figure 3.** The co-ions density profiles at different surface charge densities in double-logarithmic coordinates for (a) full-atoms surface and (b) uniformly charged surface. The theoretical prediction  $c_i(z) = [2\pi l_B(z + \Lambda)^2]^{-1}$  is shown by the solid line.

To compare with the theoretical results, we first reviewed the Poisson–Boltzmann theory for a charged flat surface in the presence of co-ions.<sup>21,22</sup> Considering a uniformly charged infinite flat surface with charge density  $\rho$ , the distribution of the dimensionless electrostatic potential in the direction  $z$  perpendicular to the surface can be represented as  $\Psi(z) \equiv e\psi(z)/k_B T$  from the electrochemical potential of any ion. Since the electrochemical potential must be constant throughout the solution, we can get the Boltzmann distribution of the neutralizing monovalent co-ions at any position in the solvent. The density profile of the co-ions,  $c_i(z)$ , in the case of no added electrolyte solution, can be obtained on the basis of the Poisson–Boltzmann equation

$$\frac{d^2\Psi}{dz^2} = -4\pi l_B c_i(z) \quad (23)$$

$$c_i(z) = c_i(0) \exp(-\Psi(z)) \quad (24)$$

where  $l_B$  is the Bjerrum length  $l_B = e^2/k_B T \epsilon$  and  $\epsilon$  is the dielectric permittivity of the solvent. With the boundary



**Figure 4.** The comparison of force in the  $z$  direction with a charged smooth surface between simulation and theory for different surface charge densities. The solid line is the theoretical value calculated via eq 13.

conditions  $\Psi(0) = 0$  and  $(d\Psi(z)/dz)_{z=0} = 2/\Lambda$ , because of the zero surface potential at electric field and the requirement of overall electro neutrality, the solution is given by

$$\Psi(z) = 2 \ln(1 + z/\Lambda) \quad (25)$$

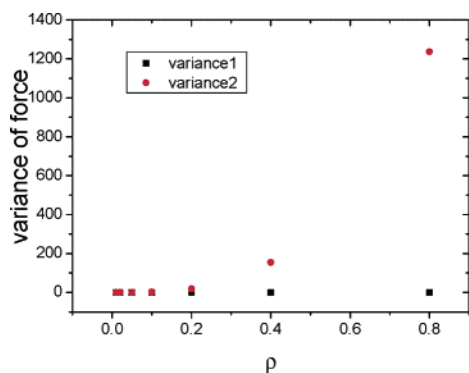
$$c_i(z) = [2\pi l_B(z + \Lambda)^2]^{-1} \quad (26)$$

where  $\Lambda = (2\pi l_B \rho)^{-1}$  is the Gouy–Chapman length, which characterizes the thickness of the co-ions cloud above a homogeneously charged planar surface.

From the analysis above, we also plotted the theoretical curves of the co-ions density distribution for different surface charge densities in the same figures. The graph exhibits that they agree with the simulation results especially at a small surface charge density. As far as the co-ions density distribution at high surface charge density is concerned, since the Poisson–Boltzmann equation is no longer available, the computing results deviate from the theoretical value.

**B. Forces between Charged Surface and Co-Ions.** The most accurate and direct means to affirm the validity of the methodology is the calculation of the interaction force. The same geometry of the simulation cell with uniformly charged surface and parameters mentioned in simulation II have been used here. The co-ions are arranged symmetrically on a plate that is parallel with the charged surface because we want to model the system of the parallel plate capacitor. Theoretically, eq 13 shows that the  $z$  component of the force exerted on any particle is independent of distance away from the surface in the parallel plate capacitor system if the charge density of the surface is constant. We calculated the force of the particles located at  $z = -58.0 \sigma$  contributed from the surface with different charged density after one MD step, which are presented in Figure 4. The simulation results are in excellent agreement with theoretical values obtained from eq 13. This implies that applying the parallel plate capacitor model to deal with the long-range interaction for a uniformly charged surface is feasible.

We also computed the force on particles for simulation I and for the same model except for replacing the full-atoms surface with a uniformly charged surface. At the same time the force exerted on particles for the simulation with full-atoms charged surface was calculated by EW3DC and EW3D, respectively. Figure 5 shows the comparison of the



**Figure 5.** The comparison for the variance of force in the z direction for the full-atoms surface with the smooth surface (variance 1) and with the full-atoms surface using no correction term (variance 2).

variance of forces for each case at different charge densities. The variance of forces is calculated as follows:  $\Delta f = \sum_{i=1}^{\text{ions}} (f_i^1 - f_i^2)^2$ , where the superscript represents two systems which are to be compared. The plot clearly implies that the correction term is necessary to be considered for computing long-range Coulombic interactions in the interfacial system, and it is also reported in Yeh's work.<sup>9</sup> After comparing with the data obtained from the calculation for full-atoms surface, we also conclude that the algorithm of the electrostatic force for a uniformly charged surface gives good numerical results.

#### 4. Conclusion

In this study, we introduced a modification of the previously developed Ewald summation technique to apply to systems with uniformly charged surfaces. The Yeh's algorithm, which just adds a correction term to 3D Ewald summation, could be benefited from the development of the Ewald summation method. Molecular dynamics simulations of co-ions adsorbed on the charged surface and detailed comparison of forces were performed to test the validity of the algorithm. Comparing the simulation results with the theoretical prediction, we clearly demonstrated that the method we adopted to treat the electrostatic interaction is suitable to study systems with a smoothly charged surface.

**Acknowledgment.** This work was financially supported by NFSC Grant no. 20474075 and no. 20490220, by Innovation Funding No. KJCX2-SW-H07 from the Chinese

Academy of Sciences, and 973 Grant no. 2003CB615604 from the Chinese Ministry of Science and Technology.

#### References

- (1) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University: New York, 1987; pp 155–166.
- (2) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (3) Parry, D. E. *Surf. Sci.* **1975**, *49*, 433–440.
- (4) Heyes, D. M.; Barber, M.; Clarke, J. H. R. *J. Chem. Soc., Faraday Trans.* **1977**, *73*, 1485–1496.
- (5) de Leeuw, S. W.; Perram, J. W. *Mol. Phys.* **1979**, *37*, 1313–1322.
- (6) Liem, S. Y.; Clarke, J. H. R. *Mol. Phys.* **1997**, *92*, 19–25.
- (7) Widmann, A. H.; Adolf, D. B. *Comput. Phys. Commun.* **1997**, *107*, 167–186.
- (8) Shelley, J. C.; Patey, G. N. *Mol. Phys.* **1996**, *88*, 385–398.
- (9) Yeh, I. C.; Berkowitz, M. L. *J. Chem. Phys.* **1999**, *111*, 3155–3162.
- (10) Smith, E. R. *Proc. R. Soc. London A* **1981**, *375*, 475–505.
- (11) Lekner, J. *Physica A* **1991**, *176*, 485–498.
- (12) Clark, A. T.; Madden, T. J.; Warren, P. B. *Mol. Phys.* **1996**, *87*, 1063–1069.
- (13) Gronbech-Jensen, N. *Int. J. Mod. Phys. C* **1997**, *8*, 1287–1297.
- (14) Gronbech-Jensen, N.; Hummer, G.; Beardmore, K. M. *Mol. Phys.* **1997**, *92*, 941–945.
- (15) Grzybowski, A.; Brodka, A. *Mol. Phys.* **2002**, *100*, 1017–1023.
- (16) Juffer, A. H.; Shepherd, C. M.; Vogel, H. J. *J. Chem. Phys.* **2001**, *114*, 1892–1905.
- (17) Mazars, M. *J. Chem. Phys.* **2002**, *117*, 3524–3525.
- (18) Spohr, E. *J. Chem. Phys.* **1997**, *107*, 6342–6348.
- (19) Feynman, R. P.; Leighton, R. B.; Sands, M. *The Feynman Lectures on Physics*; Addison-Wesley: London, 1964; Vol. 2.
- (20) Forester, T. R.; Smith, W. *The DL\_POLY\_2 Reference Manual*; Daresbury Laboratory: Daresbury, 2000; pp 1–144.
- (21) Israelachvili, J. N. *Intermolecular and Surface Forces*; Academic: London, 1991; pp 213–227.
- (22) Borisov, O. V.; Leermakers, F. A. M.; Fleer, G. J.; Zhulina, E. B. *J. Chem. Phys.* **2001**, *114*, 7700–7712.

CT600083S

## Simulation of Interaction Forces between Nanoparticles: End-Grafted Polymer Modifiers

Krishna Tej Marla and J. Carson Meredith\*

*School of Chemical and Biomolecular Engineering, Georgia Institute of Technology,  
Atlanta, Georgia 30332-0100*

Received May 9, 2006

**Abstract:** The interaction forces between nanoscale colloidal particles coated with end-grafted Lennard-Jones homopolymers are calculated using off-lattice Monte Carlo simulations in the NVT ensemble. The focus of this work is on grafted polymers that are of approximately the same size as the nanoparticle, a regime intermediate to the star-polymer and Derjaguin limits. The effects of chain length ( $N$ ), nanoparticle diameter ( $\sigma_c$ ), grafting density ( $\rho_a$ ), and colloid–polymer and polymer–polymer interaction energies ( $\epsilon_{cp}$  and  $\epsilon_{pp}$ ) on the polymer-induced force between the nanoparticles are explored. The inclusion of attractive dispersion interactions between the particle and polymeric modifier results in either long-ranged attraction and short-ranged repulsion or pure repulsion, depending on the molecular parameters. The polymer-induced attraction occurs even under good solvent conditions below a threshold grafting density ( $\rho_a$ ) and chain length ( $N$ ) and could be attributed to both bridging (colloid–polymer) and intersegmental (polymer–polymer) attraction. Above the threshold  $\rho_a$  and  $N$  values, chain entropy and excluded volume effects begin to dominate and lead eventually to polymer-induced repulsion and, consequently, nanoparticle stabilization. These results point to the importance of considering dispersion attractions between grafted segments and the nanoparticle surface in modeling these high-curvature colloid interactions.

### 1. Introduction

We previously used Monte Carlo simulation to examine homopolymer adsorption on nanoscale colloidal particles and the associated polymer-induced forces between them.<sup>1</sup> However, polymers irreversibly *end-grafted* to the surface of nanoparticles are technologically more significant for passivating, stabilizing, or directing the assembly of nanoparticle dispersions. These systems have important applications in areas such as photonics and electronics, chemical and biological sensing, and energy storage. In addition, organically modified nanoparticles also serve as functionalized building blocks for programmed assembly into supramolecular entities. Models of the effect of molecular parameters (polymer chain length, graft density, nanoparticle diameter, and polymer–particle interaction energy) on the resulting interaction forces are critical for the rational design of modifiers for these purposes.

**Planar Surface Models.** The structure and physical properties of end-grafted polymers and the interactions between polymer-grafted surfaces have been studied extensively in recent years. Models of nonadsorbing end-grafted polymers (no segment–surface attraction beyond the tethered segment) predict the well-known mushroom-to-brush transition as the graft density increases. When two such polymer brushes approach in a good solvent, unfavorable steric overlaps between the polymer segments give rise to an entropic repulsion, commonly called “polymeric stabilization”. These phenomena have been described successfully via scaling analysis,<sup>2,3</sup> self-consistent field (SCF) theory,<sup>4–9</sup> and Monte Carlo<sup>10,11</sup> and molecular dynamics simulations.<sup>12,13</sup> In addition, many experimental studies of the interactions between planar polymer brushes<sup>14–19</sup> report a monotonically increasing repulsive force, concurrent with models that neglect polymer–particle attractive interactions.

Alternately, attractive interactions between the surface and the tethered polymer segments result in a more complex

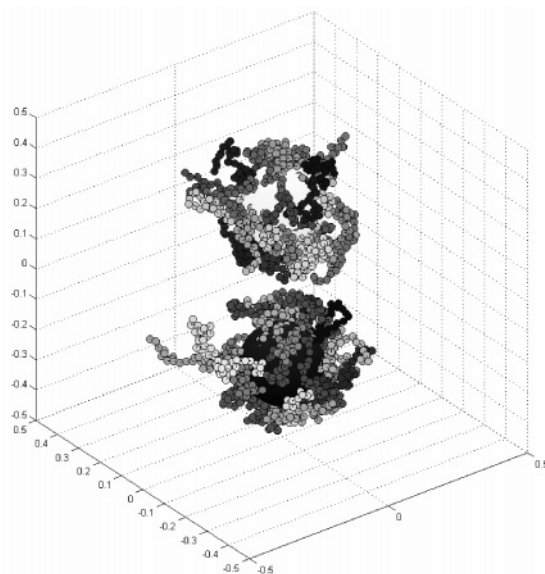
\* Corresponding author e-mail: carson.meredith@chbe.gatech.edu.

situation. At low grafting densities, the polymers adopt a flat, pancakelike conformation to maximize the favorable surface–segment contacts. As the grafting density is increased, the polymer chains form brushes, but at a relatively higher grafting density compared to the nonadsorbing surfaces. Attractive interactions between the grafted polymer and the surface can result in bridging attraction between the polymer-coated surfaces. In a recent study, the force between a *spherical* polymer brush in a good solvent and a coated atomic force microscope (AFM) tip indicated bridging attraction even at relatively high grafting density.<sup>20</sup> One important question is whether the curvature of the brush influences the importance of segment–surface attractions.

**Curvature Effects.** SCF and scaling methods have also been extended to curved surfaces, mostly in either the star-polymer limit, where the polymers are very long compared to the particle diameter,<sup>21–24</sup> or the Derjaguin limit, where the range of the interaction is small compared to the particle radius (small curvature effects). Nanoscale colloidal particles are often tethered to oligomers (short alkyl chains) of approximately the same size range as the nanoparticle. This regime, intermediate to the star-polymer and the Derjaguin limits, has been relatively unexplored, particularly with regard to effects of polymer–surface attraction. The primary effect of curvature, relative to flat surfaces, is that the volume available to the grafted chains increases as  $\sim r^3$  rather than  $\sim r$  ( $r$  = radial distance from the surface). As a result, entropy loss from overlapping chains is not as much of a penalty when curved brushes approach one another. The diminished role of overlap entropy points to the possibility that polymer–surface attraction might become significant, which would lead to differences in the interactions between highly curved brushes (nanoparticles) versus low-curvature or flat brushes (conventional colloid sizes  $\sim 100$  nm to  $10 \mu\text{m}$ ).

Wijmans et al.<sup>25</sup> used a two-dimensional lattice SCF to calculate the interaction between two polymer-coated particles whose radii of curvature are of the same order of magnitude as the polymer layer thickness. They found the repulsive interactions to be *far less* than those predicted by the Derjaguin approximation. Recent Monte Carlo simulations in this regime<sup>26</sup> predicted a lower repulsive force at short distances than the Wijmans et al. results,<sup>27</sup> attributed to the neglect of monomer correlations in the SCF theory. The interaction forces obtained from the simulations were also modeled by a combination of the Witten–Pincus approach<sup>26</sup> and the Flory theory for dilute polymer solutions.<sup>22</sup> SCF studies of grafted-polymer nanoparticles in a good solvent<sup>28</sup> show that the interaction can become *attractive* when the segment “clouds” of the two spheres just begin to overlap and can become purely attractive at low grafting densities. The reason for these contrasting experimental and modeling results is not apparent and reflects the need for further study.

In this paper, we consider the effect of end-grafted polymeric modifiers on the interaction forces between nanoparticles that (1) experience Lennard-Jones (LJ) attraction with the end-grafted polymer segments and (2) are of the same dimensions as the polymers; for example, curvature effects are important. Using Monte Carlo simulations, we



**Figure 1.** Configuration snapshot of the simulation box showing two colloidal nanoparticles with end-grafted polymeric modifiers ( $\sigma_c = 10\sigma_p$ ,  $N = 30$ ,  $\rho_a^* = 3.0$ ,  $T^* = 3.0$ ).

study the effect of nanoparticle diameter ( $\sigma_c$ ), polymer chain length ( $N$ ), grafting density ( $\rho_a$ ), and colloid–polymer and polymer–polymer interaction energies ( $\epsilon_{cp}$  and  $\epsilon_{pp}$ ) on the polymer-induced force profiles between nanoparticles. We identify the physical mechanisms that lead to attraction or repulsion in the force profiles on the basis of interplay between the various molecular parameters described above.

## 2. Simulation Method

**Parameters and Model.** We have simulated the interaction between two spherical nanoparticles grafted with LJ polymer chains by using three-dimensional off-lattice Monte Carlo simulations in the canonical (NVT) ensemble. Figure 1 contains a schematic of the simulation box with two nanoscale colloidal particles that are fixed along the  $z$  axis. Each of the nanoparticles has  $n_c$  chains of length  $N$  that are grafted at random positions on the nanoparticle surface. In most cases,  $N = 10$ – $30$ , while a few results are also presented for longer chains of  $N = 50$ . To study the effect of grafting density, the number of grafting chains is varied from  $n_c = 6$  to  $n_c = 55$ , which correspond to grafting densities ranging from 0.02 to 0.18. The grafting density ( $\rho_a$ ) is given by  $\rho_a = n_c/\pi\sigma_c^2$ . Grafting density is normalized by dividing by the bulk radius of gyration,  $R_g$ , with  $\rho_a^* = \rho_a\pi R_g^2$ . A “mushroom” (or “pancake”, depending on the polymer–particle interactions) to “brush” transition occurs at a reduced grafting density  $\rho_a^* \approx 1$ , where chains grafted to a single surface begin to overlap laterally. The grafting densities studied in this work correspond to  $\rho_a^*$  values of  $0.5 \leq \rho_a^* \leq 5$ , depending on chain length and nanoparticle diameter. Experimental values of  $\rho_a^*$  commonly range between 1 and 20.<sup>29</sup>

The polymers are modeled as fully flexible chains with a bond length of  $\sigma_p$ . The polymer segments interact via the Lennard-Jones potential, cut and shifted to zero at  $r_c = 2.5\sigma_p$ , given in eq 1

$$U(r_{ij}) = \begin{cases} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 - \left( \frac{\sigma_{ij}}{r_c} \right)^{12} + \left( \frac{\sigma_{ij}}{r_c} \right)^6 \right] & r \leq r_c \\ 0 & r > r_c \end{cases} \quad (1)$$

where  $\sigma_{ij} = (\sigma_i + \sigma_j)/2$  is the size parameter and  $\epsilon_{ij}$  is the LJ interaction energy parameter. The cut-and-shifted LJ potential, for which phase behavior and structure have been characterized well, is used as a computational simplification for evaluating the polymer segment–segment interactions. However, the colloid–colloid and the colloid–polymer interactions are modeled using the “full” LJ potential; for example, the potential was evaluated at all interparticle distances without a cutting off the potential. This was done to avoid computational artifacts and difficulty in interpreting results because the cutoff value,  $r_c$ , would vary widely as the colloid size is varied. The interaction between the colloid and the first polymer segment that is grafted permanently to the colloid surface is not considered, as it would contribute only a constant to the energy and cancel in calculations of energy change. The temperature was reduced with the LJ parameters in the usual manner,  $T^* = Tk_B/\epsilon_{pp}$ , where  $\epsilon_{pp}$  is the polymer segment–segment interaction parameter and  $k_B$  is the Boltzmann constant. All of the simulations were performed at  $T^* = 3.0$  (good solvent conditions),  $\epsilon_{cc} = 1$ , and  $\epsilon_{pp} = 1$ , unless specified otherwise. The polymer-coated nanoparticle interaction forces were calculated for two different particle diameters,  $\sigma_c = 10\sigma_p$  and  $5\sigma_p$ . The effect of the colloid–polymer interaction strength ( $\epsilon_{cp}$ ) on the force profiles was studied by varying  $\epsilon_{cp}$  from  $\epsilon_{cp} = 2\epsilon_{pp}$  to a purely soft-repulsive interaction between the colloid and polymer segments. The dimensions of the simulation box are  $50\sigma_p$  in the  $x$  and  $y$  directions and  $60\sigma_p$  in the  $z$  direction, large enough to prevent artificial interactions between periodic images.

It is important to understand the relevance of the chosen model to realistic experimental systems. We have not accounted for atomic or electronic detail because this work does not aim to account for the effects of chemistry or chemical differences between polymers and surfaces. Rather, the focus of this work is physical effects of curved particles that interact attractively with short polymer chains. The freely jointed LJ model is a coarse-grained approach chosen for its simplicity, its ability to capture physical behavior, and the large amount of information available for LJ model systems. The LJ potential is one of the most studied attractive potentials and has been used for decades to examine the physical behavior of pure fluids, mixtures of small molecules, polymer solutions, melts, and blends, as well as colloidal systems. The equations of state properties of LJ fluids and polymers are well-known, allowing us to accurately set the solvent quality, temperature, grafting densities, and other parameters. It is likely that our results would be most appropriately compared to experiments where oligomers are irreversibly grafted to random sites on colloidal nanoparticles, in which there is a net attraction between the chain segments and the particle. In particular, systems in which isotropic dispersion (London) forces dominate would be most similar to the model, for example, polystyrene grafted to silica in a toluene solvent. The use of freely jointed chain segments is

an extreme simplification that artificially increases the number of available chain conformations (entropy) relative to more realistic bead-spring models.<sup>30</sup> However, the inclusion of restrictions to segmental rotation tends to result in quantitative corrections to calculated properties but often does not change the underlying physical behavior.

**Computational Details.** The simulation was initialized by fixing the two nanoparticles at a specified distance from each other along the  $z$  axis. Grafting sites were then chosen at random on each surface, and the first monomer of each polymer chain was fixed permanently at those sites. The remaining  $N - 1$  segments of each chain were then grown in a stretched-out conformation avoiding overlaps with neighboring chains. The initially stretched conformation facilitates faster equilibrium and helps avoid “locked” configurations. Equilibration of the grafted chains is achieved by two methods for generating chain conformations. Chain segments 2 through  $N/2$  were moved by using the “crankshaft” method, which performs a random rotation about the axis passing through the center of the neighboring segments. The remaining part of the chain farthest from the particle surface,  $N/2 + 1$  to  $N$ , was moved using the continuum configurational bias (CCB) algorithm. The CCB method consists of cutting a chain at a randomly selected site and regrowing it site by site until the original chain length is restored. The regrowth process involves scouting various trial positions on a spherical surface centered at the previous existing site, favoring nonoverlapping low-energy configurations at each stage. Details on the computer implementation of the CCB method and its applicability and limitations have been outlined in the literature.<sup>31,32</sup> After equilibrium is achieved, the forces on each particle [ $F_{AP}(r)$  and  $F_{BP}(r)$ ] due to the polymer chains were calculated. Such individual simulations were performed from the nearest center-to-center separation ( $r_{c-c}$ ) between the nanoparticles of  $r_{c-c} = \sigma_c + 2\sigma_p$  in increments of  $1\sigma_p$  up until approximately  $r_{c-c} = 2.5\sigma_c$ . The method and equations used for the force calculation are the same as those described in part I of this series. The force profiles reported in this work are normalized by subtracting the force at “infinite” separation, where the nanoparticle surfaces are far enough not to affect one another significantly.

The reaction-coordinate approach used here, where particles are constrained to approach one another along a collision path, is a common geometry in free energy and force calculations, for example, free energy perturbation and thermodynamic integration approaches. The approach has been shown to be accurate for calculating equilibrium properties of reacting species, fluid structure, and forces between submerged colloidal particles.<sup>33</sup> The potential disadvantages include the fact that experimental systems may not really achieve equilibrium chain conformations as particles approach and collide, although in general, the time scale for chain rearrangement ( $\sim$ nanoseconds) is usually several orders of magnitude faster than the particle–particle collision time scale ( $\sim$ microseconds). The other possibility is that of frozen configurations trapped in a local energy minimum as the particles approach closely. We have

**Table 1.** Simulation Conditions Explored in This Study<sup>a</sup>

case number	$\sigma_c$	$N$	$\epsilon_{cp}$	$\epsilon_{pp}$	$\rho_a^* = \frac{\rho_a \pi R_g^2}{\rho_a \pi \sigma_c^2}$	$\rho_a = \frac{\rho_a \pi R_g^2}{n_c \pi \sigma_c^2}$	$n_c$	$R_g^2$	$R_{EED}^2$
1	10	30	1.0	1.0	0.5	0.02	6	8.05	45.23
2					1.0	0.04	11	8.15	47.33
3					2.0	0.07	22	8.16	48.16
4					3.0	0.11	33	8.33	50.63
5					5.0	0.18	55	8.86	57.90
6	10	30	2.0	1.0	5.0	0.18	55	8.16	48.91
7			Rep <sup>b</sup>	1.0	5.0	0.18	55	9.95	72.11
8			1.0	Rep	5.0	0.18	55	15.82	127.71
9			Rep	Rep	5.0	0.18	55	16.44	137.02
10	10	10	1.0	1.0	0.8	0.11	33	2.42	15.12
11	10	50	1.0	1.0	5.0	0.11	33	14.74	86.19
12	5	10	1.0	1.0	0.8	0.11	8	2.42	14.94
13	5	30	1.0	1.0	3.0	0.11	8	8.54	51.55

<sup>a</sup>  $\rho_a$  ( $= n_c/\pi\sigma_c^2$ ) is the grafting density, and  $\rho_a^*$  ( $= \rho_a \pi R_g^2$ ) is the reduced grafting density.  $n_c$  is the number of chains grafted on each colloidal particle.  $R_g$  and  $R_{EED}$  are the radius of gyration and end-to-end distance of the end-grafted polymer chains. <sup>b</sup> “Rep” indicates a purely repulsive interaction obtained by using only the repulsive term in the Lennard-Jones 6–12 potential.

attempted to circumvent this by running each simulation multiple times from different initial configurations.

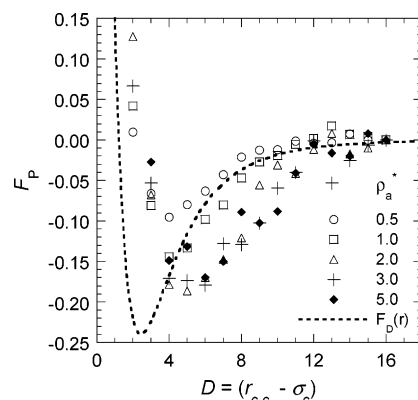
Equilibration required  $(20-40) \times 10^6$  Monte Carlo steps (where each step corresponds to one attempted move) depending on the grafting density and chain length. “Equilibrium” was defined as occurring when the energy approached an asymptote and fluctuated less than 2% and did so in a repeatable manner from different initial states. Average forces, chain conformational properties, and segment density profiles were collected over  $(100-160) \times 10^6$  attempted moves at each separation distance. The statistical error for the conformational properties of the chains was taken as the root-mean-square fluctuation divided by the square root of the number of independent blocks. Statistical error in the forces was based on the statistical inefficiency parameter discussed by Allen and Tildesley.<sup>34</sup> Error bars were omitted from the figures if they were smaller than the symbols representing the data points. Table 1 gives the details of all of the cases that were investigated in this study.

### 3. Results and Discussion

**Effect of Grafting Density ( $\rho_a^*$ ).** Figure 2 shows the effect of reduced grafting density ( $\rho_a^*$ ) on the polymer-induced forces ( $F_p$ ) between the nanoparticles at  $\sigma_c = 10\sigma_p$ ,  $N = 30$ ,  $\epsilon_{cp} = \epsilon_{pp} = 1$ . The direct LJ dispersion force between the uncoated nanoparticles obtained by

$$F_D(r) = -\frac{\partial U_{cc}(r)}{\partial r} = \frac{48}{r} \left[ \left( \frac{\sigma_c}{r} \right)^{12} - 0.5 \left( \frac{\sigma_c}{r} \right)^6 \right] \quad (2)$$

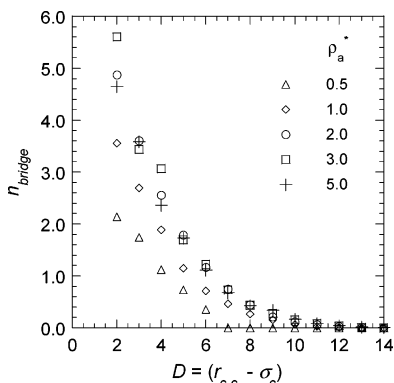
is also shown for comparison. In all force profiles reported in this paper, the  $x$  coordinate is taken to be  $D = (r_{c-c} - \sigma_c)$ , which is the nearest surface-to-surface distance between the nanoparticles. For all grafting densities, the force–distance plots show long-ranged attraction and short-ranged repulsion, similar in form and comparable in magnitude to the direct nanoparticle force. The attractive minimum grows deeper and shifts to larger separations with increasing



**Figure 2.** Force–distance plots [polymer-mediated force  $F_p(r)$  vs  $D$ ] at different grafting densities ranging from  $\rho_a^* = 0.5$  to  $\rho_a^* = 5.0$ .  $\sigma_c = 10\sigma_p$ ,  $N = 30$ , and  $\epsilon_{cp} = \epsilon_{pp} = 1$ . The direct force between the bare nanoparticles ( $F_D$ ) calculated from eq 2 is also shown for comparison.

grafting densities up to  $\rho_a^* = 3.0$ , beyond which the minimum shifts back upward (for  $\rho_a^* = 5.0$ ). The nature and magnitude of the polymer-induced forces between the nanoparticles are a consequence of the competition between attraction due to bridging and intersegmental interactions and repulsion due to steric interactions between the grafted chains. Bridging occurs because of attractive segment–surface interactions and the presence of exposed adsorption sites on the nanoparticle surface. At  $\rho_a^* = 0.5$ , the polymer is in a pancakelike conformation wherein the average distance between the grafting sites is more than the  $R_g$  of the polymer chains and the grafted polymers tend to lie flat on the particle surface because of favorable particle–polymer contacts. Although the surface coverage is low (which is favorable for bridging), the small number of grafted chains results in a very low density of bridges formed. In addition, the flattened conformations of the grafted polymers also constrain the amount of bridging that can occur. As  $\rho_a^*$  is increased beyond  $\rho_a^* = 1.0$ , the polymer chains undergo a gradual transition to the brush regime, wherein the excluded volume considerations cause the polymer chains to extend outward from the surface. It is noteworthy that the pancake (or mushroom) to brush transition for short chains (low molecular weight polymers) is not a very sharp transition and occurs over a broad range of grafting densities.<sup>35</sup> With increasing  $\rho_a^*$ , there are more chains available that can form bridges between the two particles. However, there is a progressive decrease in the availability of binding sites on the surface and also an increase in the monomer density in the interparticle region. These competing effects suggest that there is a threshold value of grafting density beyond which the number of bridging chains would be limited by entropic effects.

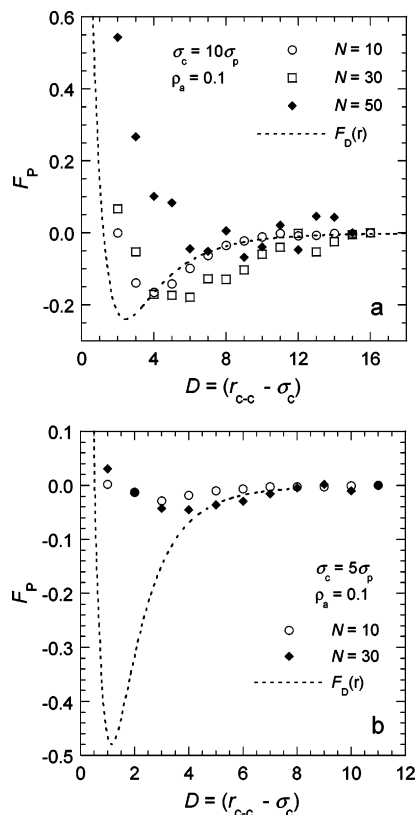
Figure 3 shows the average number of bridging chains ( $n_{\text{bridge}}$ ) versus the separation distance ( $D$ ) at different grafting densities corresponding to Figure 2. The plots indicate that there is a maximum in  $n_{\text{bridge}}$  at a grafting density of  $\rho_a^* = 3.0$ , and a further increase in the grafting density results in a slight decrease in  $n_{\text{bridge}}$ . The grafting density at which the polymer-induced attraction between the nanoparticles is strongest coincides with the maximum in  $n_{\text{bridge}}$  (at  $\rho_a^* =$



**Figure 3.** Number of bridging chains versus surface-to-surface separation distance ( $D = r_{c-c} - \sigma_c$ ) at different grafting densities ranging from  $\rho_a^* = 0.5$  to  $\rho_a^* = 5.0$ .  $\sigma_c = 10\sigma_p$ ,  $N = 30$ , and  $\epsilon_{cp} = \epsilon_{pp} = 1$ .

3.0), suggesting that bridging is responsible primarily for the attraction in the force profiles. Also, at a constant grafting density, the number of bridging chains increases with decreasing interparticle separation, suggesting a monotonic attraction in the force profiles. However, at close separations, steric effects dominate the bridging attraction resulting in polymer-induced repulsion between the nanoparticles. While the intersegmental interactions between the grafted polymer chains also contribute to the polymer-induced attraction, it is difficult to quantify exactly the relative contributions of bridging and segmental interactions to the overall attraction. The effect of polymer segment–segment interactions on the polymer-induced forces is explored in greater detail later in this section. The decrease in the magnitude of the attractive minimum as  $\rho_a^*$  is increased from  $\rho_a^* = 3.0$  to  $\rho_a^* = 5.0$  suggests that further increasing the grafting density may eventually result in the repulsive entropic effects dominating the attraction leading to monotonic repulsion in the force profiles. The separation distance between the nanoparticles at which  $F_P$  changes sign (i.e.,  $F_P = 0$ ) and the location of the attractive minimum are related closely to the conformations of the end-grafted polymers. The mean-squared end-to-end distance ( $R_{EED}^2$ ) and the radius of gyration squared ( $R_g^2$ ) of the polymer chains at different grafting densities are shown in Table 1. The reported values of  $R_{EED}^2$  and  $R_g^2$  are obtained by averaging over all nanoparticle separation distances at each grafting density. The bulk values of  $R_{EED}^2$  and  $R_g^2$  (for  $N = 30$ ) are 53.2 and 8.96, respectively. For grafting densities up to  $\rho_a^* = 3.0$  (at  $\epsilon_{cp} = \epsilon_{pp} = 1$ ),  $R_{EED}^2$  and  $R_g^2$  are less than the bulk values due to the attractive interactions between the nanoparticle and the grafted chains, which cause the polymers to adopt more compact conformations near the particle. However at  $\rho_a^* = 5.0$ ,  $R_{EED}^2$  is more than the bulk value, strengthening the argument that the excluded volume effects begin to dominate at this grafting density. The monotonic increase in  $R_{EED}^2$  and  $R_g^2$  with increasing  $\rho_a^*$  is responsible for the shift in the location of the attractive minimum and the thermodynamic minimum (location at which  $F_P = 0$ ) to larger separation distances.

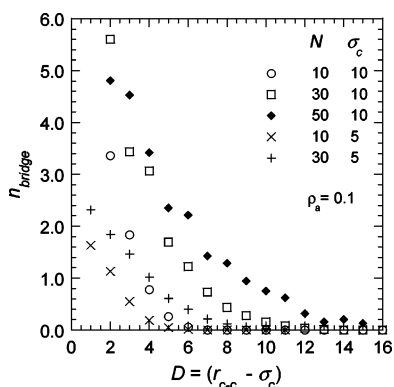
**Effect of Particle Diameter ( $\sigma_c$ ) and Chain Length ( $N$ ).** Figure 4 shows a plot of force due to the grafted chains ( $F_P$ ) between two nanoparticles for diameters of  $\sigma_c = 10\sigma_p$  (Figure



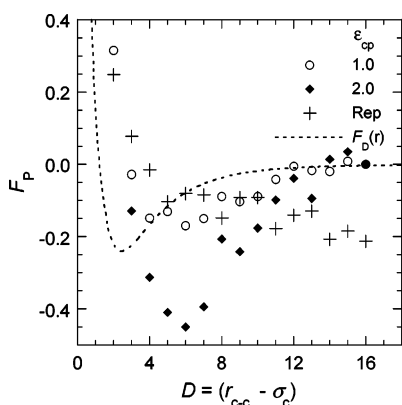
**Figure 4.** Force-distance plots [polymer mediated force  $F_P(r)$  vs  $D$ ] for diameters of  $\sigma_c = 10\sigma_p$  (a) and  $5\sigma_p$  (b) at different chain lengths.  $\epsilon_{cp} = \epsilon_{pp} = 1$ . The grafting density is constant at  $\rho_a = 0.1$  for all cases. Note however that  $\rho_a^*$  varies with  $\sigma_c$  and  $N$  (see Table 1). The direct force between the bare nanoparticles ( $F_D$ ) calculated from eq 2 is also shown for comparison.

4a) and  $5\sigma_p$  (Figure 4b) at different chain lengths and  $\epsilon_{cp} = \epsilon_{pp} = 1$ . The effect of chain length was studied by keeping the number of grafted chains per particle constant at all chain lengths. This implies that, although the grafting density  $\rho_a$  ( $= n_c/\pi\sigma_c^2$ ) remained constant at each particle size ( $\rho_a = 0.1$ ), the reduced grafting density  $\rho_a^*$  ( $= \rho_a\pi R_g^2$ ) varied depending on the chain length. At  $\sigma_c = 10\sigma_p$ , the attractive minimum becomes deeper and shifts to a larger distance as  $N$  increases from  $N = 10$  to  $N = 30$ . Further increasing the chain length to  $N = 50$  results in a polymer-induced force that oscillates between attraction and repulsion for  $D > 6\sigma_p$  and that is purely repulsive for  $D < 6\sigma_p$ . In addition, the magnitude of the attractive minimum in  $F_P$  for  $N = 50$  is less compared to that for  $N = 30$ . Figure 5 shows  $n_{\text{bridge}}$  versus the separation distance for all of the  $N$  and  $\sigma_c$  values corresponding to the force–distance curves in Figure 4. For  $\sigma_c = 10\sigma_p$ ,  $n_{\text{bridge}}$  is higher for longer chains, suggesting that attraction due to bridging increases monotonically with chain length. However, as chain length is increased at constant  $\rho_a$ , there is a larger number of polymer segments in the region between the two particles, resulting in increased excluded volume effects. The force plots suggest that, at  $N = 50$ , these entropic effects tend to dominate the increased attraction due to bridging and intersegmental interactions, resulting in relatively long-range repulsion and a higher attractive minimum. At  $\sigma_c = 5\sigma_p$ , the force due to the polymer is very





**Figure 5.** Number of bridging chains versus surface-to-surface separation distance ( $D = r_{c-c} - \sigma_c$ ) at different  $\sigma_c$  and  $N$  values corresponding to the force profiles in Figure 4. The grafting density is constant at  $\rho_a = 0.1$  and  $\epsilon_{cp} = \epsilon_{pp} = 1$ .

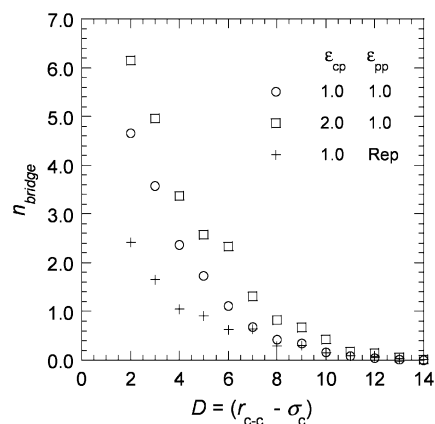


**Figure 6.** Force-distance plots [polymer-mediated force  $F_P(r)$  vs  $D$ ] for different values of the colloid–polymer interaction strength ( $\epsilon_{cp}$ ).  $\sigma_c = 10\sigma_p$ ,  $N = 30$ ,  $\rho_a^* = 5.0$ , and  $\epsilon_{pp} = 1$  for all cases. “Rep” stands for the purely repulsive interaction obtained by using only the repulsive term in the Lennard-Jones 6–12 potential. The direct force between the bare nanoparticles ( $F_D$ ) calculated from eq 2 is also shown for comparison.

small compared to the direct nanoparticle force for both chain lengths ( $N = 10$  and  $N = 30$ ). For a constant grafting density ( $\rho_a = 0.1$ ), the number of grafting chains decreases with decreasing particle size, and hence, at  $\sigma_c = 5\sigma_p$ , the polymer-induced attraction is weaker compared to that at  $\sigma_c = 10\sigma_p$ . In addition, the decrease in the range of the LJ potential energy with decreasing  $\sigma_c$  results in a lesser number of bridging chains (shown in Figure 5) and consequently a weaker polymer-induced attraction.

#### Effect of Colloid–Polymer Interaction Parameter ( $\epsilon_{cp}$ ).

Figure 6 explores the effect of tuning the strength of the colloid–polymer interaction on the force–distance plots at  $\sigma_c = 10\sigma_p$ ,  $N = 30$ ,  $\epsilon_{pp} = 1$ , and  $\rho_a^* = 5.0$ . To ensure that the polymers are in the brush regime, we use the highest grafting density explored in our simulations ( $\rho_a^* = 5.0$ ). In addition, we also ensured that the initial grafting sites were the same for all of the cases studied here. As  $\epsilon_{cp}$  is increased from  $\epsilon_{cp} = 1\epsilon_{pp}$  to  $\epsilon_{cp} = 2\epsilon_{pp}$ , the attractive minimum grows deeper because of an increase in the number of bridging chains. The increased colloid–polymer interaction strength causes an increase in  $n_{\text{bridge}}$ , shown in Figure 7. Also,  $R_{\text{EED}}^2$

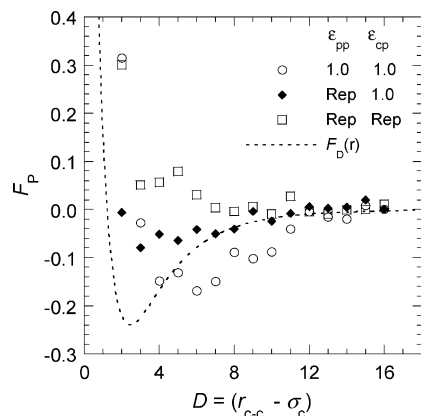


**Figure 7.** Number of bridging chains versus surface-to-surface separation distance ( $D = r_{c-c} - \sigma_c$ ) at  $\sigma_c = 10\sigma_p$ ,  $N = 30$ , and different values of  $\epsilon_{cp}$  and  $\epsilon_{pp}$ . The grafting density is constant at  $\rho_a^* = 5.0$ . “Rep” stands for the purely repulsive interaction obtained by using only the repulsive term in the Lennard-Jones 6–12 potential.

and  $R_g^2$  decrease (brush contracts) with increasing  $\epsilon_{cp}$ , which results in the transition from attractive to repulsive force occurring at shorter separations. There is no significant difference, however, in the location of the attractive minimum.

We also studied the effect of turning off the colloid–polymer attraction by using only the repulsive part of the LJ 6–12 potential ( $\epsilon_{cp} = \text{“rep”}$ ) to model the colloid–polymer interactions. This allows us to eliminate attraction due to bridging and isolate the effect of polymer intersegmental attractions on the force profiles. Figure 6 shows that the force due to the polymers oscillates slightly (with a periodicity of  $\approx 3\sigma_p$ ) but grows increasingly repulsive as the particles approach each other. It is noteworthy that, for all other cases explored in this paper (wherein  $F_P$  is attractive at some distance), the force due to the polymers passes through a minimum. The occurrence of negative force at a large distance (for  $\epsilon_{cp} = \text{“rep”}$ ) is due to the repulsive colloid–polymer interactions, which drives the grafted chains to extend out as far as possible from the particle surface. Note that the  $R_{\text{EED}}^2$  and  $R_g^2$  values for this case are larger than those for any of the attractive  $\epsilon_{cp}$  cases (Table 1). In addition, no bridging attraction was observed for this case ( $n_{\text{bridge}} = 0$ ). The only contribution to the negative force is from the polymer intersegmental attraction. (It must be noted that the repulsion between the colloid and the polymer segments enhances the range and magnitude of the intersegmental attractions compared to the case when  $\epsilon_{cp} = \epsilon_{pp} = 1$ .) However, at  $\epsilon_{cp} = \epsilon_{pp}$ , the polymer segments experience favorable attractive interactions with both the colloid and each other. This reduces the relative contribution of intersegmental attractions to the overall attraction.

**Effect of Polymer–Polymer Interaction Parameter ( $\epsilon_{pp}$ ).** Figure 8 shows the force profiles at  $\epsilon_{pp} = 1$  and  $\epsilon_{pp} = \text{“rep”}$  (soft-repulsive polymer segment interactions). The colloid–polymer interaction strength is kept constant at  $\epsilon_{cp} = 1$  for both cases. When the polymer–polymer interactions are purely repulsive, the magnitude of polymer-induced attraction is lower compared to that in the case of



**Figure 8.** Force-distance plots [polymer mediated force  $F_p(r)$  vs  $D$ ] for different values of the polymer–polymer interaction strength ( $\epsilon_{pp}$ ).  $\sigma_c = 10\sigma_p$ ,  $N = 30$ , and  $\rho_a^* = 5.0$  for all cases. “Rep” stands for the purely repulsive interaction obtained by using only the repulsive term in the Lennard-Jones 6–12 potential. The direct force between the bare nanoparticles ( $F_D$ ) calculated from eq 2 is also shown for comparison.

$\epsilon_{cp} = \epsilon_{pp} = 1$  because there is no longer any contribution of intersegmental attractions to the overall force profile. The attraction that is observed in this case is solely due to bridging (because  $\epsilon_{cp} = 1$ ). However, the bridging attraction observed here is less than that in the case of  $\epsilon_{cp} = \epsilon_{pp} = 1$  (see Figure 7) because the polymer–polymer repulsion limits the number of polymer chains that interpenetrate the opposite grafted layer. These results suggest that there is a complex interplay of the effects of  $\epsilon_{cp}$  and  $\epsilon_{pp}$ , and both these parameters need to be accounted for explicitly in order to accurately model organically modified nanoparticle systems. We also calculated the force profiles for the limiting case of repulsive interactions between both colloid–polymer and polymer–polymer interactions. This represents the condition of athermal polymer chains grafted to a hard colloid surface. As expected, the polymer-induced forces are purely repulsive, in agreement with previous simulation results of interactions between spherical brushes modeled using the hard-sphere potential.<sup>36</sup> In addition, the polymer chains are in the most stretched-out conformations as the  $R_{EED}^2$  and  $R_g^2$  values for this case are larger than those for any other condition at  $\rho_a^* = 5.0$  (see Table 1). Also, comparing this athermal system with the case of  $\epsilon_{cp} = \epsilon_{pp} = 1$  at the same grafting density ( $\rho_a^* = 5.0$ ) suggests that relatively higher grafting densities would be required for the steric stabilization of nanoparticles when the colloid–polymer and polymer–polymer dispersion interactions are taken into account.

Attractive forces between polymer brushes are not usually observed in experimental studies. Most previous experiments report a monotonically increasing repulsive force which decreases exponentially with increasing distance. However, bridging attraction is sometimes observed in the case of *telephonic* (functionalized at both ends) polymers at low surface coverage.<sup>27</sup> These experimental studies are invariably in the limit of flat planar brushes. Recently, Goodman and co-workers<sup>20</sup> used the AFM to measure the interactions between spherical polymer brushes and a silicon nitride tip in aqueous media. They observed a long-ranged bridging

force at a low grafting density and a monotonically increasing repulsive force at higher grafting densities. In addition, the grafting density at which the bridging force was observed also varied with the polymer chemistry. The authors suggest that, in addition to the polydispersity of the polymer that causes the AFM tip to infrequently sample long polymer chains, the curvature of the latex particles may also be partially responsible for the anomalous bridging attraction. While the experimental system is different from our simulations in that we calculate the interaction force between two brushes as compared to the AFM experiments which measure the force between a brush and a surface, a qualitative comparison could still be made. The effect of different polymers used in the experiments could be compared to the effect of varying the colloid–polymer interaction parameter ( $\epsilon_{cp}$ ) in our simulations. From Figure 6, at a constant grafting density, the attractive minimum in the force profile grows deeper with increasing  $\epsilon_{cp}$ , suggesting that a higher grafting density would be required for stabilization. A similar observation has been reported in the AFM study wherein poly(*N*-isopropyl acrylamide) exhibited a stronger affinity for the AFM tip than poly(*N,N*-dimethyl acrylamide) at the same grafting density. These results underscore the importance of incorporating van der Waals dispersion interactions in models and theories for studying nanoparticle–polymer systems.

#### 4. Conclusions

The force between two nanoscale colloidal particles coated with end-grafted polymeric modifiers has been calculated using continuum Monte Carlo simulations in the NVT ensemble. The inclusion of LJ dispersion interactions results in long-ranged attraction as well as short-ranged repulsion in the  $F_p$  profiles. Polymer-induced attraction was observed under good solvent (or temperature) conditions for the polymeric modifier and when there was no energetic preference for segment–segment versus segment–surface interactions, for example, LJ terms  $\epsilon_{cp} = \epsilon_{pp}$ . This is also before the effect of particle–particle (bare surface) attraction is considered. Apparently, the polymer-induced attraction is attributed to both bridging and intersegmental interactions, the relative contributions of which can be adjusted with  $\epsilon_{cp}$  and  $\epsilon_{pp}$ . The increased volume available to chain segments at highly curved surfaces may lead to an increased significance of polymer–particle attraction. A minimum  $N$  and  $\rho_a^*$  is required for nanoparticle stabilization even under good solvent conditions. Beyond the minimum  $N$  and  $\rho_a^*$ , excluded volume effects begin to dominate the attractive forces, leading ultimately to purely repulsive forces (and thereby stabilization of the nanoparticles).

These results contrast with what has been observed previously in grafted-chain simulations and models for flat or curved surfaces neglecting attractive interactions between the grafted polymer and the particle surface. The grafted-polymer dispersion attraction with the surface of highly curved nanoparticles is physically significant because it causes polymer-induced attraction in situations where larger, lower-curvature colloids would remain repulsive. We note that this result has been shown here only for the LJ attraction,

which certainly is far from reality in many experimental systems. However, the results nevertheless should motivate further study of the effects of grafted modifier–particle attraction on particle stability models, and the applicability of flat-brush models to nanoparticle stabilization.

The ability of end-grafted polymers to induce both attractive and repulsive interactions between the nanoparticles makes them particularly useful for either directed-assembly or stabilization purposes. The location and depth of the attractive minimum in the force profiles can be controlled by varying the polymer chain length, grafting density, and the relative values of colloid–polymer and polymer–polymer interaction strength. In addition, these parameters also control the particle equilibrium spacing which corresponds to  $F_P = 0$ . Thus, by tuning the various molecular parameters ( $\rho_a^*$ ,  $N$ ,  $\epsilon_{cp}$ , and  $\epsilon_{pp}$ ) of end-grafted modifiers, it is perhaps possible to achieve precise control over nanoparticle alignment and spacing in thermodynamically driven assembly processes.

Freely adsorbing homopolymer modifiers studied in part I of this series also showed a similar dependence of the polymer-induced forces on particle diameter, polymer chain length, and colloid–polymer interaction strength. However, over the range of chain lengths,  $\epsilon_{cp}$  and  $\epsilon_{pp}$ , that were investigated, it was found that obtaining repulsion with homopolymer modifiers was significantly more difficult. Hence, adsorbing homopolymers are not suited ideally for stabilization purposes.

### References

- (1) Marla, K. T.; Meredith, J. C. *Langmuir* **2005**, *27*, 487–497.
- (2) Alexander, S. *J. Phys. (Paris)* **1977**, *38*, 983–987.
- (3) de Gennes, P.-G. *Macromolecules* **1980**, *13*, 1069–1075.
- (4) Milner, S. T.; Witten, T. A. *J. Phys. (Paris)* **1988**, *49* (11), 1951–62.
- (5) Milner, S. T.; Witten, T. A.; Cates, M. E. *Macromolecules* **1988**, *21* (8), 2610–19.
- (6) Milner, S. T.; Witten, T. A.; Cates, M. E. *Europhys. Lett.* **1988**, *5* (5), 413–18.
- (7) Milner, S. T.; Witten, T. A.; Cates, M. E. *Macromolecules* **1989**, *22* (2), 853–61.
- (8) Zhulina, E. B.; Borisov, O. V.; Priamitsyn, V. A. *J. Colloid Interface Sci.* **1990**, *137* (2), 495–511.
- (9) Wijmans, C. M.; Scheutjens, J. M. H. M.; Zhulina, E. B. *Macromolecules* **1992**, *25*, 2657–2665.
- (10) Toral, R.; Chakrabarti, A.; Dickman, R. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1994**, *50* (1), 343–8.
- (11) Lai, P. Y.; Binder, K. *J. Chem. Phys.* **1991**, *95* (12), 9288–99.
- (12) Murat, M.; Grest, G. S. *Phys. Rev. Lett.* **1989**, *63* (10), 1074–7.
- (13) Murat, M.; Grest, G. S. *Macromolecules* **1989**, *22* (10), 4054–9.
- (14) O’Shea, S. J.; Welland, M. E.; Rayment, T. *Langmuir* **1993**, *9*, 1826–1835.
- (15) Yamamoto, S.; Ejaz, M.; Tsujii, Y.; Fukuda, T. *Macromolecules* **2000**, *33* (15), 5608–5612.
- (16) Yamamoto, S.; Ejaz, M.; Tsujii, Y.; Matsumoto, M.; Fukuda, T. *Macromolecules* **2000**, *33* (15), 5602–5607.
- (17) Yamamoto, S.; Tsujii, Y.; Fukuda, T. *Macromolecules* **2000**, *33* (16), 5995–5998.
- (18) Taunton, H. J.; Toprakcioglu, C.; Fetters, L. J.; Klein, J. *Nature* **1988**, *332* (6166), 712–714.
- (19) Taunton, H. J.; Toprakcioglu, C.; Fetters, L. J.; Klein, J. *Macromolecules* **1990**, *23*, 571–580.
- (20) Goodman, D.; Kizhakkedathu, J. N.; Brooks, D. E. *Langmuir* **2004**, *20*, 2333–2340.
- (21) Daoud, M.; Cotton, J. P. *J. Phys. (Paris)* **1982**, *43* (3), 531–8.
- (22) Witten, T. A.; Pincus, P. A. *Macromolecules* **1986**, *19* (10), 2509–13.
- (23) Toral, R.; Chakrabarti, A. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1993**, *47* (6), 4240–6.
- (24) Lin, E. K.; Gast, A. P. *Macromolecules* **1996**, *29* (1), 390–7.
- (25) Wijmans, C. M.; Leermakers, F. A. M.; Fleer, G. J. *Langmuir* **1994**, *10* (12), 4514–16.
- (26) Cerda, J. J.; Sintes, T.; Toral, R. *Macromolecules* **2003**, *36*, 1407–1413.
- (27) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press: London, 1953.
- (28) Roan, J.-R. *Phys. Rev. Lett.* **2001**, *86* (6), 1027–1030.
- (29) Baranowski, R.; Whitmore, M. D. *J. Chem. Phys.* **1995**, *103* (6), 2343–2353.
- (30) Kreer, T.; Metzger, S.; Muller, M.; Binder, K.; Baschnagel, J. *J. Chem. Phys.* **2003**, *120*, 4012–4023.
- (31) Siepmann, J. I.; Frenkel, D. *Mol. Phys.* **1992**, *75*, 59–70.
- (32) de Pablo, J. J.; Laso, M.; Suter, U. W.; Cochran, H. D. *Fluid Phase Equilib.* **1993**, *83*, 323–331.
- (33) Mezei, M.; Beveridge, D. L. *Ann. N. Y. Acad. Sci.* **1986**, *482*, 1–21.
- (34) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, U. K., 1987.
- (35) Carignano, M. A.; Szleifer, I. *Macromolecules* **1995**, *28* (9), 3197–3204.
- (36) Eiser, E.; Klein, J.; Witten, T. A.; Fetters, L. J. *Phys. Rev. Lett.* **1999**, *82*, 5076–5079.

## Radical Addition to Thioketones: Computer-Aided Design of Spin Traps for Controlling Free-Radical Polymerization

Hugh Chaffey-Millar,<sup>†</sup> Ekaterina I. Izgorodina,<sup>‡</sup> Christopher Barner-Kowollik,<sup>\*,†</sup> and Michelle L. Coote<sup>\*,‡</sup>

*Centre for Advanced Macromolecular Design, School of Chemical Sciences and Engineering, The University of New South Wales, Sydney, New South Wales 2052, Australia, and ARC Centre of Excellence in Free-Radical Chemistry and Biotechnology, Research School of Chemistry, Australian National University, Canberra, Australian Capital Territory 0200, Australia*

Received April 6, 2006

**Abstract:** An extensive study has been undertaken of the radical affinity of a number of thioketones ( $S=C(X)(Y)$ ) with the aim of selecting combinations of X and Y that render the substrate suitable for the mediation of free radical polymerizations. Using high level ab initio molecular orbital calculations, enthalpies at 0 K were determined for the reactions  $R\cdot + S=C(X)(Y) \rightarrow R-S-C\cdot(X)(Y)$  for  $R\cdot = CH_3, CH_2OH, CH_2CN$ , and benzyl, in reactions with a variety of thioketones, including various combinations of X and Y taken from H,  $CH_3$ , Ph, CN,  $OCH_3$ ,  $C(CH_3)_3$  and *para*-CN-Ph as well as several compounds in which the X and Y are bonded, namely xanthene-9-thione, fluorine-9-thione, and cyclopenta[*def*]phenanthrene-4-thione. The radical affinities of the various thioketones has been discussed in terms of the radical stabilization energies (RSEs) of the adduct radicals and stabilities of the  $S=C$  bonds. From these studies, the two thioketones  $S=C(CN)(Ph)$  and fluorene-9-thione were selected as being potentially suitable candidates for use in controlling free radical polymerizations due to their high radical affinities. However, based on transition state theory calculations of the rate coefficients for homo/copolymerization of  $S=C(CN)(Ph)$  with itself and styrene at 333.15 K, this substrate was deemed to be unsuitable, as it was likely to undergo side reactions. Instead, the more-hindered fluorine-9-thione was identified as the ideal thioketone, and the equilibrium constants at 333.15 K for the reactions of the styryl and vinyl acetate dimer radicals with fluorine-9-thione were made. These two reactions, at 333.15 K, displayed equilibrium constants in the vicinity of  $10^{14} \text{ L mol}^{-1}$  and  $10^{16} \text{ L mol}^{-1}$ , respectively, indicating that there is significant scope within the thioketone class of compounds to mediate free radical polymerizations via radical stability alone.

### Introduction

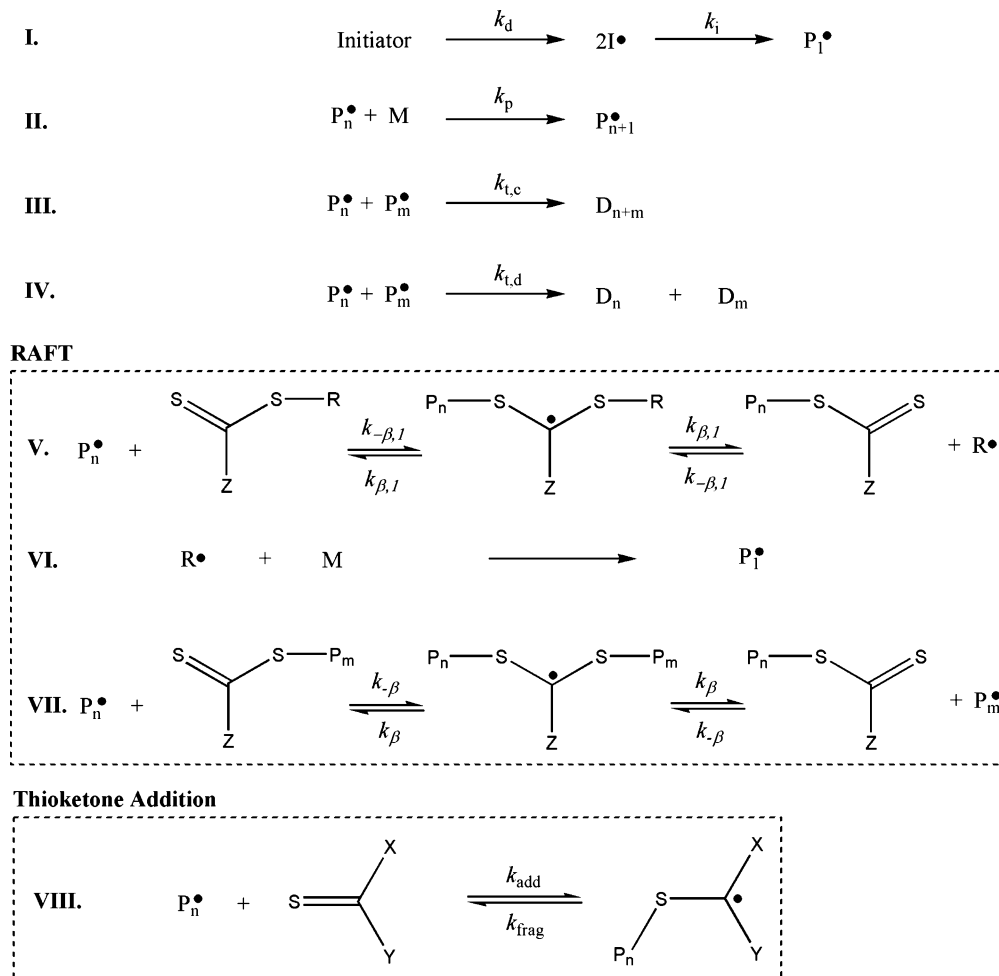
During the past two decades there has been much interest in the use of free radicals as a reactive center for the polymerization of a variety of monomers. Conventionally, this

process has resulted in average molecular weights that are approximately constant with conversion. More recently, a number of methods have been devised to control free-radical polymerization processes such that (in the ideal case) the average molecular weight grows linearly with conversion. An additional benefit of these controlled radical polymerizations (CRP) is the ability to tailor the chemical structure of chain ends and to reduce the polydispersity index (PDI) from

\* Corresponding author e-mail: c.barner-kowollik@unsw.edu.au (C.B.K.) and mcoote@rsc.anu.edu.au (M.L.C.).

<sup>†</sup> The University of New South Wales.

<sup>‡</sup> Australian National University.

**Scheme 1.** Reactions of a Generic Radical Polymerization as Well as Those of the RAFT and Thioketone Addition Processes

values of 1.5–2.0 (for conventional radical polymerization) down to values approaching 1.05 (for CRP).

Three of the most successful techniques which have emerged are nitroxide mediated polymerization (NMP),<sup>1</sup> atom transfer radical polymerization (ATRP),<sup>2,3</sup> and reversible addition-fragmentation chain-transfer polymerization (RAFT).<sup>4,5</sup> All of these techniques afford living behavior by maintaining a population of dormant chains that is high in concentration relative to the number of radicals and hence the number of terminated chains that accumulate. The first two methods, NMP and ATRP, achieve this by reduction of the radical concentration. RAFT, however, can operate successfully with a radical concentration that is the same as a conventional radical polymerization—control is instead achieved through the creation of a population of dormant chains that is superimposed on that of its radical counterparts. The technique can, however, benefit from a reduction in radical concentration, and this is brought about with the use of RAFT agents that form an intermediate that is relatively stable toward fragmentation. More specifically, and with reference to Scheme 1, this stability is reflected by a high equilibrium constant  $K = k_{-\beta}/k_{\beta}$  in reaction VII. It should be pointed out that this stability has been the subject of considerable scientific debate;<sup>6,7,10</sup> however, there is a body of consistent experimental and theoretical evidence that

suggests that high radical stability is a physical reality when the RAFT agent is substituted with radical-stabilizing Z-groups, such as phenyl.<sup>8–11</sup>

In a recent communication,<sup>12</sup> it was shown that molecules with radical stabilizing (but not chain transfer) features in their substrate can lead to controlled behavior in the polymerization of styrene and that a reversible radical trapping process is a possible mechanism for the origin of the characteristic increase of average molecular weight with conversion. The experiments featured a large initial increase in molecular weight followed by a highly linear subsequent increase in molecular weight with conversion. This was akin to what is termed hybrid behavior in the RAFT process and, in the reversible trapping mechanism, would be expected to occur if the equilibrium constant ( $K = k_{\text{add}}/k_{\text{frag}}$ ) for the reversible trapping reaction (Scheme 1.VIII) fell into the approximate range of  $10^5$ – $10^6$  L mol<sup>-1</sup>. High level ab initio calculations for the addition of model polystyryl radicals to the di-*tert*-butyl thioketone agent (Scheme 1.VIII) indicated that the equilibrium constant ( $7.0 \cdot 10^4$  L mol<sup>-1</sup>) did indeed fall into this approximate range, providing some support for the proposed mechanism.<sup>12</sup>

Although controlled radical polymerization has been achieved using thioketone spin traps, for a practical polymerization process it would be desirable to eliminate the hybrid

behavior and in this way improve the molecular weight control. To this end, it is necessary to identify thioketone spin traps for which the equilibrium constants are significantly higher. However, the equilibrium constant should not be too large, since this can lead to severe retardation/inhibition phenomena. Kinetic studies of the related nitroxide mediated free-radical polymerizations (NMP, which is based on the same key idea as the present process but uses a radical spin trap) indicate that  $K$  should be around  $10^9 \text{ L mol}^{-1}$  in order to effect an efficient living process.<sup>6</sup> There are various other characteristics that are desirable if thioketones are to be used in controlling polymerizations. As well as a high equilibrium constant for the reversible radical trapping process, chain transfer to and copolymerization of the thioketone compounds must be absent. For this reason it is likely that substrates with bulky substituents (such as phenyl or *tert*-butyl) might be practically suitable. Additionally, since the dormant species bears a radical, it is essential that termination of two dormant species does not occur to a significant extent. This is another reason bulky substituents are likely to be advantageous. Finally, we note that it is important that the rate coefficient for addition of the propagating radical to the thioketone is substantially higher than that of the monomer. As in the case of RAFT polymerization, thioketones are suitable control agents in this respect because C=S  $\pi$  bonds are considerably weaker than C=C bonds, due to the poorer overlap between the p-orbitals on first row carbon and second row sulfur atoms, and consequently their singlet-triplet excitation energies are considerably smaller.<sup>13,14</sup>

The fact that the dormant species in thioketone-mediated polymerizations bear radicals also brings about the complication that at the conclusion of the polymerization, these species may still be "active". It might be possible in the future to devise some chemical means to terminate these active species, analogous to the quenching of anionic polymerization mediums with a carboxylic acid to give all chains a hydrogen end group. In this regard there is a greater degree of complication as compared to RAFT. Another complicating factor in the radical spin trapping process is that all polymerizing chains originate from initiator fragments. This is in contrast to the RAFT process, one of whose crucial features is the fragmenting and reinitiating R-group. As a result, rather than most chains having nearly the same length, molecular weights will be more evenly distributed between the maximum and minimum values present in the system, meaning that with a slowly decomposing initiator, very low polydispersities (<1.3) cannot be achieved.

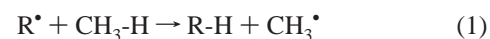
Despite these challenges, there are several reasons for interest in this polymerization system. First it may in the future be possible to provide some degree of control over previously difficult to control monomers since the presence of only a single thiyl group provides greater flexibility for fine-tuning the reactivity of the molecular substrate. Second, there is academic interest in that it offers an experimental scenario in which it should be possible to elucidate relative radical stabilities of propagating and trapped radicals. Third, with a greater understanding of the effect of the above two variables, it might be possible to adjust the RAFT process,

extending its control to a wider variety of systems, such as those involving monomers with very high or very low reactivities.

To facilitate the further experimental study of this new process, the purpose of the current paper is to use high-level ab initio molecular orbital calculations to design improved thioketone spin traps for controlling free-radical polymerization. We initially survey the effects of substituents on the enthalpies for the trapping of various radicals by a range of thioketone molecules, so as to identify combinations of substituents that are likely to maximize the equilibrium constant of the trapping reaction. This complements and extends our earlier studies of substituent effects in the RAFT process, for which the presence of the additional thiyl substituent was found to dominate the structure-reactivity trends.<sup>11-17</sup> In the present work it is possible to vary both thiocarbonyl substituents and in this way target a wider range of stabilities and reactivities. Having identified potential combinations of substituents, we then select the thioketone spin traps most likely to resist copolymerization and radical-radical termination on the basis of steric factors. We then evaluate their likely performance in a polymerizing system through direct calculation of their equilibrium constants for trapping of model propagating radicals and also their copolymerization propagation rate coefficients.

## Theoretical Procedures

**Definitions.** In the present work we study the effects of substituents on the stabilities of the adduct-radicals using the standard radical stabilization energies (RSEs).<sup>18,19</sup> For a radical  $R^\bullet$ , it is defined as the enthalpy of the hydrogen abstraction reaction

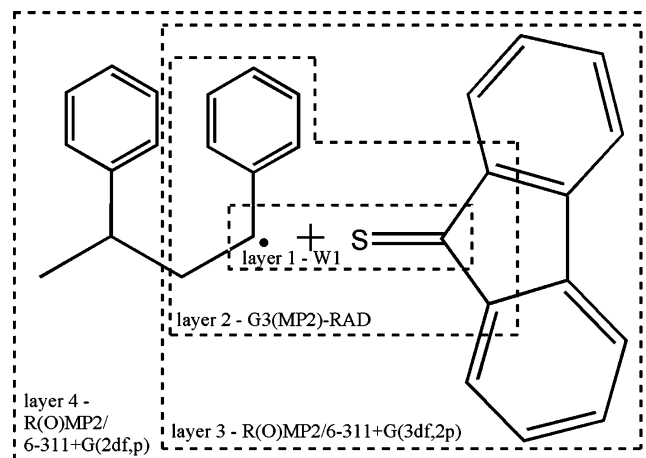


As discussed previously,<sup>15,20,21</sup> in using the RSE to measure the stability of the radical we make the implicit assumption that the effect of the substituent on the stability of the closed shell compound is negligible and therefore cancels. While this assumption can be invalid in some situations (such as the study of phosphoranyl radicals<sup>20</sup>), it has been found to provide a reasonable qualitative guide to the stability of carbon-centered radicals.<sup>15,21</sup> Within this framework, a positive value indicates that  $R^\bullet$  is more stable than  $\text{CH}_3^\bullet$  and vice versa.

The thiocarbonyl bond stability in a compound  $\text{S}=\text{CXY}$  can be estimated relative to that of thioformaldehyde ( $\text{S}=\text{CH}_2$ ) as the enthalpy of the following isodesmic reaction, analogous to that defined previously for RAFT agents.<sup>15</sup>



Once again, a positive value of the enthalpy for the isodesmic reaction indicates that the X/Y substituents stabilize the thiocarbonyl bond relative to that in thioformaldehyde. As in the case of the RSEs, in using the enthalpy to study the stability of the thiocarbonyl compound it is assumed that the X,Y substituents do not destabilize the reference compound  $\text{CH}_2\text{XY}$ . As a result, this measure will at best offer only a qualitative guide to relative stabilities but has



**Figure 1.** Example of the use of the ONIOM method to approximate a high level of theory energy calculations. In this method, only the core is calculated at the highest level of theory, which the effect of the substituents being accounted for at lower levels of theory. This allows approximations of higher levels of theory for which computations of the full system would render the problems intractable.

nonetheless proven useful in predicting the behavior of RAFT agents toward radical addition.<sup>15</sup>

**Computational Methods.** Ab initio molecular orbital theory and DFT calculations were carried out using GAUSSIAN 98,<sup>22</sup> GAUSSIAN 03,<sup>23</sup> GAMESS-US,<sup>24</sup> and MOLPRO 2000.6.<sup>25</sup> Unless noted otherwise, calculations on radicals were performed with an unrestricted wave function. In cases where a restricted-open-shell wave function was used, it is designated with an R prefix. All enthalpies, RSEs, and agent stabilities were calculated at 0 K; all equilibrium constants and kinetic parameters were evaluated at 333.15 K.

Calculations were performed at a high level of theory, selected on the basis of previous assessment studies for radical addition to C=S double bonds.<sup>26,27</sup> Geometries and vibrational frequencies were obtained at the B3-LYP/6-31G(d) level of theory, and conformational searches for the global minimum energy structure were also performed at this level of theory. Improved energies were then calculated, using an ONIOM method<sup>27</sup> to approximate the W1<sup>28,29</sup> level of theory. This operated by calculating only the core reaction  $\text{CH}_3\cdot + \text{S}=\text{CH}_2 \rightarrow \text{CH}_3-\text{S}-\text{CH}_2\cdot$  at the W1 level of theory and then correcting the effect of the substituents using a less computationally intensive procedure, G3(MP2)-RAD.<sup>30</sup> When even this method proved too computationally expensive, a three-layer ONIOM method was used in which the energies of the final layer of substituents were calculated at the RMP2/6-311+G(3df,2p) level of theory. We have previously shown that, provided all forming and breaking bonds and all  $\alpha$ -substituents on the attacking radical are included in the core (as in Figure 1), this approach provides excellent approximations to high-level calculations on the full system.<sup>27</sup> To assist in the qualitative rationalization of the results, charge and (where relevant) spin density distributions within selected thioketones and their adduct radicals were calculated on the basis of a natural bond orbital (NBO) population

analysis, carried out in GAUSSIAN at the B3-LYP/6-311+G(3df,2p) level of theory using the density = current keyword.

For the agents selected as being potentially good radical trapping agents in an experimental scenario, the equilibrium constants were calculated for their reaction with model propagating species at the experimentally realistic temperature of 333 K. Rate coefficients for propagation of the thioketone adduct radicals (via addition to monomer and addition to thioketone agents) were also calculated to assess whether such side reactions were likely to be operative. Equilibrium constants ( $K$ ) and rate coefficients ( $k$ ) were calculated using the optimized geometries, frequencies, and W1-corrected energies in conjunction with the following standard textbook formulas.<sup>31</sup>

$$K(T) = (c^\circ)^{\Delta n} e^{(-\Delta G/RT)} = (c^\circ)^{\Delta n} \left( \frac{\prod_{\text{products}} Q_j}{\prod_{\text{reactants}} Q_i} \right) e^{(-\Delta E/RT)} \quad (3)$$

$$k(T) = \kappa(T) \frac{k_B T}{h} (c^\circ)^{1-m} e^{(-\Delta G^\ddagger/RT)} = \kappa(T) \frac{k_B T}{h} (c^\circ)^{1-m} \frac{Q^\ddagger}{\prod_{\text{reactions}} Q_i} e^{(-\Delta E^\ddagger/RT)} \quad (4)$$

In these formulas,  $\kappa(T)$  is the tunneling correction factor,  $T$  is the absolute temperature,  $k_B$  is Boltzmann's constant,  $h$  is Planck's,  $c^\circ$  is the standard unit of concentration ( $\text{mol L}^{-1}$ ),  $R$  is the universal gas constant,  $m$  is the molecularity of the reaction, and  $\Delta n$  the change in moles upon reaction,  $Q^\ddagger$ ,  $Q_i$ , and  $Q_j$  are the molecular partition functions of the transition structure, reactant  $i$  and product  $j$ , respectively,  $\Delta G^\ddagger$  is the Gibb's free energy of activation,  $\Delta G$  is the Gibb's free energy of reaction,  $\Delta E^\ddagger$  is the 0 K, zero-point energy corrected energy barrier for the reaction, and  $\Delta E$  is the 0 K, zero-point energy corrected energy change for the reaction. The value of  $c^\circ$  depends on the standard-state concentration assumed in calculating the thermodynamic quantities (and translational partition function). The tunneling coefficient  $\kappa(T)$  corrects for quantum effects in motion along the reaction path and is close to unity in the systems of interest.<sup>14</sup> Molecular partition functions ( $Q_i$ ) were calculated using the B3-LYP/6-31G(d) geometries and frequencies in conjunction with the standard textbook formulas, based on the statistical thermodynamics of an ideal gas.<sup>31</sup>

As outlined previously,<sup>11,14</sup> for these types of reactions (and indeed other radical addition reactions<sup>32</sup>), the harmonic oscillator model is not an adequate approximation for many modes of vibration that correspond to rotations (typically those less than  $300 \text{ cm}^{-1}$ ). Therefore, for the calculations of the equilibrium constants, all low frequency torsional modes were treated separately as hindered internal rotations, using the standard procedure described elsewhere.<sup>11,14</sup> In the case of the propagation rate calculations, the errors in the harmonic oscillator approximation would be expected to

**Table 1.** Radical Stabilization Energies (RSEs) for Thioketone Adduct Radicals, Enthalpies for the Methyl Trapping Reaction ( $\text{CH}_3^\bullet + \text{S}=\text{CXY} \rightarrow \text{CH}_3\text{S}^\bullet\text{CXY}$ ),<sup>a</sup> and Thioketone Bond Stability

S=CXY <sup>b</sup>			RSE <sup>c</sup> (kJ mol <sup>-1</sup> )	$\Delta H$ (kJ mol <sup>-1</sup> )	S=C stability (kJ mol <sup>-1</sup> )
no.	X	Y			
1	H	H	40.8	-123.7	0.0
2	Me	Me	10.1	-94.7	60.3
3	<i>tert</i> -butyl	H	44.2	-106.7	38.1
4	CN	H	88.1	-168.1	-1.7
5	OMe	H	56.6	-53.4	103.3
6	OMe	OMe	42.3	-12.1	140.4
7	phenyl	H	82.7	-137.0	49.9
8	<i>tert</i> -butyl	<i>tert</i> -butyl	39.9	-91.3	37.3
9	CN	CN	107.3	-193.3	7.7
10	CN	phenyl	115.0	-169.7	47.8
11	OMe	phenyl	82.3	-75.9	127.4
12	phenyl	phenyl	102.7 <sup>e</sup>	-139.8 <sup>e</sup>	71.7
13	(xanthene-9-thione)		125.2 <sup>e</sup>	-138.3 <sup>e</sup>	101.7 <sup>f</sup>
14	phenyl	Me	81.3	-124.5	64.5
15	(fluorene-9-thione)		108.5 <sup>e</sup>	-151.3 <sup>e</sup>	76.0
16	p-CN-phenyl	p-CN-phenyl	112.0 <sup>e</sup>	-149.0 <sup>e</sup>	65.9 <sup>f</sup>
17	(phenanthrene-4-thione)		110.0 <sup>e</sup>	-156.0 <sup>e</sup>	69.0 <sup>f</sup>

<sup>a</sup> Enthalpies were calculated at an approximate W1 level obtained using an ONIOM approach in conjunction with the G3(MP2)-RAD level of theory and (where necessary) RMP2/6-311+G(3df,2p). The W1 core used for all structures was  $\text{CH}_3^\bullet + \text{S}=\text{CH}_2$ , and unless noted otherwise G3(MP2)-RAD was used for the full system. <sup>b</sup> Optimized geometries were obtained using B3-LYP/6-31G(d), and energies include scaled B3-LYP/6-31G(d) zero-point energy corrections. See Figures 2 and 3 for structures. <sup>c</sup> Defined as the enthalpy change of the reaction given in eq 1. <sup>d</sup> Defined as the enthalpy change of the reaction given in eq 2. <sup>e</sup> Substrate used as the G3(MP2)-RAD outer core when applying the ONIOM method was  $\text{S}=\text{C}(\text{Ph})(\text{CH}_3)$ . <sup>f</sup> Substrate used as the G3(MP2)-RAD outer core was  $\text{S}=\text{C}(\text{Ph})_2$ .

cancel to some extent (due to the early transition structures).<sup>14</sup> Moreover, for the propagation calculations we required only order-of-magnitude accuracy, to assess the likely interference of copolymerization side reactions. For these reasons, for the propagation rate calculations, we adopted the less computationally intensive harmonic oscillator model. However, the more accurate hindered rotor model was used for the calculation of the equilibrium constants.

We have recently shown that the above methodology (when using the hindered rotor model) is capable of achieving chemical accuracy for radical addition to thioketones. For example, Scaiano and Ingold have reported an experimental equilibrium constant (based on photolysis studies) of  $1.2 \times 10^6 \text{ L mol}^{-1}$  for the addition of *tert*-butyl radicals to di-*tert*-butyl thioketone at 25 °C. The theoretical equilibrium constant of  $7.9 \times 10^5 \text{ L mol}^{-1}$  obtained using the theoretical procedure outlined above is in good agreement with experiment, the theoretical number being slightly lower possibly due to the neglect of solvent effects.<sup>12</sup>

## Results and Discussion

Enthalpies for  $\bullet\text{CH}_3$  radical addition to a wide range of thioketones ( $\text{S}=\text{CXY}$ ; X,Y = H,H;  $\text{CH}_3,\text{CH}_3$ ;  $\text{C}(\text{CH}_3)_3,\text{H}$ ; CN,H;  $\text{OCH}_3,\text{H}$ ;  $\text{OCH}_3,\text{OCH}_3$ ; Ph,H;  $\text{C}(\text{CH}_3)_3,\text{C}(\text{CH}_3)_3$ ; CN,-CN; CN,Ph,  $\text{OCH}_3,\text{Ph}$ ; Ph-O-Ph; Ph-Ph; *para*-CN-Ph,*para*-CN-Ph; phenanthrene; see also Table 1) have been calculated at 0 K and placed in Table 1. Corresponding enthalpies for the addition of substituted radicals ( $\text{CN}-\text{CH}_2^\bullet$ ,  $\text{HO}-\text{CH}_2^\bullet$ ,  $\text{Bz}^\bullet$ ) are shown in Table 2. To assist in the qualitative rationalization of the results, radical stabilization energies (RSEs) of the adduct radicals and C=S bond stabilities of thioketones are also included in Table 1. Geometries of the

adduct radicals for each series of reaction are shown in Figures 2 and 3, respectively; complete geometries of all species in the form of Gaussian archive entries are provided in the Supporting Information.

The exothermicities for  $\bullet\text{CH}_3$  addition to thioketones, together with stabilities of the thioketones and their adduct radicals, are plotted in order of increasing exothermicity in Figure 4. Inspection of Figure 4 shows that the exothermicity for the methyl trapping reaction depends on the stabilities of both the adduct radical and the thioketone agent, with the most effective agents combining high adduct stabilities with low agent stabilities. In what follows we first explore how to optimize each of these properties in turn for the case of  $\bullet\text{CH}_3$  radical addition, and we then consider the additional effect on the enthalpy when the electronic properties of the attacking radical are varied. Finally, we use the results of Tables 1 and 2 to design improved thioketone agents for controlling the polymerization and then test them computationally for the polymerization of styrene and also vinyl acetate.

**Radical Stability.** As expected, the radical stabilization energy of an adduct-radical is particularly sensitive to the type of substituents, with RSEs ranging from 10.1 kJ mol<sup>-1</sup> (adduct of thioacetone) to well over 100 kJ mol<sup>-1</sup> (in the case of many of the phenyl and/or cyano-substituted species). It can be further noted that the substrates with at least one substituent capable of delocalizing the unpaired electron all have an RSE in excess of 80 kJ mol<sup>-1</sup>, while in most cases those without such a substituent have an RSE falling below 40 kJ mol<sup>-1</sup>. In the four simplest compounds (i.e. those in which one of the substituents is a hydrogen), the largest stabilization is afforded by the cyano-substituent, and cyano-



**Table 2.** Enthalpies at 0 K for the Methyl Trapping Reaction ( $R^{\bullet} + S=CXY \rightarrow RS^{\bullet}CXY$ )<sup>a</sup> and Radical Stabilization Energies (RSEs) for the Adduct Radicals of  $R^{\bullet}$  and  $S=C(CN)_2$ 

no.	structure <sup>b</sup>	Me $\Delta H$ (kJ mol <sup>-1</sup> )	CN-CH <sub>2</sub> $\Delta H$ (kJ mol <sup>-1</sup> )	HO-CH <sub>2</sub> $\Delta H$ (kJ mol <sup>-1</sup> )	Bz $\Delta H$ (kJ mol <sup>-1</sup> )
3	S=C(C(CH <sub>3</sub> ) <sub>3</sub> ) <sub>2</sub>	-91.3	-65.9	-91.0	-43.6 <sup>c</sup>
10	S=C(CN)(Ph)	-169.7	-111.5	-166.3	-128.3 <sup>c</sup>
11	S=C(OMe)(Ph)	-75.9	-48.2	-76.2	-41.8 <sup>c</sup>
12	S=C(Ph)(Ph)	-139.8 <sup>e</sup>	-114.1 <sup>d</sup>	-137.6 <sup>d</sup>	-95.0 <sup>c</sup>
9	S=C(CN) <sub>2</sub>	-193.3	-136.7	-186.0	-153.1
9 (RSE)	R-SC <sup>•</sup> (CN) <sub>2</sub>	107.3	109.4	108.1	114.0

<sup>a</sup> Enthalpies were calculated at an approximate W1 level obtained using an ONIOM approach in conjunction with the G3(MP2)-RAD level of theory and (where necessary) RMP2/6-311+G(3df,2p). The W1 core used for all structures was  $CH_3^{\bullet} + S=CH_2$ , and unless noted otherwise G3(MP2)-RAD was used for the full system. <sup>b</sup> Optimized geometries were obtained using B3-LYP/6-31G(d), and energies include scaled B3-LYP/6-31G(d) zero-point energy corrections. See Figures 2 and 3 for structures. <sup>c</sup> The G3(MP2)-RAD outer core used was  $Bz^{\bullet} + S=CH_2$ , with the effect of the outer layer of substituents calculated at RMP2/6-311+G(3df,2p). <sup>d</sup> The G3(MP2)-RAD outer core used was  $R^{\bullet} + S=C(CH_3)_2$ , with the effect of the outer layer of substituents calculated at RMP2/6-311+G(3df,2p). <sup>e</sup> The G3(MP2)-RAD outer core used was  $CH_3^{\bullet} + S=C(Ph)(CH_3)$ , with the effect of the outer layer of substituents calculated at RMP2/6-311+G(3df,2p).

substituted substrates prove in general to have the highest stabilities. This is due to the  $\pi$ -accepting ability of the cyano group, which affords a captodative interaction with the lone pair donor thiy group.

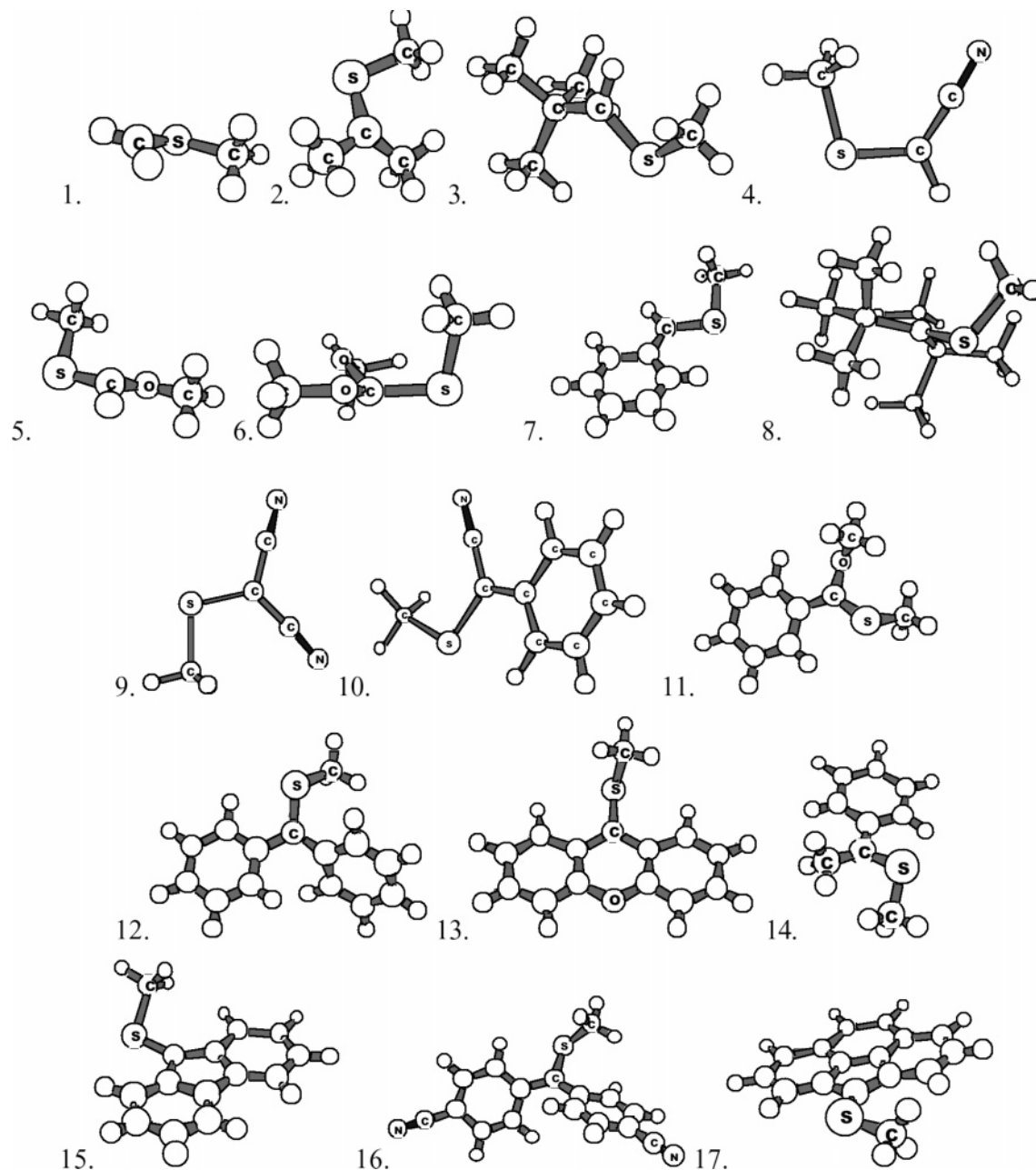
In contrast to the  $\pi$ -accepting groups, the lone pair donor methoxy substituents are less stabilizing, despite the fact that they show similar radical stabilizing properties in isolation. This feature was noted previously in our study of RAFT-adduct radicals<sup>15</sup> and arises because the lone pair donor thiy group, though stabilizing, results in the unpaired electron occupying a higher energy orbital than  $2p(C^{\bullet})$ .<sup>33</sup> As a result, the stabilizing interaction with subsequent lone pair donor groups is progressively weakened due to the increasing the energy gap between the lone pair and the unpaired electron. In the present work, we note that the first lone pair donor group (the thiy substituent in the adduct of  $CH_3SCH_2^{\bullet}$ ) affords a stabilization of 40.8 kJ mol<sup>-1</sup>, the second (the methoxy group in  $CH_3SCH(OCH_3)^{\bullet}$ ) affords an additional stabilization of only 15.8 kJ mol<sup>-1</sup>, and the third (the additional methoxy group in  $CH_3SC(OCH_3)_2^{\bullet}$ ) affords no additional stabilization and actually destabilizes the radical, relative to  $CH_3SCH(OCH_3)^{\bullet}$ , due to the increasing sigma withdrawal.

Comparing the  $^{\bullet}CH_3$  adducts of  $S=CH(Ph)$  with  $S=C(Ph)_2$ , a somewhat surprising effect of the second phenyl group is a reduction in the RSE. First, it might be possible that there is a reduced  $\pi$ -accepting stabilizing effect due to steric factors causing a reduction in the planarity of the phenyl groups at the radical center. Second, close inspection of the optimized geometry of  $CH_3-S^{\bullet}(Ph)_2$  reveals that in contrast to that for  $CH_3-S^{\bullet}H(Ph)$ , the methyl group, rather than adopting a well spaced trans conformation, is angled considerably out of the plane formed at the radical center. As a result, there is less efficient overlap of the p-type lone pair on the sulfur with the  $2p(C^{\bullet})$  and hence reduced stabilization from this substituent. In support of this we note that when we compare the RSE of the  $S=C(CN)_2$  adduct with that of its monosubstituted analogue  $S=CH(CN)$ , it can be seen that the second substituent does now increase the stability of the adduct radical though it still does not afford the same increase in stabilization as was provided by the first. In this case the CN substituent is much smaller than

the phenyl group and therefore much less affected by steric interactions.

We also note that this steric interaction may also help to explain why the thioacetone adduct has a lower radical stability than the thioformaldehyde adduct, despite the presence of additional radical-stabilizing methyl groups in the former case, and why the adduct of the di-*tert*-butyl thioketone has a lower radical stability than the corresponding monosubstituted thioketone. For example, we consider the geometries of the adduct radicals of both di-*tert*-butyl thione and *tert*-butyl thione. In the former, the newly formed C-S bond is nearly at right angles to the plane at the radical center, whereas for the latter it is essentially planar. It can be said, therefore, that the radical on the methyl adduct of *tert*-butyl thione is experiencing stabilization from lone pair donation of the adjacent sulfur atom and that this is absent in the radical adduct of di-*tert*-butyl thione. This is clearly seen in a reduction in the spin density at sulfur, which decreases from -0.23 in the  $CH_3SCH_2^{\bullet}$  radical and -0.19 in the  $CH_3SCH(C(CH_3)_3)^{\bullet}$  radical to only -0.01 in the  $CH_3SC(C(CH_3)_3)_2^{\bullet}$  radical. This loss of the lone pair donation effect is partially (but not wholly) compensated for by the increased hyperconjugative stabilization from the additional *tert*-butyl group, resulting in a small net destabilization of the adduct radical. This is evident in the fact that the spin density at the formal radical center increases only marginally from -0.79 for  $CH_3SCH_2^{\bullet}$  and -0.78 for  $CH_3SCH(C(CH_3)_3)^{\bullet}$  to -0.87 for  $CH_3SC(C(CH_3)_3)_2^{\bullet}$ .

Given these steric effects, it is interesting therefore to look at the conformation of  $CH_3-S^{\bullet}(Ph-O-Ph)$ . Here, the phenyl groups are held in plane by a linking oxygen atom. The attacking methyl radical is almost perpendicular to the plane of the phenyls, meaning that the p-type lone pair on the sulfur has very little overlap with and stabilization of the radical. This represents a compromise between the energetic cost of steric repulsion of the methyl by the phenyl group and that of the reduced stabilization of the radical by the p-type sulfur lone pair, it now being perpendicular to the  $2p(C^{\bullet})$  orbital. In forming this linked compound, we have strengthened the electron accepting ability of the phenyls at the expense of the electron donating ability of the sulfur lone pair. Nonetheless, the former effect dominates and leads to an increase of

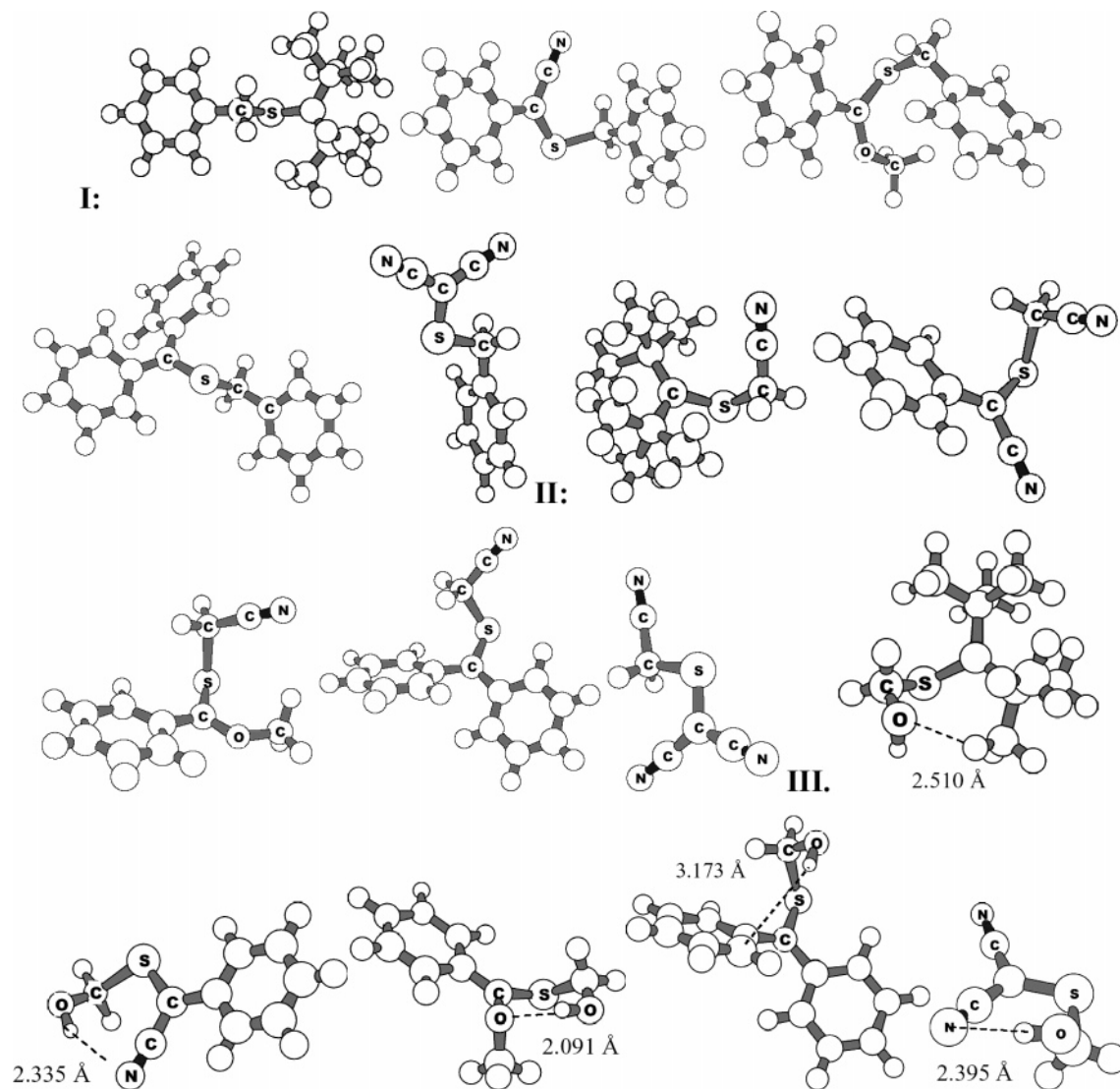


**Figure 2.** Optimized geometries of methyl radical adducts for each substrate. Geometry optimizations and conformation scans were carried out at the level of theory B3-LYP/6-31G(d).

approximately  $22 \text{ kJ mol}^{-1}$  in the RSE of the adduct radical of xanthene-9-thione, compared with the other twin-phenyl bearing compound, thiobenzophenone. However, the enthalpies for methyl trapping of these two compounds are similar in both cases because the increased stability of the adduct radical is countered by a concurrent increase in the stability of the thiocarbonyl compound due to the presence of the lone pair donor oxygen atoms (see below). To probe this further, calculations were also made for fluorene-9-thione and cyclopenta[*def*]phenanthrene-4-thione, for which the lone pair donor oxygen atoms were absent. These also showed an increase in the RSE of the adduct radicals (as in the case of the xanthene-9-thione) but without the significant increase in the stability of the thiocarbonyl compounds. As a result, they showed significant improvements in their methyl radical

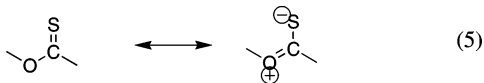
trapping capabilities, with cyclopenta[*def*]phenanthrene-4-thione showing the greatest improvement.

In summary, all substrates with  $\pi$ -accepting groups (such as cyano, phenyl, and other aromatic ring systems) demonstrate radical stabilities that are high relative to compounds without  $\pi$ -accepting features in their substrate. Differences within this series appear to be due to the degree of delocalization and how well the sulfur lone pair can be utilized to stabilize the radical in the adduct. It seems likely that the best results are attained either when the  $\pi$  accepting group is small and compact (as in the cyano substituted analogues) and/or the radical center is incorporated into an aromatic ring system that enforces planarity and thereby improves the ability of the  $\pi$ -system to accept the unpaired electron.

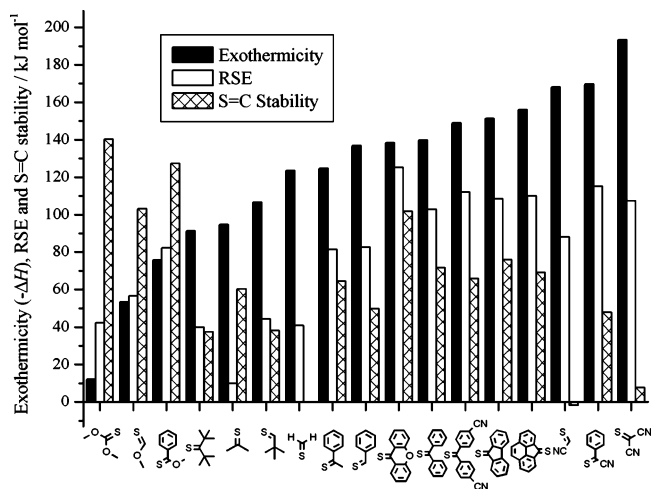


**Figure 3.** Optimized geometries of benzyl (I), cyanomethyl (II), and hydroxymethyl (III) radical adducts for selected substrates. Geometry optimizations and conformation scans were carried out at the level of theory B3-LYP/6-31G(d).

**Thioketone Stability.** As in the case of our earlier study of RAFT agents,<sup>15</sup> we find that the C=S bond of the thioketone agents is strongly stabilized by lone pair donating groups (such as OCH<sub>3</sub>) and weakly stabilized by  $\pi$  or pseudo- $\pi$  donor groups (such as phenyl or alkyl groups). The stabilizing effect of lone pair donor groups is readily understood in terms of the possible resonance contributors to the thiocarbonyl compound as follows.<sup>5</sup>



The  $\pi$  or pseudo- $\pi$  donor groups are less stabilizing because they are much weaker electron donors. As in the case of RAFT agents,<sup>15</sup> we also find that the thioketones are destabilized by steric interactions. Thus for example, the stabilizing effect of a single (electron donating) *tert*-butyl group is 38.1 kJ mol<sup>-1</sup>; however, the addition of the second group actually slightly destabilizes the thioketone (by approximately 1 kJ mol<sup>-1</sup>). In a similar manner, the stabilizing



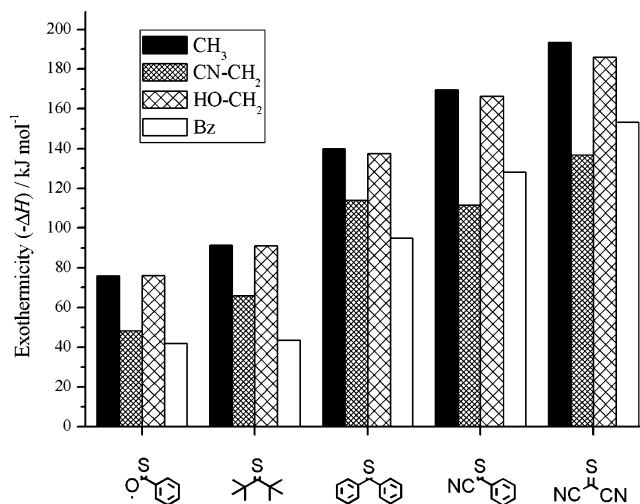
**Figure 4.** Exothermicities (i.e.,  $-\Delta H$ ) for trapping of the methyl radical at 0 K by all substrates studied (black bars); radical stabilization energies (RSEs) for the adduct radicals (CH<sub>3</sub>-S-C\*(X)(Y)) (white bars); and stability of the S=C bond (hatched bars). The values in this figure directly reflect those reported in Table 1.

effect of two phenyl substituents ( $71.7 \text{ kJ mol}^{-1}$ ) is much less than that sum of their isolated effects ( $99.8 \text{ kJ mol}^{-1}$ ).

In our previous study of RAFT agents,<sup>15</sup> we found that  $\sigma$ -withdrawing substituents (such as CN or  $\text{CF}_3$ ) destabilize the C=S bond by inhibiting resonance with the lone pair donor thiyl group of the dithioester. In the thioketones of the present work, this resonance is already absent (as there is no thiyl substituent), and their effect is much smaller (in some cases negligible). Thus, for example,  $\text{S}=\text{C}(\text{CN})\text{H}$  is only  $1.7 \text{ kJ mol}^{-1}$  less stabilized than  $\text{S}=\text{CH}_2$ ,  $\text{S}=\text{C}(\text{CN})\text{Ph}$  is only  $2.1 \text{ kJ mol}^{-1}$  less stabilized than  $\text{S}=\text{C}(\text{H})\text{Ph}$ , and the para-cyano-substituted thioketone,  $\text{S}=\text{C}(\text{p-CN-Ph})$ , is only  $5.8 \text{ kJ mol}^{-1}$  less stabilized than the corresponding unsubstituted analogue,  $\text{S}=\text{C}(\text{Ph})_2$ . However, one might expect stronger destabilizing effects (more in line with those seen in RAFT agents) if the cyano substituents were included in the presence of a second lone pair donor group, such as  $\text{OCH}_3$ .

Interestingly, the di-cyano-substituted thioketone appears to be weakly stabilized (by  $7.7 \text{ kJ mol}^{-1}$ ) relative to  $\text{S}=\text{CH}_2$ , despite the fact that the monosubstituted analogue  $\text{S}=\text{CH}(\text{CN})$  was not. This apparently unphysical result would seem to indicate that the isodesmic reaction used to measure stabilities of the thioketones contains an influence from the stabilities of the reference compounds (in this case  $\text{CH}_4$ ,  $\text{CH}_3\text{-CN}$ , and  $\text{CH}_2\text{CN}_2$ ). Indeed the reaction,  $2\text{CH}_3\text{CN} \rightarrow \text{CH}_2\text{CN}_2 + \text{CH}_4$ , is actually endothermic by  $31.1 \text{ kJ mol}^{-1}$ , implying that the observed increase in the stabilization energy of  $\text{S}=\text{C}(\text{CN})_2$  is probably due to the lower inherent stability of  $\text{CH}_2\text{CN}_2$ . This serves to illustrate the cautionary remarks made in the Theoretical Procedures section concerning the use of isodesmic reactions to measure stabilization energies. Such measures always contain an influence from the closed shell reference species and will at best provide a qualitative guide to relative stabilities. As was clear from Figure 3, the isodesmic reaction used in the present work to assess thiocarbonyl stability is a reasonable qualitative predictor of reaction exothermicity, provided the differences in stabilities are large ( $>15 \text{ kJ mol}^{-1}$ ). Thus for example the exothermicities for the lone pair donor substituted thioketones (such as  $\text{S}=\text{C}(\text{OCH}_3)\text{H}$ ) are considerably smaller than those for  $\text{S}=\text{CH}_2$ , despite the fact that the adduct radicals are actually more stabilized, and this is in accord with their significantly greater thioketone stabilities. However, it is also clear that one should be cautious in attributing physical significance to more minor differences in compound stability.

Based on these trends, the most active thioketone agents should contain at least one cyano-substituent in preference to  $\pi$ , pseudo- $\pi$ , or lone pair donor groups. However, since this not very bulky, it may not offer sufficient protection from chain termination and propagation reactions. Of the more practical possibilities, the di-*tert*-butyl thioketone shows the next lowest stability, which helps to explain its recent success in controlling styrene polymerization.<sup>12</sup> Other agents with relatively low stabilities would include the diphenyl substituted agents, especially the para-cyano-substituted ones, and also fluorene-9-thione and phenanthrene-4-thione. The other cage-like diphenyl substituted system, xanthene-9-



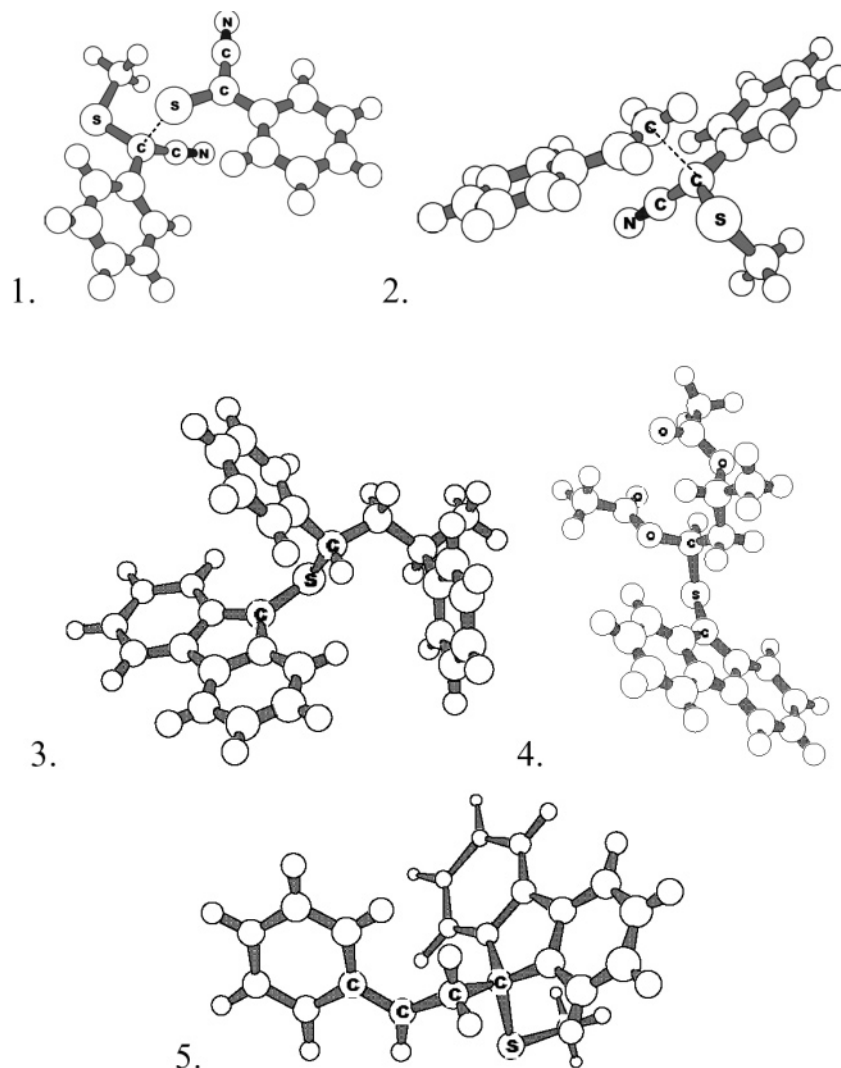
**Figure 5.** Exothermicities (i.e.,  $-\Delta H$ ) for trapping of several radicals by several thioketones at 0 K by all substrates studied. The values in this figure directly reflect those reported in Table 2.

thione, is much more stable than the others, due to the presence of the lone pair donating oxygen atom.

**Effect of Attacking Radical.** Figure 5 shows the exothermicities for addition of  $\cdot\text{CH}_3$ ,  $\cdot\text{CH}_2\text{CN}$ ,  $\cdot\text{CH}_2\text{OH}$ , and  $\cdot\text{CH}_2\text{Ph}$  to the various substrates. In general, it appears that the trends in the exothermicities for  $\cdot\text{CH}_3$  addition to the various substrates provide a reasonable qualitative guide to the exothermicity trends for the other radicals. Thus, for example, of the agents considered, the methoxy-substituted agent shows the lowest exothermicities for trapping of the various radicals, the dicyano-substituted agent shows the highest, and the other substrates show intermediate behavior. Within these broad trends there are minor variations among the different attacking radicals. In particular, although the trapping of all radicals is promoted when the thioketone bears at least one cyano-substituent, the relative trapping efficiencies are reversed for the benzyl and  $\cdot\text{CH}_2\text{CN}$  radicals. This may be related to the fact that the RSE of the adduct of benzyl and  $\text{S}=\text{C}(\text{CN})_2$  is higher than the adducts of this substrate and the other radicals (see the bottom row of Table 2). Despite these minor variations, it appears that it may be possible to select optimal thioketone agents using calculations of their  $\cdot\text{CH}_3$  trapping abilities, rather than calculations on larger polymeric radicals.

Comparing next the exothermicities for addition of the various radicals to a specific substrate, we note that, if we were to consider only the stability of the attacking radical, then the reactions of the least stable radical  $\cdot\text{CH}_3$  (i.e. RSE = 0) should be most exothermic, followed by those of  $\cdot\text{CH}_2\text{OH}$  and  $\cdot\text{CH}_2\text{CN}$  (RSE = 31.9 and 31.6  $\text{kJ mol}^{-1}$ , respectively),<sup>33</sup> with those of the most stable radical  $\cdot\text{CH}_2\text{-Ph}$  (RSE = 58.9  $\text{kJ mol}^{-1}$ )<sup>33</sup> expected to be the least exothermic. If we examine Table 2, we find that these trends generally hold but with two important exceptions.

First, we note that  $\cdot\text{CH}_2\text{OH}$  additions are more exothermic than one would have expected on the basis of radical stability alone. This feature has been noted previously in addition of the same radicals to alkenes,<sup>34</sup> and its generality would seem



**Figure 6.** B3LYP/6-31G(d) optimized geometries of the transition states of (1)  $S=C(CN)(ph)$  homopolymerizing (Scheme 2.I) and (2)  $S=C(CN)(ph)$  copolymerizing with styrene (Scheme 2.II) as well as those of the adducts of fluorene-9-thione with (3) styryl dimer radical and (4) VA dimer radical. Structure (5) is the B3LYP/6-31G(d) optimized geometry of the adduct formed upon cross-propagation of fluorene-9-thione with styrene (Scheme 2.III).

to indicate that the  $\bullet CH_2OH$  attacking radical may actually be less stabilized than  $\bullet CH_2CN$ , despite their almost identical RSEs. As explained in the Theoretical Procedures section, the RSE reflects not only the relative stabilities of the attacking radicals but also those of the corresponding closed shell compounds, and these can sometimes affect results. However, this is probably only part of the explanation: although the  $\bullet CH_2OH$  radical could easily be less stabilized than  $\bullet CH_2CN$ , it is unlikely that it could be as destabilized as  $\bullet CH_3$ . Instead, it appears that the stability of the product also plays a role in the greater exothermicity for  $\bullet CH_2OH$  addition. This enhanced stability may arise through a number of factors. For example, the presence of the lone pair donor OH group may enhance the lone pair donation of the sulfur atom and its consequent ability to stabilize the unpaired electron of the adduct radical. To test this idea, we calculated RSEs of the various adduct radicals for a typical substrate (see Table 2) but found that those for the  $\bullet CH_3$ ,  $\bullet CH_2OH$ , and  $\bullet CH_2CN$  adducts were all very similar to one another. Another possible explanation for the increased exothermicity

in  $\bullet CH_2OH$  addition is the formation of stabilizing H-bonding interactions in the adduct-radical. These occur either between the O of the OH group and H on the substrate (in the case of  $S=C(C(CH_3)_3)_2$ ) or between the H of the OH group and the N of the CN-group or O of the methoxy group in the substrate (in the case of the  $S=C(CN)_2$ ,  $S=C(CN)(Ph)$ , and  $S=C(OCH_3)(Ph)$ ). For depictions of all of these hydrogen bonds, see Figure 3. To establish the importance of these interactions we took selected adduct radicals and recalculated the exothermicities for corresponding conformations of the adduct radicals in which the OH group angled away from the substrate (and hence the H-bond was inhibited). These were typically higher in energy by a small but significant amount (for example,  $5.2 \text{ kJ mol}^{-1}$  for the  $\bullet CH_2OH$  adduct of  $S=C(C(CH_3)_3)_2$ ), suggesting that H-bonding does play a role. Finally, the adduct radicals of  $\bullet CH_2OH$  with the thioketones are substituted with two lone pair donor groups (instead of one, as in the case of the other attacking radicals), raising the possibility that an anomeric effect on the strength

of the formed R $\cdots$ S bond also contributes to their enhanced stability.

Second, the relative exothermicities for addition of the  $\cdot\text{CH}_2\text{CN}$  and  $\cdot\text{CH}_2\text{Ph}$  radicals vary somewhat according to the substrate. In particular, the exothermicity for the (more stable)  $\cdot\text{CH}_2\text{Ph}$  radical is considerably higher than that for the (less stable)  $\cdot\text{CH}_2\text{CN}$  radical for those substrates bearing cyano substituents. This may be explained as follows. In the  $\cdot\text{CH}_2\text{CN}$  adducts for which the substrate contained no CN group ( $\text{S}=\text{C}(\text{Ph})_2$ ,  $\text{S}=\text{C}(\text{C}(\text{CH}_3)_2)$ , and  $\text{S}=\text{C}(\text{OCH}_3)(\text{Ph})$ ), the lone pair on the sulfur was not parallel to the spin unpaired orbital, implying that it is not contributing to the radical stabilization significantly. The reasons for this are steric in the case of  $\text{S}=\text{C}(\text{Ph})_2$  and  $\text{S}=\text{C}(\text{C}(\text{CH}_3)_2)$  but electronic in the case of  $\text{S}=\text{C}(\text{OCH}_3)(\text{Ph})$ , where there is competing lone pair donation from the oxygen. In the case of  $\text{S}=\text{C}(\text{CN})_2$  and  $\text{S}=\text{C}(\text{CN})(\text{Ph})$  however, there exists the possibility for stabilization of the radical by lone pair donation from the sulfur, since the steric and competitive donation mentioned in the previous sentence are absent. Further, we note that in the geometries of all adducts coming from each possible adduct of  $\cdot\text{CH}_2\text{Ph}$  or  $\cdot\text{CH}_2\text{OH}$  with  $\text{S}=\text{C}(\text{CN})_2$  or  $\text{S}=\text{C}(\text{CN})(\text{Ph})$ , the forming bond is roughly coplanar with the radical center, meaning that stabilization of the radical by the sulfur lone pair is occurring. In systems where the attacking radical is  $\cdot\text{CH}_2\text{CN}$ , it seems likely that the  $\text{CH}_2\text{CN}$  group is undergoing  $\pi$ -acceptance of the p-type lone pair on the sulfur, thereby reducing the sulfur's lone pair stabilization of the radical. This is a phenomenon which cannot exist in adducts of other attacking radicals and hence explains the reversed order of reactivity of CN bearing substrates toward  $\cdot\text{CH}_2\text{CN}$  and  $\cdot\text{CH}_2\text{Ph}$ .

**Thioketone Design.** As discussed above, the requirements of a thioketone to be able to control a radical polymerization are that it (a) has a high equilibrium constant for the trapping process  $K = k_{\text{add}}/k_{\text{frag}}$  (see Scheme 1.VIII) and (b) forms an adduct radical with a high stability toward copolymerization or cross-termination. While  $\text{S}=\text{C}(\text{CN})_2$  might appear to be the best candidate based on the first criterion (see Figure 4), its nonbulky substituents most likely preclude satisfaction of the second. The agent cyanophenyl thione,  $\text{S}=\text{C}(\text{CN})(\text{Ph})$ , also performs well with respect to the first criterion and is slightly more bulky. However, although more hindered than  $\text{S}=\text{C}(\text{CN})_2$  it is still unclear whether the substituents are likely to prevent the adduct radical from undergoing propagation or termination reactions.

Apart from  $\text{S}=\text{C}(\text{CN})_2$  and  $\text{S}=\text{C}(\text{CN})(\text{Ph})$ , the next two most promising reagents on the basis of methyl radical trapping exothermicities appear to be fluorene-9-thione and cyclopenta[def]phenanthrene-4-thione. These also have the desirable feature of a closed ring structure, meaning that termination or copolymerization of the thioketones would be (at least somewhat) hindered. In deciding which of these two substrates to choose for further investigation, an additional criterion has been invoked—that of synthetic feasibility. The most common methods of synthesizing thioketones involves conversion of the ketone precursor.<sup>35</sup> Both fluorene-9-one and cyclopenta[def]phenanthrene-4-one are both available commercially, however the latter only at a synthetically

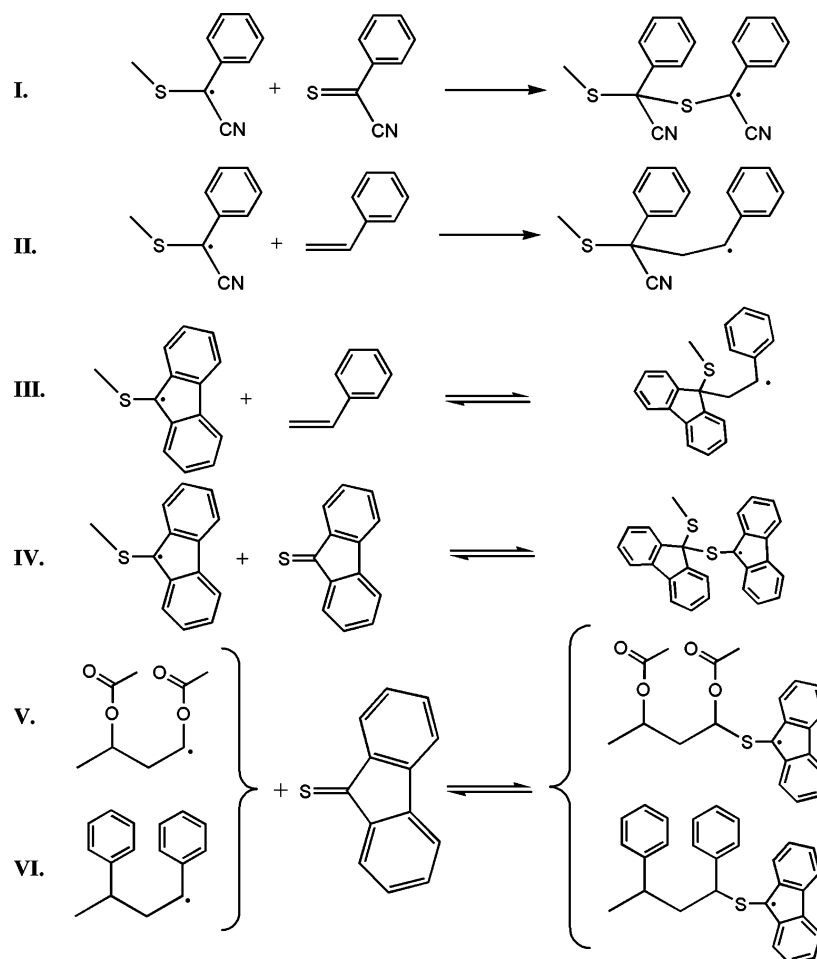
infeasible expense.<sup>36</sup> For this reason, fluorene-9-thione has been chosen for further evaluation. However, calculations on the more thermodynamically optimal agent  $\text{S}=\text{C}(\text{CN})(\text{Ph})$  were also undertaken with a view to establishing whether it is sufficiently hindered to resist propagation reactions.

**Computational Evaluation.** Having selected two potential thioketone agents,  $\text{S}=\text{C}(\text{CN})(\text{Ph})$  and fluorene-9-thione, it is now necessary to evaluate their potential for controlling free-radical polymerization. In the first instance, we need to establish whether the trapped radical is sufficiently hindered to prevent propagation and termination reactions from occurring. Accurate calculations of radical–radical termination reactions (including termination in situations where the radical can be delocalized into e.g. a phenyl ring), which would require expensive multireference methods, are not currently feasible on systems of this size. However, we can study the propagation behavior. For each substrate, there are two types of propagation reaction to consider: addition of the adduct radical to the thioketone (“homopropagation”) and addition of the adduct radical to the monomer (“cross-propagation”). Models of these two reactions for each of the two substrates are illustrated in Scheme 2.

In the case of the smaller substrate,  $\text{S}=\text{C}(\text{CN})\text{Ph}$  thioketone, the calculated rate coefficients for these two reactions are  $1.9 \cdot 10^8 \text{ L mol}^{-1} \text{ s}^{-1}$  and  $11 \text{ L mol}^{-1} \text{ s}^{-1}$  for homopropagation (I) and cross-propagation (II), respectively. Based on these results, it is clear that homopropagation of the adduct radical of cyanophenyl thione ( $\text{S}=\text{C}(\text{CN})(\text{Ph})$ ) is a significant reaction. Since the steric bulk at the radical center was insufficient to prevent this side reaction, it also seems likely that there will be insufficient steric bulk to prevent biradical termination and that this particular thioketone is unlikely to be suitable for the mediation of free radical polymerizations. For this reason, no further calculations were performed for cyanophenyl thione. In the case of the larger substrate, fluorene-9-thione, it seems reasonable to assume that the closed ring structure could be too hindered to undergo significant propagation reactions. Indeed, when we calculated the equilibrium constant for the addition to the smaller of the two substrates (styrene, reaction III in Scheme 2), we found that the reaction was not very thermodynamically favorable ( $K_{333} = 3.3 \times 10^1 \text{ L mol}^{-1}$ ;  $\Delta H_{333} = -63.8 \text{ kJ mol}^{-1}$ ;  $\Delta S_{333} = -190.0 \text{ J mol}^{-1} \text{ K}^{-1}$ ), having a ceiling temperature of only 336 K. In other words, propagation, if it occurred at all, would be reversible and merely contribute to radical storage. We therefore selected fluorene-9-thione for further evaluation.

The next step in the evaluation is to investigate the reaction of fluorene-9-thione with the propagating radicals of common radical polymerizations and thereby determine whether the adduct radical is sufficiently stabilized (toward  $\beta$ -scission) to provide sufficient control. As noted in the Introduction, kinetic studies of the related living radical polymerization process, nitroxide mediated polymerization, indicate that the equilibrium constant for the addition reaction  $K$  should be around  $10^9 \text{ L mol}^{-1}$  in order to effect an efficient living process.<sup>6</sup> Values lower than this can lead to hybrid behavior and loss of control; values significantly higher than this can

**Scheme 2.** (I–IV) Model System Used in Calculations Relating to Copolymerization Reactions<sup>a</sup> and (V and VI) Model System Used in Calculation of an Addition/Fragmentation Equilibrium Constant for Mediation of a Polymerization by Fluorene-9-thione



<sup>a</sup> Which must occur to a negligible extent for the thioketone to exert effective control over the polymerization.

lead to rate retardation effects. We therefore evaluated the equilibrium constant for addition of the propagating radical to fluorene-9-thione in both styrene and vinyl acetate polymerization (see reactions V and VI of Scheme 2), the former being chosen as a prototypical stable monomer, the latter as a prototypical unstable monomer. In each case, the propagating radical was modeled as a dimer as we have previously shown that, for propagation reactions at least, this model includes the most important substituent effects on radical reactivity and provides a reasonable guide to the propagation kinetics of the corresponding long chain radical.<sup>37</sup>

Using these model reactions, it is found that the theoretical predictions for the equilibrium constant of a propagating poly(styrene) and poly(vinyl acetate) radical adding to fluorene-9-thione are  $4.8 \cdot 10^{14} \text{ mol L}^{-1}$  and  $2.9 \cdot 10^{16} \text{ mol L}^{-1}$ , respectively. These equilibrium constants are considerably higher than those calculated previously for the reaction of a styryl dimer radical with di-*tert*-butyl thione (DTBT,  $4.6 \cdot 10^4 \text{ mol L}^{-1}$ ), under the same conditions.<sup>12</sup> Hence the objective of the present work, to identify a thioketone agent with radical trapping abilities superior to the existing DTBT, has indeed been met, and it is clear that the radical-trapping abilities of thioketones span a very wide range of values. Nonetheless, it is also clear that the equilibrium constants

for these two systems are so large as to render it likely that they would induce rate retardation at 60 °C. However, it is important to question the physical realization of this equilibrium constant as forward and reverse rate coefficients. Although rate coefficients were not calculated for this system, it seems likely that in this highly exothermic reaction, the addition reaction is likely to be barrierless and occur with a chemically controlled rate coefficient above the diffusion-controlled limit. Hence in the context of a polymerizing system, the steady-state equilibrium concentrations of the propagating and adduct radicals may not be achievable on the time-scale of the process, and the effective kinetically limited concentrations (and associated apparent equilibrium constant) may be quite different. It would therefore be worth investigating the behavior of fluorene-9-thione as a potential control agent, despite its high  $K$  value. If however, rate retardation and inhibition effects are observed, hindered thioketones showing intermediate radical affinities between di-*tert*-butyl-thione and fluorene-9-thione, such as xanthene-9-thione (see Figure 4), could be adopted instead.

## Conclusion

The enthalpies at 0 K for radical trapping by a number of thioketones ( $S=C(X)(Y)$ ) have been evaluated via high level ab initio molecular orbital calculations. By altering the X

and Y, a degree of control can be exerted over the radical affinity. In general terms, the adduct radical is favored by  $\pi$  accepting substituents, which delocalize and therefore stabilize the unpaired electron, and sigma accepting substituents, which weaken the C=S bond. In contrast, due to the pre-existing lone pair on the sulfur, additional lone pair substituents (such as X = OCH<sub>3</sub>) do not afford as significant a stabilization as might have been expected, had they been in isolation. Other features of the thioketone substrates, such as how phenyl groups are attached to the rest of the molecule, can also affect the radical affinities of these compounds. In particular, twin phenyl bearing substrates (such as fluorene-9-thione), for which the phenyls were linked and held in a plane, form a radical adduct in which a contiguous region of spin delocalization is spread over most of the molecule and are able to strongly stabilize the radical. As well as radical affinity, additional molecular features were highlighted which might enhance or make less suitable a particular thioketone for use in controlling a radical polymerization. Of particular importance is the steric bulk at the adduct radical center, which provides a barrier toward biradical termination of two thioketone terminated polymer chains or co/homopolymerization of the thioketone molecules.

On the basis of radical affinities, steric demands, and synthetic feasibility, fluorene-9-thione was selected as being a potentially suitable candidate for use in controlling free radical polymerizations. Cyanophenyl thione also displayed excellent radical trapping abilities but was not sufficiently hindered to prevent homopropagation side reactions. For the fluorene-9-thione, equilibrium constants (at 333.15 K) in the vicinity of 10<sup>14</sup> L mol<sup>-1</sup> and 10<sup>16</sup> L mol<sup>-1</sup> were obtained for trapping of dimeric styryl and vinyl acetate radicals, respectively. Although these may be too high to achieve nonretarded polymerizations at this temperature, they indicate that there is sufficient scope within the thioketone class of compounds to mediate free radical polymerizations via radical stability alone, and other thioketones with intermediate radical trapping abilities (such as xanthene-9-thione) should be suitable for controlling free-radical polymerization.

**Acknowledgment.** This work was produced with the support of the Australian Research Council under the ARC Centres of Excellence program. M.L.C. also gratefully acknowledges generous allocations of computing time from the National Facility of the Australian Partnership for Advanced Computing and the Australian National University Supercomputing Facility and helpful discussions with Professor Leo Radom and Professor Chris Easton. H.C.M. acknowledges a University of New South Wales Post-Graduate Award.

**Supporting Information Available:** B3-LYP/6-31G-(d) optimized geometries (in the form of GAUSSIAN archive entries) for the radicals, alkanes, and thioketone compounds used in the calculation of the radical stabilization energies, enthalpies, and equilibrium constants (Table S1) and all fragments which were the subject of hindered rotor calculations and the rotational potentials of all rotating fragments

and their assigned frequencies (Table S2). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Hawker, C. J.; Bosman, A. W.; Harth, E. *Chem. Rev.* **2001**, *101*, 3661–3688.
- (2) Matyjaszewski, K. *Controlled/Living Radical Polymerisation*; American Chemical Society: Washington, DC, 2000.
- (3) Wang, J.-S.; Matyjaszewski, K. *J. Am. Chem. Soc.* **1995**, *117*, 5614–5615.
- (4) Le, T. P. T.; Moad, G.; Rizzardo, E.; Thang, S. H. PCT Int Appl WO 9801478 A1 980115, 1998; *Chem. Abstr.* **1998**, *128*, 115390. (b) Charmot, D.; Corpart, P.; Michelet, D.; Zard, S. Z.; Biadatti, T. PCT Int Appl WO 9858974, 1998; *Chem. Abstr.* **1999**, *130*, 82018.
- (5) Moad, G.; Rizzardo, E.; Thang, S. H. *Aust. J. Chem.* **2005**, *58*, 379–410.
- (6) Goto, A.; Fukuda, T. *Prog. Polym. Sci.* **2004**, *29*, 329–385.
- (7) Wang, A. R.; Zhu, S.; Kwak, Y.; Goto, A.; Fukuda, T.; Monteiro, M. S. *J. Polym. Sci., Part A: Polym. Chem.* **2003**, *41*, 2833–2839.
- (8) Barner-Kowollik, B.; Vana, P.; Quinn, J. F.; Davis, T. P. *J. Polym. Sci. A* **2002**, *40*, 1058.
- (9) Coote, M. L.; Radom, L. *J. Am. Chem. Soc.* **2003**, *125*, 1490–1491.
- (10) Feldermann, A.; Coote, M. L.; Stenzel, M. H.; Davis, T. P.; Barner-Kowollik, C. *J. Am. Chem. Soc.* **2004**, *126*, 15915–15923.
- (11) Coote, M. L. *Macromolecules* **2004**, *37*, 5023–5031.
- (12) Ah Toy, A.; Chaffey-Millar, H.; Davis, T. P.; Stenzel, M. H.; Coote, M. L.; Barner-Kowollik, C. *Chem. Commun. (Cambridge)* **2006**, *8*, 835–837.
- (13) Henry, D. J.; Coote, M. L.; Gómez-Balderas, R.; Radom, L. *J. Am. Chem. Soc.* **2004**, *126*, 1732–1740.
- (14) Coote, M. L. *J. Phys. Chem. A* **2005**, *109*, 1230–1239.
- (15) Coote, M. L.; Henry, D. J. *Macromolecules* **2005**, *38*, 1415–1433.
- (16) Krenske, E. H.; Izgorodina, E. I.; Coote, M. L. In *Advances in Controlled/Living Radical Polymerization*; Matyjaszewski, K., Ed.; American Chemical Society: Washington, DC, 2006; in press.
- (17) Coote, M. L.; Krenske, E. H.; Izgorodina, E. I. *Macromol. Rapid Commun.* **2006**, *27*, 473–497.
- (18) Griller, D.; Ingold, K. *Acc. Chem. Res.* **1976**, *9*, 13–19.
- (19) Benson, S. W. *Thermochemical Kinetics Methods for the Estimation of thermochemical data and rate parameters*; John Wiley & Sons: 1976.
- (20) Hodgson, J. L.; Coote, M. L. *J. Phys. Chem. A* **2005**, *109*, 10013–10021.
- (21) These issues are also discussed in a recent review article: Zipse, H. *Top. Curr. Chem.* **2006**, *263*, 163–189.
- (22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.;



- Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; 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 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
- (23) 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 B03*; Gaussian, Inc.: Pittsburgh, PA, 2003.
- (24) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (25) Werner, H.-J.; Knowles, P. J.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Korona, T.; Lindh, R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schütz, M.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO 2000.6*; University of Birmingham: Birmingham, 1999.
- (26) Coote, M. L.; Wood, G. P. F.; Radom, L. *J. Phys. Chem. A* **2002**, *106*, 12124–12138.
- (27) Izgorodina, E. I.; Coote, M. L. *J. Phys. Chem. A* **2006**, *110*, 2486–2492.
- (28) Martin, J. M. L.; de Oliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843–1856.
- (29) Parthiban, S.; Martin, J. M. L. *J. Phys. Chem.* **2001**, *114*, 6014–6029.
- (30) Henry, D. J.; Sullivan, M. B.; Radom, L. *J. Chem. Phys.* **2003**, *118*, 4849–4860.
- (31) These formulas are described in full in the following: Coote, M. L. In *Encyclopedia of Polymer Science and Technology*; Kroschwitz, J. I., Ed.; Wiley: New York, 2004; pp 319–371.
- (32) Heuts, J. P. A. In *Handbook of Radical Polymerization*; Matyjaszewski, K., Davis, T. P., Eds.; John Wiley and Sons: New York, 2002; pp 1–76.
- (33) Henry, D. J.; Parkinson, C. J.; Mayer, P. M.; Radom, L. *J. Phys. Chem. A* **2001**, *105*, 6750–6756.
- (34) Wong, M. W.; Pross, A.; Radom, L. *J. Am. Chem. Soc.* **1994**, *116*, 6284–6292.
- (35) A typical conversion employs Lawesson's reagent, for a review, see: Jesberger, M.; Davis, T. P.; Barner, L. *Synthesis-Stuttgart* **2003**, *13*, 1929–1958.
- (36) As listed at <http://www.sigmaaldrich.com> (accessed Jan 9, 2006).
- (37) Izgorodina, E. I.; Coote, M. L. *Chem. Phys.* **2006**, *324*, 96–110.

CT600128T

## Particle-Based Implicit Solvent Model for Biosimulations: Application to Proteins and Nucleic Acids Hydration

Nathalie Basdevant, Tap Ha-Duong,\* and Daniel Borgis\*

*Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement – UMR 8587, Bâtiment Maupertuis, Université d'Evry-Val-d'Essonne, Bd François Mitterrand, 91025 Evry Cedex, France*

Received February 3, 2006

**Abstract:** In addition to the simulation of two proteins described previously, we report on the application of our recently developed particle-based implicit solvent model to the simulations of four nucleic acid molecules, the 17 bases anticodon hairpin of the Asp-tRNA, the decamer d(CCGCCGGCGG) in both A and B form, and the containing EcoRI restriction site dodecamer d(CGCGAATTCGCG). The solvent is represented by a fluid of Lennard-Jones polarizable pseudoparticles of molecular size, the induced dipoles of which are sensitive to the solute electric field but not to each other. When implemented in a molecular dynamics algorithm with the Amber94 force field, the model allows to simulate efficiently the conformational evolution of the nucleic acids, yielding stable three-dimensional structures in agreement with experiments and other simulations in explicit solvent. In the same run, it is also able to provide estimations of the electrostatic solvation free energy within short time windows which correlate well with the Poisson–Boltzmann calculations. In addition, the molecular aspect of the solvent model allows for the reproduction of the highly localized water molecules in the major or minor grooves of the nucleic acid double helices, despite the absence of explicit water hydrogen bonds.

### Introduction

Long and stable molecular dynamics simulations of nucleic acids in explicit solvent, such as TIP3 or SPC models, allow nowadays to analyze in detail the role of water in the stabilization of their three-dimensional structures and in the recognition processes by ligands or proteins. Several simulations in the nanosecond range have emphasized the crucial long-lived hydration patterns in both the minor and major grooves of DNA and RNA helices.<sup>1–5</sup> Other theoretical studies have confirmed the importance of water molecules at DNA or RNA-protein interfaces for the stability and/or the selectivity of such complexes.<sup>6–10</sup> Only a few studies have been interested in quantitative results of energetics and electrostatics in these complexes (see, e.g., refs 11–15). In this respect, however, the computational cost of simulations

in explicit solvent makes it difficult to estimate solvation free energies, and explicit solvent models do not allow for the estimation of proteins solvation free energy accurately, even with nanosecond-long simulations, because of the protein landscape which causes a prohibitive solvent penetration time. One possible approach is to combine molecular mechanics (MM) and continuum models by applying implicit solvent methods, such as Poisson–Boltzmann (PB) or generalized Born (GB) – solvent accessible surface area (SA) calculations, to structural snapshots extracted from MD simulations in explicit solvent.<sup>16–18</sup> It should be noticed that this method has some connections to the earlier PDL/D/S-LRA method of Warshel and collaborators (semimicroscopic protein dipoles Langevin dipoles/linear response approximation),<sup>19</sup> which has the merit to take additionally into account the protein reorganization effect. These so-called MM-PBSA or MM-GBSA calculations allow for relating molecules conformational changes to their free energy

\* Corresponding author e-mail: daniel.borgis@univ-evry.fr (D.B.) and thaduong@univ-evry.fr (T.H.-D.).

variations and understand in detail the physical factors that favor or disfavor the stability and specificity of biomolecular systems. To avoid to generate computationally expensive molecular trajectories in explicit solvent, the GBSA solvent model can be implemented directly in an MD algorithm in order to provide in the same calculation both stable dynamics simulations and reliable solvation free energy estimations.<sup>20–22</sup> This all-implicit solvation approach is much faster than the former MM-GBSA combination, but it neglects the molecular aspect of solvent, and it is therefore unable to provide structural information about the macromolecule first solvation shells that can be related or compared to observations from NMR or X-ray experiments.

We propose here an alternative solvent model that combines advantages of both explicit solvent (molecular aspect for structural information) and implicit solvent (efficient and rapid evaluation of solvation free energy). This model, which theoretical foundations are summarized in the first section, considers each solvent molecule as a Lennard-Jones particle that embeds a polarizable electric dipole (polarizable pseudoparticles or PPP model). The polarization of solvent dipoles obeys the macroscopic electrostatic laws of dielectric materials and provides a simple expression of the electrostatic solvation free energy. It should be noted that other promising methods tend to combine an efficient evaluation of solvation free energies with a characterization of the averaged solvation structure at a molecular level. Among these models, one should mention the density functional theory (DFT) approach that describes the aqueous environment around a solute in terms of both molecular density and polarization density<sup>23–25</sup> and the three-dimensional integral equation theory which is able to provide averaged positional and orientational distribution function of the water molecules around a solute.<sup>26–28</sup> The practicability and accuracy of these elaborated methods have been shown for various compounds in solution, but they still seem computationally too expensive for long molecular dynamics simulations of large biomolecules. As it will be shown in the methodology section, our original hydration model can be seen essentially as a particular DFT formulation for the solvent polarization density but projected on moving particles instead of fixed grid points. It can also be related to the iterative and noniterative grid-based Langevin dipole (LD) model,<sup>29–31</sup> the surface constrained soft sphere dipoles model of Warshel,<sup>32</sup> and to the more general class of dipolar solvent models.<sup>23–25,33–37</sup>

Our solvent model has already been parametrized, tested, and applied to small peptides and proteins with satisfactory results, particularly in terms of CPU gain, since our model is about 5-fold faster than an explicit solvent model for simulations of solutes with around 1000 atoms.<sup>38,39</sup> It can provide stable MD trajectories in the nanosecond range, yielding averaged structures in accordance with experimental observations as well as reasonable estimations of solvent electrostatic free energies with good correlations with Poisson–Boltzmann calculations.<sup>39</sup> In this paper, we report the results and analyses of nanoseconds long and stable molecular dynamics simulations of four nucleic acids molecules dissolved in our PPP semi-implicit solvent model. Trajec-

tories obtained in our previous work for two proteins, the bovine pancreatic trypsin inhibitor (5pti) and the B1 immunoglobulin-binding domain of streptococcal protein (1pgb), will be analyzed too for a broader perspective.

## Methodology

**Summary of the PPP Solvent Model.** According to a macroscopic electrostatic theory, as formulated by Marcus,<sup>33</sup> when a solute built up with atomic point charges is immersed in a dielectric medium, a nonequilibrium electrostatic solvation free energy functional can be written as

$$\Delta F_{\text{pol}} = \int \frac{\mathbf{P}^2}{2\epsilon_0(\epsilon - 1)} dv - \int \mathbf{P} \cdot \mathbf{E}_0 dv - \frac{1}{2} \int \mathbf{P} \cdot \mathbf{E}_p dv \quad (1)$$

where  $\epsilon_0$  is the vacuum permittivity and  $\epsilon$  is the solvent dielectric constant,  $\mathbf{P}$  is the nonequilibrium polarization density vector in a solvent elementary volume  $dv$ ,  $\mathbf{E}_0$  is the vacuum electrostatic field created by the solute atoms, and  $\mathbf{E}_p$  denotes the polarization electric field created by the polar medium surrounding the volume  $dv$ . The first integral is the free energy necessary to polarize the solvent volume  $dv$ . The second and third integrals account for the solute–solvent and solvent–solvent electrostatic interactions, respectively. Introducing into the Marcus functional the so-called “Coulomb field” or “local” approximation,<sup>40,37</sup> which assumes that the polarization field remains longitudinal, yields the simplified solvation free energy functional:<sup>35</sup>

$$\Delta F_{\text{pol}} = \int \frac{\epsilon \mathbf{P}^2}{2\epsilon_0(\epsilon - 1)} dv - \int \mathbf{P} \cdot \mathbf{E}_0 dv \quad (2)$$

The local approximation has been tested carefully for the solvation free energy of different solutes of complex shape in comparison to the Poisson–Boltzmann model.<sup>37</sup> In the case considered here of nonpolarizable solutes, the correlation with respect to Poisson–Boltzmann calculations is good, with correlation coefficients very close to 1, but the slope of the correlation line departs slightly from 1. It was shown that the agreement can be greatly improved by defining a systematic renormalization of the solute atomic charges in order to increase slightly the local dipoles. This feature will arise later in the computational details and parametrization section.

To calculate numerically the two integrals involved in eq 2, one could classically discretize the solvent region with a regular grid and consider at each point a dipole  $\mu_i = \mathbf{P} dv = \mathbf{P}/\rho$ , where  $\rho$  is the grid density. Then the electrostatic solvation free energy can be evaluated as

$$\Delta F_{\text{pol}} = \sum_i \frac{\mu_i^2}{2\alpha} - \sum_i \mu_i \cdot \mathbf{E}_{0i} \quad (3)$$

where  $\mathbf{E}_{0i}$  is the vacuum electrostatic field created by the solute atoms at the dipole  $\mu_i$  position. The constant  $\alpha$  is defined as  $\alpha = \epsilon_0(\epsilon - 1)/(\epsilon\rho)$ . It should be noted that  $\alpha$  arises as a microscopic polarizability constant that is related to the medium macroscopic dielectric constant  $\epsilon$ , and, therefore, it includes both the electronic and orientational polarization. The electrostatic solvation free energy at thermodynamic

equilibrium is obtained by minimizing  $\Delta F_{\text{pol}}$  relative to all the  $\mu_i$  which yields the following two simple expressions:

$$\mu_i^{\text{eq}} = \alpha \mathbf{E}_{0i} \quad (4)$$

$$\Delta F_{\text{pol}}^{\text{eq}} = -\frac{1}{2} \sum_i \mu_i^{\text{eq}} \cdot \mathbf{E}_{0i} \quad (5)$$

At this point, using eqs 4 and 5, the central idea of our method is to consider that the punctual dipoles  $\mu_i^{\text{eq}}$ , are carried by discrete and mobile Lennard-Jones particles of molecular size, instead of the fixed nodes of a regular grid. Two particular features arising from our model should be emphasized: First, due to the local approximation, the induced dipoles do not interact with each other, and only Lennard-Jones interactions between solvent particles have to be calculated. Second, since the solvent pseudoparticles have translational degrees of freedom, the electrostatic solvation free energy  $\Delta F_{\text{pol}}^{\text{eq}}$  calculated from eq 5 is in fact an “instantaneous” polarization free energy that has to be averaged over different particles positions, generated for example by molecular dynamics simulations. In practice, the electrostatic solvation free energy  $\Delta F_{\text{pol}}^{\text{eq}}$  is included in the total Hamiltonian of the system, yielding the following expression:

$$H = K + V_{\text{Bond}}(\{\mathbf{R}_k\}) + V_{\text{Lj+Coul}}(\{\mathbf{R}_k\}) + V_{\text{Lj}}(\{\mathbf{r}_i\}) + V_{\text{Lj}}(\{\mathbf{R}_k - \mathbf{r}_i\}) + \Delta F_{\text{pol}}^{\text{eq}} \quad (6)$$

Here,  $K$  is the system kinetic energy,  $\mathbf{R}_k$  are the atomic positions of the solute,  $V_{\text{Bond}}(\{\mathbf{R}_k\})$  denotes the solute bonded potential energy, including the stretching, bending, and torsion interactions, and  $V_{\text{Lj+Coul}}(\{\mathbf{R}_k\})$  is nonbonded Lennard-Jones and Coulomb energy. The solvent pseudoparticles are characterized by a position  $\mathbf{r}_i$  and a solvent–solvent (possibly truncated) Lennard-Jones potential  $V_{\text{Lj}}(\{\mathbf{r}_i\})$ . The solute–solvent interactions are captured by the Lennard-Jones function  $V_{\text{Lj}}(\{\mathbf{R}_k - \mathbf{r}_i\})$  and the electrostatic term  $\Delta F_{\text{pol}}^{\text{eq}}$  described in eq 5. From the total potential energy of the system, the forces formulations are the same as in explicit models, except for the electrostatic solute–solvent forces which can be expressed in terms of partial derivatives of the vacuum electrostatic field  $\mathbf{E}_{0i}$ , on a solute atom ( $\mathbf{k}$ ) and a solvent pseudoparticle ( $\mathbf{i}$ ), respectively, by the following expressions:

$$\mathbf{F}_k = -\frac{\partial \Delta F_{\text{pol}}^{\text{eq}}}{\partial \mathbf{R}_k} = \sum_i \mu_i^{\text{eq}} \cdot \frac{\partial \mathbf{E}_{0i}}{\partial \mathbf{R}_k} \quad (7)$$

$$\mathbf{F}_i = -\frac{\partial \Delta F_{\text{pol}}^{\text{eq}}}{\partial \mathbf{r}_i} = \mu_i^{\text{eq}} \cdot \frac{\partial \mathbf{E}_{0i}}{\partial \mathbf{r}_i} \quad (8)$$

As shown in a previous paper, this solvation model leads to a potential of mean force for opposite charge ion pairs that is slightly too repulsive and without the expected two minima corresponding to the contact ion pair and the solvent separated ion pair.<sup>38</sup> This failure arises mainly from the local approximation, and in order to cure this deficiency, we have added another level of phenomenology to our solvation

model: We imposed that the solvent dipoles  $\mu_i^{\text{eq}}$  can saturate when the electrostatic field  $\mathbf{E}_{0i}$  becomes too strong, instead of supposing a simple proportionality. As in the Langevin dipole model,<sup>31</sup> this can be done by expressing the induced dipoles via a Langevin function

$$\bar{\mu}_i^{\text{eq}} = \frac{\mu_{\text{sat}}}{E_{0i}} L\left(3\alpha \frac{E_{0i}}{\mu_{\text{sat}}}\right) \mathbf{E}_{0i} \quad (9)$$

where  $L(x) = \coth(x) - 1/x$ , and  $\mu_{\text{sat}}$  is the value of the dipoles saturation. Finally, to keep the simplicity of the solute–solvent electrostatic forces expressed in eqs 7 and 8, it has been demonstrated that the electrostatic solvation free energy  $\Delta F_{\text{pol}}^{\text{eq}}$  has just to be reformulated as<sup>38</sup>

$$\Delta F_{\text{pol}}^{\text{eq}} = -\frac{\mu_{\text{sat}}^2}{3\alpha} \sum_i \ln \left[ \frac{\sinh\left(3\alpha \frac{E_{0i}}{\mu_{\text{sat}}}\right)}{3\alpha \frac{E_{0i}}{\mu_{\text{sat}}}} \right] \quad (10)$$

**Computational Details and Parametrization.** In this work, the force field used for nucleic acids molecules is the second generation Amber94 force field.<sup>41</sup> As in our previous study on proteins,<sup>39</sup> the solvent pseudoparticles are characterized by the Lennard-Jones parameters  $\sigma_{\text{LJ}} = 2.88 \text{ \AA}$  and  $\epsilon_{\text{LJ}} = -3.197 \text{ kJ/mol}$  and a polarizability  $\alpha = 2.33 \text{ \AA}^3$  calculated from the water dielectric constant  $\epsilon = 80$  and a solvent density  $\rho = 0.0337 \text{ \AA}^{-3}$ . After trial and error, the value of the dipole saturation was fixed at an optimum value  $\mu_{\text{sat}} = 1.5 \text{ D}$ .

Since our solvent model is based on continuum electrostatics, we had to calibrate the atomic effective radii and partial charges in order to reproduce the known Born and Kirkwood electrostatic solvation free energies of isolated charged and neutral dipolar atoms. To fit to the Born solvation free energy,<sup>42</sup>  $\Delta G_{\text{Born}} = -1/2 (1-1/\epsilon)q^2/r$ , we proceeded in the following way: For ions of different charge  $q$  and radius  $r$ , we first determined the solvation energy  $\Delta F_{\text{pol}}^{\text{eq}}$  given by the PPP model, and we then reassigned to each of them an effective radius  $r_{\text{eff}} = f_{\text{rd}} r$  where the rescaling factor is defined as  $f_{\text{rd}} = \Delta G_{\text{Born}} / \Delta F_{\text{pol}}^{\text{eq}}$ . For further studies with any force field, the atomic radii used for solute–solvent Lennard-Jones interactions have been rescaled systematically with the factor  $f_{\text{rd}}(r_i, q_i)$ .

As mentioned in our previous paper,<sup>39</sup> the local approximation in the PPP solvent model introduces a systematic error in the electrostatic solvation free energy of small neutral dipolar molecules. In the case of a point dipole  $\mu$  embedded in a hard sphere of radius  $R$  and surrounded by a high dielectric material, one can show that the relative error with respect to the exact Kirkwood solvation free energy<sup>43</sup>  $\Delta G_{\text{Kirkwood}} = -1/2(\epsilon - 1)/(\epsilon + 1/2)\mu^2/R^3$  amounts to 1/3. One way to correct these errors is to assign to the small molecules an effective dipole  $\mu_{\text{eff}} = (3/2)^{1/2}\mu$  without changing their total charge. This can be done by rescaling the solute atomic partial charges  $q_i$  using the formula  $q_i^{\text{eff}} = f_q q_i + (1 - f_q)Q_{\text{tot}}/N$ , where  $Q_{\text{tot}}$  is the total charge of the small molecules,  $N$  is its number of atoms, and  $f_q = (3/2)^{1/2}$

$\approx 1.225$ . It should be noted that the factor  $f_q = 1.225$  which applies strictly to a spherical dipole with sharp boundaries should be slightly adjusted to accommodate for smooth and continuous Lennard-Jones boundaries. Numerous tests have led to the conclusion that the best results in terms of both structural stability and free energy correlations with Poisson–Boltzmann calculations are obtained with the factor  $f_q = 1.185$ .<sup>39</sup>

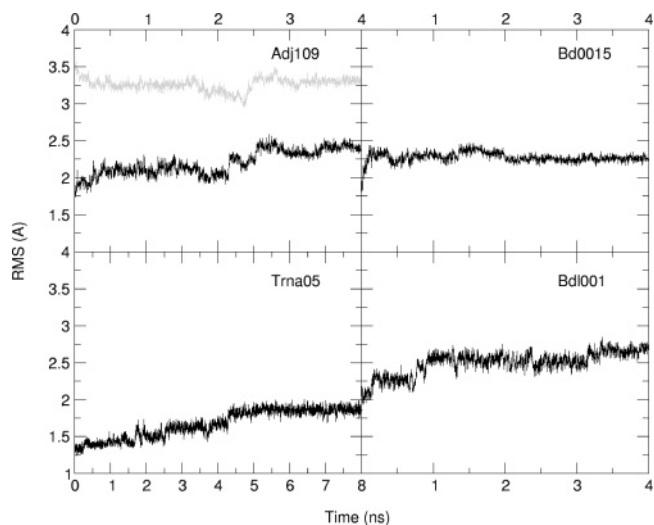
Our solvation model has been implemented into the molecular dynamics program Orac<sup>44</sup> which uses a multiple time step r-RESPA algorithm.<sup>45</sup> In this work, three different time steps were used: a 4 fs time step for the integration of long-range nonbonded interactions (greater than 8 Å), a 2 fs time step for medium range nonbonded interactions (between 5 and 8 Å), and a 1 fs time step for short-range nonbonded (lower than 5 Å) and bonded interactions. The nonbonded interactions were smoothly cut with a switching function from 11 to 12 Å. All simulations were performed in the canonical NVT ensemble, using a Nosé-Hoover thermostat algorithm.<sup>46,47</sup> All systems have been equilibrated progressively from 220 to 300 K during 228 ps and simulated at this last temperature for 4 ns without any constraints except those on bonds involving hydrogen, using the SHAKE algorithm.<sup>48</sup>

Poisson–Boltzmann calculations of electrostatic solvation free energies of molecules were performed using the APBS software<sup>49</sup> with no ionic strength (explicit ions are included in the simulations) using a dielectric constant of  $\epsilon_i = 1$  inside the solute and  $\epsilon = 80$  outside and a solvent radius of 1.4 Å for probing the solute surface. The cubic grid for the finite differences algorithm of APBS was set to 141 nodes in each dimension for a rectangular box of the same size as the one used for the MD simulations, so that the grid size was below 0.5 Å.

## Results

**Stability of Proteins and Nucleic Acids Simulations.** The PPP solvent model has been used previously to simulate two solvated proteins, the bovine pancreatic trypsin inhibitor (PDB code: 5pti) and the B1 immunoglobulin-binding domain of streptococcal protein G (1pgb). The 2-ns long production trajectories were proved stable and meaningful, with RMSD comparable to all-atom simulations and structural characteristics in overall agreement with X-rays and NMR measurements.<sup>39</sup>

In this paper we substantiate our study with the structure, dynamics, and solvation properties of four different nucleic acid molecules. The starting structures were taken from the crystallographic structures available in the Nucleic Acid Database:<sup>50</sup> the 17 bases anticodon hairpin of the Asp-tRNA (trna05),<sup>51</sup> the decamer d(CCGCCGGCGG) in both A form (adj109)<sup>52</sup> and B form (bd0015),<sup>53</sup> and the containing EcoRI restriction site dodecamer d(CGCGAATTCGCG) (bdl001).<sup>54</sup> These four negatively charged molecules have been neutralized with an appropriate number of sodium counterions initially placed at 6 Å from the phosphate groups, except for bd0015 for which ions were placed at 3 Å. All the molecular systems considered, proteins and nucleic acids were simulated in a  $48 \times 48 \times 70.7$  Å<sup>3</sup> box, containing about

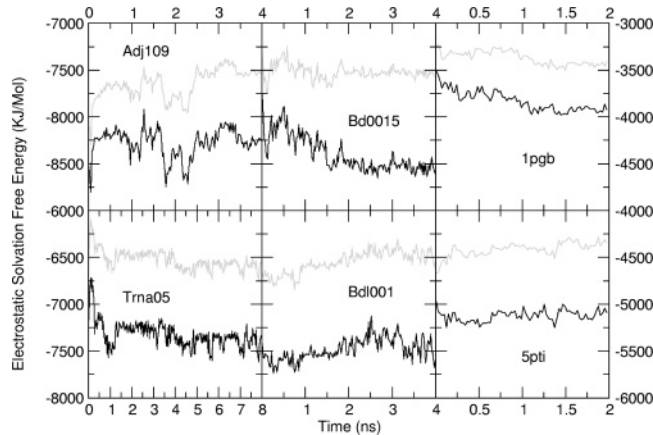


**Figure 1.** Time evolutions of the RMS deviation calculated over nucleic acids all non-hydrogen atoms from their initial structures (black lines). For the d(CCGCCGGCGG) decamer in A form, the RMS deviation from the B conformation is also plotted in gray line.

5200 solvent pseudoparticles. Molecular dynamics trajectories are 4-ns long for the three DNA molecules, whereas the trna05 trajectory was carried on for a total of 8 ns.

Figure 1 represents the RMS deviations (calculated over all non-hydrogen atoms) from the starting structures as a function of time for the four studied nucleic acids. This global structural criterion shows that all nucleic acid molecules are stable in a PPP solvent and remain close to their initial conformation during the 4 ns trajectories, the maximum RMS deviation being observed around 2.7 Å for the dodecamer bdl001, which is very close to the value obtained for a 5 ns MD simulation using a TIP3P water model.<sup>55</sup> The relatively small RMS deviation observed for the tRNA loop during the whole length of the 8-ns trajectory (around 1.9 Å at the end of the trajectory) is comparable to that obtained in a series of 3-ns long MD simulations performed in an explicit SPC/E solvent (1.7 Å on average) by Auffinger and Westhof.<sup>56</sup> It reveals that the tRNA structure is globally very stable with the PPP solvent, although it has been previously shown that this tRNA loop is particularly sensitive to the representation of solvent.<sup>56</sup> It reflects smaller deformations of RNA compared to DNA molecules, even in A form, for which the RMS deviation is stabilized around 2.5 Å. In Figure 1 is also plotted the RMS deviation of the decamer d(CCGCCGGCGG) in A form relative to the crystallographic B conformation. These two structures being initially distant by a RMSD of about 5.4 Å, it appears that the simulated A-DNA slightly drifts toward a B-like conformation, with a final RMSD around 3.3 Å away from the experimental B form. As it will be detailed below, this A to B partial transition, which is not observed for RNA, has been characterized by numerous molecular dynamics simulations on various DNA sequences with different force fields and water models.<sup>2,57–60</sup>

The partial conclusion to be drawn here is that, as for proteins, the PPP solvent model allows for the reproduction of the global structural stability of different kinds of nucleic

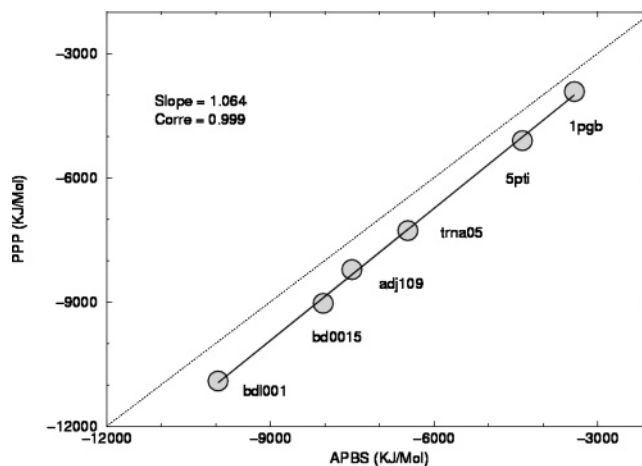


**Figure 2.** Time evolution of the nucleic acids and proteins electrostatic solvation free energies. Calculations with the PPP solvent model and with the Poisson–Boltzmann equation are plotted respectively with black lines and gray lines. For clarity and convenience, all plots have been translated into the same energy range.

acids molecules in comparison to available MD simulations using explicit solvent models.

**Electrostatic Solvation Free Energies for Proteins and Nucleic Acids.** As mentioned in the Introduction, the PPP solvent model provides an efficient way to estimate “on-the-fly” electrostatic solvation free energies during molecular dynamics simulations. Nevertheless, it should be mentioned that, whereas the solvent dipole orientations are in equilibrium at each simulation step, the pseudoparticles translational positions have to be sampled in order to yield an averaged solvation free energy corresponding to a real equilibrium state. Previous tests on small peptides have shown that these translational degrees of freedom are rapidly equilibrated within a few dozen picoseconds of simulation.<sup>39</sup> Thus, to compare the electrostatic solvation free energies estimated with the PPP model and by the Poisson–Boltzmann equation resolution, we have proceeded in the following way: Over time-windows of 12 ps, the average value of the PPP electrostatic solvation free energy was compared to the electrostatic solvation free energy calculated using the APBS software<sup>49</sup> to solve the Poisson–Boltzmann equation for an averaged molecular structure determined over the same time-window (see Methodology section).

In Figure 2 we have plotted the time evolution of the electrostatic solvation free energies calculated “on-the-fly” with the PPP solvent model and a-posteriori with the APBS algorithm. The figure includes the four nucleic acid molecules and the two proteins (5pti and 1pgb). These two charged proteins were simulated without neutralizing counterions and using a nonbonded interaction cutoff of 12 Å. Therefore, to account for the long-range solute–solvent electrostatic contribution of the solvation free energy (which is negligible for neutral solutes), we have added to the PPP solvation free energy a bulk correction evaluated by a continuum approximation using the Born formula  $\Delta G_{\text{Bulk}} = -1/2(1-1/\epsilon)Q^2/R_{\text{Bulk}}$ , where  $Q$  is the solute total charge and  $R_{\text{Bulk}}$  is the distance beyond which the solvent dipoles were supposed insensitive to the solute electrostatic field. In our simulations, the  $R_{\text{Bulk}}$  distance is physically defined as the

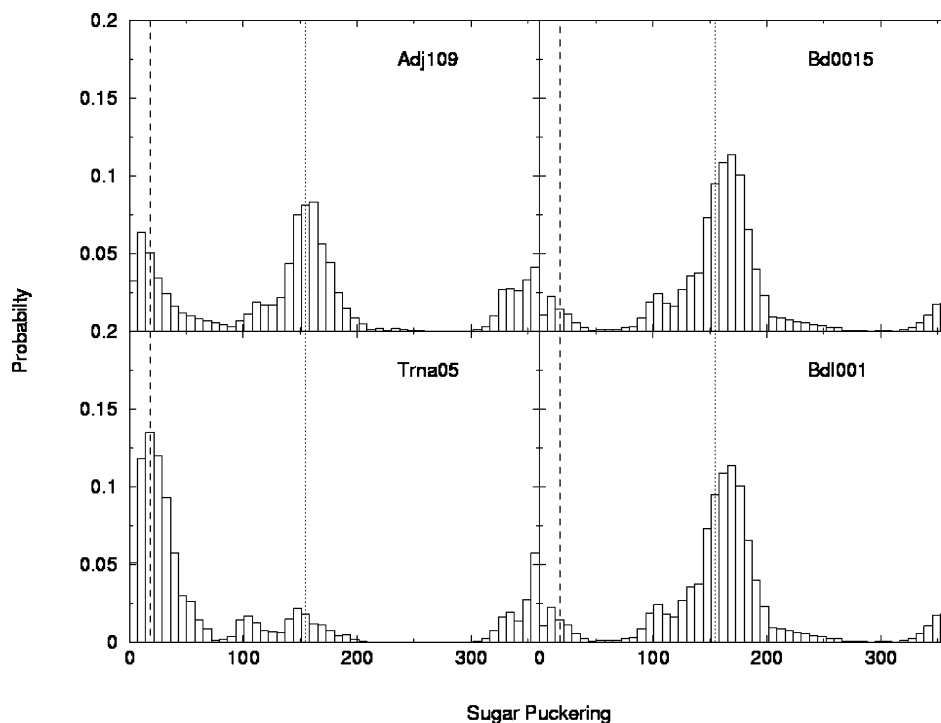


**Figure 3.** Comparison between the PPP and Poisson–Boltzmann electrostatic solvation free energies for the four nucleic acid molecules and the two proteins described in the text.

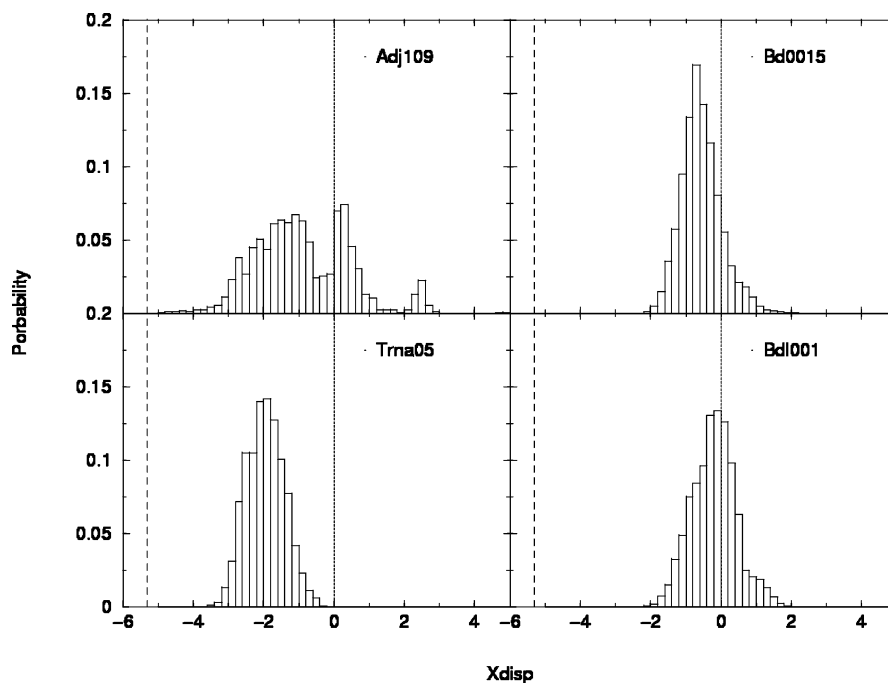
sum of the protein averaged gyration radius (about 14 Å for the two proteins) and the nonbonded interaction cutoff distance.

Despite the fact that we do not compare completely identical physical quantities (an average of free energies versus a free energy of an averaged structure), the plots in Figure 2 indicate visually that the variations of the PPP electrostatic solvation free energy due to the solutes conformational changes correlate well with those calculated with the Poisson–Boltzmann equation. Despite a systematic shift between the two values, correlation coefficients between the two methods range from 0.83 to 0.89 for nucleic acids (calculated over the last 2 ns of the trajectories) and from 0.76 to 0.86 for proteins<sup>39</sup> (calculated over the last 1.75 ns of the trajectories). This trend is particularly noteworthy for the A-DNA molecule which undergoes a significant conformational transition to a mixed A and B form. Similar good correlations have been previously observed for two small 10 residues peptides that undergo large conformational variations.<sup>39</sup> These encouraging results show that the PPP solvent model can efficiently reproduce “on-the-fly” the fine structure dependent solvation free energy of flexible solutes, at the expense of a relatively short time average over the translational degrees of freedom of the pseudoparticles.

We have summarized the applicability of our PPP model to the estimation of electrostatic free energies in Figure 3. There are displayed the averages over the 2 last ns of simulations of the PPP and APBS electrostatic solvation free energies (except for the proteins, for which averages were taken on the last equilibrated nanosecond of the trajectories). This plot reveals an excellent correlation between the two approaches for all studied biomolecules, the correlation coefficient and the slope of the linear relationship being equal to 0.99 and 1.06, respectively. Reminded that the PPP solvent model can be used for rapid and long molecular dynamics simulations, this result opens prospects for future studies focusing on the solvent influence on the DNA-protein or protein–protein association specificity, for which both the solutes deformation and dehydration process play a important role. For this purpose, however, further validation is needed,



**Figure 4.** Probability distributions of the sugar pucker phases ( $^{\circ}$ ) for the four studied nucleic acid molecules. The canonical values of the A- and B-DNA are represented respectively with dashed and dotted lines.

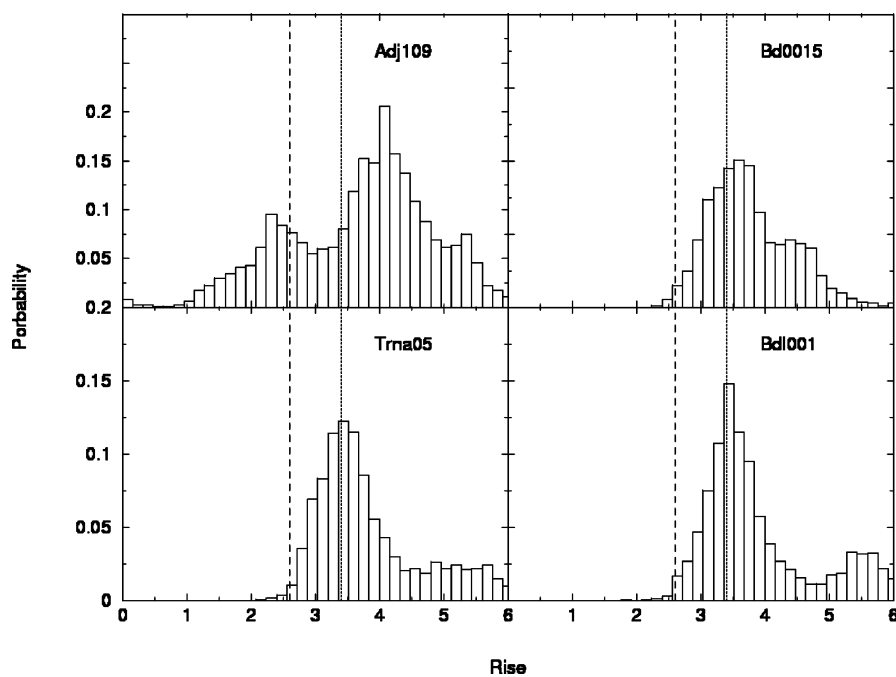


**Figure 5.** Probability distributions of the base pairs  $X_{\text{disp}}$  ( $\text{\AA}$ ) parameters, relative to the best curved helices axis.

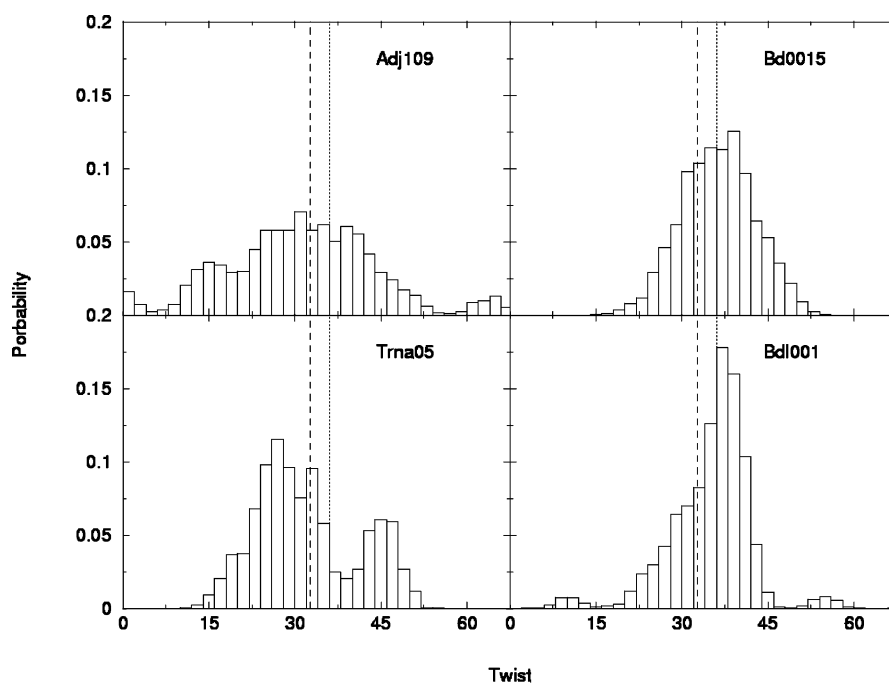
like the prediction of mutational effects on protein–protein association<sup>61</sup> or the comparison to experimental affinities. Provided this validation, the PPP solvent model can provide an efficient and self-consistent alternative to MM-PB or PDL/D/S-LRA calculations.

**Structural Analysis of the Nucleic Acids.** Since the structural properties of the proteins simulated in the PPP solvent were previously analyzed in details,<sup>39</sup> we focus here on the four nucleic acid molecules. We checked earlier in the paper the global stability of the molecules by looking at

time evolution of RMS deviations (see Figure 1), but more detailed analysis is needed in order to check the fine structure of nucleic acid molecules. We studied various helicoidal parameters of the DNA and RNA molecules, and we compared them with explicit simulations or experimental data. Various structural probability distributions pertinent to nucleic acids symmetries are displayed in Figures 4–7. These structural features were calculated on snapshots extracted every 12 ps from the 4 ns trajectories using the program CURVES (the first 4 ns in the case of trna05), which



**Figure 6.** Probability distributions of the interbase pairs rise ( $\text{\AA}$ ) parameters.



**Figure 7.** Probability distributions of the interbase pairs twist ( $^{\circ}$ ) parameters.

calculates the helicoidal parameters relative to an optimized nonlinear axis of the nucleic acid helices.<sup>62</sup> Regarding the sugar pucker phase (Figure 4), the distribution peaks of the two B-DNA are close to the canonical *B* value, in the typical C2'-endo conformation. Moreover, as expected from canonical values of RNA molecules, the probability distribution of the tRNA has one maximum at the canonical *A* value, in the C3'-endo conformation. In contrast, the sugar pucker phase probability of the A-DNA decamer has clearly a bimodal distribution with one peak around the canonical *A* value, the other around the *B* one. This clearly reflects a partial transition of the molecule toward an intermediate structure between an *A* and a *B* conformation. This inter-

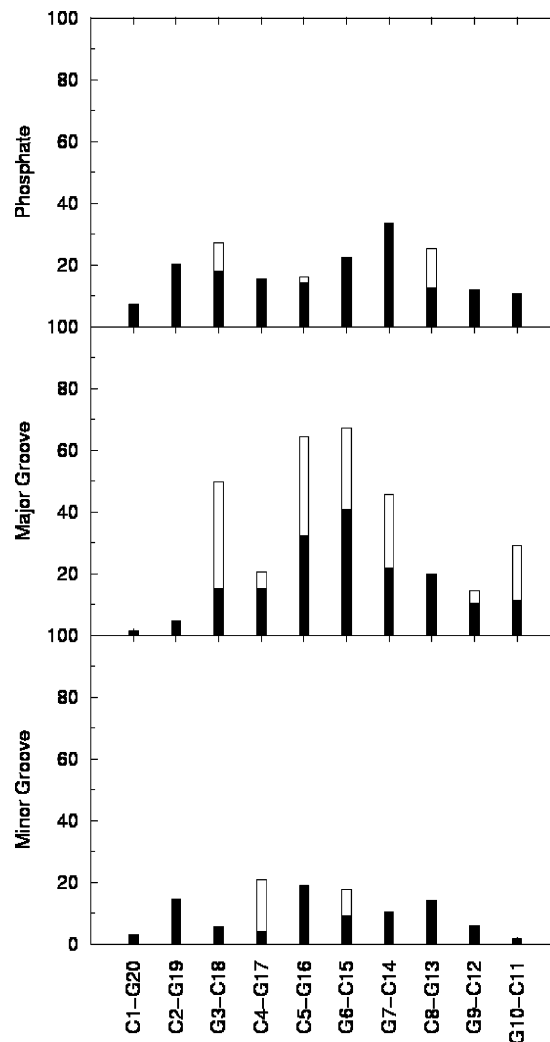
mediate structure is also confirmed by the probability distribution of the *Xdisp* parameter (Figure 5), which describes the displacement of the base pairs from the double helix axis. For the A-DNA, the *Xdisp* parameter also presents a bimodal distribution, whereas for the others nucleic acids, the distributions have only one maximum. This maximum is close to 0  $\text{\AA}$ , the canonical *B* value, for the two B-DNA, or has a negative value, characteristic of the *A*-form, for the tRNA molecule.

Figures 6 and 7 show the probability distributions of the rise and twist parameters, respectively, for the four studied nucleic acids helices. For the two B-DNA, the distributions are centered on the canonical *B* value. However, Figure 6



shows that the dodecamer bd1001 simulation yields some unusual high values of the rise parameter. A detailed analysis revealed that this is due to some incursions at intervals of two sodium ions between the base pairs C3/G4 and C9/G10 (data not shown). Similar incursions of counterions have also been observed between the two base pairs G3/C4 and G6/G7 of the A-DNA decamer adj109, for which the rise probability presents a bimodal distribution reflecting again a mixed form between A and B. The tRNA molecule simulation also yields high values of the rise parameter for the C3/G4 step (data not shown), but no incursion of counterions was observed in this case. This high rise is probably inherent to the presence of the wobble base pair G4-U14, which induces an unstacking and an underwinding of the RNA double helices as reviewed by Masquida and Westhof.<sup>63</sup> It should be noted that the rise parameters for the other tRNA base pairs are higher than the canonical A value, but they actually remain very close to the initial values of the crystallographic structure. The special geometry of the tRNA double helix due to the wobble base pair is also revealed by its twist parameter distribution (Figure 7), which shows a peak around  $45^\circ$  that arises from a high twist value for the G4/C5 step following the wobble pair. This overwound step is preceded by an underwound and high rise for the C3/G4 step, in agreement with experimental observations.<sup>63</sup> For both B-DNA molecules, as expected, the twist probability is sharply distributed around the canonical B value, whereas the A-DNA decamer distribution has a broad profile covering both the characteristic A and B values. Overall, despite some incursions of counterions between a few base pairs which can distort the double helices geometry at some intervals, the simulations of the four nucleic acids in the PPP solvent yield on average fine structures in good agreement with previous experimental and theoretical studies of the same molecules. This accomplishment should be stressed particularly for t-RNA, whose stability in silico was shown to be extremely sensitive to the solvent model employed, even in an explicit approach.<sup>56</sup>

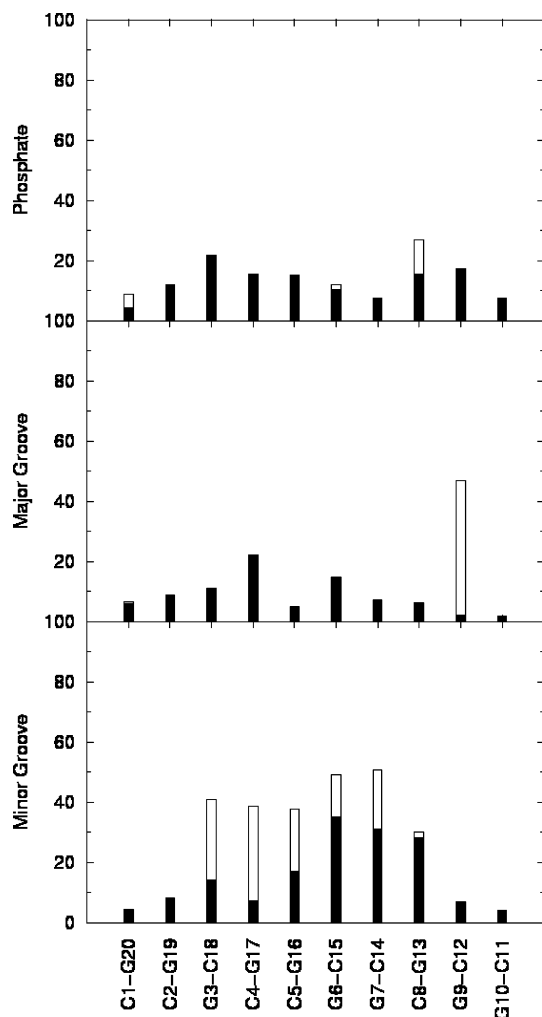
**Nucleic Acids Hydration Sites.** In contrast to a purely implicit solvent model such as the Poisson–Boltzmann equation or generalized Born, the PPP solvent is a particle-based model and thus keeps the molecular aspect of the solvent. Therefore, the PPP model allows for identifying the highly localized water molecules on the surface of biomolecules which are assumed to have a determinant structural influence upon their associations, as emphasized by NMR or crystallographic studies as well as MD simulations in explicit solvent.<sup>64</sup> The plots in Figures 8–10 represent a hydration index, the maximum residence time of a single solvent pseudoparticle close to different atom groups of the studied nucleic acids (defined in percentage with respect to the total simulation time). This quantity is similar to the “water residence time” usually measured or calculated from simulations. Since sodium counterions are able to occupy hydration sites, we have added to our “water residence time” the maximum residence time of sodium ions near the same sites. Three different groups of hydration sites were studied for each nucleic acid molecule: “Phosphate” hydration sites were considered as the four oxygen atoms bonded to the



**Figure 8.** Maximum residence time (in % of the simulation length) of water molecules (black bars) and sodium ions (additional white bars) near the A-DNA d(CCGCCGCGG) phosphate group, major groove and minor groove, as a function of the base pairs number.

phosphate atom. The “major groove” sites taken into account were the polar atoms N7 and O6 for guanine, the N7, N6, H61, and H62 atoms for adenine, the N4, H41, and H42 atoms for cytosine, and the O4 atom for thymine or uracil. The “minor groove” hydration sites were the N3, N2, H21, and H22 atoms for G, the N3 atom for A, and the O2 atom for C, T, or U. We specified that a solvent pseudoparticle or a sodium ion “occupies a hydration site” when its distance to a nitrogen or an oxygen site and to a hydrogen site were smaller than 3.0 and 2.5 Å, respectively. The results presented in Figures 8–10 are averages of the solvent and sodium maximum residence times over all the potential hydration sites for the three groups “phosphate”, “major groove”, and “minor groove” of each base (for tRNA) or base-pair (for all DNA molecules).

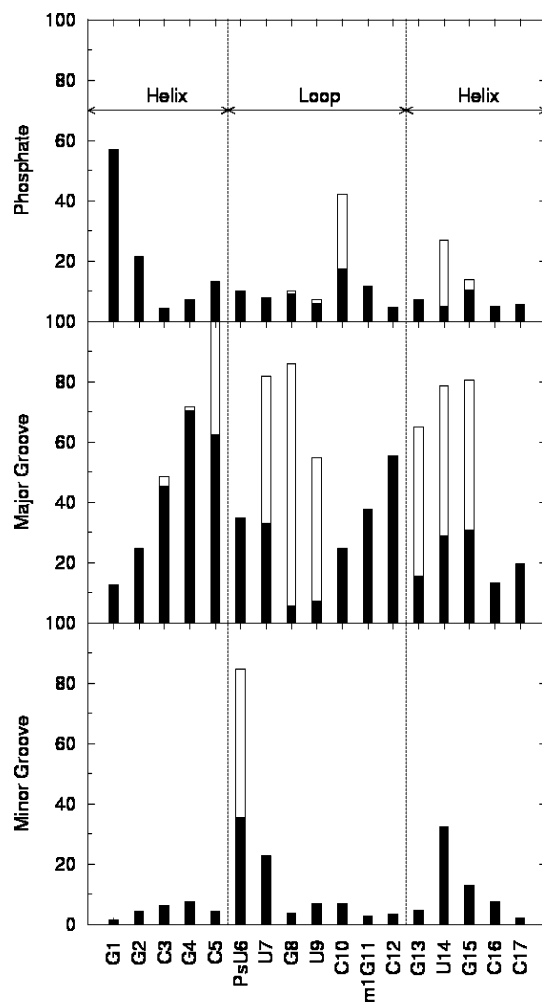
The first well-known result on nucleic acid hydration patterns that we reproduced is that the major grooves of A form double helices are more hydrated by the solvent pseudoparticles or sodium counterions than their minor grooves, whereas the longest water or ion residence times are found for the B-DNA minor groove rather than for the



**Figure 9.** Solvent particles and ions maximum residence time for the B-DNA d(CCGCCGGCGG).

major one. Despite the absence of explicit hydrogen bonds, our solvent model allows for the reproduction of the preferential hydration sites on the nucleic acids double helix surfaces in general agreement with experimental observations.<sup>3,52,65–67</sup> Concerning the decamer d(CCGCCGGCGG), it is remarkable in Figures 8 and 9 that several sodium counterions are highly localized into the major groove of the A conformation and into the B form minor groove, apparently replacing some long residency waters at these hydration sites. On the other hand, similar long-life sodium ions were not observed neither in the minor nor in the major groove of the B dodecamer d(CGCGAATTCGCG), especially in the AATT central region (data not shown). The presence of counterions in DNA minor groove and its sequence dependence is still theoretically unclear, especially by MD simulations, since the equilibration of ionic atmospheres around charged solutes in aqueous solvent turns out in the multnanoseconds range.<sup>1,2,4,55,68</sup>

Similarly to the A-DNA, the double-helix part of the tRNA hairpin is preferentially hydrated (by PPP solvent molecules or ions) in the major groove, as shown in Figure 10. We found that the base pairs around the wobble G4-U14 pair are particularly well hydrated. This result is in agreement with previous results showing that the intrabase pair hydrogen



**Figure 10.** Solvent particles and ions maximum residence time for the tRNA-Asp anticodon hairpin.

bonds within the wobble pair are weak and more accessible to the solvent.<sup>69</sup> Our results show that the bases of the tRNA loop are preferentially bound to waters or ions in their major groove side, except for the pseudouridine ( $\Psi$ 6). In this case, the polar atoms of the base in the minor groove side appears to be better hydration sites, in agreement with the theoretical study of the same hairpin hydration by Auffinger and Westhof.<sup>69</sup> As these authors also noticed, the three anticodon bases GUC are not particularly hydrated by water molecules. However, in our simulation, this base triplet seems to be strongly bound to sodium counterions, which has not been observed previously in other simulations.

Despite the absence of explicit hydrogen bonds in the PPP solvent model, the general hydration patterns of nucleic acids previously determined by experiments or simulations in explicit solvent were correctly reproduced. The PPP model can thus provide some molecular information about hydration which is not accessible by classical implicit models such as generalized Born or Poisson–Boltzmann. One of the future challenges will be to obtain reliable energetic results for, e.g., nucleic acids/protein association and compare carefully with experimental data or available all-atom simulations.<sup>15</sup>

**Computation Times.** To conclude this study, we should recall the numerical performances of the PPP solvent model that have been previously presented.<sup>39</sup> We showed that, in

our present (not fully optimized) implementation, the CPU gain of the PPP solvent model with respect to a typical 3-sites explicit water model (TIP3) is around 12.5 for pure solvent and drops to about 5 for solutes of about 800 atoms (like tRNA, DNA, and 1pgb) surrounded by ~5000 solvent molecules. Moreover, the computational overhead of PPP compared to vacuum simulations is about 8 for a solute of more than 800 atoms (for which solute–solvent interactions and solute–solute interactions become comparable), which makes the PPP solvent model competitive with respect to, e.g., the most reliable implementations of the generalized Born model.

## Conclusion

The polarizable pseudoparticle solvent model can be considered as the simplest possible particle-based model of a polar solvent. The particles induced dipoles do not see each other so that no self-consistent resolution is needed, and the solvent–solvent interactions are reduced to short-range (Lennard-Jones-like) interactions. When implemented in a molecular dynamics algorithm, this solvent model provides a fast and efficient way to simulate proteins and nucleic acids structures and dynamics as well as both the thermodynamic and structural aspects of their hydration. It makes it possible to reproduce the fine three-dimensional structures of biomolecules in agreement with experimental observations. For all studied biomolecules we observed a very good correlation between the electrostatic solvation free energy estimated with the PPP model and those accurately calculated with the Poisson–Boltzmann equation. This result allows for confidence for future investigations with the PPP model of the hydration influence on protein–protein or DNA-protein association specificity. Here, however, a careful comparison to experimental data and to available explicit water MD simulations results will be needed. Moreover, despite the absence of hydrogen bonds in the PPP solvent model, preferentially hydration sites of nucleic acids experimentally identified were correctly reproduced with our particle-based model, therefore adding some molecular information about hydration.

This study also reveals the unclear behavior of sodium counterions during the nucleic acids simulations due to their slow equilibration. To investigate more accurately the particular role of counterions in the structure and dynamics of highly charged solutes, an extension of our PPP model to include the ions implicitly in the model is currently ongoing. This will be done by adding on each solvent pseudoparticle a partial charge that will be determined at each MD step based on the minimization of a polarization and ionic density free energy functional equivalent to the Poisson–Boltzmann equation.

At the end, our PPP model can provide an interesting alternative to MM-PBSA, MM-GBSA, or PDL/D/S-SLRA calculations since the electrostatic free energies are estimated “on-the-fly”, as the MD trajectories of the system are propagated. The hydrophobic surface free-energy contribution can be easily evaluated on the fly too, using standard (and fast) algorithms. It could be also estimated by defining an average of the solvent particle density over short time

windows and an appropriate density free-energy functional. This approach is under study.

**Acknowledgment.** N.B. is supported by a grant from Genopole, which is gratefully acknowledged.

## References

- (1) Feig, M.; Pettitt, M. *Biophys. J.* **1999**, *77*, 1769–1781.
- (2) Auffinger, P.; Westhof, E. *J. Mol. Biol.* **2000**, *300*, 1113–1131.
- (3) Stefl, R.; Koca, J. *J. Am. Chem. Soc.* **2000**, *122*, 5025–5033.
- (4) McConnell, K. J.; Beveridge, D. L. *J. Mol. Biol.* **2000**, *304*, 803–820.
- (5) Schneider, C.; Brandl, M.; Suhnel, J. *J. Mol. Biol.* **2001**, *305*, 659–667.
- (6) Sen, S.; Nilsson, L. *Biophys. J.* **1999**, *77*, 1782–1800.
- (7) Komeiji, Y.; Uebayasi, M. *Biophys. J.* **1999**, *77*, 123–138.
- (8) Tang, Y.; Nilsson, L. *Biophys. J.* **1999**, *77*, 1284–1305.
- (9) Tsui, V.; Radhakrishnan, I.; Wright, P. E.; Case, D. A. *J. Mol. Biol.* **2000**, *302*, 1101–1117.
- (10) Paulino, M.; Esperon, P.; Vega, M.; Scazzocchio, C.; Tapia, O. *J. Mol. Struct.* **2002**, *580*, 225–242.
- (11) Paillard, G.; Lavery, R. *Structure* **2004**, *12*, 113–122.
- (12) Lavery, R. *Q. Rev. Biophys.* **2006**, in press.
- (13) Florian, J.; Goodman, M. F.; Warshel, A. *J. Phys. Chem. B* **2000**, *104*, 10092–10099.
- (14) Florian, J.; Goodman, M. F.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 5739–5753.
- (15) Florian, J.; Goodman, M. F.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6819–6824.
- (16) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.
- (17) Kollman, P. A.; Massova, I.; Reynes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III *Acc. Chem. Res.* **2000**, *33*, 889–897.
- (18) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891–913.
- (19) Muegge, I.; Tao, H.; Warshel, A. *Protein Eng.* **1998**, *10*, 1363–1372.
- (20) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
- (21) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275–291.
- (22) Calimet, N.; Schaefer, M.; Simonson, T. *Proteins* **2001**, *45*, 144–158.
- (23) Ramirez, R.; Gebauer, R.; Mareschal, M.; Borgis, D. *Phys. Rev. E* **2002**, *66*, 031206.
- (24) Ramirez, R.; Borgis, D. *J. Phys. Chem. B* **2005**, *109*, 6754–6763.
- (25) Ramirez, R.; Mareschal, M.; Borgis, D. *Chem. Phys.* **2005**, *319*, 261–272.
- (26) Pettitt, B. M.; Karplus, M.; Rossky, P. J. *J. Phys. Chem.* **1986**, *90*, 6335–6345.
- (27) Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 7821–7826.

- (28) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1–20.
- (29) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (30) Warshel, A.; Russel, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283–422.
- (31) Florian, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583–5595.
- (32) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640–1652.
- (33) Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 979–989.
- (34) Pollock, E. L.; Alder, B. *J. Phys. A* **1980**, *102*, 1–21.
- (35) Calef, D. F.; Wolynes, P. G. *J. Phys. Chem.* **1983**, *87*, 3387–3399.
- (36) Frodl, P.; Dietrich, S. *Phys. Rev. A* **1992**, *45*, 7330–7354.
- (37) Borgis, D.; Levy, N.; Marchi, M. *J. Chem. Phys.* **2003**, *119*, 3516–3528.
- (38) Ha-Duong, T.; Phan, S.; Marchi, M.; Borgis, D. *J. Chem. Phys.* **2002**, *117*, 541–556.
- (39) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Comput. Chem.* **2004**, *25*, 1015–1029.
- (40) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578–1599.
- (41) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (42) Born, M. Z. *Phys.* **1920**, *1*, 45–48.
- (43) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351–361.
- (44) Procacci, P.; Darden, T. A.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848–1862.
- (45) Humphreys, D. D.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem.* **1994**, *98*, 6885–6892.
- (46) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (47) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (48) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Chem.* **1977**, *23*, 327–341.
- (49) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.
- (50) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S. H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751–759.
- (51) Comarmond, M. B.; Giege, R.; Thierry, J. C.; Moras, D.; Fischer, J. *Acta Crystallogr. B* **1986**, *42*, 272–280.
- (52) Mayer-Jung, C.; Moras, D.; Timsit, Y. *Embo J.* **1998**, *17*, 2709–2718.
- (53) Timsit, Y.; Moras, D. *Embo J.* **1994**, *13*, 2737–2746.
- (54) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179–2183.
- (55) Young, M. A.; Ravishanker, G.; Beveridge, D. L. *Biophys. J.* **1997**, *73*, 2313–2336.
- (56) Auffinger, P.; Westhof, E. *Biophys. J.* **1996**, *71*, 940–954.
- (57) Cheatham, T. E., III; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434–444.
- (58) Mackerell, A. D.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105–120.
- (59) Trantirek, L.; Stefl, R.; Vorlickova, M.; Koca, J.; Sklenar, V.; Kypr, J. *J. Mol. Biol.* **2000**, *297*, 907–922.
- (60) Cheatham, T. E., III; Young, M. *Biopolymers* **2001**, *56*, 232–256.
- (61) Muegge, I.; Schweins, T.; Warshel, A. *Proteins* **1998**, *30*, 407–423.
- (62) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dynam.* **1988**, *6*, 63–91.
- (63) Masquida, B.; Westhof, E. *RNA* **2000**, *6*, 9–15.
- (64) Schwabe, J. W. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 126–134.
- (65) Shotton, M. W.; Pope, L. H.; Forsyth, T.; Langan, P.; Denny, R. C.; Giesen, U.; Dauvergne, M. T.; Fuller, W. *Biophys. Chem.* **1997**, *69*, 85–96.
- (66) Guerri, A.; Simpson, I. J.; Neidle, S. *Nucleic Acids Res.* **1998**, *26*, 2873–2878.
- (67) Makarov, V. A.; Pettitt, B. M.; Feig, M. *Acc. Chem. Res.* **2002**, *35*, 376–384.
- (68) Rueda, M.; Cubero, E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2004**, *87*, 800–811.
- (69) Auffinger, P.; Westhof, E. *J. Mol. Biol.* **1997**, *269*, 326–341.

# JCTC

Journal of Chemical Theory and Computation

## Computational Study of the Effects of Mutations A156T, D168V, and D168Q on the Binding of HCV Protease Inhibitors

Zhuyan Guo,<sup>\*,†</sup> Andrew Prongay,<sup>†</sup> Xiao Tong,<sup>‡</sup> Thierry Fischmann,<sup>†</sup> Stephane Bogen,<sup>§</sup> Francisco Velazquez,<sup>§</sup> Srikanth Venkatraman,<sup>§</sup> F. George Njoroge,<sup>§</sup> and Vincent Madison<sup>†</sup>

*Departments of Structural Chemistry, Medicinal Chemistry, and Antiviral Therapy, Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, New Jersey 07033*

Received April 25, 2006

**Abstract:** The effect of the resistance mutations A156T, D168V, and D168Q in HCV protease on the binding of SCH 6, SCH 503034, VX-950, BILN-2061, and compound 1 was evaluated using the free energy perturbation (FEP) approach. All the inhibitors are highly potent against the wild-type enzyme, but their activity was affected differently by the mutants. A156T reduced the activity of SCH 503034, BILN-2061, and VX950 drastically (200–1000-fold) but that of SCH 6 only moderately (27-fold). SCH 503034, SCH 6, and VX-950 were not affected by either mutation D168V or D168Q, but these mutations conferred a high level of resistance to BILN-2061. Comparison of BILN-2061 with its acyclic analogue compound 1 emphasized the importance of inhibitor flexibility in overcoming drug resistance arising from the D168Q mutation. The results from FEP calculations compared well with experimental binding potencies within an error of <1 kcal/mol. Structural analysis was carried out to relate the resistance profiles to the atomic changes in the mutants.

### Introduction

Hepatitis C virus (HCV) is a positive-strand RNA virus of the Flaviviridae family.<sup>1</sup> Its genome is translated into a polyprotein of ~3000 amino acids in infected cells.<sup>2–4</sup> HCV is a major cause of non-A-non-B hepatitis, which can lead to mortality through either cirrhosis-induced hepatic failure or hepatocellular carcinoma. About 170 million individuals worldwide are chronically infected. Currently, the only therapeutic regimens are subcutaneous interferon- $\alpha$  or PEG-interferon- $\alpha$  alone or in combination with oral ribavirin.<sup>5</sup>

As in the case of human immunodeficiency virus (HIV), most efforts to develop antiviral agents for HCV have focused on the inhibition of the key viral protease, helicase,

and polymerase enzymes.<sup>6</sup> The serine protease of HCV comprises the N-terminal domain of the NS3 protein and the NS4A cofactor. It is responsible for proteolytic cleavage at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A, and NS5A/NS5B sites of the nonstructural region of the encoded polyprotein and is essential for viral replication.<sup>7</sup> The NS3/NS4A serine protease has been the most extensively studied HCV target.<sup>8</sup>

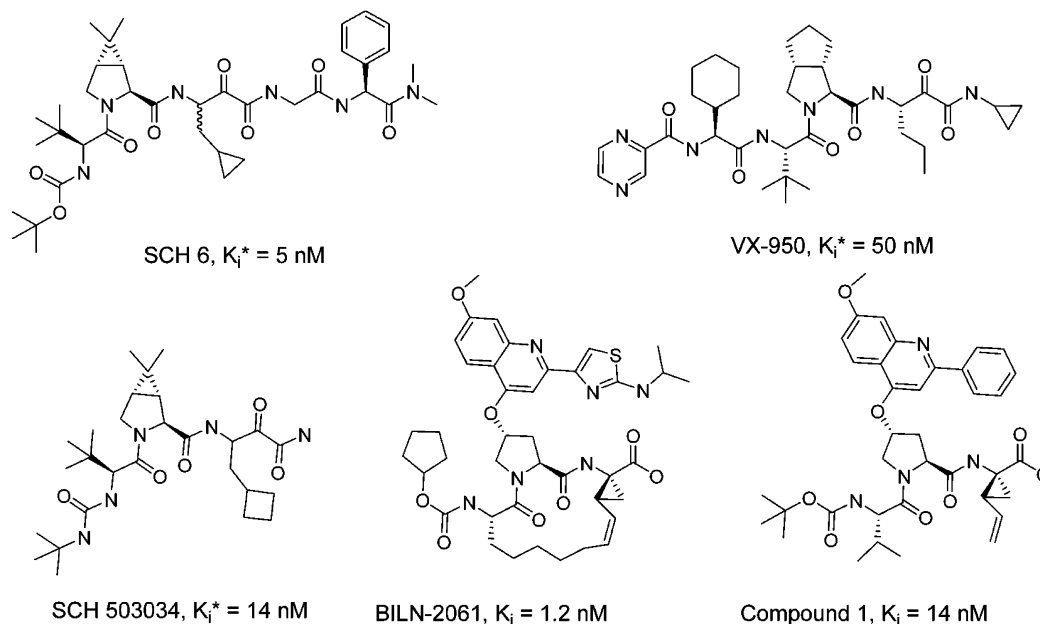
As in the case of other viruses, the emergence of drug resistance is a major concern in the development of HCV antiviral therapy. During emergence of escape variants, pre-existing minor viral species resistant to the selected drug will gain a growth advantage over the existing wild-type viral population and rapidly become the dominant genotype. Recently the NS3 protease inhibitors BILN-2061 and VX-950 have been reported to reduce viral loads in proof-of-concept clinical trials.<sup>9–11</sup> SCH 503034 has also advanced to clinical studies.<sup>12</sup>

\* Corresponding author phone: (908)740-3796; fax: (908)740-4640; e-mail: zhuyan.guo@spcorp.com.

<sup>†</sup> Department of Structural Chemistry.

<sup>‡</sup> Department of Antiviral Therapy.

<sup>§</sup> Department of Medicinal Chemistry.

**Table 1.** Chemical Structure and Experimental Fold Increase in Binding Activities  $K_i$  (or  $K_i^*$  To Indicate Covalent Bond) of HCV Protease Inhibitors upon Mutations A156T, D168V, and D168Q<sup>a</sup>

mutation	SCH 6	SCH 503034	VX-950	BILN-2061	compd1
A156T	27	400	390	1000	
D168V	0.4	1	0.4	260	
D168Q		1	1	70	2

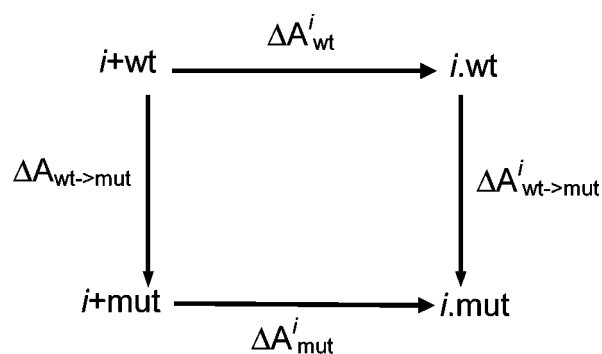
<sup>a</sup> The data were obtained in-house using the assay described.<sup>31</sup>

Protease resistance mutations have been identified by culturing replicon cells in the presence of these and additional inhibitors.<sup>11,13–16</sup> Losses of inhibitor potency have been quantitated,<sup>17,18</sup> and these are summarized in Table 1. The A156T mutation is a major determinant of resistance to BILN-2061, VX-950, and SCH 503034 with increases in  $K_i$  or  $K_i^*$  of 200–1000-fold. In contrast, SCH 6 was still relatively potent against this mutant with only a 27-fold increase in  $K_i^*$ . Only BILN-2061 was sensitive to mutations of D168 with increases of 260-fold for D168V and 70-fold for D168Q.

In an attempt to elucidate the variations in drug resistance for different HCV protease inhibitors, free energy perturbation (FEP) simulations<sup>19,20</sup> were carried out to study the effects of the two dominant mutations A156T and D168V on the binding of SCH 6, VX-950, BILN-2061, and SCH 503034. In addition, FEP simulations were performed to compare the effects of the D168Q mutation on the binding of BILN-2061 and its acyclic analogue compound 1.<sup>15</sup> This variant occurs in the genotype 3 virus and has been found to have a major impact on the binding of BILN-2061 but does not affect compound 1 or VX-950.<sup>13</sup> Structures from the simulations were analyzed to investigate relationships between the atomic changes of the mutations and the different resistance profiles of these inhibitors.

## Materials and Methods

**Theory.** The FEP method is a general computational approach to determine relative binding free energies.<sup>19–21</sup> It is a thermodynamically rigorous method that is capable of fine structural and energetic distinctions (e.g.,  $\Delta A < 1$



**Figure 1.** Schematic representation of the thermodynamic cycle. *i.wt* and *i.mut* represent inhibitor *i* bound to the wild-type and mutant enzymes, respectively. *i+wt* and *i+mut* indicate inhibitor *i* and the enzyme in the unbound state.

kcal/mol) in favorable cases. FEP is based on a thermodynamic cycle as depicted in Figure 1 for our mutational studies. Since free energy is a function of state only, the following relation holds for inhibitor *i*

$$\Delta A_{\text{mut}}^i - \Delta A_{\text{wt}}^i = \Delta A_{\text{wt} \rightarrow \text{mut}}^i - \Delta A_{\text{wt} \rightarrow \text{mut}} \quad (1)$$

Here  $\Delta A_{\text{wt}}^i$  and  $\Delta A_{\text{mut}}^i$  are the free energy of binding of inhibitor *i* to the wild-type and mutant enzymes, respectively, and  $\Delta A_{\text{wt} \rightarrow \text{mut}}^i$  and  $\Delta A_{\text{wt} \rightarrow \text{mut}}$  are the changes in free energy upon mutation in the presence and absence of inhibitor *i*. When applying eq 1 to both inhibitors *i* and *j*, we obtain the following equation

$$(\Delta A_{\text{mut}}^i - \Delta A_{\text{wt}}^j) - (\Delta A_{\text{mut}}^i - \Delta A_{\text{wt}}^i) = \Delta A_{\text{wt} \rightarrow \text{mut}}^j - \Delta A_{\text{wt} \rightarrow \text{mut}}^i \quad (2)$$

The left-hand side of eq 2, defined to be the relative resistance binding energy  $\Delta\Delta A$ , is the difference in binding energy between inhibitors  $i$  and  $j$  upon mutation and is related to the experimentally observed  $K_D$  values as

$$\Delta\Delta A = RT[\ln K_D(j_{\text{mut}})/K_D(j_{\text{wt}})] - RT[\ln K_D(i_{\text{mut}})/K_D(i_{\text{wt}})] \quad (3)$$

where  $K_D(i_{\text{mut}})$  and  $K_D(i_{\text{wt}})$  are the experimentally observed binding constant for inhibitor  $i$  with mutant and wild-type enzyme, respectively. In this work, the binding constant is denoted as  $K_i$  and  $K_i^*$ .  $K_i$  is the dissociation constant for the inhibitor–enzyme complex exactly analogous to  $K_D$  as the dissociation constant for the protein–ligand complex.  $K_i^*$  is simply an indication that there are two apparent stages to inhibitor binding: one fast and a second slower.  $K_i^*$  represents the overall dissociation constant for the inhibitor–enzyme complex.

From eq 2, the relative resistance binding energy can also be expressed as

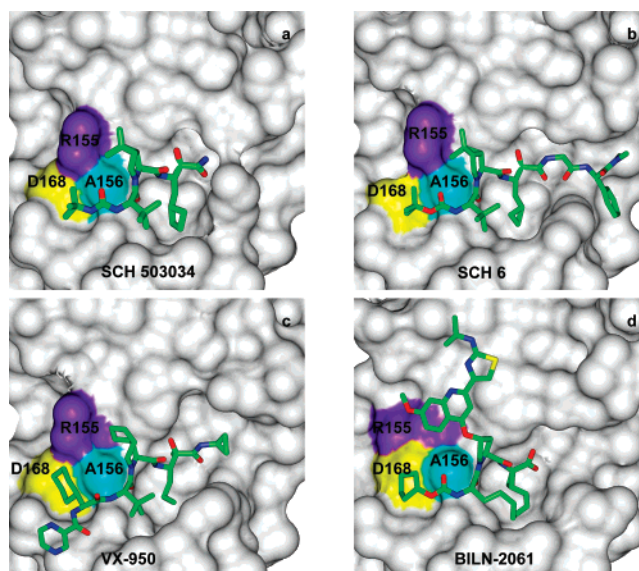
$$\Delta\Delta A = \Delta A_{\text{wt} \rightarrow \text{mut}}^j - \Delta A_{\text{wt} \rightarrow \text{mut}}^i \quad (4)$$

where  $\Delta A_{\text{wt} \rightarrow \text{mut}}^i$  and  $\Delta A_{\text{wt} \rightarrow \text{mut}}^j$  represent the free energy change upon mutation of the wild-type to mutant enzymes in the presence of inhibitors  $i$  and  $j$ , respectively. While the quantities on the left side of eq 2 are computationally challenging,  $\Delta A_{\text{wt} \rightarrow \text{mut}}^i$  and  $\Delta A_{\text{wt} \rightarrow \text{mut}}^j$  can be obtained through FEP calculations by slowly transforming the wild-type to the mutant enzyme or vice versa.

It is noteworthy that computationally one could mutate either the inhibitor or the protein. However, since the structure of the inhibitors under study differ significantly, we have chosen to mutate the protein side chains; specifically, T156 (mutant) was mutated to Ala (wild-type) and D168 (wild-type) was mutated to either Val or Gln (mutant), in the presence of the inhibitor. In general, mutation from a bulkier to a smaller side chain is preferred computationally, as in the case of T156 to Ala. However, since D168 is situated at the surface of the protein and is solvent exposed, it is less important as to the order of the mutations.

To perform FEP calculations, the wild-type (or mutant) enzyme is converted to the mutant (or wild-type) through a series of theoretical intermediate states by using a coupling parameter  $\lambda$  ( $0 < \lambda < 1$ ). To overcome sampling problems, simulations with different values of  $\lambda$  have to be performed. The free energy changes for all the  $\lambda$  intervals are added together to obtain the overall free energy change. As a rule of thumb, the free energy change for each  $\lambda$  interval should be less than  $2RT$  ( $R$  = gas constant,  $T$  = temperature) to ensure adequate sampling.<sup>19</sup> Of course, this also depends on the length of the simulation.

**Model Building and Simulation Details.** The X-ray coordinates for the A156T mutant complexed with SCH 503034 are available in-house (PDB file in preparation). Compared with the inhibitor binding to the wild-type enzyme,<sup>22</sup> the P2 and P3 backbone atoms of the inhibitor



**Figure 2.** Three-dimensional structures of SCH 503034, SCH 6, VX-960, and BILN-2061 binding to the HCV NS3-NS4A protease. The inhibitors are represented as stick models. Two resistance mutations at A156 (cyan) and D168 (yellow) are colored. R155 which shows a large conformational change upon binding of BILN-2061 is colored purple.

have moved away from T156. This is caused by the crowding in that region due to the mutation of A156 to the bulkier Thr residue. This A156T mutant protein structure was used as a starting point for all the A156T calculations. The coordinates of inhibitor SCH 6 were built using SCH 503034 as template. The remaining P1' and P2' residues of the inhibitor were modeled based on the crystal structure of this compound binding to the wild-type enzyme (PDB ID 2FM2). The coordinates were then optimized in the A156T mutant protein active site. Similarly, VX-950 was modeled based on SCH 503034, incorporating the binding mode information of an available peptide compound which has the same N-terminal cap (PDB ID 1RGQ). The coordinates of BILN-2061 were obtained from the crystal structure of BILN-2061 complexed with the wild-type enzyme determined in-house. For the D168V and D168Q mutations, the initial coordinates of SCH 503034 were taken from the crystal structure of the inhibitor complexed with the wild-type enzyme.<sup>22</sup> The coordinates of compound 1 were modeled based on the crystal structure of BILN-2061 by removing the P1–P3 macrocycle, modifying the P1 and P3 side chains, and modifying the N-terminal cap. Figure 2 illustrates the binding modes of the inhibitors.

The force field parameters for standard protein residues available in Quanta CHARMM<sup>23</sup> were used. For the non-standard residues of the inhibitors, atomic charges were assigned consistent with comparable standard residues. Other missing parameters such as bond and torsion angles were assigned by comparison with the parameters involving similar atom types.

The X-ray structures of protease–inhibitor complexes indicate that there is no notable movement in the active site upon inhibitor binding except for a few flexible side chains. Therefore, harmonic restraints were imposed on selected

atoms to prevent unrealistic movement during long dynamics simulations. For residues that are more than 10 Å away from the mutation center, a harmonic restraint with force constant 1.0 kcal/mol-Å<sup>2</sup> (mass weighted) was applied to both the side chain and backbone heavy atoms of the protein. The same harmonic restraint was also applied to the protein backbone atoms within 10 Å of the inhibitor. No restraint was imposed on those side chains that have at least one atom within 10 Å of the inhibitor heavy atoms. A weak harmonic restraint (0.05 kcal/mol-Å<sup>2</sup>, mass weighted) was found to be necessary to maintain the observed binding mode for the P2 group of BILN-2061. This mobility of the P2 group may be due to the shallowness of the S2 binding pocket. For consistency, the same restraint was imposed on all the inhibitor heavy atoms.

FEP simulations of the solvated protein–ligand complex were carried out using CHARMM.<sup>24</sup> A total of 11  $\lambda$  windows for A156T and 38 windows for either D168V or D168Q were used with double-wide sampling. The free energy change within each  $\lambda$  interval was less than  $2RT$  ( $\sim 1.5$  kcal/mol). The protein–inhibitor complex contained 181 residues from NS3 and 14 residues from NS4A plus the inhibitor. The system was centered at the site of mutation. A cap of TIP3P waters was added to fill the space in a 24 Å sphere.<sup>25</sup> The hydrogen atoms were built using CHARMM. During the dynamics simulations, the system temperature was maintained at 300 K. The velocities were reassigned every 200 steps using a Gaussian distribution if the temperature during this period was outside the target 300 K  $\pm$  10 K. Bonds involving hydrogen atoms were constrained using the SHAKE algorithm.<sup>26</sup> The time step of the simulation was 1.5 fs. A cutoff of 11 Å was used for nonbonding interactions (electrostatic and van der Waals). A solvent boundary force was used to retain the water molecules.<sup>27</sup> The system was equilibrated for 30 ps followed by 45 ps of data collection for each  $\lambda$  window, and the data were recorded every 10 time steps.

## Results and Discussion

FEP simulations on four inhibitors were performed for the A156T mutation. The computational results are reported as relative resistance binding energies  $\Delta\Delta A$  using SCH 6 as the reference inhibitor (Table 2). SCH 6 was predicted to be least sensitive to the mutation followed by VX-950, BILN-2061, and SCH 503034 with  $\Delta\Delta A$  values ranging from 0.8 to 1.3 kcal/mol. Multiple simulations performed on SCH 6 and VX-950 indicate that the computational statistical error is about 0.5 kcal/mol. For comparison with the computations, experimental fold increase ratios were converted to  $\Delta\Delta A$ s using eq 3. For SCH 503034, experimental and FEP  $\Delta\Delta A$ s agree within 0.3 kcal/mol, while the experimental values were underestimated by 0.8 and 0.9 kcal/mol for VX-950 and BILN-2061, respectively. All three values agree within the generally accepted precision estimate for FEP of  $\sim 1$  kcal/mol.

For mutations of D168, efforts were focused on BILN-2061 because of its sensitivity to these mutations, while none of the other inhibitors are affected significantly. For D168Q, compound 1 was used as reference to permit a direct

**Table 2.** Relative Resistance Binding Energies (kcal/mol) upon Mutations A156T, D168V, and D168Q<sup>b</sup>

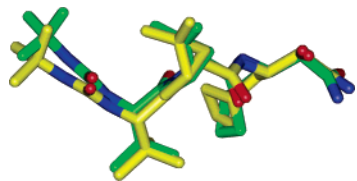
mutation	SCH 6	SCH 503034	VX-950	BILN-2061	compd 1
A156T					
experiment	0.0	1.6	1.6	2.1	
FEP	0.0 <sup>a</sup>	1.3	0.8 <sup>a</sup>	1.2	
D168V					
experiment		0.0		3.3	
FEP		0.0		2.8	
D168Q					
experiment				2.1	0.0
FEP				2.4	0.0

<sup>a</sup> For SCH 6 and VX-950, multiple simulations were performed, and the average values were used. The original data are SCH 6 = (−0.9, 0.1, 0.3, 0.5), VX-950 = (0.4, 1.2). <sup>b</sup> SCH 6, SCH 503034, and compound 1 are used as references for the A156T, D168V, and D168Q mutations, respectively.

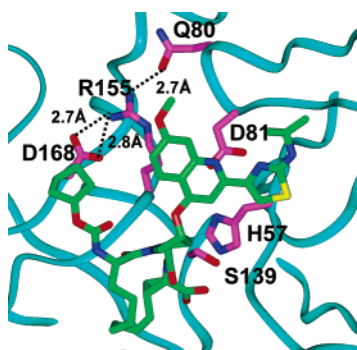
comparison of the impact of the macrocycle. For D168V, lacking experimental data for compound 1, SCH 503034 was used as reference due to the availability of its crystal structure. Note that the mutations D168V and D168Q are computationally more challenging because the perturbation of the side chain involves a net change in charge and therefore requires the use of a large number of intermediate states ( $\lambda$  windows) for good convergence. Thus, a total of 38 windows were needed to map the wild-type to the mutant enzyme, compared to 11 windows for the A156T mutation. Nevertheless, computed  $\Delta\Delta A$ s agree well with experimental values differing by 0.5 and 0.3 kcal/mol for D168V and D168Q, respectively.

The structural models suggest some factors that contribute to different resistance profiles for the various inhibitors. As in the case of all peptidic, active site inhibitors of HCV protease, there is a network of hydrogen bonds between the backbone atoms of the protease and the inhibitors connecting antiparallel  $\beta$ -strands. Three hydrogen bonds are common to the inhibitors of this study, namely, P1 NH–R155 CO, P3 CO–A157 NH, and P3 NH–A157 CO. Additional hydrogen bonds are formed between the P4 cap C=O of VX-950 and the NH of C159 and P2' NH of SCH 6 and the carbonyl of T42. Bordering the S2 and S4 pockets, the A156 side chain is in van der Waals contact with the P2 and P3 cap (or P4 for VX-950) groups of the inhibitors. Mutation of A156 to the bulkier Thr causes crowding in that region. As observed in the crystal structure of SCH 503034 complexed with the A156T mutant, the P2–P3 cap backbone atoms shifted away from T156 to avoid crowding (Figure 3). This backbone movement of the inhibitors was also observed in our simulations. This results in weaker hydrogen bonding interactions for both P3 CO–A157 NH and P3 NH–A157 CO. The inhibitor side chains in this region are also crowded. For example, T156 would crowd the P4 cyclohexyl group of VX-950. In turn, the movement of this group would affect inhibitor binding to the S4 and S5 subsites. For BILN-2061 backbone crowding at P2 by T156 would also shift the large P2 side chain, interfering with cation- $\pi$  interactions between the P2 group and the S2 region of the enzyme to reduce binding affinity. The rigidity of the BILN-2061 macrocycle may also reduce the inhibitor's





**Figure 3.** Superimposition of the structures of SCH503034 binding to the HCV NS3-NS4A wild-type and A156T mutant enzymes to illustrate the backbone movement. The carbon atoms are colored green in the wild-type and yellow in the mutant.



**Figure 4.** Binding of BILN-2061 to the HCV NS3-NS4A protease. The protein is shown as a ribbon drawing (cyan). The inhibitor is represented as a stick model (green). The side chains of R155, D168, and Q80 are shown (magenta) to illustrate the interactions that stabilize the R155 binding conformation. The catalytic triad S139, H57, and D81 are also displayed.

ability to adapt to changes in the enzyme surface. On the other hand, crowding with the N-terminal cyclopentyl cap could be minimized because of the flexibility of this end of the inhibitor. For SCH 6, decreased binding in the nonprime regions as in SCH 503034 is partially compensated by binding of the C-terminal extension in the P1' and P2' sites.

D168 forms salt bridges with the side chains of R123 and R155. As part of the S4 pocket, its methylene group makes van der Waals contacts with the *tert*-butyl group of SCH 503034, cyclopentyl group of BILN-2061, and the P4 cyclohexyl-glycine of VX-950. Based on the structural models, the mutation D168V is not expected to cause steric conflicts. However, as observed in an earlier report,<sup>12</sup> this D168V substitution results in the loss of salt bridge interactions with the R155 side chain which in turn makes extensive contacts with the large P2 group of BILN-2061 (Figure 4). The conformation of R155 in the crystal structure of the BILN-2061-protease complex is no longer energetically favored in the D168V mutant due to the lack of stabilizing interactions. Furthermore, the P2 quinoline group of BILN-2061 would conflict with an alternative conformation of R155 that features a R155-D81 salt bridge as observed in the structures of other inhibitors and the apoprotease.<sup>28</sup> In contrast, SCH 6, SCH 503034, and VX-950 are compatible with this conformation of R155 in the D168V mutation.

Compared to D168V, the D168Q mutation affects the binding of BILN-2061 to a lesser extent, decreasing the binding affinity by only 70 versus 260-fold. This could be

explained by the fact that the side-chain carbonyl oxygen of Q168 can still hydrogen bond with the guanidinium group of R155, therefore partially compensating for the loss of the D168-R155 salt bridge interactions to maintain a population of the R155 similar to that in the complex of BILN-2061 and the wild-type enzyme. However, either the difference in this R155 conformation or its lower population causes a 70-fold loss of potency in the D168Q mutant. Apparently, the greater flexibility of the acyclic analogue compound 1 allows it to compensate for these differences. Similar to the D168V mutation, SCH 503034, SCH 6, and VX-950 would not be affected by the mutation D168Q. In another study of genotype specificity, it was observed that the D168Q mutation has no impact on the binding of a set of peptide inhibitors.<sup>29</sup>

In terms of simulations, it is noteworthy that in an earlier work evaluating the effect of P1 substitutions of a product inhibitor on the binding affinity to the HCV protease,<sup>30</sup> no restraint was imposed on the inhibitor, unlike the current work in which a weak constraint was necessary to restrain the inhibitor near its binding conformation. Generally speaking, adding restraints to the inhibitor limits the exploration of alternative binding modes, if any. For the current inhibitors, alternative conformations are unlikely since the backbone conformation is defined by the consensus network of hydrogen bonds. Fixing the backbone leaves little conformational freedom for the side chains except for the P2 groups of BILN-2061 and compound 1.

## Conclusions

Employing an all-atom molecular representation of the system with explicit solvent, the FEP method was used to quantify the effect of mutations A156T, D168V, and D168Q in HCV protease on the binding of SCH 6, SCH 503034, VX-950, BILN-2061, and compound 1. All the inhibitors display high potency against wild-type enzyme, but the mutant enzymes are resistant to one or more of the inhibitors. The computed relative resistance energies for the inhibitors agree well with experimental values. Structural analysis identified atomic interactions that allowed each mutant enzyme to be resistant to one or more inhibitors. These results coupled with further FEP calculations and analysis could guide design of inhibitors less susceptible to escape mutants.

**Acknowledgment.** The authors thank J. Wright-Minogue and E. Xia for generating and assaying the D168Q mutant protease. Z. Guo thanks Drs. J. Duca, J. Voigt, H. Wang, and L. Xiao for helpful discussions and Dr. E. Zaborowski for providing software support. Special thanks to Dr. B. A. Malcolm for careful reading of the manuscript and providing valuable input.

**Supporting Information Available:** Partial atomic charges of SCH 503034, SCH 6, VX-950, BILN-2061, and compound 1. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Choo, Q. L.; Kuo, G.; Weiner, A. J.; Overby, L. R.; Bradley, D. W.; Houghton, M. Isolation of a cDNA clone derived

- from a blood-borne non-A, non-B viral hepatitis genome. *Science* **1989**, *244*, 359–362.
- (2) Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimotohno, K. Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9524–9528.
- (3) Choo, Q. L.; Richman, K. H.; Han, J. H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coit, D.; Medina-Selby, R.; Barr, P. J.; Weiner, A. J.; Bradley, D. W.; Kuo, G.; Houghton, M. Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2451–2455.
- (4) Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; Andoh, T.; Yoshida, I.; Okayama, H. Structure and organization of the hepatitis C virus genome isolated from human carriers. *J. Virol.* **1991**, *65*, 1105–1113.
- (5) Houghton, M. Hepatitis C viruses. In *Fields Virology*; 3rd ed.; Fields, B. N., Knipe, D. M., Howley, P. M., Eds.; Lippincott-Raven Publishers: Philadelphia, PA, 1996; pp 1035–1058.
- (6) Bartenschlager, R. Candidate targets for hepatitis C virus-specific antiviral therapy. *Intervirology* **1997**, *40*, 378–393.
- (7) Kolykhalov, A. A.; Mihalik, K.; Feinstone, S. M.; Rice, C. M. Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo. *J. Virol.* **2000**, *74*, 2046–2051.
- (8) Bartenschlager, R. The NS3/4A proteinase of the hepatitis C virus: unravelling structure and function of an unusual enzyme and a prime target for antiviral therapy. *J. Viral Hepat.* **1999**, *6*, 165–181.
- (9) Lamarre, D.; Anderson, P. C.; Bailey, M.; Beaulieu, P.; Bolger, G.; Bonneau, P.; Bos, M.; Cameron, D. R.; Cartier, M.; Cordingley, M. G.; Faucher, A. M.; Goudreau, N.; Kawai, S. H.; Kukolj, G.; Lagace, L.; LaPlante, S. R.; Narjes, H.; Poupard, M. A.; Rancourt, J.; Sentjens, R. E.; St George, R.; Simoneau, B.; Steinmann, G.; Thibeault, D.; Tsantrizos, Y. S.; Weldon, S. M.; Yong, C. L.; Llinas-Brunet, M. An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **2003**, *426*, 186–189.
- (10) Perni, R. B.; Chandorkar, G.; Chaturvedi, P. R.; Courtney, L. F.; Decker, C. J.; Gates, C. A.; Harbeson, S. L.; Kwong, A. D.; Lin, C.; Lin, K.; Luong, Y. P.; Markland, W.; Rao, B. G.; Tung, R. D.; Thompson, J. A. *Hepatology* **2003**, *38*, Abstr. 972.
- (11) Lin, C.; Lin, K.; Luong, Y.; Rao, B. G.; Wei, Y.; Brennan, D. L.; Fulghum, J. R.; Hsiao, H.; Ma, S.; Maxwell, J. P.; Cottrell, K. M.; Perni, R. B.; Gates, C. A.; Kwong, A. D. In vitro resistance studies of hepatitis C virus serine protease inhibitors, VX-950 and BILN 2061. *J. Biol. Chem.* **2004**, *279*, 17508–17514.
- (12) Malcolm, B. A.; Liu, R.; Lahser, F.; Agrawal, S.; Belanger, B.; Butkiewicz, N.; Chase, R.; Gheyas, F.; Hart, A.; Hesk, D.; Ingravallo, P.; Jiang, C.; Kong, R.; Lu, J.; Pichardo, J.; Prongay, A.; Skelton, A.; Tong, X.; Venkatraman, S.; Xia, E.; Girijavallabhan, V.; Njoroge, F. G. SCH 503034, a Mechanism-based Inhibitor of Hepatitis C Virus NS3 Protease Suppresses Polyprotein Maturation and Enhances the Antiviral Activity of Interferon  $\alpha$ . *Antimicrob. Agents Chemother.* **2006**, *50*, 1013–1020.
- (13) Tong, X.; Guo, Z.; Wright-Minogue, J.; Xia, E.; Madison, V.; Qiu, P.; Venkatraman, S.; Velazquez, F.; Njoroge, F. G.; Malcolm, B. A. Impact of Naturally Occurring Variants of HCV Protease on the Binding of Different Classes of Protease Inhibitors. *Biochemistry* **2006**, *45*, 1353–1361.
- (14) Lin, K.; Kwong, A. D.; Lin, C. Combination of a hepatitis C virus NS3–NS4A protease inhibitor and alpha interferon synergistically inhibits viral RNA replication and facilitates viral RNA clearance in replicon cells. *Antimicrob. Agents Chemother.* **2004**, *48*, 4784–4792.
- (15) Trozzi, C.; Bartholomew, L.; Ceccacci, A.; Biasiol, G.; Pacini, L.; Altamura, S.; Narjes, F.; Muraglia, E.; Paonessa, G.; Koch, U.; De Francesco, R.; Steinkuhler, C.; Migliaccio, G. In vitro selection and characterization of hepatitis C virus serine protease variants resistant to an active site peptide inhibitor. *J. Virol.* **2003**, *77*, 3669–3679.
- (16) Lin, C.; Gates, C. A.; Rao, B. G.; Brennan, D. L.; Fulghum, J. F.; Luong, Y.; Frantz, J. D.; Lin, K.; Ma, S.; Wei, Y.; Perni, R. B.; Kwong, A. D. In vitro studies of cross-resistance mutations against two hepatitis C virus serine protease inhibitors, VX-950 and BILN 2061. *J. Biol. Chem.* **2005**, *280*, 36784–36791.
- (17) Yi, M.; Tong, X.; Skelton, A.; Chase, R.; Chen, T.; Pyles, R.; Bourne, N.; Malcolm, B. A.; Lemon, S. M. Mutations conferring resistance to SCH 6, a novel hepatitis C virus NS3/4A protease inhibitor: Reduced RNA replication fitness and partial rescue by second-site mutations. *J. Biol. Chem.* **2006**, *281*, 8205–8215.
- (18) Tong, X.; Chase, R.; Skelton, A.; Chen, T.; Wright-Minogue, J.; Malcolm, B. A. SCH 503034 Resistance Mutations Reduce Fitness of HCV Replicon. *Antiviral Res.* **2006**, in press.
- (19) Beveridge, D. L.; DiCapua, F. M. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.
- (20) Kollman, P. Free energy calculations: application to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (21) Udier-Blagovicia, M.; Tirado-Rives, J.; Jorgensen, W. L. Structural and Energetic Analyses of the Effects of the K103N Mutation of HIV-1 Reverse Transcriptase on Efavirenz Analogues. *J. Med. Chem.* **2004**, *47*, 2389–2392.
- (22) Prongay, A. J.; Guo, Z.; Fischmann, T.; Strickland, C.; Myers, J. J.; Yao, N.; Weber, P. C.; Malcolm, B. A.; Beyer, B. M.; Ingram, R.; Pichardo, J.; Hong, Z.; Prorise, W. W.; Ramanathan, L.; Taremi, S. S.; Yarosh-Tomaine, T.; Zhang, R.; Arasappan, A.; Bennett, F.; Bogen, S. L.; Chen, K.; Jao, E.; Liu, Y.; Lovey, R. G.; Saksena, A. K.; Venkatraman, S.; Girijavallabhan, V.; Njoroge, F. G.; Madison, V. Discovery of the HCV NS3/4A Protease Inhibitor SCH503034 II. Key Steps in Structure-Based Optimization. *J. Med. Chem.* **2006**, unpublished results.
- (23) Momany, F.; Rone, R. Validation of the general purpose QUANTA 3.2/CHARMM force field. *J. Comput. Chem.* **1992**, *13*, 888–900.
- (24) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

- (25) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (26) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (27) Brooks, C. L., III.; Karplus, M. Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.* **1983**, *79*, 6312–6325.
- (28) Di Marco, S.; Rizzi, M.; Volpari, C.; Walsh, M. A.; Narjes, F.; Colarusso, S.; De Francesco, R.; Matassa, V. G.; Sollazzo, M. Inhibition of the hepatitis C virus NS3/4A protease. *J. Biol. Chem.* **2000**, *275*, 7152–7157.
- (29) Beyer, B. M.; Zhang, R.; Hong, Z.; Madison, V.; Malcolm, B. A. Effect of naturally occurring active site mutations on hepatitis C virus NS3 protease specificity. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 82–88.
- (30) Guo, Z.; Durkin, J.; Fischmann, T.; Ingram, R.; Prongay, A.; Zhang, R.; Madison, V. Application of the  $\lambda$ -dynamics method to evaluate the relative binding free energies of inhibitors to HCV protease. *J. Med. Chem.* **2003**, *46*, 5360–5364.
- (31) Zhang, R.; Beyer, B. M.; Durkin, J.; Ingram, R.; Njoroge, F. G.; Windsor, W. T.; Malcolm, B. A. A Continuous Spectrophotometric Assay for the Hepatitis C Virus Serine Protease. *Anal. Biochem.* **1999**, *270*, 268–275.

CT600151Y

## Stationary Points on the PES of *N*-Methoxy Peptides and Their Boron Isosteres: An Ab Initio Study

Alpeshkumar K. Malde, Santosh A. Khedkar, and Evans C. Coutinho\*

Department of Pharmaceutical Chemistry, Bombay College of Pharmacy, Kalina, Santacruz (E), Mumbai 400 098, India

Received June 6, 2006

**Abstract:** The conformational space of *N*-methoxy-*N*-methylacetamide [ $\text{CH}_3\text{—CO—N}(\text{OCH}_3)\text{—CH}_3$ , NMA-NOM] and its boron isostere [ $\text{CH}_3\text{—CO—B}(\text{OCH}_3)\text{CH}_3$ , BMA-BOM] has been studied at the HF, B3LYP, and MP2 levels of theory with the 6-31+G\* basis set. The minima, saddle points, and rotation barriers on the PES of these molecules have been located, and the energy barriers estimated. The *omega* rotation barrier is relatively lower in the boron isostere than in NMA-NOM. The difference in the rotation barrier has been attributed to second-order orbital interactions, like negative hyperconjugation, as revealed by NBO calculations. As an extension, *N*-acetyl-*N*-methoxy-*N*-methylamide of alanine (Ala-NOM) and its boron isostere (B-Ala-BOM) have been adopted as model peptides to study the conformational preferences about the  $\phi$  and  $\psi$  torsion angles. The study reveals a strong preference for conformations of type-V beta turn and left-handed  $\alpha$ -helix for Ala-NOM. B-Ala-BOM, on the other hand, favors conformations of type-Va beta turn, mirror image of Poly-L-Pro II helix, and structures with positive  $\phi$  and extended  $\psi$ . The replacement of nitrogen by boron changes the electronic and conformational properties of the peptide, extends greater flexibility around the *omega* angle, induces a strong preference for positive *phi* values, and shifts the site of nucleophilic attack from the carbonyl group to boron.

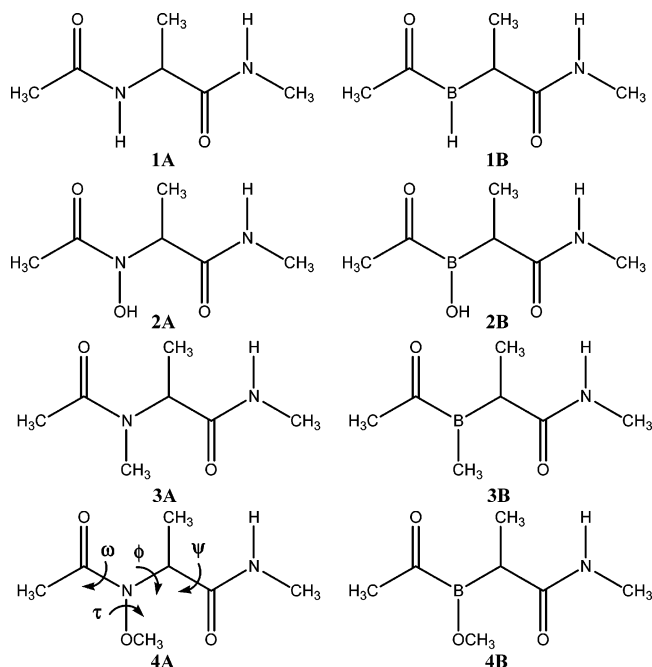
### Introduction

The biological effects of peptides and proteins are governed by their conformations. The values of the three backbone torsion angles—*omega* ( $\omega$ ), *phi* ( $\phi$ ), and *psi* ( $\psi$ )—dictate the secondary structure and hence the conformations of peptides.<sup>1</sup> Most natural peptides adopt  $\omega$  with 180° (trans), and occasionally,  $\omega$  assumes 0° (cis) for peptides with the Xxx-Pro and Xxx-Gly motifs.<sup>2</sup> The  $\phi$  and  $\psi$  values in natural peptides and proteins are restricted to the allowed regions of the Ramachandran map.<sup>1</sup> Peptides form an important area of therapeutics,<sup>3</sup> e.g. insulin, gastric inhibitory polypeptide, substance P, thyrotropin releasing hormone, gastrin, growth hormone, bradykinin, etc. have important therapeutic applications. The advantages of using peptides as therapeutic agents are their high potency, selectivity, and specificity. The major obstacle is oral delivery due to degradation at the “scissile” amide bond. This is the reason peptides like insulin

and thyrotropin cannot be given orally and have to be administered parenterally.

The major challenges in peptide therapeutics are conferring potency, specificity, and selectivity to peptides designed from natural analogues for certain biological end points. Modifications of the amide bond and isosteric/bioisosteric replacements have been explored to address issues like selectivity toward a particular protein target, improving binding affinity, and stability toward proteolytic enzymes. N-Methylation;<sup>4</sup> N-hydroxylation;<sup>5</sup> replacement of the amide bond by sulfonamide, phosphoramidate, and carbamate;<sup>6,7</sup> reduction of the amide bond; inversion of stereochemistry at the alpha carbon; isosteric replacement of the carbonyl carbon with boron (peptide boronic acid<sup>8,9</sup>); and isosteric replacement of the alpha carbon with boron (ammonia-carboxyboranes<sup>10–12</sup>) have been reported as techniques to explore new peptide conformations and to design proteolytically stable, “druglike” molecules.

\* Corresponding author phone: +91-22-26670871; fax: +91-22-26670816; e-mail: evans@bcplindia.org.

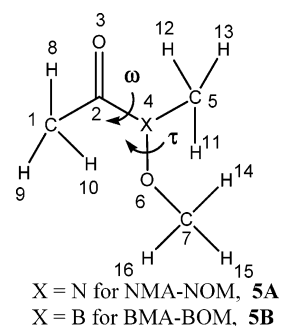


**Figure 1.** Structures of natural and the corresponding boron peptides.

We had reported for the first time a boron isostere of the amide nitrogen in peptides (Figure 1, **1A** and **1B**) and had studied the  $\omega$ ,  $\phi$ , and  $\psi$  preferences by ab initio and density functional methods.<sup>13,14</sup> These molecules were designed as plausible serine protease inhibitors. The replacement of nitrogen with boron leads to two new characteristics: a preference of the  $\omega$  angle for  $90^\circ$ , in contrast to  $180^\circ$  or  $0^\circ$  for natural peptides; and second, conformations that lie in the “disallowed regions” (positive  $\phi$  angles) of the Ramachandran plot. These peptides also exhibit greater flexibility around the  $\omega$  angle. However, these boron analogues are synthetically challenging. In a subsequent paper,<sup>15</sup> we had studied hydroxy derivatives of boron (Figure 1, **2B**) which are counterparts of the *N*-hydroxy derivatives of natural peptides (Figure 1, **2A**). The acylhydroxyboranes (Figure 1, **2B**) are more accessible synthetically than the acylboranes<sup>13</sup> (Figure 1, **1B**) designed earlier.

Another group of peptides are the *N*-methyl analogues (Figure 1, **3A**) which are seen naturally in sarcosine, actinomycin D, cyclosporin, etc.<sup>16,17</sup> and have also been exploited synthetically with the aim of imparting specific secondary structures ( $\alpha$ -helix) to peptides.<sup>18</sup> The boron analogues of the *N*-methyl peptides are represented by **3B** in Figure 1. However, from a synthetic standpoint, the methoxy derivatives (Figure 1, **4B**) are much more accessible and stable over **3B**. *N*-methoxy peptides (Figure 1, **4A**) have received little attention from theoretical and experimental chemists and could offer new approaches of modulating the peptide backbone in novel folds.

In this paper, we look at the  $\omega$ ,  $\phi$ , and  $\psi$  preferences of *N*-methoxy peptides (Figure 1, **4A**) and their boron isosteres (Figure 1, **4B**), by ab initio and density functional methods. To understand the preferences for the ‘ $\omega$  angle’ in such peptides, *N*-methyl-*N*-methoxyacetamide (NMA-NOM, Figure 2 **5A**) and acetylmethylmethoxyborane (BMA-BOM, Figure 2 **5B**) have been established as models for **4A** and



**Figure 2.** Structures and atomic numbering of NMA-NOM and BMA-BOM.

**4B**, respectively. *N*-Acetyl-*N'*-methoxy-*N'*-methylamide of alanine (Ala-NOM, Figure 1 **4A**) and its boron isostere (B-Ala-BOM, Figure 1 **4B**) have been adopted as models to study the  $\phi$  and  $\psi$  distribution of such peptides. The hypersurfaces of NMA-NOM (Figure 2, **5A**) and BMA-BOM (Figure 2, **5B**), with their associated ground and transition states, and the ground states of Ala-NOM (Figure 1, **4A**) and B-Ala-BOM (Figure 1, **4B**) have been mapped by ab initio Hartree–Fock (HF), density functional, and post-HF methods. Second-order orbital interactions by natural bond orbitals (NBO) method was also carried out to understand the fundamental differences in the structures of the *N*-methoxy peptides and their boron isosteres.

## Computational Details

Ab initio molecular orbital<sup>19</sup> and density functional theory<sup>20</sup> calculations have been carried out using the Gaussian03W<sup>21</sup> (revision C.01) package running on a Pentium III processor with 512 MB RAM. The stability of all wave functions was checked at the HF,<sup>22</sup> Becke’s three parameter exchange functional and the gradient corrected functional of Lee, Yang, and Paar (B3LYP)<sup>23–25</sup> and second-order Møller–Plesset MP2 (full)<sup>26,27</sup> level of theory using the 6-31+G\* basis set.

The atom labels for NMA-NOM (Figure 2, **5A**) and BMA-BOM (Figure 2, **5B**) are listed in Figure 2, and the two torsion angles,  $\omega$  and  $\tau$ , are defined as shown in Figure 1. In NMA-NOM, the methoxy moiety can adopt two conformations around the N–O bond. In the first, the two lone pairs of electrons of O are both *syn-clinal* and in the second, both *anti-clinal* with respect to the lone pair of electrons on N. This has been observed from a conformational search of methoxyamine by ab initio calculations. These initial two conformations around the N–O bond in NMA-NOM were chosen, and for each such arrangement of  $\tau$ , a scan in increments of  $30^\circ$  of the  $\omega$  torsion angle was carried out at the HF/6-31+G\* level of theory. Conformations with an  $\omega$  value of  $30^\circ$  and  $210^\circ$  were found to be the lowest in energy. Now, for each conformation with  $\omega$  value of  $30^\circ$  and  $210^\circ$ , respectively, a  $\tau$  scan in increments of  $30^\circ$  was run at the HF/6-31+G\* level of theory. The minima and saddle points for rotations about the  $\omega$  and  $\tau$  torsions were thus identified. All these conformations were further optimized at the B3LYP and MP2 levels of theory with the same basis set, and the conformations were confirmed by frequency calculations, which returned one imaginary frequency for each transition state and all positive frequencies for each ground state. The

**Table 1.** Energies (au) and Relative Energies (kcal/mol) of Various Minima and Transition States on PES of NMA-NOM (5A)<sup>d</sup>

		NIMAG	PG	HF/6-31+G*		B3LYP/6-31+G*		MP2/aug-cc-pVDZ//MP2/6-31+G*			
				au <sup>a</sup>	rel <sup>b</sup>	au <sup>a</sup>	rel <sup>b</sup>	au	rel <sup>b</sup>	$\omega^c$	$\tau^c$
minima	GM	0	C <sub>1</sub>	-360.692065	0.0	-362.8731751	0.0	-362.0266299	0.0	202	121
	LM	0	C <sub>1</sub>	-360.686262	3.6	-362.8679049	3.3	-362.0222521	2.7	34	92
$\omega$ rotation transition state (TS)	$\omega$ TS1	1	C <sub>1</sub>	-360.670192	13.7	-362.8497351	14.7	-362.0053375	13.4	124/-124	127/-127
	$\omega$ TS2	1	C <sub>1</sub>	-360.662019	18.8	-362.8434716	18.6	-361.9990177	17.3	117/-117	-74/74
	$\omega$ TS3	1	C <sub>1</sub>	-360.669763	14.0	-362.8499436	14.6	-362.0062633	12.8	41/-41	-104/104
	$\omega$ TS4	1	C <sub>1</sub>	-360.649892	26.5	-362.8330564	25.2	-361.9889636	23.6	49/-49	88/-88
$\tau$ rotation TS	$\tau$ TS1	1	C <sub>1</sub>	-360.674862	10.8	-362.8597467	8.4	-362.010931	9.9	-165	-151
	$\tau$ TS2	1	C <sub>1</sub>	-360.661238	19.3	-362.8492256	15.0	-362.0017625	15.6	179	21
	$\tau$ TS3	1	C <sub>1</sub>	-360.663241	18.1	-362.8492949	15.0	-362.0028588	14.9	32	-148
	$\tau$ TS4	1	C <sub>1</sub>	-360.669256	14.3	-362.8552055	11.3	-362.0073941	12.1	19	6

<sup>a</sup> Zero-point vibrational energy corrected values. <sup>b</sup> Relative energy in kcal/mol. <sup>c</sup> Torsion angle in degrees. <sup>d</sup> NIMAG = number of imaginary frequency, PG = point group, GM = global minimum, LM = local minimum.

MP2/aug-cc-pVDZ calculations were carried out on all minima and transition states optimized at the MP2/6-31+G\* level of theory.

A similar strategy was adopted for probing the conformational space of BMA-BOM. The methoxyborane moiety has a planar conformation, and the resulting  $\tau$  angles in BMA-BOM are either 0° or 180°. The two BMA-BOM conformations with  $\tau$  values of 0° and 180° were each examined by an  $\omega$  scan in increments of 30° at the HF/6-31+G\* level of theory. With a  $\tau$  value of 0°, structures with  $\omega$  of 60°/-60° were found to be the minima, while with a  $\tau$  value of 180°, the structure with  $\omega$  of 180° was found to be the minimum. The two conformations with  $\omega$  values of 60° and 180° were then evaluated by a  $\tau$  scan in increments of 30° at the HF/6-31+G\* level of theory. The minima and saddle points for rotations around the  $\omega$  and  $\tau$  angles were thus located. All these structures were further optimized at the B3LYP and MP2 levels of theory using the 6-31+G\* basis set and the minima and saddle points confirmed by frequency calculations. The MP2/aug-cc-pVDZ calculations were carried out on all minima and transition states optimized at the MP2/6-31+G\* level of theory.

The NBO<sup>28-30</sup> analysis was carried out on the minimum energy structures of NMA-NOM (Figure 2, 5A) and BMA-BOM (Figure 2, 5B), optimized at the MP2(full)/6-31+G\* level, to quantitatively estimate the second-order interactions as  $E_{ij} = -2F_{ij} / \Delta E_{ij}$ , where  $E_{ij}$  is the energy of the second-order interaction;  $\Delta E_{ij} = E_i - E_j$  is the energy difference between the interacting molecular orbitals  $i$  and  $j$ ; and  $F_{ij}$  is the Fock matrix element for the interaction between orbitals  $i$  and  $j$ . The “atomic partial charges” of the global minimum of NMA-NOM (5A) and BMA-BOM (5B), optimized at the MP2(full)/6-31+G\* level, were calculated using natural population analysis (NPA) as implemented in NBO and additionally by the ‘ESP fit’ method formulated by Merz, Singh, and Kollman.<sup>31</sup>

For Ala-NOM (Figure 1, 4A), the minima in the  $\phi$  and  $\psi$  space was searched starting with two different conformations for  $\omega$  and  $\tau$  as identified (Table 1) previously for NMA-NOM (5A). This corresponds to structures with  $\omega = 202^\circ$ ;  $\tau = 120^\circ$  and  $\omega = 34^\circ$ ;  $\tau = 92^\circ$ . For each ( $\omega$ ,  $\tau$ ) pair, 144 conformations were generated with 30° increments of the  $\phi$ ,  $\psi$  dihedrals. Each conformation was geometry optimized

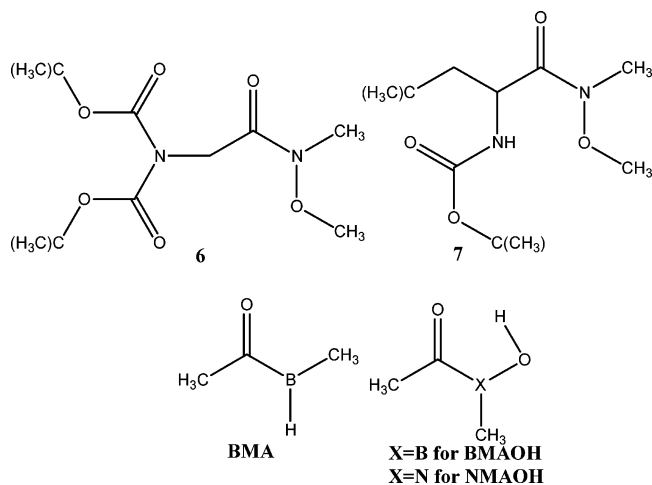
first at the HF/3-21G level of theory with “constraints” on the initial  $\phi$ ,  $\psi$  angles. A Ramachandran map of the 144 conformations was constructed, and conformations within 5.0 kcal/mol of the global minimum were identified. These low-energy conformations were further optimized without constraints at the B3LYP/6-31+G\* level of theory. A similar study was carried out for B-Ala-BOM (4B) with the starting ( $\omega$ ,  $\tau$ ) pairs of (67°, 0°) and (165°, 181°).

## Results and Discussion

All wave functions for molecules 4A, 4B, 5A, and 5B (Figures 1 and 2) were found to be stable under the perturbations considered at the HF, B3LYP, and MP2 levels of theory.

**Minima and Saddle Points of NMA-NOM (5A).** For NMA-NOM (5A) besides the global minimum (GM), there is also a local minimum (LM) at 2.7 kcal/mol of the GM. For each structure, several transition states (TS) for rotation about the  $\omega$  angle exists. The geometries of these TS depend on the state of the pyramidal amide nitrogen, i.e. the lone pair of electrons on nitrogen may either be directed downward which is labeled as ‘pyramidal up’, or the lone pair of electrons on nitrogen may be positioned upward which is labeled as ‘pyramidal down’. This is further complicated by the orientation of the two lone pairs of electrons on the methoxy oxygen relative to the lone pair on the amide nitrogen. In all, four transition states can be identified for ‘ $\omega$  rotation’ taking into consideration all positions of the lone pair of electrons on the amide nitrogen and methoxy oxygen atoms.

Further, proceeding from the GM and LM structures four TS corresponding to rotation about the  $\tau$  angle have been identified. The transition states  $\tau$ TS1 and  $\tau$ TS2 are the rotation barriers for the  $\tau$  angle in the GM, while the transition states  $\tau$ TS3 and  $\tau$ TS4 are for the LM. In summation, a total of eight TS have been identified on the potential energy surface of NMA-NOM (5A). The energies of the minima and TS at the HF, B3LYP, and MP2 levels of theory are listed in Table 1. The geometries of the minima and TS have been pictorially depicted in Figure 4, and the geometrical data (bond lengths, bond angles, and torsion angles) are given in Table 3.

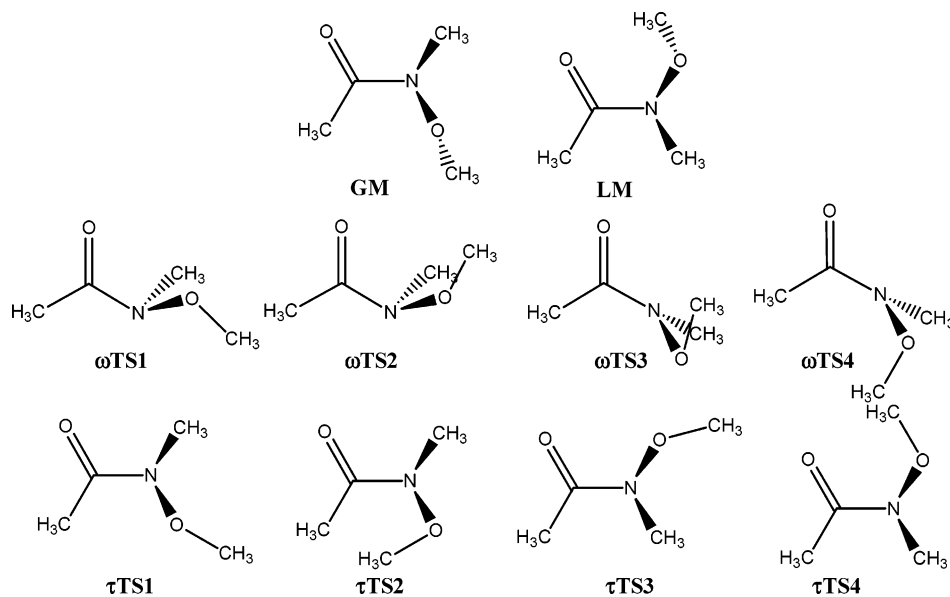


**Figure 3.** Structures of BMA, BMAOH, NMAOH, and *N*-methoxy peptides (**6** and **7**).

All structures exhibit  $C_1$  symmetry. There is a small increase of about 0.06–0.07 Å in the C(O)–N bond length in the transition states for rotation about the  $\omega$  angle compared to the two ground states (GM and LM). Kosturkiewicz et al.<sup>32</sup> have reported the X-ray crystal structure of *N,N*-di-*tert*-butyloxycarbonyl-glycin-*N'*-methoxy-*N'*-methylamide (Figure 3, **6**), and Kolter et al.<sup>33</sup> have reported the X-ray crystal structure of 2(*S*)-*N*-*tert*-butoxycarbonylamino-*N*-methoxy-*N*-methylbutanamide (Figure 3, **7**). In both these structures the NMA-NOM segment has close resemblance to the GM of NMA-NOM. The geometric parameters of these two structures have been compared in Table 4 with the calculated parameters for NMA-NOM (**5A**). Thus, the bond lengths and bond angles of NMA-NOM at the MP2(full)/6-31+G\* level of theory are close to those in the crystal structures mentioned above around the *N*-methoxy amide region. The  $\omega$  angle is found to be 202° and 34° in the global and local minimum structures of NMA-NOM (**5A**), while this torsion in *N,N*-dimethylacetamide<sup>34</sup> (a model for *N*-methyl peptides), *N*-methylacetamide<sup>35</sup> (a model for natural peptides), and in natural peptides and proteins is found to

be ca. 180° and ca. 0° for the corresponding global and local energy minimum structures, respectively. The nitrogen in NMA-NOM is pyramidal, while that in natural peptide is planar. The deviation of amide bond from planarity in NMA-NOM is a consequence of the electronegativity of the hydroxy group attached to nitrogen in addition to the steric effect of the methoxy group. The deviation in the  $\omega$  angle in NMA-NOM (**5A**), from that found in natural peptides, subsequently affects the  $\phi$ ,  $\psi$  distributions in *N*-methoxy peptides. The transition states for  $\omega$  rotation in NMA-NOM (**5A**) has a structural pattern in the amide bond (pyramidal nitrogen with the lone pair of electrons either *syn*- or *anti*-periplanar to the carbonyl group) similar to that seen in the transition states for  $\omega$  rotation of *N*-methylacetamide.<sup>36</sup>

**Minima and Saddle Points of BMA-BOM (5B).** The potential energy surface of BMA-BOM (**5B**) is characterized by two minima—the global minimum (GM) and a local minimum (LM); three transition states for rotation about the  $\omega$  angle ( $\omega$ TS1 to  $\omega$ TS3); and one transition state for rotation about the  $\tau$  angle ( $\tau$ TS). Boron adopts a planar structure unlike pyramidal for N in NMA-NOM (**5A**). The conformations of the ground and TS of BMA-BOM are shown in Figure 5, and the absolute and relative energies at the HF, B3LYP, and MP2(full) levels of theory with the 6-31+G\* basis set are given in Table 2. The GM, LM, and the structure corresponding to the transition state for  $\tau$  rotation ( $\tau$ TS) exhibit a  $C_1$  symmetry, while all the three transition states for  $\omega$  rotation ( $\omega$ TS1 to  $\omega$ TS3) exhibit a  $C_s$  symmetry. In the ground-state structures, the  $\omega$  and  $\tau$  values in the GM are 165° and 181°, while in the LM they are 67° and 0°, respectively. In comparison, the  $\omega$  and  $\tau$  values in the GM of NMA-NOM (**5A**) are 202° and 121° and in the LM they are 34° and 92°, respectively. The minima of NMA-NOM and BMA-BOM exhibit significant conformational differences. In the transition state for  $\omega$  rotation ( $\omega$ TS1) the  $\omega$  and  $\tau$  angles have values of 0° and 180°; in  $\omega$ TS2 they are 0° and 0°, while in  $\omega$ TS3 they are 180° and 0°, respectively. In the case of the transition state corresponding to rotation



**Figure 4.** Ground and transition states of NMA-NOM (**5A**).

**Table 2.** Energies (au) and Relative Energies (kcal/mol) of Various Minima and Transition States on PES of BMA-BOM (5B)<sup>d</sup>

		NIMAG	PG	HF/6-31+G*		B3LYP/6-31+G*		MP2/aug-cc-pVDZ//MP2/6-31+G*			
				au <sup>a</sup>	rel <sup>b</sup>	au <sup>a</sup>	rel <sup>b</sup>	au	rel <sup>b</sup>	$\omega^c$	$\tau^c$
minima	GM	0	C <sub>1</sub>	-331.036343	0.0	-333.0720115	0.0	-332.2595155	0.0	165	181
	LM	0	C <sub>1</sub>	-331.0346427	1.1	-333.0709604	0.7	-332.2583229	0.7	67	0
$\omega$ rotation TS	$\omega$ TS1	1	C <sub>s</sub>	-331.029499	4.3	-333.0660019	3.8	-332.2536856	3.7	0	180
	$\omega$ TS2	1	C <sub>s</sub>	-331.0321876	2.6	-333.0686721	2.1	-332.256573	1.8	0	0
	$\omega$ TS3	1	C <sub>s</sub>	-331.0269482	5.9	-333.0637057	5.2	-332.251972	4.7	180	0
$\tau$ rotation TS	$\tau$ TS	1	C <sub>1</sub>	-331.0202402	10.1	-333.0542493	11.1	-332.237326	13.9	-98/98	85/-85

<sup>a</sup> Zero-point vibrational energy corrected values. <sup>b</sup> Relative energy in kcal/mol. <sup>c</sup> Torsion angle in degrees. <sup>d</sup> NIMAG = number of imaginary frequency, PG = point group, GM = global minimum, LM = local minimum.

**Table 3.** Bond Length (Å) and Bond Angles (deg) of NMA-NOM (5A) Optimized at the MP2(full)/6-31+G\* Level

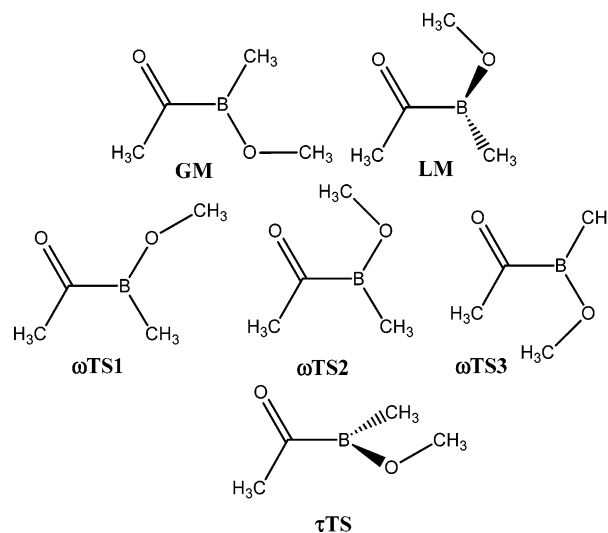
parameter	GM	LM	$\omega$ TS1	$\omega$ TS2	$\omega$ TS3	$\omega$ TS4	$\tau$ TS1	$\tau$ TS2	$\tau$ TS3	$\tau$ TS4
CC (1,2)	1.506	1.513	1.498	1.497	1.506	1.514	1.505	1.512	1.517	1.519
CO (2,3)	1.233	1.230	1.218	1.223	1.220	1.219	1.238	1.240	1.230	1.238
CN (2,4)	1.393	1.394	1.473	1.461	1.464	1.461	1.380	1.389	1.390	1.369
NC (4,5)	1.457	1.457	1.467	1.461	1.460	1.453	1.452	1.457	1.456	1.452
NO (4,6)	1.421	1.416	1.459	1.445	1.464	1.448	1.443	1.446	1.438	1.429
OC (6,7)	1.434	1.435	1.430	1.443	1.435	1.437	1.427	1.425	1.425	1.442
CCO (1,2,3)	123.5	122.7	125.3	124.7	123.5	121.2	123.3	119.7	122.1	121.0
CCN (1,2,4)	116.3	115.3	112.4	112.3	118.2	120.7	116.3	121.3	115.0	114.5
CNC (2,4,5)	118.6	122.4	109.6	112.0	114.4	116.3	119.6	117.1	121.4	125.1
CNO (2,4,6)	113.2	113.2	103.6	108.0	101.1	106.8	112.0	126.2	109.6	124.7
CNO (5,4,6)	111.1	108.3	105.7	112.5	105.4	113.5	120.1	104.2	118.2	105.0
NOC (4,6,7)	108.7	109.1	107.2	117.5	107.4	116.7	114.7	119.0	115.6	119.0
CCNC ( $\omega$ ) (1,2,4,5)	202.5	33.6	124.1	116.7	41.0	48.4	195.0	179.4	31.7	18.6
CCNO (1,2,4,6)	335.3	166.0	236.5	241.0	288.3	280.7	343.5	315.1	175.3	169.0
CNOC ( $\tau$ ) (2,4,6,7)	121.1	92.2	127.0	285.8	255.7	88.3	209.3	20.9	212.0	6.4
OCNC (3,2,4,5)	25.2	209.3	304.0	296.0	218.8	227.0	18.4	8.9	207.7	294.8

**Table 4.** Geometric Parameters of GM of NMA-NOM at the MP2(full)/6-31+G\* Level of Theory and X-ray Crystal Structures of *N,N*-Di-*tert*-butyloxycarbonyl-glycin-*N*-methoxy-*N*-methylamide and 2(*S*)-*N*-*tert*-Butoxycarbonylamino-*N*-methoxy-*N*-methylbutanamide

parameter	GM	<i>N,N</i> -di- <i>tert</i> -butyloxycarbonyl-glycin- <i>N</i> -methoxy- <i>N</i> -methylamide (6)	2( <i>S</i> )- <i>N</i> - <i>tert</i> -butoxycarbonylamino- <i>N</i> -methoxy- <i>N</i> -methylbutanamide (7)
CC (1,2)	1.506	1.505	1.530
CO (2,3)	1.233	1.212	1.223
CN (2,4)	1.393	1.351	1.357
NC (4,5)	1.457	1.450	1.453
NO (4,6)	1.421	1.390	1.404
OC (6,7)	1.434	1.418	1.453
CCO (1,2,3)	123.5	123.2	123.2
CCN (1,2,4)	116.3	115.3	116.4
CNC (2,4,5)	118.6	124.0	123.5
CNO (2,4,6)	113.2	115.7	117.8
CNO (5,4,6)	111.1	115.4	114.3
NOC (4,6,7)	108.7	110.7	109.7
CCNC ( $\omega$ ) (1,2,4,5)	202.5	191.6	200.61
CCNO (1,2,4,6)	335.3	344.7	354.3
CNOC ( $\tau$ ) (2,4,6,7)	121.1	114.1	109.7
OCNC (3,2,4,5)	25.2	12.1	17.6

of the  $\tau$  angle ( $\tau$ TS), the values of both the torsion angles ( $\omega$  and  $\tau$ ) are  $\sim 90^\circ$ .

The geometric parameters of the minima and all transition states of BMA-BOM (5B) at the MP2(full)/6-31+G\* level

**Figure 5.** Ground and transition states of BMA-BOM (5B).

of theory are given in Table 5. The changes in the bond lengths from the ground to the TS are relatively small. Some geometric parameters for alkylboranes, arylboranes, and borane complexes have been reported, but there are no experimental data for acylboranes such as BMA<sup>14</sup> (Figure 3) and BMAOH<sup>15</sup> (Figure 3). We had earlier reported the geometry of BMA (Figure 3) which is the boron isostere of *N*-methylacetamide, at the QCISD/6-31G\* level of theory.<sup>14</sup> The B–O bond length in BMAOH (Figure 3) is 1.356 Å in the GM and 1.369 Å in the LM at the MP2(full)/6-31+G\*



**Table 5.** Bond Length (Å) and Bond Angles (deg) of BMA-BOM (**5B**) Optimized at the MP2(full)/6-31+G\* Level

parameter	GM	LM	$\omega$ TS1	$\omega$ TS2	$\omega$ TS3	$\tau$ TS
CC (1,2)	1.509	1.508	1.518	1.511	1.514	1.502
CO (2,3)	1.244	1.245	1.199	1.244	1.244	1.249
CB (2,4)	1.612	1.605	1.639	1.627	1.632	1.592
BC (4,5)	1.568	1.569	1.588	1.573	1.566	1.567
BO (4,6)	1.362	1.358	1.334	1.353	1.361	1.381
OC (6,7)	1.437	1.441	1.405	1.442	1.433	1.429
CCO (1,2,3)	120.7	120.9	119.9	119.7	118.5	121.9
CCB (1,2,4)	120.3	121.8	119.9	119.3	125.9	127.4
CBC (2,4,5)	122.0	120.7	121.0	120.8	117.5	122.5
CBO (2,4,6)	113.4	120.3	114.1	121.8	125.6	115.5
CBO (5,4,6)	124.5	118.9	124.9	117.3	116.9	121.8
BOC (4,6,7)	121.4	121.9	125.9	126.8	128.2	123.8
CCBC ( $\omega$ ) (1,2,4,5)	165.0	66.7	0.0	0.0	180.0	262.8
CCBO (1,2,4,6)	345.3	244.4	180.0	180.0	0.0	83.7
CBOC ( $\tau$ ) (2,4,6,7)	180.8	0.6	180.0	0.0	0.0	84.7
OCBC (3,2,4,5)	344.3	249.8	180.0	180.0	0.0	85.6

level of theory.<sup>15</sup> The B–O bond length in BMA-BOM (**5B**) is found to be 1.362 Å in the GM and 1.358 Å in the LM structures (Table 5) at the MP2(full)/6-31+G\* level of theory. The B–O bond length, reported in the literature for a range of organic and inorganic boron containing molecules, varies from 1.34 to 1.42 Å (average 1.38 Å) for boron with trigonal planar geometry and 1.39 to 1.52 Å (average 1.48 Å) for boron with tetrahedral geometry.<sup>37,38</sup> The B–O bond length for BMA-BOM calculated in this study is in the range of the experimental values. The NBO calculations on the minimum energy structure of BMA-BOM at the MP2(full)/6-31+G\* level of theory reveals a double bond character for the B–O bond, and the second B–O bond has an occupancy of 1.99 (~2.0) electrons with contribution from one of the lone pairs of electrons of oxygen. In alkylboranes, the B–C (aliphatic carbon) bond length is about 1.590 Å as in for e.g. dimethylborane,<sup>39</sup> 1.596 Å in dimesitylborane,<sup>40</sup> 1.570 Å in ditriptylborane,<sup>41</sup> 1.571 Å in BMA,<sup>14</sup> and 1.564 Å in BMAOH.<sup>15</sup> The B4–C5 bond length in BMA-BOM is 1.57 Å, which comes near to the experimental value for the aliphatic carbon–boron bond.

**Rotation Barrier in NMA-NOM (5A) and BMA-BOM (5B).** The barrier to rotation about the  $\omega$  angle in the natural peptide ranges from 16.0 to 25.0 kcal/mol,<sup>36</sup> while that for the boron isostere is only about 5.0 kcal/mol.<sup>14</sup> The boron analogues are thus relatively more flexible than the natural peptides. In case of the *N*-methoxy peptides and the corresponding boron isosteres, there are two rotation barriers governed by the  $\omega$  and  $\tau$  angles. In the example of NMA-NOM (**5A**), the  $\omega$  rotation barrier is relatively higher (12.8–23.6 kcal/mol) than the  $\tau$  rotation barrier (9.9–15.6 kcal/mol). In BMA-BOM (**5B**), the  $\tau$  rotation barrier is comparatively higher (13.9 kcal/mol) than its  $\omega$  rotation barrier (1.8–4.7 kcal/mol). The relative higher  $\tau$  rotation barrier in boron peptides is a consequence of the B–O double bond character as revealed by NBO calculations. The heights of the rotation barriers seen for the methoxy peptides (NMA-NOM and BMA-BOM) is similar to that for the hydroxyl peptides (NMAOH and BMAOH, Figure 3) reported earlier.<sup>15</sup>

The rotation barrier in amide systems (like peptides, urea, guanidine, etc.) has been attributed to delocalization of the lone pair of electrons on nitrogen onto the C–N bond as explained by the classical resonance model.<sup>42</sup> This explanation imparts a partial double bond character to the C–N bond. But recent experimental and theoretical studies<sup>43–47</sup> reveal that the electron delocalization in the amide system has been attributed to second-order orbital interactions namely,  $n_{\text{O}} \rightarrow \sigma^*_{\text{C-N}}$  (delocalization from lone pairs on carbonyl oxygen into the sigma antibonding orbital of the C–N bond i.e. negative hyperconjugation) and  $n_{\text{N}} \rightarrow \pi^*_{\text{C=O}}$  (delocalization from the lone pair on amide nitrogen to the pi antibonding orbital of the carbonyl group). The energy  $E^{(2)}$  associated with negative hyperconjugation i.e.  $n_{\text{O}} \rightarrow \sigma^*_{\text{C-N}}$  is 34.8 kcal/mol (occupancy of  $n_{\text{O}}$  is 1.906 and  $\sigma^*_{\text{C-N}}$  is 0.070) and that with  $n_{\text{N}} \rightarrow \pi^*_{\text{C=O}}$  is 58.0 kcal/mol (occupancy of  $n_{\text{N}}$  is 1.827 and  $\pi^*_{\text{C=O}}$  is 0.170) for the global minimum of NMA-NOM at the MP2(full)/6-31+G\* level. In case of BMA-BOM, the energy associated with negative hyperconjugation i.e.  $n_{\text{O}} \rightarrow \sigma^*_{\text{C-B}}$  is only 13.4 kcal/mol (occupancy of  $n_{\text{O}}$  is 1.941 and  $\sigma^*_{\text{C-B}}$  is 0.036), indicating that the C–B bond delocalization is insignificant, as a result of which, the rotation barrier in boron peptide is very small. Thus, in BMA-BOM, the C–B bond has an essentially single bond character, while the C–N bond in NMA-NOM has a larger double bond character. The boron analogues are thus far more flexible than the *N*-methoxy peptides. The second-order orbital interactions closely parallel that for the corresponding *N*-hydroxy peptides reported earlier.<sup>15</sup> The energy associated with negative hyperconjugation i.e.  $n_{\text{O}} \rightarrow \sigma^*_{\text{C-N}}$  is 32.8 kcal/mol and that with  $n_{\text{N}} \rightarrow \pi^*_{\text{C=O}}$  is 98.5 kcal/mol for the global energy minimum structure of *N*-methylacetamide.<sup>13</sup> Thus, the  $n_{\text{N}} \rightarrow \pi^*_{\text{C=O}}$  electron delocalization is reduced significantly when N–H bond in natural peptide (*N*-methylacetamide) is replaced by N–OCH<sub>3</sub> (NMA-NOM, **5A**). This is probably due to the fact that the electronegative oxygen withdraws electrons from the nitrogen atom.

**Partial Atomic Charges of NMA-NOM (5A) and BMA-BOM (5B).** The “natural charges” derived from NPA for the global minimum energy structure of NMA-NOM (**5A**)

**Table 6.** Partial Atomic Charges of NMA-NOM (**5A**) and BMA-BOM (**5B**) Calculated Using NPA and the 'ESP Fit' as per Merz–Singh–Kollman Scheme at the MP2(full)/6-31+G\* Level

atom	atom no.	natural charges		ESP fitted charges	
		NMA-NOM ( <b>5A</b> )	BMA-BOM ( <b>5B</b> )	NMA-NOM ( <b>5A</b> )	BMA-BOM ( <b>5B</b> )
C	1	-0.73111	-0.73103	-0.605573	-0.293486
C	2	0.83787	0.33658	0.904859	0.394028
O	3	-0.71280	-0.63221	-0.650117	-0.553132
X	4	-0.29335	1.08601	-0.260075	0.711483
C	5	-0.41351	-1.07467	-0.318880	-0.587438
O	6	-0.49098	-0.83597	-0.244952	-0.519024
C	7	-0.23095	-0.23724	-0.119883	0.032558
H	8	0.25241	0.24135	0.153101	0.113004
H	9	0.24073	0.23748	0.180700	0.067244
H	10	0.25758	0.24544	0.180506	0.069023
H	11	0.21366	0.24032	0.149242	0.126094
H	12	0.21753	0.24612	0.155416	0.127817
H	13	0.25216	0.27020	0.159344	0.114528
H	14	0.19358	0.19508	0.075715	0.050418
H	15	0.19566	0.19846	0.117315	0.064449
H	16	0.21151	0.21406	0.123283	0.082434

and BMA-BOM (**5B**) are given in Table 6. Replacement of nitrogen by boron decreases the positive charge on the carbonyl carbon and increases the negative charge on the C5 methyl carbon. In BMA-BOM (**5B**), the boron atom has a much greater positive charge than the carbonyl carbon (1.086 vs 0.336). The site for nucleophilic attack in case of NMA-NOM (**5A**) is normally the carbonyl carbon. In BMA-BOM (**5B**), a nucleophile will preferentially attack boron rather than the carbonyl group. This preference for boron as the site for nucleophilic attack is also evident in the 'ESP fitted charges', even though the partial charges differences are of a smaller magnitude. This was the basis of our hypothesis, used to design boron peptides<sup>13,15</sup> as  $k_{\text{cat}}$  inhibitors of the enzyme serine protease. The hydroxyl group of serine in the active site is the nucleophile which attacks the carbonyl carbon of the amide of the substrate peptide, leading to a final hydrolysis of the substrate. With the boron peptide in the active site, the hydroxyl group of serine would preferentially attack boron instead of the carbonyl carbon and form a tetrahedral covalent complex leading to irreversible inhibition of the enzyme, thus acting as plausible inhibitor of serine protease enzyme. The inhibitors of serine proteases have great potential in therapeutics.<sup>48–54</sup> When the electrophilicity of the carbonyl carbon in NMA-NOM vs that in *N*-methylacetamide is compared,<sup>13</sup> the differences are insignificant based on the NPA charges on carbonyl carbon (0.837 vs 0.862). However, in case of the boron analogues, the electrophilic character of the boron atom is 1.5-fold higher in BMA-BOM than BMA (Figure 3) (the NPA charges of 1.086 vs 0.669) accentuating the attack of the nucleophile.

**PES of Ala-NOM (4A).** The preferred values of the  $\omega$  and  $\tau$  angles found in NMA-NOM (**5A**) were introduced into Ala-NOM (**4A**) and the  $(\phi, \psi)$  space of Ala-NOM scrutinized (Table 7). With an  $\omega$  value of 30° and a  $\tau$  value of 90°, the global minimum corresponds to a structure with  $\phi$

= 55° and  $\psi$  = 35° (Figure 6a). These are close to the values for a left-handed alpha helix ( $\phi$  = 57°,  $\psi$  = 47°). The local minimum energy structure has  $\phi$  = -150° and  $\psi$  = 30° (Figure 6b). These two structures fall in the "allowed regions" of the Ramachandran map. All minimum energy structures are distinguished by the presence of an intramolecular hydrogen bond (Figure 6a,b). The GM structure is characterized by a hydrogen bond between the methoxy O (as acceptor) and the amide NH (as donor), forming a six-membered ring, while the LM structure has a H-bond between the *N*-methoxy N (as acceptor) and the amide NH (as donor), forming a five-membered ring. With an  $\omega$  angle of 200° and a  $\tau$  angle of 120°, there is only one favored structure for Ala-NOM (**4A**) with  $\phi$  = -90° and  $\psi$  = 80°. These values are close to that for a residue at the  $i+1$  position in a type V  $\beta$ -turn ( $\phi$  = -80°,  $\psi$  = 80°). This structure is characterized by only one intramolecular hydrogen bond (Figure 6c) between the carbonyl O (as acceptor), near the methoxy group, and the amide NH (as donor) outlining a seven-membered ring. Thus, all the preferred conformations of Ala-NOM are characterized by the presence of one intramolecular hydrogen bond and are conformationally rigid.

**PES of B-Ala-BOM (4B).** In a similar manner, the  $\phi, \psi$  preferences of B-Ala-BOM (**4B**) were investigated, and the results are shown in Table 7. With an  $\omega$  value of 165° and a  $\tau$  value of 180°, one conformation is observed within 5.0 kcal/mol of the global minimum energy conformer. The global minimum corresponds to a structure with  $\phi$  = 90° and  $\psi$  = -60° (Figure 7a), while the local minimum relates to a structure with  $\phi$  = -100° and  $\psi$  = 70° (Figure 7b). The global minimum shows a strong preference for a positive  $\phi$  value. The GM corresponds to the mirror image of a 2.27 ribbon ( $\phi$  = 78°,  $\psi$  = -59°), while the LM corresponds to a 2.27 ribbon ( $\phi$  = -78°,  $\psi$  = 59°) structure. The two structures exhibit an intramolecular hydrogen bond (Figure 7a,b) between the carbonyl O (as acceptor) next to the B-methoxy group and the amide NH (as donor), forming a seven-membered ring. With an  $\omega$  angle of 60° and a  $\tau$  value of 10°, the global minimum corresponds to a structure with  $\phi$  = -40° and  $\psi$  = 120° (Figure 7c). These are close to the values adopted by an  $i+1$  residue in a type Va  $\beta$ -turn ( $\phi$  = -60°,  $\psi$  = 120° with 'cis'  $\omega$ ). The local minimum corresponds to a structure with  $\phi$  = 85° and  $\psi$  = 180° (Figure 7d). These are near to the values of a poly-L-Pro II helix ( $\phi$  = 79°,  $\psi$  = 210°). The next local minimum energy structure has  $\phi$  = 90° and  $\psi$  = 100° (Figure 7e). The structure is again characterized by a positive  $\phi$  value which is 'disallowed' for natural peptides.

**Overview of Boron Peptides.** In an earlier paper,<sup>13,14</sup> we had studied the geometries and transition states of boron analogues of the natural peptides. The natural peptides have  $\omega$  values either 180° or 0°, while the  $\phi, \psi$  torsion angles fall in the 'allowed' regions of the Ramachandran map. The boron analogue (BMA, Figure 3) exhibits an unusual  $\omega$  value of 90°, and the corresponding Ala-dipeptide exhibits a unique preference for positive  $\phi$  values in the 'disallowed' region of the Ramachandran map but with no regular secondary structure motif. In a subsequent paper,<sup>15</sup> we had investigated the conformational space of the B-OH analogue (BMAOH,

**Table 7.** Conformations and Energies of Ala-NOM (4A) and B-Ala-BOM (4B)

	$\omega$	$\tau$	$\phi$	$\psi$	rel <i>E</i> (kcal/mol)	structural feature (ideal values of the torsion angles)
Ala-NOM	30°	90°	55°	35°	0.0	left-handed $\alpha$ -helix (57°, 47°)
			-150°	30°	0.3	random
B-Ala-BOM	200°	120°	-90°	80°	0.0	type V $\beta$ -turn <i>i</i> +1 residue (-80°, 80°)
			90°	-60°	0.0	positive $\phi$ , mirror image of 2.27 ribbon (78°, -59°)
	165°	180°	-100°	70°	0.3	2.27 ribbon (-78°, 59°)
			-40°	120°	0.0	type Va $\beta$ -turn <i>i</i> +1 residue (-60°, 120°, <i>cis</i> $\omega$ )
			85°	180°	1.4	positive $\phi$ , mirror image of Poly-L-Pro II helix (79°, 210°)
60°	10°	90°	100°	2.2	positive $\phi$ , folded	

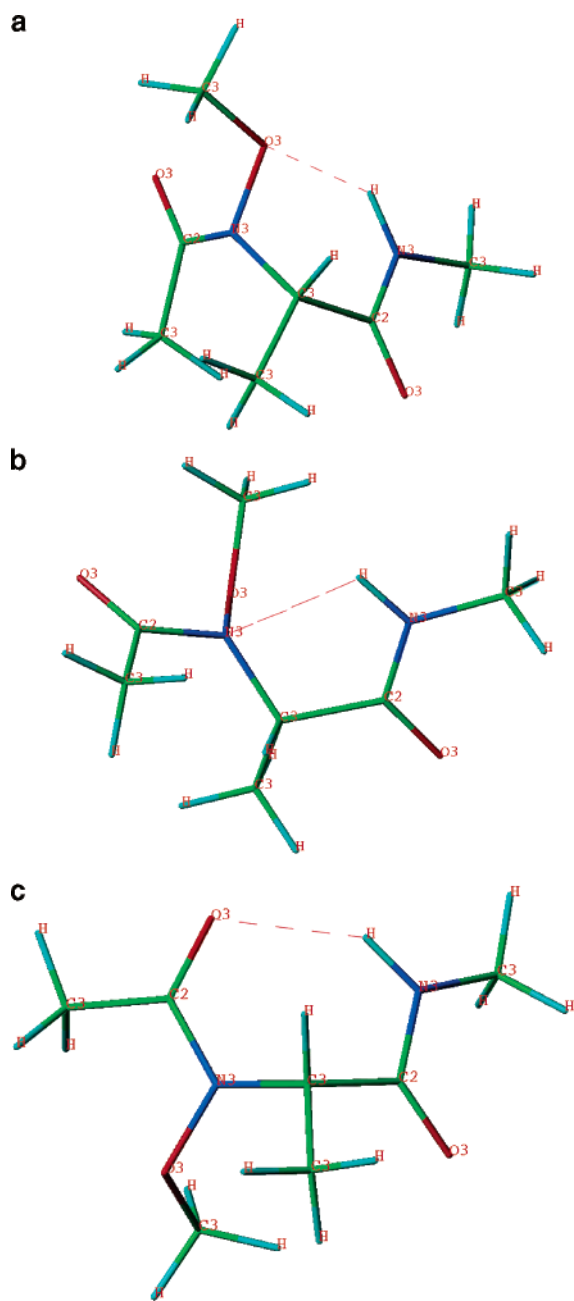
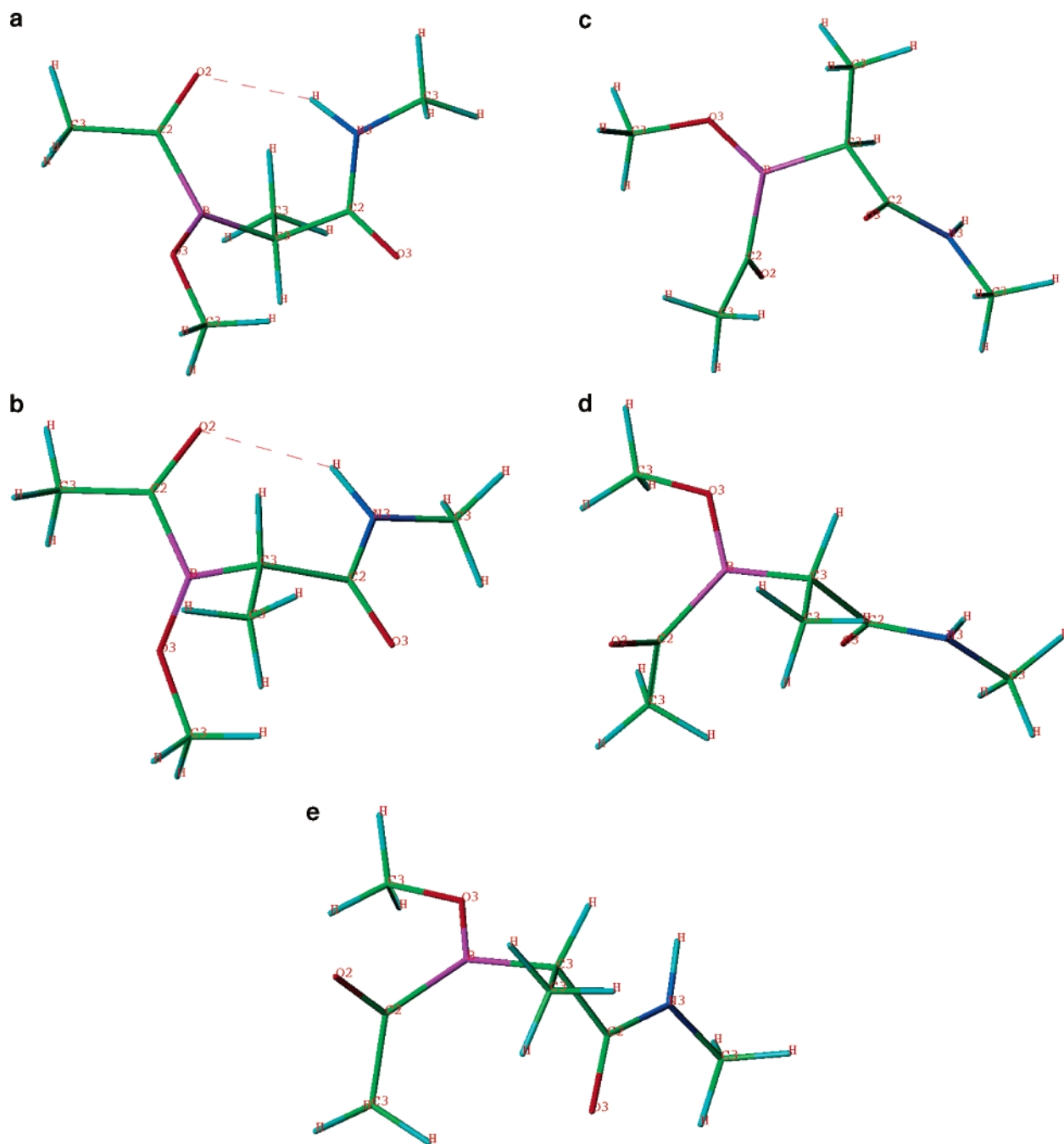
**Figure 6.** Preferred conformations of Ala-NOM (4A): (a)  $\omega = 30^\circ$ ,  $\tau = 90^\circ$ ,  $\phi = 55^\circ$ ,  $\psi = 35^\circ$ , (b)  $\omega = 30^\circ$ ,  $\tau = 90^\circ$ ,  $\phi = -150^\circ$ ,  $\psi = 30^\circ$ , and (c)  $\omega = 200^\circ$ ,  $\tau = 120^\circ$ ,  $\phi = -90^\circ$ ,  $\psi = 80^\circ$ .

Figure 3) and its corresponding peptide with the N–OH moiety. The replacement of the N–H bond (*N*-methylacetamide) by N–OH (NMAOH, Figure 3) shifts the preference

for the  $\omega$  torsion from  $180^\circ$  to  $32^\circ$  due to a strong tendency to form a hydrogen bond within the amide segment. The local energy minimum structure has an  $\omega$  value of  $202^\circ$ . Similarly modifying the boron analogue by replacement of B–H (BMA, Figure 3) with B–OH (BMAOH, Figure 3) leads to a change in the  $\omega$  torsion from  $90^\circ$  to  $0^\circ$  and again this is driven by the predisposition to form an intramolecular hydrogen bond. The local energy minimum structure of BMAOH has an  $\omega$  value of  $153^\circ$ . The intramolecular hydrogen bond imparts conformational rigidity to these peptides. The Ala-dipeptide derived from NMAOH assumes either a type-I beta-turn or a left-handed alpha-helix. The Ala-dipeptide derived from BMAOH shows a leaning toward positive  $\phi$ , and extended  $\psi$  values, with no regular secondary structure motifs. In the present class of molecules where the N–H bond is replaced by N–OCH<sub>3</sub> (NMA-NOM, 5A) and the B–H bond is replaced by B–OCH<sub>3</sub> (BMA-BOM, 5B), the conformational space of the corresponding peptides is different from the previously reported natural peptides and their hydroxy derivatives. In NMA-NOM (5A) the  $\omega$  values are  $202^\circ$  and  $32^\circ$  in the GM and LM structures, which are near to the structures for the natural peptides. A similar shift in the torsion is also seen in BMA-BOM with a value of  $165^\circ$  for the  $\omega$  angle. The dipeptides bearing the N–OCH<sub>3</sub> and B–OCH<sub>3</sub> (Figure 1, 4A and 4B) moieties have in common the type-V  $\beta$ -turn motif beside adopting several of the regular secondary structures. Thus, it has been observed that modifying the B–H bond as B–OH and further as B–OCH<sub>3</sub> is able to induce a change in the peptide backbone from a random coil-like state to a more ordered secondary structure.

## Conclusions

The boron isosteres of the amino acids were designed by replacement of the amide nitrogen with boron,<sup>13</sup> with the intention of developing an irreversible inhibitor of the enzyme serine protease. The synthesis of such a molecule is a big challenge. Based on the strong affinity of boron for oxygen, these analogues were modified to B–OH which are easy to synthesize, and this derivative is echoed in the *N*-hydroxy amides which are well-known. However, from a synthetic standpoint, the methoxy analogues (B–OCH<sub>3</sub>) are much more accessible. The conformational space of *N*-methoxy peptides and their boron isosteres has been the focus of investigation in this paper. The minimum in the  $\omega$  torsion space of such molecules has been identified using *N*-methoxy-*N*-methylacetamide (NMA-NOM, 5A) and acetyl-methylmethoxyborane (BMA-BOM, 5B) as model peptides.



**Figure 7.** Preferred conformations of B-Ala-BOM (**4B**): (a)  $\omega = 165^\circ$ ,  $\tau = 180^\circ$ ,  $\phi = 90^\circ$ ,  $\psi = -60^\circ$ , (b)  $\omega = 165^\circ$ ,  $\tau = 180^\circ$ ,  $\phi = -100^\circ$ ,  $\psi = 70^\circ$ , (c)  $\omega = 60^\circ$ ,  $\tau = 10^\circ$ ,  $\phi = -40^\circ$ ,  $\psi = 120^\circ$ , (d)  $\omega = 60^\circ$ ,  $\tau = 10^\circ$ ,  $\phi = 85^\circ$ ,  $\psi = 180^\circ$ , and (e)  $\omega = 60^\circ$ ,  $\tau = 10^\circ$ ,  $\phi = 90^\circ$ ,  $\psi = 100^\circ$ .

The ground and various transition states for **5A** and **5B** have been calculated at the HF, B3LYP, and MP2(full) levels of theory with the 6-31+G\* basis set. The  $\omega$  rotation barrier is seen to be 12.8–23.6 kcal/mol for the *N*-methoxy peptide (NMA-NOM, **5A**) and 1.8–4.7 kcal/mol for its corresponding boron isostere BMA-BOM (**5B**). The difference in the rotation barriers has been attributed to second-order orbital interactions, mainly negative hyperconjugation. The barrier for rotation about the torsion angle  $\tau$  i.e. rotation about the N–O and B–O bonds is 9.9–15.6 kcal/mol for the *N*-methoxy peptide and 13.9 kcal/mol for the boron isostere. The replacement of nitrogen by boron also significantly changes the charge distribution in these molecules.

A relatively greater positive charge on the boron atom over the carbonyl carbon makes boron the preferential site of attack by a nucleophile in boron peptides, which otherwise occurs on the carbonyl carbon in the natural peptides. The minimum energy structures of NMA-NOM and BMA-BOM were then used to study the  $\phi$  and  $\psi$  preferences in *N*-acetyl-*N'*-methoxy-*N'*-methylamide of alanine (Ala-NOM) and its boron isostere (B-Ala-BOM). Ala-NOM demonstrates conformations with a type-V  $\beta$ -turn and a left-handed  $\alpha$ -helix. B-Ala-BOM, on the other hand, favors conformations with a 2.27 ribbon, the mirror image of 2.27 ribbon, a type-Va  $\beta$ -turn, the mirror image of poly-Pro-II helix, and positive  $\phi$  and extended  $\psi$  values. In previous work on natural peptides

and their boron isosteres, we had noticed a much lower barrier to rotation about the  $\omega$  angle, a unique preference for positive  $\phi$  values in the boron analogues and no regular secondary structure observed. The *N*-methoxy peptides and their boron analogues abound with a variety of secondary structural elements. In conclusion, the replacement of nitrogen by boron in natural and *N*-methoxy peptide causes significant changes in the conformational space and electronic properties, and these features can be profitably exploited to design peptides with specific geometries and chemical attributes.

**Acknowledgment.** This work is supported by the Department of Science and Technology, New Delhi through their FIST program (SR/FST/LS1-163/2003). A.K.M. thanks both the University Grants Commission, New Delhi and the Council of Scientific and Industrial Research, New Delhi, while S.A.K. thanks the Lady Tata Memorial Trust, Mumbai for financial support.

### References

- Ramachandran, G. N.; Sasisekharan, V. Conformation of Polypeptides and Proteins. *Adv. Protein Chem.* **1968**, *28*, 283–437.
- Hagler, T. A.; Leiserowitz, L.; Tuval, M. Experimental and Theoretical Studies of the Barrier to Rotation about N–C $\alpha$  and C $\alpha$ –C' Bonds ( $\phi$  and  $\psi$ ) in Amides and Peptides. *J. Am. Chem. Soc.* **1976**, *98*, 4600–4612.
- Hruby, V. J. Designing Peptide Receptor Agonists and Antagonists. *Nat. Rev. Drug Discovery* **2002**, *1*, 847–858.
- Vogen, S. M.; Paczkowski, N. J.; Kimarsky, L.; Short, A.; Whitmore, J. B.; Sherman, S. A.; Taylor, S. M.; Sanderson, S. D. Differential Activities of Decapeptide Agonists of Human C5a: The Conformational Effects of Backbone N-Methylation. *Int. Immunopharmacol.* **2001**, *12*, 2151–62.
- Ye, Y.; Liu, M.; Kao, J. L.; Marshall, G. R. Peptide-bond Modification for Metal Coordination: Peptides Containing Two Hydroxamate Groups. *Biopolymers* **2003**, *71*, 489–515.
- Fischer, P. M. The design, synthesis and application of stereochemical and directional peptide isomers: a critical review. *Curr. Protein Pept. Sci.* **2003**, *4*, 339–356.
- Baldauf, C.; Günther, R.; Hofmann, H. J. Conformational Properties of Sulphonamido Peptides. *J. Mol. Struct. (THEOCHEM)* **2004**, *675*, 19–28.
- Kettner, C. A.; Shenvi, A. B. Inhibition of Serine Protease Leukocyte Elastase, Pancreatic Elastase, Cathepsin G, and Chymotrypsin by Peptide Boronic Acids. *J. Biol. Chem.* **1984**, *259*, 15106–15114.
- Kettner, C. A.; Bone, R.; Agard, D. A.; Bachovchin, W. W. Kinetic Properties of the Binding of  $\alpha$ -Lytic Protease to Peptide Boronic Acids. *Biochemistry* **1988**, *27*, 7682–7688.
- Spielvogel, B. F.; Wojnowich, L.; Das, M. K.; McPhail, A. T.; Hargrave, K. D. Boron Analogs of Amino Acids. Synthesis and Biological Activity of Boron Analogs of Betaine. *J. Am. Chem. Soc.* **1976**, *98*, 5702–5703.
- Spielvogel, B. F.; Das, M. K.; McPhail, A. T.; Onam, K. D.; Hall, I. H. Boron Analogs of the  $\alpha$ -Amino Acids. Synthesis, X-Ray Crystal Structure, and Biological Activity of Ammonia-Carboxyborane, the Boron Analog of Glycine. *J. Am. Chem. Soc.* **1980**, *102*, 6343–6344.
- Miller, M. C.; Sood, A.; Spielvogel, B. F.; Hall, I. H. Synthesis and Antitumor Activity of Boronated Dipeptides containing Aromatic Amino Acids. *Anticancer Res.* **1997**, *5A*, 3299–3306.
- Datar, P. A.; Coutinho, E. C. The  $\phi$ ,  $\psi$  Space of Boron Isosteres of Amino Acids: An Ab Initio Study. *J. Theor. Comput. Chem.* **2004**, *3*, 189–202.
- Malde, A. K.; Khedkar, S. A.; Coutinho, E. C.; Saran, A. Geometry, Transition States, and Vibrational Spectra of Boron Isostere of *N*-methylacetamide by Ab Initio Calculations. *Int. J. Quantum Chem.* **2005**, *102*, 734–742.
- Malde, A. K.; Khedkar, S. A.; Coutinho, E. C. The  $\omega$ ,  $\phi$  and  $\psi$  Space of *N*-hydroxy-*N*-methylacetamide and *N*-acetyl-*N'*-hydroxy-*N'*-methylamide of Alanine and their Boron Isosteres. *J. Chem. Theory Comput.* **2006**, *2*, 312–321.
- Tsai, G.; Lane, H. Y.; Yang, P.; Chong, M. Y.; Lange, N. Glycine transporter I inhibitor, *N*-methylglycine (sarcosine), added to antipsychotics for the treatment of schizophrenia. *Biol. Psychiatry* **2004**, *55*, 452–456.
- Pritchard, D. L. Sourcing a chemical succession for cyclosporin from parasites and human pathogens. *Drug Discovery Today* **2005**, *10*, 688–691.
- Lee, C. H.; Zimmerman, S. S. Calculations of the  $\phi$ - $\psi$  conformational contour maps for *N*-acetyl alanine *N'*-methyl amide and of the characteristic ratios of poly-L-alanine using various molecular mechanics forcefields. *J. Biomol. Struct. Dyn.* **1995**, *13*, 201–218.
- Hehre, W. J.; Random, L.; Schleyer, P. V. R.; Pople, J. A. In *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1985.
- Parr, R. G.; Yang, W. In *Density Functional Theory of Atoms and Molecules*; O.U.P.: New York, 1989.
- 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.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.01*; Gaussian, Inc.: Wallingford, CT, 2004.
- Roothan, C. C. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69–89.
- Becke, A. D. Density functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev.* **1988**, *37B*, 785–789.

- (25) Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev.* **1992**, *45B*, 13244–13249.
- (26) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
- (27) Martin-Head, G.; Pople, J. A.; Frisch, M. J. MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988**, *153*, 503–506.
- (28) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO Version 3.1*.
- (29) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *J. Chem. Phys.* **1985**, *83*, 735–746.
- (30) Reed, A. E.; Weinhold, F.; Curtiss, L. A. Intermolecular interactions from a natural bond orbital, donor–acceptor viewpoint. *Chem. Rev.* **1988**, *88*, 899–926.
- (31) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.
- (32) Lubkowski, J.; Lankiewicz, L.; Kosturkiewicz, Z. Structure of *N*, *N*-Di-*tert*-butyloxycarbonyl-glycine *N'*-methoxy-*N'*-methylamide. *Acta Crystallogr.* **1992**, *C48*, 1616–1619.
- (33) Sawatzki, P.; Mikeska, T.; Nieger, M.; Hupfer, H.; Kolter, T. 2(*S*)-*N*-*tert*-butoxycarbonylamino-*N*-methoxy-*N*-methylbutanamide. *Acta Crystallogr.* **2002**, *E58*, o1415–o1417.
- (34) Suzuki, H.; Ishiguro, S. *N,N*-Dimethylacetamide complex of aluminium(III) perchlorate. *Acta Crystallogr.* **2006**, *E62*, m576–m578.
- (35) Kang, Y. K. Ab initio MO and density functional studies on *trans* and *cis* conformers of *N*-methylacetamide. *J. Mol. Struct. (THEOCHEM)* **2001**, *546*, 183–193.
- (36) Villani, V.; Alagona, G.; Ghio, C. Ab Initio Studies on *N*-Methylacetamide. *Mol. Eng.* **1999**, *8*, 135–153.
- (37) Filatov, S.; Shepelev, Y.; Bubnova, R.; Sennova, N.; Egorysheva, A. V.; Kargin, Y. F. The study of  $\text{Bi}_3\text{B}_5\text{O}_{12}$ : synthesis, crystal structure and thermal expansion of oxoborate  $\text{Bi}_3\text{B}_5\text{O}_{12}$ . *J. Solid State Chem.* **2004**, *177*, 515–522.
- (38) Shishkov, I. F.; Khristenko, L. V.; Rudakov, F. M.; Vilkov, L. V.; Karlov, S. S.; Zaitseva, G. S.; Samdal, S. The molecular structure of boratrane determined by gas electron diffraction and quantum mechanical calculations. *J. Mol. Struct.* **2002**, *641*, 199–205.
- (39) Vijay, A.; Sathyanarayana, D. N. Effects of Basis Set and Electron Correlation on the Structure and Vibrational Spectra of Diborane. *J. Mol. Struct.* **1995**, *351*, 215–229.
- (40) Entwistle, C. D.; Marder, T. B.; Smith, P. S.; Howard, A. K.; Fox, M. A.; Mason, S. A. Dimesitylborane monomer-dimer equilibrium in solution, and the solid-state structure of the dimer by single-crystal neutron and X-ray diffraction. *J. Organomet. Chem.* **2003**, *680*, 165–172.
- (41) Bartlett, R. A.; Rasikadis, H. V.; Olmstead, M. M.; Power, P. P.; Weese, K. J. Synthesis of the Monomeric HBtrip<sub>2</sub> (trip-2,4,6-iso-Pr<sub>3</sub>C<sub>6</sub>H<sub>2</sub>) and the X-ray Crystal Structures of [HBMes<sub>2</sub>]<sub>2</sub> (Mes = 2,4,6-Me<sub>3</sub>C<sub>6</sub>H<sub>2</sub>) and HBtrip<sub>2</sub>. *Organometallics* **1990**, *9*, 146–150.
- (42) Pauling, L. *In The Nature of Chemical Bond*; Cornell University Press: Ithaca, NY, 1960.
- (43) Bharatam, P. V.; Iqbal, P.; Malde, A.; Tiwari, R. Electron Delocalization in Aminoguanidines: A Computational Study. *J. Phys. Chem.* **2004**, *108*, 10509–10517.
- (44) Bharatam, P. V.; Moudgil, R.; Kaur, D. Electron Delocalization in Isocyanates, Formamides, and Ureas: Importance of Orbital Interactions. *J. Phys. Chem.* **2003**, *107*, 1627–1634.
- (45) Glendening, E. D.; Hrabal, J. A., II. Resonance in Formamide and Its Chalcogen Replacement Analogues: A Natural Population Analysis/Natural Resonance Theory Viewpoint. *J. Am. Chem. Soc.* **1997**, *119*, 12940–12946.
- (46) Lauvergnat, D.; Hiberty, P. C. Role of Conjugation in the Stabilities and Rotational Barriers of Formamide and Thioformamide. An ab initio Valence-Bond Study. *J. Am. Chem. Soc.* **1997**, *119*, 9478–9482.
- (47) Wiberg, K. B.; Rush, D. J. Solvent Effects on the Thioamide Rotational Barrier: An Experimental and Theoretical Study. *J. Am. Chem. Soc.* **2001**, *123*, 2038–2046.
- (48) Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **2002**, *102*, 4501–4524.
- (49) Lebeurrier, N; Vivien, D.; Ali, C. The complexity of tissue-type plasminogen activator: can serine protease inhibitors help in stroke management? *Expert Opin. Ther. Targets* **2004**, *8*, 309–320.
- (50) Syrovets, T.; Simmet, T. Novel aspects and new roles for the serine protease plasmin. *Cell Mol. Life Sci.* **2004**, *61*, 873–885.
- (51) Simonet, G.; Claeys, I.; Franssens, V; De Loof, A.; Broeck, J. V. Genomics, evolution and biological functions of the pacifastin peptide family: a conserved serine protease inhibitor family in arthropods. *Peptides* **2003**, *24*, 1633–1644.
- (52) Roesenblum, J. S.; Kozarich, J. W. Prolyl peptidases: a serine protease subfamily with high potential for drug discovery. *Curr. Opin. Chem. Biol.* **2003**, *7*, 496–504.
- (53) Waxler, B.; Rabito, S. F. Aprotinin: a serine protease inhibitor with therapeutic actions: its interaction with ACE inhibitors. *Curr. Pharm. Des.* **2003**, *9*, 777–787.
- (54) Narjes, F.; Koch, U.; Steinkuhler, C. Recent developments in the discovery of hepatitis C virus serine protease inhibitors—towards a new class of antiviral agents? *Expert Opin. Invest. Drugs* **2003**, *12*, 153–163.

CT600192G

## The Low-Barrier Double-Well Potential of the $O^{\delta 1}-H-O^{\delta 1}$ Hydrogen Bond in Unbound HIV Protease: A QM/MM Characterization

Melissa A. Porter and Pablo A. Molina\*

Department of Chemistry, Murray State University, Murray, Kentucky 42071

Received June 14, 2006

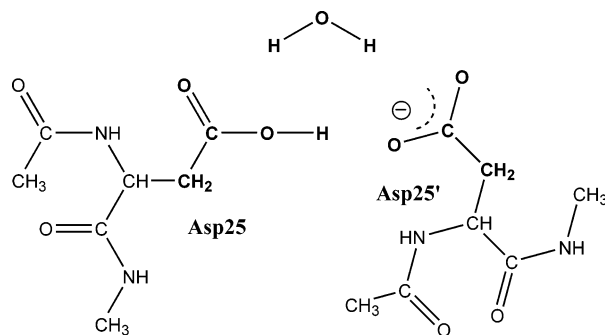
**Abstract:** The presence of a low-barrier hydrogen bond (LBHB) in aspartyl proteases and its implications in drug design have been the subject of intense study. Here, we present a combined quantum mechanical/molecular mechanical (QM/MM)–Numerov procedure and use it to characterize the  $O^{\delta 1}-H-O^{\delta 1}$  hydrogen bond (HB) in unbound HIV protease. The QM/MM scheme fully traces the shape of the HB's potential energy curve. The potential is used to obtain numerical solutions to the wave functions and vibrational energies of hydrogen, deuterium, and tritium. The vibrational eigenfunctions are used to compute expectation values for interatomic distances and vibrationally and thermally averaged spectroscopic properties of the  $O^{\delta 1}-H-O^{\delta 1}$  HB. Our work corroborates previous results by Piana and Carloni who found a LBHB via an ab initio molecular dynamics simulation (Piana, S.; Carloni, P. *Proteins* **2000**, 39, 26–36). Our predictions of isotope effects on the chemical shift of unbound HIV protease are consistent with experimental measurements in similar HBs. These results support the predictive power of this method and its potential use in screening inhibitors of aspartyl proteases.

### Introduction

Twelve years ago, Frey and others postulated that low-barrier hydrogen bonds (LBHBs) are key components of enzymatic acceleration.<sup>1,2</sup> They proposed that the transformation of a weak hydrogen bond (HB) in the ground state into a LBHB in the transition state or in a transient intermediate could significantly contribute to catalytic efficiency. The structural characteristics of a LBHB are a short heteroatom separation and a highly deshielded proton occupying a central position in a double well where the zero-point vibrational energy is close to the energy barrier.<sup>3</sup> Yet, its physicochemical nature, requirements for existence, and actual importance in catalysis have been subject to an intense debate.<sup>4–6</sup> The  $N^{\delta 1}-H-O^{\delta 1}$  HB of serine proteases was the first protein system in which the existence of this type of HB was postulated. The LBHB was then defined in terms of measured spectroscopic properties: an extreme downfield hydrogen chemical shift ( $16.0 \text{ ppm} \leq \delta_H \leq 21.0 \text{ ppm}$ ), a low fractionation factor ( $\theta < 1$ ), and large differences in the chemical shifts upon

isotopic substitution. However, as noted early by Perrin and Nielson, the ambiguity in the range of values given for these properties creates some difficulty in predicting the exact circumstances under which a LBHB will be observed.<sup>3</sup> In particular,  $\delta_H$  and  $\theta$  measurements cannot be used as the sole evidence for the presence of a LBHB: both a  $\delta_H > 16.0 \text{ ppm}$  and a  $\theta < 1$  can correspond to LBHBs, but they can also correspond to “strong” (asymmetric double-well) or “very strong” (symmetric single-well) HBs.<sup>7,8</sup>  $\Delta\delta_{H-D}$  measurements are not accurate because of the large errors arising from  $\delta_D$ 's broad band in the NMR spectrum. For instance, the  $\Delta\delta_{H-D}$  of the chymotrypsin–AcL–CF<sub>3</sub> complex is reported as  $1.1 \pm 0.5 \text{ ppm}$ .<sup>9</sup> In contrast,  $\Delta\delta_{H-T}$ —because of  $\delta_T$ 's narrow line width in the spectrum—is a much more precise measure of HB strength and provides reliable experimental evidence for LBHBs. Indeed,  $\Delta\delta_{H-T}$  measurements suggest that, while the  $N^{\delta 1}-H-O^{\delta 1}$  HB in the low-pH form of unbound serine proteases is “strong”, it becomes a “LBHB” in transition-state analogues. Among chymotrypsin–inhibitor complexes,  $\Delta\delta_{H-T}$  varies from 0.63 to 0.68.<sup>10</sup>

\* Corresponding author fax: (270) 809-6474; e-mail: pablo.molina@murraystate.edu.



**Figure 1.** Small model used for the initial optimization. Protomer Asp25-H is shown. The “QM” atoms relaxed during the optimization appear in bold.

LBHBs have also been reported in many other systems, for instance, aspartyl proteases. This family of enzymes is widely distributed in nature, and a number of aspartyl proteases have been identified as useful targets for chemotherapeutic intervention in human diseases.<sup>11–13</sup> Two notable members are memapsin 2—an enzyme associated with the pathogenesis of Alzheimer’s disease—and HIV protease (Hp), a target for AIDS. Hp is a homodimer that cleaves the polyprotein products of the HIV-1 genome, producing several proteins necessary for viral growth and cellular infection. The active site of Hp consists of a monoprotonated aspartate dyad with a water molecule bridging the two Asp residues (Figure 1). The Hp–substrate complex exhibits  $C_2$  symmetry, and numerous structure-based  $C_2$ -symmetric inhibitors have been designed to exploit this feature.<sup>14</sup> The symmetric nature of the  $O^{\delta 1}$ –H– $O^{\delta 1}$  HB is suggestive of the presence of a LBHB. Indeed, a computational study by Piana and Carloni on unbound Hp—via *ab initio* molecular dynamics simulations—found a LBHB between the  $O^{\delta 1}$  atoms that compensates for their negative charge repulsion and holds them within a distance of 2.5 ( $\pm 0.1$  Å).<sup>15</sup> Following this study, LBHB-based reaction mechanisms for Hp have been postulated and their implications in drug design have been discussed.<sup>16</sup> Northrop proposed that “looking to new designs, one might expect to achieve tighter binding by developing inhibitors that do not break the short, strong LBHB”.<sup>17</sup> Given the incidence of aspartyl proteases as targets of major human diseases, efforts taken in Northrop’s direction (i.e., relating “LBHB character” to inhibitory efficacy) could prove to be tremendously beneficial. As discussed above, experimental measurements of  $\delta_H$ , and more particularly  $\Delta\delta_{H-T}$ , can serve as indicators of HB strength. Yet, in aspartyl proteases, the absence of experimental data of these NMR properties for the  $O^{\delta 1}$ –H– $O^{\delta 1}$  HB hinders this line of study. In general,  $\delta_H$  measurements of carboxylic acids are complicated because of rapid proton exchange with water; moreover, in the case of Hp, autolysis might create an additional problem.<sup>15,18,19</sup> In this paper, we introduce a computational model that can correct this experimental inadequacy. Our model can be used to investigate low-barrier double-well hydrogen bonds of biological importance such as the  $O^{\delta 1}$ –H– $O^{\delta 1}$  HB in aspartyl proteases but should also be transferable to other HB systems of a similar nature. The model is used here to characterize the  $O^{\delta 1}$ –H– $O^{\delta 1}$  HB of unbound Hp. Results are consistent with Piana and Carloni’s benchmark work and

offer the first predictions of NMR isotope effects in aspartyl proteases. To our knowledge, this is the first quantum mechanical/molecular mechanical (QM/MM) study that traces the entire shape of a double-well “LBHB” potential energy curve. The paper is organized as follows. First, we present the computational methodology in a stepwise manner. Next, we discuss our computed structural and spectroscopic properties in light of available experimental results and theoretical considerations. We finish by discussing the implications of our work and the usefulness of the methodology.

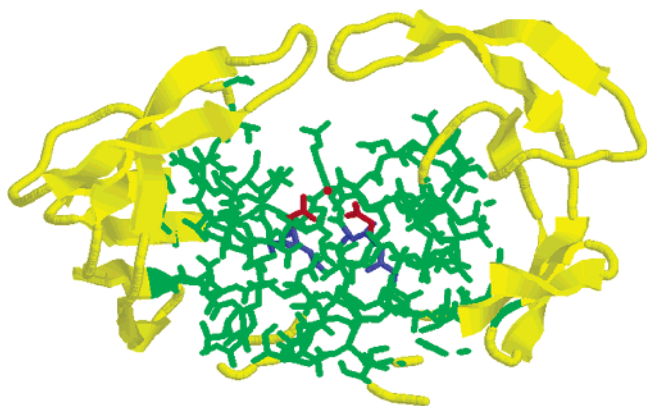
## Method

The model builds upon the method used by Molina and Jensen to characterize strong hydrogen bonds in serine proteases.<sup>8</sup> We trace the potential energy curve for proton transfer at the B3LYP/6-31G+(d,p) level of theory using the effective fragment potential (EFP) methodology, a QM/MM scheme in which the QM and MM regions are separated by a “buffer region” consisting of localized molecular orbitals.<sup>20</sup> We then obtain numerical solutions of the 1-D Schrödinger equation via the Numerov method.<sup>21</sup> The vibrational eigenfunctions are used to compute expectation values for interatomic distances and vibrationally and thermally averaged spectroscopic properties of the  $O^{\delta 1}$ –H– $O^{\delta 1}$  HB.

**A. Geometry Optimization.** The model for the  $O^{\delta 1}$ –H– $O^{\delta 1}$  HB is based on the same crystal structure utilized by Piana and Carloni (PDB code: 3PHV, 2.7 Å resolution).<sup>22</sup> A small 52-atom model of the active site (Figure 1) is built from the crystal structure coordinates; hydrogens are added with the online utility WhatIf Web Interface.<sup>23</sup> A water molecule is added manually and preoptimized semiempirically at the PM3 level. The small model is optimized at the B3LYP/6-31G(d) level of theory. When two different conformations with opposite protonation states are started with, two minima which we will call protomers Asp25-H and Asp25’-H are obtained. During the optimization, the Asp side chains and water molecule are relaxed while coordinates for the rest of the atoms remain frozen. This geometry setup leaves the carboxylates flexible and free to rotate, whereas their surrounding protein frame is kept fixed, thus mimicking the actual structural conditions of that protein region.<sup>15</sup> For practical purposes, the minima are equivalent with regard to energy and interatomic distances, as one could expect from the symmetry of the unbound Hp homodimer. The protomers differ in energy by just  $\Delta E = 2.16 \times 10^{-2}$  kcal/mol, explicable within computational error;  $\Delta r_{OH}$  is only  $7.40 \times 10^{-4}$  Å, and differences in other internal coordinates are also very small. The protomers present a short  $r_{OO}$  of 2.50 Å and an elongated  $r_{OH}$  of 1.05 Å. The carboxylates exhibit a small departure from coplanarity ( $\Omega = 32.6^\circ$ ). The external oxygens are separated by 4.59 Å. The dihedral angle between the protonated oxygen, the oxygens of the facing carboxylate, and the water oxygen is  $16.2^\circ$ .

**B. Least Linear Motion Path.** Next, a 19-point least linear motion path (LLMP) is traced along the proton-transfer coordinate that connects the two minima. Single-point energies are obtained using a large QM/MM model of the dimer at the B3LYP/6-31+G(d,p)/B3LYP/6-31G(d) level of theory. The LLMP corresponds to the potential energy





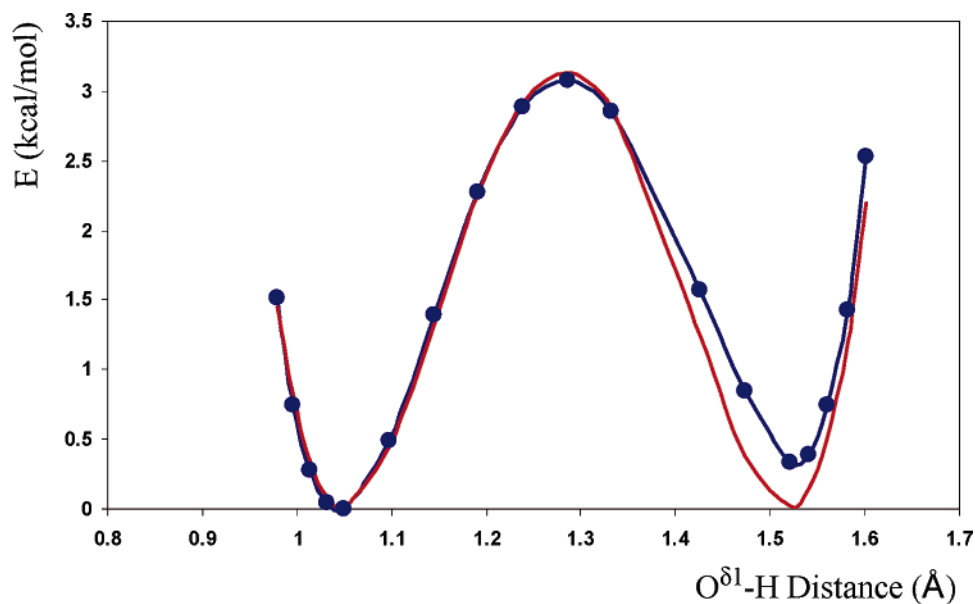
**Figure 2.** The three computational regions used in the QM/MM model. QM/buffer/EFP regions (red, blue, and green). Note: hydrogen atoms not shown for clarity; water added manually.

surface (PES) of the O<sup>δ1</sup>-H-O<sup>δ1</sup> hydrogen bond; coordinates changed are those of the 16 atoms that were allowed to relax during the optimization of the small model. The LLMP is constructed as follows: *Z* matrices with identical connectivity are built at the coordinates of the two minima and are used to calculate the changes in bond lengths, bond angles, and dihedral angles on going from Asp25-H to Asp25'-H. Then, nine intermediate points between the minima are generated by adding successive increments representing each internal coordinate change. Two additional points are created by scaling the coordinates by the factors necessary to decrease the O<sup>δ1</sup>-H bond length in both Asp residues to the O-H bond length (0.979 64 Å) of propionic acid optimized at the same level of theory. Last, using the *Z* matrix difference scheme, six additional points (three on each side) are placed between the minima and the points adjusted to the *r*<sub>OH</sub> in propionic acid.

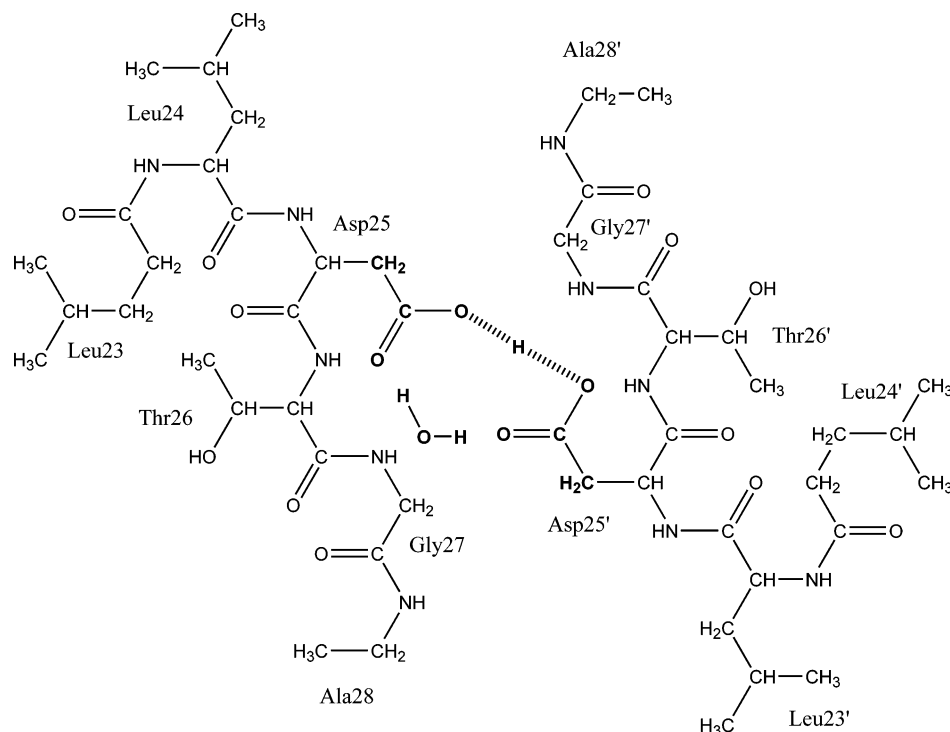
**C. QM/MM Methodology.** The QM/MM model has a total of 814 atoms. It consists of three computational regions: QM/buffer/EFP (Figure 2). The QM region contains the 16 atoms that were allowed to relax in the optimization

of the small model. The coordinates of these QM atoms are modified using the LLMP scheme and single points computed to construct the potential energy surface (see previous subsection). Note that, unlike HBs in nonhomogeneous environments, the unbound Hp homodimer affords an uncomplicated case of study of double-well potentials; the QM/MM-LLMP approach provides a sensible description of the energy barrier along proton transfer in this simple and symmetric system; structural relaxation might improve the accuracy of the computed barrier height but should not differ markedly from our estimate. The surrounding protein frame of the QM atoms is described by the buffer and EFP regions. The EFP region describes the electrostatic potential of the enzyme within roughly 12 Å of the active site by a multipole expansion (charges through octupoles at all atoms and bond midpoints and polarizable points for each valence-localized molecular orbital).<sup>24</sup> To create the EFP region, we use a “divide-and-conquer” approach.<sup>25</sup> The EFP has a total of six fragments that comprise residues 7–12, 81–91, and 21–33 (except QM and buffer atoms) in both monomers of unbound Hp. Each fragment is obtained via an RHF/6-31G(d) ab initio calculation, and then, all fragments are assembled together. The EFP and QM regions are separated by a buffer region of localized molecular orbitals obtained at the RHF/6-31G(d) level of theory.<sup>20</sup> There are a total of 28 buffer atoms comprising the backbone atoms around each Asp side chain (ending at the α carbons of neighboring residues). The geometries of the EFP and buffer regions are those of the crystal structure coordinates.

**D. Symmetrization of the Potential.** Polynomial regression of the PES traced by the computed data points yields the quartic function  $f(x) = 985.86x^4 - 5066.7x^3 + 9656.4x^2 - 8084.9x + 2510$ . Our potential is a little asymmetric to the left of the second minimum, most likely because of the sensitivity of the EFP points to small geometry changes along the LLMP. Note that point “12” had to be dropped because of an unreasonable energy. Thus, we symmetrize the PES (Figure 3) to improve the accuracy of the eigenvalues and



**Figure 3.** Original curve (blue) obtained with the LLMP points. The symmetrized potential (red) is traced with the same points.



**Figure 4.** The 162-atom model used to calculate the NMR chemical shifts. “QM” atoms are shown in bold.

the quality of the wave functions that are needed to compute structural and spectroscopic properties. Notice that unbound Hp is actually a symmetric homodimer, so no artificial correction is introduced. We construct a new symmetric function using the critical numbers of the quartic function: the  $r_{\text{OH}}$  for the two minima ( $x = 1.048\,963\,326$  and  $x = 1.518\,033\,637$ ) and the  $r_{\text{OH}}$  for the barrier maximum ( $x = 1.283\,498\,482$ ). These are set as absolute minima and the relative maximum, respectively.

Hence,

$$f'(x) = a(x - 1.048\,963\,326)(x - 1.283\,498\,482)(x - 1.518\,033\,637)$$

This implies that

$$f(x) = a(-2.043\,793\,713x + 0.250\,000\,000x^4 - 1.283\,498\,482x^3 + 2.443\,549\,159x^2) + c$$

Now we have

$$f(1.048\,963\,326) = -0.633\,903\,878a + c$$

$$f(1.283\,498\,482) = -0.633\,147\,443a + c$$

Solving for “a” and “c” with this system of linear equations, we obtain a new symmetric function:

$$V(x) = 1031.647\,795x^4 - 5296.473\,517x^3 + 10\,083.528\,41x^2 - 8433.901\,12x + 2615.878\,315$$

**E. Numerov Method.** We now proceed to obtain eigenvalues and eigenfunctions for the vibrational energy levels of H, D, and T. The one-dimensional Schrödinger equation

$$-\frac{\hbar^2}{2m} \hat{d}^2 \Psi(x) + \hat{V}(x) \Psi(x) = E \Psi(x) \quad (1)$$

is solved numerically using 10 000 grid points along the above potential via the Numerov method.<sup>6</sup> For H, the kinetic energy operator  $\hat{E}_K$  in units of kilocalories per mole is

$$\frac{\hbar^2}{2m} = 0.0478 \frac{\text{kcal}}{\text{mole}} \text{Å}^2 \quad (2)$$

and  $V(x)$  is given by our symmetrized potential. After correcting for the isotope mass,  $\hat{E}_K$  is 0.0239 for D and 0.0159 for T. The Numerov procedure uses a recursive relationship to compute the value of the wave function along the PES.<sup>21</sup> In practice, we obtain eigenvalues and eigenfunctions using the solver applet created by Schmidt.<sup>26</sup>

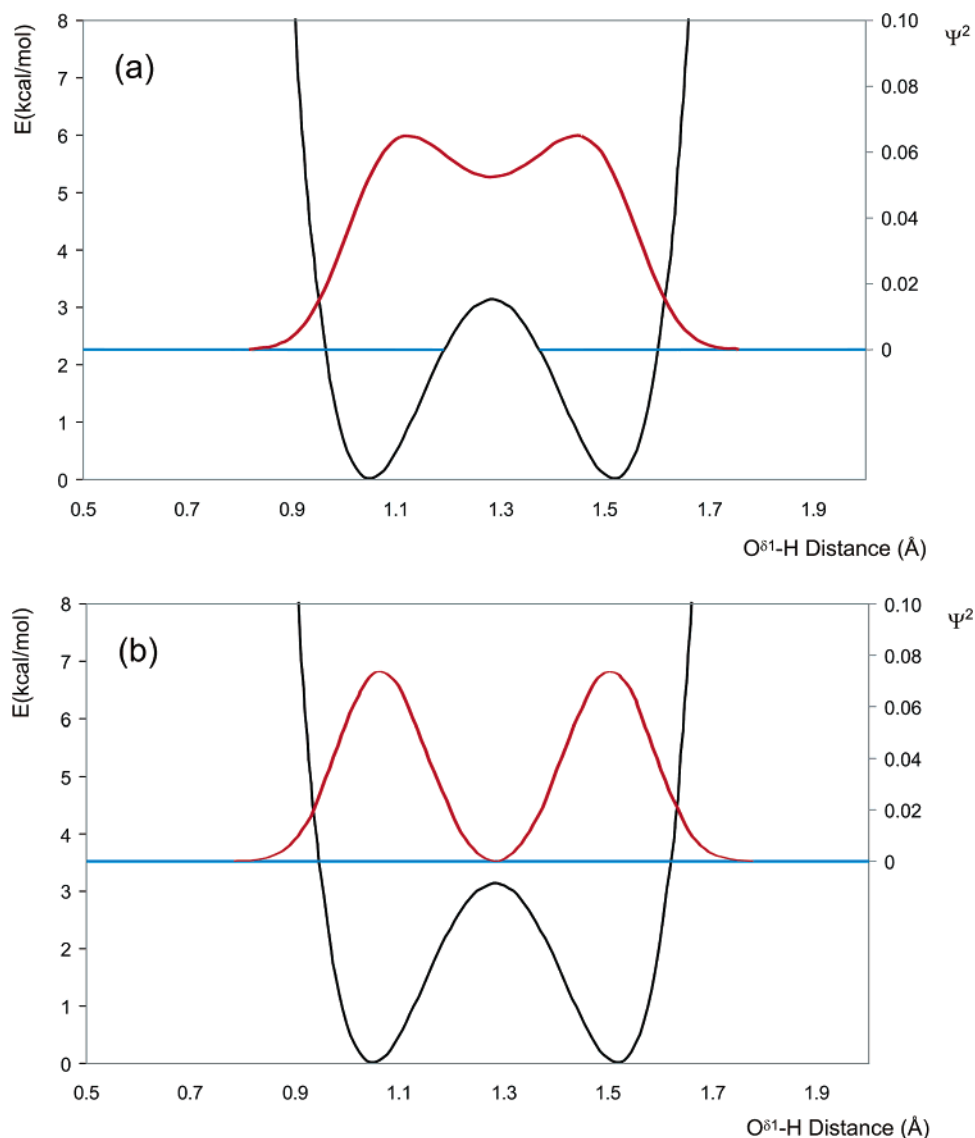
**F. Structural Properties.** The expectation value for the  $r_{\text{OH}}$  bond length is calculated with

$$\langle r_{\text{OH}} \rangle = \sum_{i=1}^{100} [\psi_0(i)]^2 r_{\text{OH}}(i) \Delta x \quad (3)$$

The square of the normalized ground-state vibrational wave function ( $\psi_0^2$ ) corresponds to the probability distribution function. The first point ( $x_{\text{min}}$ ) starts at an  $r_{\text{OH}}$  distance of 0.5 Å, and the remaining points are constructed via incremental steps ( $\Delta x = 0.015$  Å) up to  $r_{\text{OH}} = 2.0$  Å ( $x_{\text{max}}$ ).

$$\langle \pi \rangle = \sum_{i=1}^{23} [\psi_0(i)]^2 \pi(i) \Delta x \quad (4)$$

For hydrogen, the Boltzmann’s population fraction of the first excited state is only slightly bigger than 0.1 (see the next subsection) and  $\psi_1$  has two symmetric peaks around the PES midpoint. Hence, when the expectation value is averaged over the ground and first excited states,  $\langle r_{\text{OH}} \rangle$  is imperceptibly affected. Other structural properties ( $\pi$ ) are obtained via expectation values,  $\langle \pi \rangle$ , averaged over 23



**Figure 5.** Hydrogen. The energy values (horizontal lines) and the probability distribution functions (in red) of the (a) ground state and (b) first excited state.

points. The geometries of these points are built using the Z matrix difference scheme described in section B. In this case,  $x_{\min} = 0.76838 \text{ \AA}$ ,  $x_{\max} = 1.80254 \text{ \AA}$ , and  $\Delta x = 0.047 \text{ \AA}$ .

For comparison,  $\langle r_{\text{OH}} \rangle$  averaged over 23 points is only  $3.06 \times 10^{-4} \text{ \AA}$  larger than  $\langle r_{\text{OH}} \rangle$  averaged over 100 points, which shows that the use of 23 points is amply sufficient.

**G. Spectroscopic Properties.** The hydrogen NMR chemical shift is calculated using the linear scaling technique of Rablen et al.<sup>27</sup>

$$\delta_{\text{H}} = [30.60 - 0.957 \langle \sigma_{\text{H}} \rangle] + 0.33 \quad (5)$$

To their original equation (shown in brackets), we add a correction factor of 0.33 to obtain the aqueous-phase value of the chemical shift.<sup>28</sup>

$$\langle \sigma_{\text{H}} \rangle = \sum_{i=1}^{23} [\psi(i)]^2 \sigma_{\text{H}}(i) \Delta x \quad (6)$$

In eq 5,  $\langle \sigma_{\text{H}} \rangle$  is the expectation value for the chemical shielding and  $\sigma_{\text{H}}$  is computed using a 162-atom model

(Figure 4) that comprises roughly all atoms within  $7 \text{ \AA}$  of the active site. The B3LYP/6-311++G(d,p) level of theory is used for the atoms in the QM region, and STO-3G is used for the rest of the atoms.<sup>8</sup> The first excited state for hydrogen is thermally accessible at 298.15 K, yielding a Boltzmann's population ratio  $p_1/p_0$  slightly bigger than 0.1. It has a small but non-negligible contribution to the chemical shielding. Hence, we thermally average the chemical shielding over the ground and first excited states with eq 7

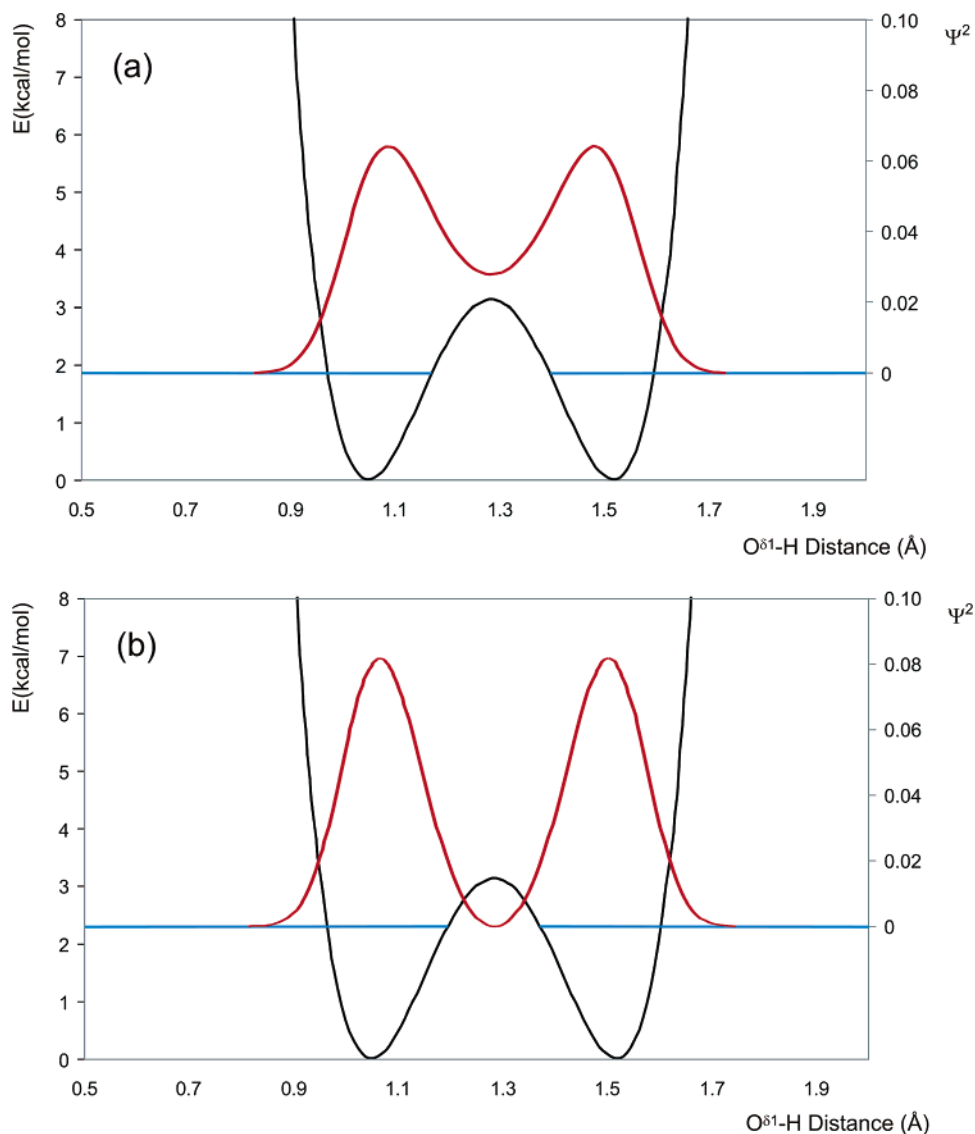
$$\langle \sigma_{\text{H}} \rangle = f_0 \sum_{i=1}^{23} [\psi_0(i)]^2 \sigma_{\text{H}}(i) \Delta x + f_1 \sum_{i=1}^{23} [\psi_1(i)]^2 \sigma_{\text{H}}(i) \Delta x \quad (7)$$

where  $f_0$  and  $f_1$  are the population fractions in the ground and first excited states, respectively.

The isotope effects on the NMR chemical shifts,  $\Delta \delta_{\text{H-D}}$  and  $\Delta \delta_{\text{H-T}}$ , are computed by

$$\langle \sigma_{\text{H}} \rangle - \langle \sigma_{\text{X}} \rangle \quad (8)$$

In eq 8, X = D or T. Once again,  $\langle \sigma_{\text{X}} \rangle$  is thermally averaged as in eq 7.



**Figure 6.** Deuterium. The energy values (horizontal lines) and the probability distribution functions (in red) of the (a) ground state and (b) first excited state.

**H. Miscellaneous.** The GAMESS<sup>29</sup> program was used for all calculations, except the NMR calculations, which were performed using Gaussian.<sup>30</sup>

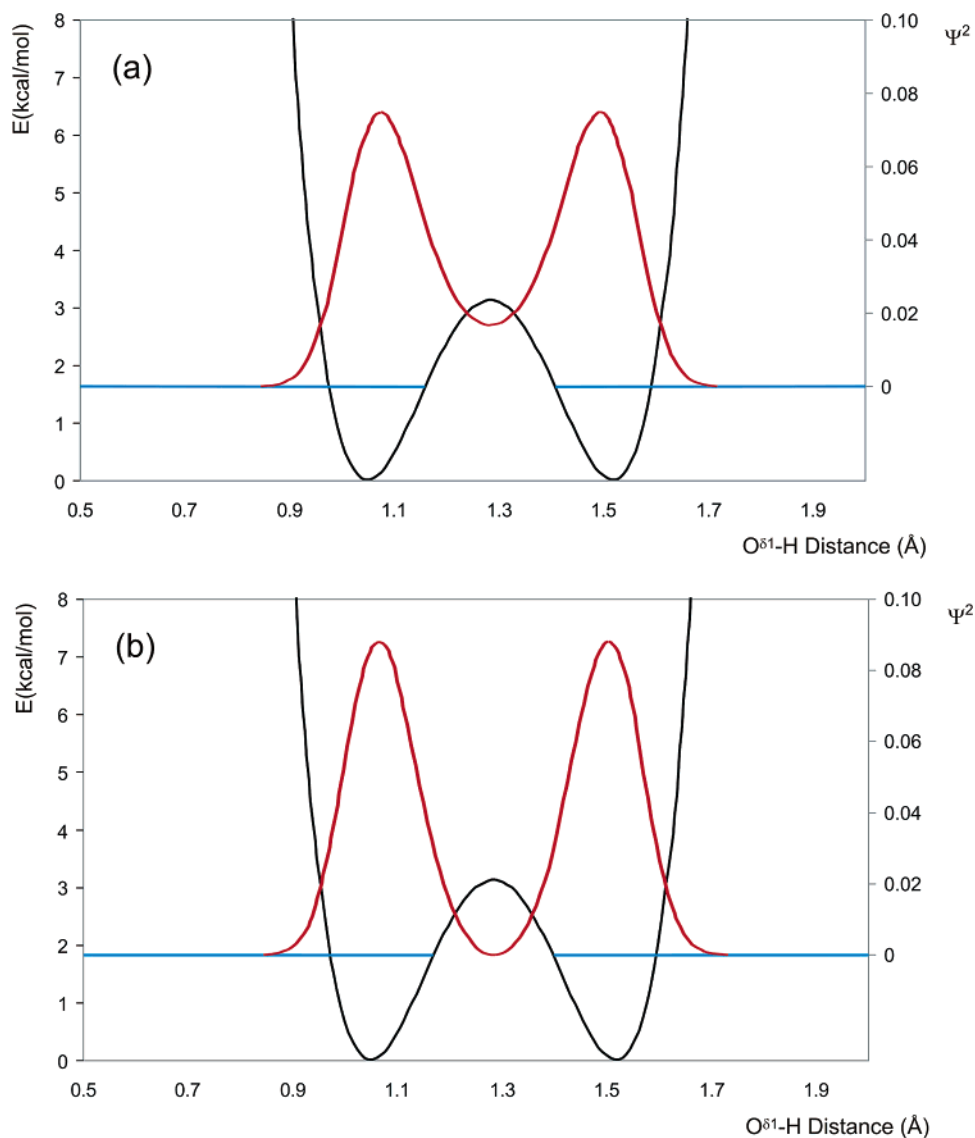
## Results and Discussion

The original quartic function  $f(x)$  yields a proton distribution function that presents irregularities, in particular, a lower “hump” above the second minimum (Figure 3). The uneven quality of the wave function somewhat affects the accuracy of our expectation values. Our symmetrized potential is used to correct this problem. The new potential  $V(x)$  is in very good agreement with both the quartic function and the curve obtained with the computed data points. The correlation coefficients are 0.995 and 0.987, respectively. Differences in values of calculated properties with each potential are small but not trivial. For comparison, the eigenvalue of the proton ground-state vibrational level is 2.27 kcal/mol using the symmetrized potential versus 2.33 kcal/mol with the unsymmetrized quartic function. Likewise,  $\langle r_{\text{OH}} \rangle$  is 1.26 Å with the quartic function versus  $\langle r_{\text{OH}} \rangle = 1.28$  Å with

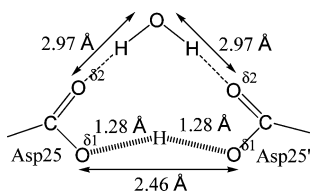
the symmetrized potential. The symmetrized potentials and the probability distribution functions for the ground-state and excited-state vibrational energies of hydrogen, deuterium, and tritium are shown in Figures 5, 6, and 7, respectively.

The symmetric double-well potential energy curve along proton transfer is consistent with the homodimeric nature of Hp. The structural properties of the HB (Figure 8) are also symmetrical. The expectation values (Table 1) reveal an equidistant water, a short  $\text{O}^{\delta 1}-\text{O}^{\delta 1}$  distance of 2.46 Å, and an  $\text{O}^{\delta 1}-\text{H}$  distance of 1.28 Å. The HB is quite nonlinear ( $\text{O}\hat{\text{H}}\text{O} = 151^\circ$ ) as also occurs in intramolecular  $\text{O}\cdots\text{H}\cdots\text{O}$  HBs. The angles of HBs commonly adjust to the local environment; notice also that, in aspartyl proteases, the proton is “sandwiched” between four lone pairs.<sup>31,32</sup>

In LBHBs, the closeness of the ground-state vibrational energy ( $E_0$ ) of the hydrogen to the barrier has been alternatively described as “at or slightly above”, “similar to”, and “below” the barrier.<sup>6,33,34</sup> Our potential energy curve places the energy barrier at 3.14 kcal/mol, 0.87 kcal/mol above the ground-state vibrational energy of the hydrogen.



**Figure 7.** Tritium. The energy values (horizontal lines) and the probability distribution functions (in red) of the (a) ground state and (b) first excited state.



**Figure 8.** Computed expectation values for interatomic distances in the active site of unbound HIV protease.

The proton is delocalized across a wide portion of the central region. Its diffuse position is revealed by the shape of the proton distribution function (Figure 5a). Notice the small dip in the probability that occurs at the top of the energy barrier.<sup>34,35</sup> The perturbation caused by the barrier has a larger effect on the probability distribution of the heavy isotopes, resulting in increasingly pronounced dips for deuterium and tritium (Figures 6 and 7). It also raises their  $E_0$  values relative to the  $E_0$  values of a symmetric, no-barrier potential; the ground states are 1.86 and 1.64 kcal/mol for D and T, respectively (Figure 9). In general, even states—having a large amplitude at the center of the potential (i.e., in the

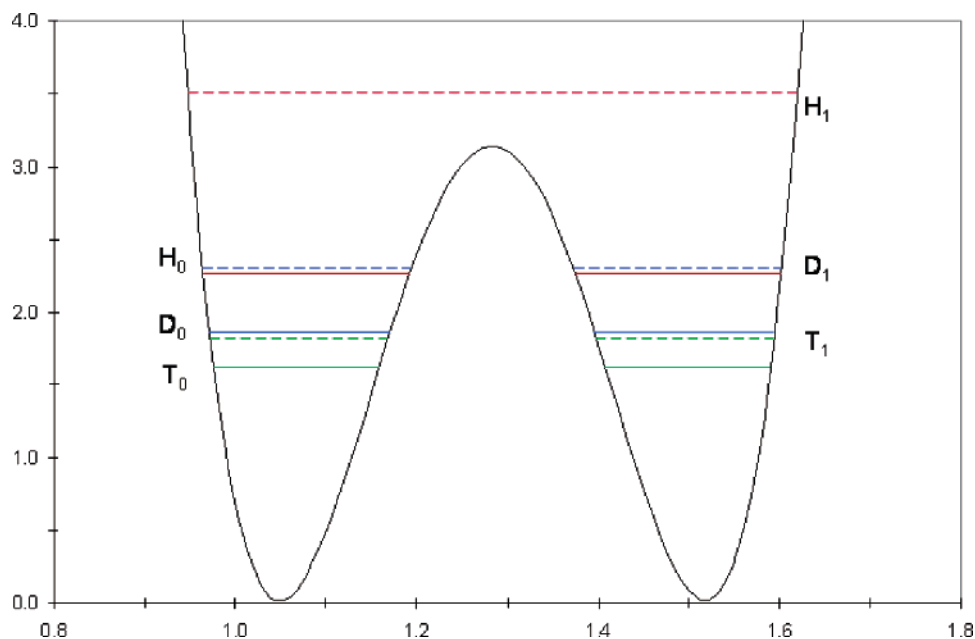
**Table 1.** Expectation Values Obtained in This Study<sup>a</sup>

property	expectation value
O <sup>δ1</sup> –H distance	1.28 Å
O <sup>δ1</sup> –O <sup>δ2</sup> distance	2.46 Å
O(water)–O <sup>δ2</sup> (Asp25) distance	2.97 Å
O(water)–O <sup>δ2</sup> (Asp25') distance	2.97 Å
O <sup>δ1</sup> –H–O <sup>δ2</sup> angle	151°
$\sigma_H$	14.9505 ppm
$\sigma_D$	15.4446 ppm
$\sigma_T$	15.6733 ppm

<sup>a</sup> The chemical shieldings are thermally averaged over the ground state and first excited state.

region of perturbation)—move up in energy, whereas odd states—with a node in the center—are much less affected (Table 2).<sup>36</sup>

This behavior is of particular importance for the ground and first excited states and needs to be considered in calculating the expectation values of the NMR chemical shift of the proton ( $\delta_H$ ) and the NMR shifts upon isotopic substitution ( $\Delta\delta_{H-D}$  and  $\Delta\delta_{H-T}$ ). Thus, the computed chemi-

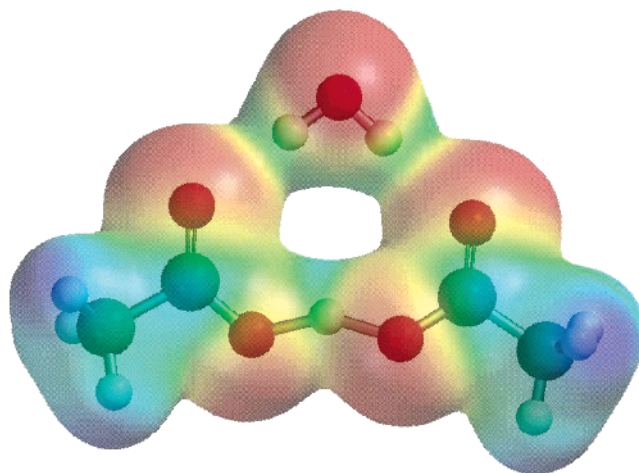


**Figure 9.** Eigenvalues of the ground state (solid lines) and excited state (dashed lines) of each isotope.

**Table 2.** The First Six Eigenvalues (in Units of kcal/mol) for Each Isotope

isotope	$E_0$	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$
H	2.27	3.52	7.47	11.67	16.66	22.21
D	1.86	2.30	4.93	7.24	10.21	13.52
T	1.64	1.83	4.02	5.55	7.73	10.15

cal shifts are averaged over the ground and first vibrational excited states (which is thermally accessible at room temperature). The first excited state for the proton is at 3.51 kcal/mol, 0.37 kcal/mol above the energy barrier. The population fraction in  $E_1$  at 298.15 K is 0.11. Hence, the first excited state has a small but non-negligible contribution to the computed spectroscopic properties. The calculated proton chemical shift is 16.6 ppm (versus 16.8 ppm if one neglects the first excited state). Note that our computed  $\delta_H$  cannot be used as sole evidence of a LBHB because  $\delta_H > 16$  ppm can also be attributed to “strong”<sup>8</sup> or “very strong”<sup>7</sup> HBs. As previously discussed, NMR measurements of  $\delta_H$  in carboxylic acids such as Hp are complicated because of rapid proton exchange with water.<sup>18</sup> Thus, experimentally, the absence of a downfield peak does not rule out a LBHB. A reasonable comparison between our computed  $\delta_H$  and an experimental aqueous  $\delta_H$  value of a system analogous to Hp is given by a study of the  $(\text{HCO}_2)\text{H}^-$  complex for which  $\delta_H = 14.1$  ppm.<sup>37</sup> This system is somewhat similar to the aspartyl dyad, and the  $r_{\text{OO}}$  distance is “expected to be ca. 2.50 Å or slightly less”. An upper limit for  $\delta_H$  is given by the intramolecular HB of maleic acid. The experimental proton chemical shift of this acid in dimethylsulfoxide was reported as  $\delta_H = 20$  ppm.<sup>31</sup> Notice that our computed  $\delta_H$  is not as downfield as  $\delta_H$ 's found in the  $\text{N}^{\delta^1}\text{H}-\text{O}^{\delta^1}$  HBs of serine proteases (which range from 16.9 to 18.9 ppm).<sup>38</sup> The proton in Hp might be somewhat shielded by the nonbonding electrons of the four surrounding oxygens. The electrostatic potential shown in Figure 10 depicts qualitatively the rich electron density areas around the carboxylic oxygens. The



**Figure 10.** Small model of the active site showing the electrostatic potential when the interatomic distances are those of the expectation values.

computed  $\delta_H$  value can also be interpreted by examining the probability distribution function (Figure 5a). The shape of  $\psi_0^2$  displays a broad region of high probability across the central region, with a dip in the center and two peaks located at about 0.05 Å from the minima. The same potential—but with a sharp peak in the center—would result in a more downfield proton chemical shift. Notice also that thermal averaging with  $\psi_1^2$  (which has a node of amplitude in the center, Figure 5b) moves  $\delta_H$  upfield by 0.2 ppm.

As aforesaid, the barrier causes the  $E_0$  of D and T to rise. Thus, the first excited states lie fairly close to the ground states;  $E_1 - E_0$  values are 0.44 and 0.19 kcal/mol for D and T, respectively, which results in first excited states having significant population fractions at 298.15 K (0.32 for D and 0.42 for T). The isotope effects on the chemical shifts are  $\Delta\delta_{\text{H-D}} = 0.49$  ppm and  $\Delta\delta_{\text{H-T}} = 0.72$  ppm. Because there is no experimental data for these spectroscopic properties in unbound HIV protease, we compare our calculations with

other relevant literature values. The NMR study of the (HCO<sub>2</sub>)H<sup>-</sup> complex in water discussed above gave a  $\Delta\delta_{\text{H-D}} = 0.64 \pm 0.14$  ppm. Bearing in mind the large error of this experimental value, our computed value of  $\Delta\delta_{\text{H-D}} = 0.49$  for unbound Hp compares quite well. Note that our calculated  $\Delta\delta_{\text{H-T}}/\Delta\delta_{\text{H-D}}$  ratio of 1.46 is in excellent agreement with the 1.44 ratio predicted by mass considerations.<sup>39</sup>  $\Delta\delta_{\text{H-T}}$  is a very accurate parameter of HB strength because of  $\delta_{\text{T}}$ 's narrow line width in the NMR spectrum. The stronger the HB, the larger  $\Delta\delta_{\text{H-T}}$  is. This parameter provides sound evidence for the presence of LBHBs in transition-state analogues of chymotrypsin. In these complexes, the experimentally measured  $\Delta\delta_{\text{H-T}}$  ranges from 0.63 to 0.68 ppm.<sup>7</sup> Thus, our calculated  $\Delta\delta_{\text{H-T}}$  of 0.72 ppm offers solid support for the presence of a LBHB in unbound Hp.

## Conclusion

We present in this work a model to characterize low-barrier double-well HBs of biological interest that should also be transferable to other HB systems of a similar nature. We apply the model to the O<sup>δ1</sup>–H–O<sup>δ1</sup> HB in unbound Hp. Our study traces the entire shape of the HB's potential energy curve and yields a double-well potential with a ground-state vibrational energy that is 0.87 kcal/mol below the barrier. Our work is in agreement with Piana and Carloni's results on unbound Hp and reveals some unexpected aspects of this HB such as the presence of the first excited states of deuterium and tritium below the energy barrier. Experimentally,  $\Delta\delta_{\text{H-T}}$  values can be used to gauge HB strength. Our computed  $\Delta\delta_{\text{H-T}}$  of 0.72 ppm is consistent with experimental measurements in transition-state analogues of chymotrypsin and strongly points toward the presence of a LBHB in unbound Hp. Northrop has proposed to relate inhibitory efficacy in aspartyl proteases to the LBHB character of the O<sup>δ1</sup>–H–O<sup>δ1</sup> HB. The driving force of this work was to develop a theoretical model to investigate such a relationship. An analysis of changes in structural and spectroscopic properties could help in the screening of drug candidate compounds. We believe that the present model can be used to study inhibitory efficacy in aspartyl proteases by evaluating how the nature of the O<sup>δ1</sup>–H–O<sup>δ1</sup> HB differs between unbound and bound enzymes. Our current efforts are pointing in that direction.

**Acknowledgment.** This work was supported by the National Science Foundation—Experimental Program to Stimulate Competitive Research and the Committee on Institutional Studies and Research—Murray State University. The authors thank Dr. John Porter and Carl Woods for their aid with the potential curves. Our study was done using computational resources at Murray State University, the University of Iowa, the University of Kentucky, and the Advanced Biomedical Computing Center.

## References

- Frey, P. A.; Whitt, S. A.; Tobin, J. B. *Science* **1994**, *264*, 1927–1930.
- Cleland, W. W.; Frey, P. A.; Gerlt, J. A. *J. Biol. Chem.* **1998**, *273*, 25529–25532.
- Perrin, C. L.; Nielson, J. B. *Annu. Rev. Phys. Chem.* **1997**, *48*, 511–544.
- Ash, E. L.; Sudmeier, J. L.; DeFabo, E. C.; Bachovchin, W. W. *Science* **1997**, *278*, 1128–1132.
- Schutz, C. N.; Warshel, A. *Proteins* **2004**, *55*, 711–23.
- Schiott, B. *Chem. Commun.* **2004**, 498–9.
- Frey, P. A. *J. Phys. Org. Chem.* **2004**, *17*, 511–520.
- Molina, P. A.; Jensen, J. H. *J. Phys. Chem. B* **2003**, *107*, 6226–6233.
- Frey, P. A. *Magn. Reson. Chem.* **2001**, *39*, S190–S198.
- Westler, W. M.; Frey, P. A.; Lin, J.; Wemmer, D. E.; Morimoto, H.; Williams, P. G.; Markley, J. L. *J. Chem. Am. Soc.* **2002**, *124*, 4196–4197.
- Coates, L.; Erskine, P. T.; Wood, S. P.; Myles, D. A.; Cooper, J. B. *Biochemistry* **2001**, *40*, 13149–57.
- Hong, L.; Turner, R. T., III; Koelsch, G.; Shin, D.; Ghosh, A. K.; Tang, J. *Biochemistry* **2002**, *41*, 10963–7.
- Ghosh, A. K.; Bilcer, G.; Harwood, C.; Kawahama, R.; Shin, D.; Hussain, K. A.; Hong, L.; Loy, J. A.; Nguyen, C.; Koelsch, G.; Ermolieff, J.; Tang, J. *J. Med. Chem.* **2001**, *44*, 2865–8.
- Andersson, H. O.; Fridborg, K.; Lowgren, S.; Alterman, M.; Muhlman, A.; Bjorsne, M.; Garg, N.; Kvarnstrom, I.; Schaal, W.; Classon, B.; Karlen, A.; Danielsson, U. H.; Ahlsen, G.; Nilroth, U.; Vrang, L.; Oberg, B.; Samuelsson, B.; Hallberg, A.; Unge, T. *Eur. J. Biochem.* **2003**, *270*, 1746–58.
- Piana, S.; Carloni, P. *Proteins* **2000**, *39*, 26–36.
- Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 11139–11149.
- Northrop, D. B. *Acc. Chem. Res.* **2001**, *34*, 790–7.
- Trylska, J.; Grochowski, P.; McCammon, J. A. *Protein Sci.* **2004**, *13*, 513–28.
- Silverstein, R.; Webster, T. Proton Magnetic Resonance Spectroscopy. In *Spectroscopic Identification of Organic Compounds*, 7th ed.; Rose, N., Ed.; John Wiley: New York, 1998; p 166.
- Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656–6665.
- Levine, I. N. The Harmonic Oscillator. In *Quantum Chemistry*, 5th ed.; Prentice Hall: New Jersey, NJ, 2000; pp 78–79.
- McKeever, B. M.; Navia, M. A.; Fitzgerald, P. M.; Springer, J. P.; Leu, C. T.; Heimbach, J. C.; Herbert, W. K.; Sigal, I. S.; Darke, P. L. *J. Biol. Chem.* **1989**, *264*, 1919–21.
- Krause, R.; Nielsen, J. E.; Vriend, G. The WHAT IF Web Interface. <http://swift.cmbi.kun.nl/WIWWWI> (accessed August 7, 2006).
- Minikis, R. M.; Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2001**, *105*, 3829–3837.
- Molina, P. A.; Li, H.; Jensen, J. H. *J. Comput. Chem.* **2003**, *24*, 1971–9.
- Schmidt, K. The Applet. <http://fermi.la.asu.edu/Schroedinger/html/node2.html#SECTION00020000000000000000> (accessed August 7, 2006)
- Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. *J. Phys. Chem. A* **1999**, *103*, 7357–7363.

- (28) Markley, J. L.; Ibanez, I. B. *Biochemistry* **1978**, *17*, 4627–40.
- (29) Schmidt, M. W.; Baldrige, 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*, 1347–63.
- (30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian* 98, revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.
- (31) Hibbert, F.; Emsley, J. *Adv. Phys. Org. Chem.* **1990**, *26*, 255–379.
- (32) Hao, M. H. *J. Chem. Theory Comput.* **2006**, *2*, 863–872.
- (33) Frey, P. A.; Whitt, S. A.; Tobin, J. B. *Science* **1994**, *264*, 1927–30.
- (34) Belot, J. A.; Clark, J.; Cowan, J. A.; Harbison, G. H.; Kolesnikov, A. I.; Kye, Y.; Schultz, A. J.; Silvernail, C.; Zhao, X. *J. Phys. Chem. B* **2004**, *108*, 6922–6926.
- (35) Benoit, M.; Marx, D. *ChemPhysChem* **2005**, *6*, 1738–1741.
- (36) Noggle, J. H. Quantum Theory. In *Physical Chemistry*, 3rd ed.; Harper Collins: New York, 1996; pp 678–679.
- (37) Fenn, M. D.; Spinner, E. *J. Phys. Chem. B* **1984**, *88*, 3993–3997.
- (38) Harris, T. K.; Mildvan, A. S. *Proteins* **1999**, *35*, 275–282.
- (39) Altman, L. J.; Laungani, D.; Gunnarsson, G.; Wennerstrom, H.; Forsen, S. *J. Am. Chem. Soc.* **1978**, *100*, 8264–8266.

CT600200S



# JCTC

Journal of Chemical Theory and Computation

## Validation of Linear Scaling Semiempirical LocalSCF Method

Victor M. Anisimov,<sup>\*,†</sup> Vladislav L. Bugaenko,<sup>‡</sup> and Vladimir V. Bobrikov<sup>‡</sup>

*Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, Maryland 21201, and Quantum Biochemistry Group, Konstantina Fedina-3/24, 105215 Moscow, Russian Federation*

Received July 5, 2006

**Abstract:** The numerical accuracy of linear scaling semiempirical methods LocalSCF and MOZYME is analyzed in comparison to conventional matrix diagonalization with respect to a variety of molecular properties including conformational energy, dipole moment, atomic charges, and bond orders. Major semiempirical MNDO, AM1, PM3, and PM5 Hamiltonians were considered in the study. As the numerical tests demonstrate, both LocalSCF and MOZYME reasonably reproduce matrix diagonalization results with the deviations being below the accuracy of semiempirical methods. However, the economical LocalSCF memory consumption and faster calculations are more beneficial for the quantum-mechanical modeling of large biological systems. The computational performance of the LocalSCF method is tested on the conformational energy calculation of a series of molecular dynamics snapshots of insulin in a large box of water.

### 1. Introduction

The linear scaling LocalSCF method was proposed recently as a computationally inexpensive alternative to matrix diagonalization for application to large biomolecular systems.<sup>1</sup> The purpose of the method is to bypass the prohibitive quadratic scaling of computer memory and cubic scaling of computer time by the size of the molecular system of the conventional diagonalization procedure, thereby allowing application to real-size biological systems at the quantum-mechanical level. Currently implemented in a semiempirical framework,<sup>2</sup> the LocalSCF method allows conducting molecular orbital computer simulations of hundreds of thousands of atoms on a personal computer. In LocalSCF, the linear scaling regimen is obtained with the help of the variational finite localized molecular orbital (VFL) approximation.<sup>1</sup> The VFL seeks a solution of the self-consistent field (SCF) task in the reduced space of compact or localized molecular orbitals (LMOs). For any LMO expansion fixed in the beginning of the SCF calculation, the resulting density matrix and the total energy are the closest possible solution to the

matrix diagonalization because of the variational principle under the constraint of the reduced LMO expansion.

Solving the SCF task using the VFL approximation does not guarantee a high-quality wave function because the initially selected fixed LMO expansion may not be optimal for the particular system of interest. Therefore, some kind of orbital expansion procedure is needed in order to account for the individual aspects of the molecular structure. The combination of the VFL method with a LMO expansion procedure creates the linear scaling LocalSCF method. LocalSCF calculations are conducted in the following way. After the SCF convergence is achieved for the initial-guess LMOs, an orbital expansion is undertaken on the basis of an expansion algorithm. This algorithm checks the atomic centers adjacent to each LMO and estimates the energy gain upon expanding the LMO on the particular center. If the expected energy gain is greater than a threshold value, the expansion is made; otherwise, the center is omitted, and another atom in the neighborhood is tested in a similar way. After the expansion step is made, the SCF refinement of linear coefficients of LMOs begins. The cycles of SCF and orbital expansion are repeated until the total energy is converged.

Another LMO-based linear scaling semiempirical method we consider in our study is MOZYME, which was developed

\* Corresponding author e-mail: victor@outerbanks.umaryland.edu.

† University of Maryland.

‡ Quantum Biochemistry Group, Konstantina Fedina-3/24.

by Stewart.<sup>3</sup> Both LocalSCF and MOZYME operate with LMOs, as their common ground. Despite this similarity, these methods are based on different principles. The atomic expansion of MOZYME LMOs is the integral part of the SCF procedure. MOZYME SCF performs occupied-virtual-orbital Jacobi rotations to achieve block diagonalization of the Fock matrix. The rotation has the purpose of orthogonalization of the occupied set of LMOs to the corresponding virtual set. Each orbital rotation combines expansion sets of occupied and virtual LMOs involved in the rotation, and as a result, the LMO size uncontrollably grows during the SCF iterations until the LMOs reach about 150 atomic centers on average. After this limit is achieved, adding new atomic centers to the LMO expansion does not change the density matrix, and such atomic centers with negligibly small linear coefficients can be effectively removed from the LMOs. For molecules consisting of several hundreds of atoms and larger, the MOZYME LMOs are considerably shorter than the canonical molecular orbitals, thereby providing a significant savings in computer resources. In contrast to MOZYME, the VFL-based SCF calculation in LocalSCF is conducted on fixed-size LMOs and the LMO expansion procedure is independent from SCF. This provides a free hand to control the LMO size, and the user may choose shorter or longer LMOs depending on the speed and desired degree of agreement with the matrix diagonalization result. A distant analogy can be portrayed between VFL and the Roothaan–Hall approximation.<sup>4,5</sup> If the Roothaan–Hall method seeks a wave function in the finite space of atomic orbitals, the VFL approximation utilizes additional degrees of freedom by constructing a wave function in the reduced space of molecular orbital expansion. A corresponding similarity may be observed between the selection of a particular basis set and the selection of the LMO size. In both cases, there is room for a rational human choice, and both solutions approach variationally the target function under elimination of the corresponding constraints. Accordingly, the VFL solution of SCF equations approaches the conventional matrix diagonalization limit when localized molecular orbitals are expanded to the size of conventional MOs. In practice, the total energy and molecular wave function converge well before the LMO expansion reaches the conventional MO limit. In most cases, the LMO expansion reaches 30 atomic centers on average. This helps to save significant computational resources when dealing with large biological molecules. Note that the wave function constructed from the short LMO expansion is a variational approximation to the wave function constructed from fully delocalized canonical molecular orbitals, and as such, the LocalSCF wave function may be reasonably good even if a very short LMO expansion is employed. The practical outcomes of the LocalSCF approximation will be studied in this work in comparison to the matrix diagonalization and the linear scaling method MOZYME. The analysis will be limited to the neglect of diatomic differential overlap (NDDO)-based semiempirical methods LocalSCF and MOZYME operating with LMOs. For the description of other linear scaling methods, readers are referred to the original articles.<sup>6–10</sup>

## 2. Methods

A small protein pro-insulin (PDB accession code 1EFE) containing 60 amino acids was downloaded from the Protein Data Bank.<sup>11</sup> Amino acid protonation was assigned on the basis of the assumption of physiological pH, and the total charge of +1 was neutralized by adding a chloride counterion. The cleaned structure was placed in an orthorhombic box of explicit TIP3P water with the solvent extending at least 10 Å from the solute molecule in each direction. The structure was equilibrated by a CHARMM force field<sup>12</sup> under a periodic boundary condition. A 200 ps molecular dynamics (MD) NPT simulation was undertaken at 300 K with the Leapfrog integrator and with a 2 fs time step. Additionally, the Langevin piston pressure control<sup>13</sup> was applied; covalent bonds to hydrogen atoms were restrained by the SHAKE algorithm,<sup>14</sup> and the particle mesh Ewald algorithm was used to treat the long-range electrostatics.<sup>15</sup> The first 10 ps of the simulation time were considered as an equilibration, and the other 190 ps were treated as a productive run out of which 20 snapshots were extracted with a 10 ps time interval for further analysis by quantum-mechanical methods. These 20 structures, each containing 20 058 atoms, were further processed by removal of the chloride counterion, leading to the molecular charge +1, and by the removal of water molecules, preserving only the 100 closest ones to the ionized amino acids. This resulted in 20 insulin conformations, each containing 1247 atoms. This size of the system was selected to allow performing matrix diagonalization calculations which fit to 1 GB of computer memory.

LocalSCF and MOZYME calculations were performed in regular and fast modes with the keyword PRECISE set on. Regular mode calculations were applied to the insulin snapshots containing 1247 atoms. The LocalSCF SCF procedure was performed by the steepest descent gradient optimization of linear coefficients of molecular orbitals, where the derivative of total energy versus the change in linear coefficients was utilized. In the regular mode LMO refining, SCF was converged to either 0.0005 eV for the gradient norm or to a value of 0.002 eV for the maximum component of the gradient for the linear coefficients of LMOs or to 0.0001 kcal/mol total energy change, whichever came first. The LMO expansion gradient threshold was set to 0.04 eV, and the expansions were performed until the total energy change was smaller than 0.5 kcal/mol between the expansions. The root-mean-square (RMS) value for LMO non-orthogonality was limited to 0.001. MOZYME regular mode calculations were conducted with the cutoff for NDDO approximation being set to 12 Å; SCF convergence criteria were set to 0.01 kcal/mol. Matrix diagonalization SCF was conducted until the total energy was converged with the 0.0001 kcal/mol threshold. Hereafter, we will use the name MOPAC as a synonym of matrix diagonalization.

LocalSCF and MOZYME in fast calculation modes were also applied to the large 20 058-atom box of insulin in water. Here, the MOZYME cutoff for NDDO approximation was reduced to 6 Å. In case of LocalSCF, the fast multipole method (FMM)<sup>16,17</sup> for the calculation of Coulomb integrals was turned on. The well-separation parameter was set to 2. The near field was set to 10 Å with a cube edge of 5 Å. The

**Table 1.** AM1 Total Energy, Dipole Moment, and Conformational Energies of Insulin Conformations Containing 1247 Atoms.

conf.	LSCFR <sup>a</sup>	MOZ12 <sup>a</sup>	MOP <sup>a</sup>	LSCFR <sup>b</sup>	MOZ12 <sup>b</sup>	MOP <sup>b</sup>	conformational energy <sup>c</sup>		
							LSCFR	MOZ12	MOP
1	-7859.73	-7866.73	-7866.30	253.33	253.34	253.28			
2	-7808.01	-7815.22	-7814.45	259.15	259.14	259.10	51.72	51.52	51.86
3	-7883.81	-7891.53	-7890.54	250.09	249.99	250.00	-75.80	-76.31	-76.10
4	-7838.84	-7846.53	-7845.51	224.07	224.02	224.02	44.97	44.99	45.04
5	-7777.99	-7786.06	-7784.70	259.18	259.24	259.13	60.85	60.48	60.81
6	-7842.35	-7849.25	-7849.19	211.81	211.76	211.75	-64.36	-63.20	-64.49
7	-7900.92	-7908.24	-7907.94	207.32	207.18	207.22	-58.57	-58.99	-58.75
8	-7871.47	-7879.05	-7878.29	247.39	247.44	247.32	29.45	29.20	29.65
9	-7864.70	-7873.22	-7871.67	275.46	275.45	275.36	6.78	5.82	6.62
10	-7889.75	-7897.31	-7896.69	225.83	225.80	225.74	-25.06	-24.08	-25.02
11	-7944.17	-7953.49	-7952.30	234.92	235.00	234.81	-54.42	-56.18	-55.62
12	-7878.29	-7886.17	-7885.26	202.73	202.75	202.67	65.89	67.32	67.05
13	-7895.03	-7902.26	-7902.03	194.59	194.55	194.51	-16.74	-16.09	-16.78
14	-7860.07	-7867.09	-7867.11	155.62	155.62	155.58	34.97	35.17	34.93
15	-7982.52	-7988.87	-7989.38	145.24	145.30	145.19	-122.46	-121.78	-122.27
16	-7970.14	-7977.65	-7976.99	151.29	151.29	151.23	12.39	11.22	12.39
17	-7903.43	-7911.07	-7910.33	174.45	174.51	174.40	66.71	66.58	66.66
18	-7848.29	-7856.44	-7855.14	161.91	161.99	161.83	55.14	54.63	55.20
19	-8005.24	-8013.12	-8012.40	197.49	197.50	197.42	-156.95	-156.68	-157.26
20	-7960.22	-7968.14	-7967.33	138.38	138.49	138.34	45.03	44.97	45.07
RMS	6.94	0.85		0.07	0.09		0.41	0.60	

<sup>a</sup> Total energy, kcal/mol; LocalSCF (LSCFR), MOZYME (MOZ12), MOPAC (MOP); LocalSCF FMM is off; MOZYME cutoff for NDDO approximation is 12 Å. <sup>b</sup> Total dipole, Debye. The dipole moment error is the module of the error vector  $|\bar{\mu}_{\text{MOPAC}} - \bar{\mu}_{\text{test}}|$  where the test represents LocalSCF or MOZYME dipole vectors. <sup>c</sup> Conformational energy, kcal/mol.  $E_i^{\text{conf}} = E_i^{\text{tot}} - E_{i-1}^{\text{tot}}$ .

series for the electrostatic potential was truncated at the fourth term, and the *B* and *C* operators were truncated at the seventh and third terms, respectively.

MNDO<sup>18</sup>, AM1<sup>19</sup>, PM3<sup>20</sup>, and PM5<sup>21</sup> semiempirical Hamiltonians were used in this work. MOZYME and matrix diagonalization calculations were performed by using the MOPAC2002 program package.<sup>21</sup> LocalSCF results were collected by using the LocalSCF computer program.<sup>2</sup> All calculations were performed on a single CPU personal computer equipped with an Intel Pentium-4 3.0 GHz processor with 1 GB of random access memory under the Microsoft Windows XP operating system.

### 3. Results and Discussion

#### 3.1. Regular Mode Calculation of Small Protein.

LocalSCF, MOZYME, and MOPAC calculations were performed on the 20 insulin conformations, each consisting of 1247 atoms. From the very beginning, matrix diagonalization calculations were unsuccessful because of the SCF convergence problem. The convergence was especially problematic on a dry protein. Preserving water molecules near ionized amino acids somewhat helped to improve the SCF convergence, but several conformations still remained problematic. A solution was found to feed the LocalSCF density matrix as an initial guess to the matrix diagonalization calculation. After this, MOPAC calculations were converging quickly in a few successive iterations, thereby confirming the good quality of the LocalSCF density matrix.

Details of the computational results for each of the 20 insulin conformations for energy and dipole moment are shown in Table 1 for the AM1 Hamiltonian. Synchronous

energy changes on the path from the first to the last conformation indicate that the protein structure is in fact in the process of equilibration. This conclusion is also supported by a synchronous change in the dipole moment. The incomplete equilibration is acceptable in this case because the purpose of the classical MD simulation was to generate realistic protein conformations for the subsequent comparison of linear scaling methods of interest with matrix diagonalization. As the data in Table 1 show, the absolute RMS deviations of the conformational energy in reproducing matrix diagonalization results are 0.41 and 0.60 kcal/mol or 0.9% and 3.8% for LocalSCF and MOZYME, respectively. RMS deviations for the dipole moment are 0.07 and 0.09 D for LocalSCF and MOZYME, respectively; both are in good agreement with matrix diagonalization. Similar studies were performed using other semiempirical Hamiltonians, and the results are presented in Table 2. Additional data include partial atomic charges, bond orders, and geometry gradients. The RMS differences for scalar properties (energy, partial atomic charges, and bond orders) are obtained from eq 1, where index *i* runs over *N* values of a property and index *c* runs over *M* protein conformations. The RMS differences for vector values are calculated according to eq 2. Insulin conformation number 1 is shown in Figure 1, illustrating the water placement in the vicinity of ionized amino acids.

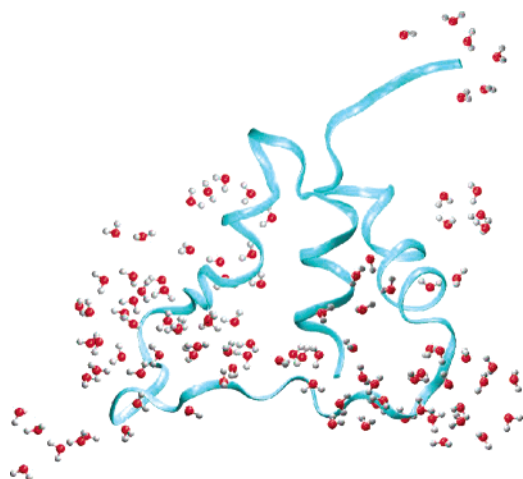
The data collected in Tables 1 and 2 demonstrate that

$$\text{RMS} = \sqrt{\frac{\sum_c \sum_i^N (V_{ic}^{\text{MOPAC}} - V_{ic}^{\text{test}})^2}{MN}} \quad (1)$$

**Table 2.** LocalSCF (LSCFR)<sup>a</sup> and MOZYME (MOZ12)<sup>b</sup> RMS Differences for 20 Insulin Conformations Containing 1247 Atoms in Comparison with Matrix Diagonalization in Their Regular Mode Settings

	MNDO		AM1		PM3		PM5	
	LSCFR	MOZ12	LSCFR	MOZ12	LSCFR	MOZ12	LSCFR	MOZ12
$E_{\text{tot}}^c$	3.50	0.61	6.94	0.85	8.22	0.45	10.50	1.98
$E_{\text{conf}}^c$	0.34	0.62	0.41	0.60	0.29	0.57	0.20	0.56
$\mu^d$	0.04	0.10	0.07	0.09	0.09	0.09	0.12	0.07
$Q_{\text{non-hydrogen}}^e$	0.0002	0.0005	0.0002	0.0005	0.0003	0.0006	0.0002	0.0007
$Q_{\text{hydrogen}}^e$	0.0001	0.0003	0.0001	0.0003	0.0001	0.0003	0.0001	0.0004
$\text{BO}_{\text{AB}}^f$	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000	0.0001	0.0001
$\nabla_{\text{non-hydrogen, \%}}^g$	0.36	0.61	0.42	0.60	0.63	0.60	0.85	1.09
$\nabla_{\text{hydrogen, \%}}^g$	0.91	0.87	1.08	0.89	1.47	0.96	1.29	1.06
CPU time, sec <sup>h</sup>	150	75	170	115	156	128	160	120

<sup>a</sup> FMM is off. <sup>b</sup> Cutoff for NDDO approximation is 12 Å. <sup>c</sup> Total and conformational energy differences, kcal/mol. <sup>d</sup> Dipole moment differences, Debye. <sup>e</sup> Partial atomic charge differences, electron units. <sup>f</sup> Bond orders; only values larger than  $10^{-6}$  were taken into analysis. <sup>g</sup> Geometry gradient differences; percent to absolute value. <sup>h</sup> Average calculation time.

**Figure 1.** Insulin conformation #1 solvated by 100 molecules of water (1247 atoms total), which are located in the vicinity of the ionized amino acids, creating the total charge of +1.

$$\text{RMS} = \sqrt{\frac{\sum_c^M \sum_i^N |\bar{V}_{ic}^{\text{MOPAC}} - \bar{V}_{ic}^{\text{test}}|^2}{MN}} \quad (2)$$

LocalSCF systematically underestimates the total energy, whereas MOZYME shows a good agreement with the matrix diagonalization. The smallest LocalSCF RMS difference of 3.50 kcal/mol is observed for the MNDO Hamiltonian. The largest difference of 10.50 kcal/mol is observed for the PM5 Hamiltonian, with AM1 and PM3 results being in the middle. This difference in total energy between LocalSCF and MOZYME is attributed to the difference in size of LocalSCF and MOZYME LMOs, which are 30 and 150, respectively. Supposedly, obtaining the energy limit requires adopting larger LMOs by LocalSCF; however, shorter LMOs are beneficial from a performance perspective. To find out whether LocalSCF is able to produce meaningful computational results with short LMOs, other molecular properties besides total energy have to be studied as well. The accuracy in the prediction of conformational energy is one such example. Here, the situation is quite different, with LocalSCF energies being systematically more accurate than the

MOZYME ones. This might be because LocalSCF LMOs are better converged. Indeed, because LocalSCF solves the task of linear scalability variationally, the produced wave function is the closest possible approximation to the target wave function under the given constraint of LMO size, whereas MOZYME LMO tail oscillations are harder to control. The ability to accurately reproduce conformational energies is extremely important when we think about the prospective performance of the molecular dynamics simulation of proteins at the semiempirical level. Overall, both methods show good agreement with the matrix diagonalization on conformational energies.

Another important aspect of comparison is to study how well linear scaling methods reproduce electrostatic properties and, namely, the electric dipole moment and partial atomic charges, in comparison to the matrix diagonalization. Because of the long-range nature of electrostatic forces, their accurate description is especially critical for large biological systems. In our analysis, the dipole moment differences are calculated in the vector form using eq 2. This approach is shown to be more sensitive to the errors in calculations of the dipole moment.<sup>22</sup> Because proteins have large dipole moments, even a 1% error in orientation of the dipole vector may have serious implications. Because insulin has a charge of +1, the center of mass was taken as the origin for the dipole moment calculation. In this study, both LocalSCF and MOZYME show very good agreement with the matrix diagonalization. LocalSCF is slightly better for the MNDO, AM1, and PM3 Hamiltonians, whereas MOZYME shows better results for the PM5 Hamiltonian. The situation with partial atomic charges is also good, although LocalSCF is in a little better agreement with the matrix diagonalization. In turn, MOZYME is slightly better on the prediction of bond orders. In all of these cases, the errors are well below the accuracy of semiempirical methods.

The next property presented in Table 2 is geometry gradients. Their accurate representation by linear scaling methods is especially important, keeping in mind forthcoming semiempirical MD simulations of proteins. Because the absolute value of the gradient strongly depends on the geometry of studied molecules, we perform an error analysis in relative units of percent value. The differences are calculated according to eq 3 in the vector form

$$\text{RMS} = \sqrt{\frac{\sum_c \sum_i^M \sum_i^N |\bar{v}_{ic}^{\text{MOPAC}} - \bar{v}_{ic}^{\text{test}}|^2 / |\bar{v}_{ic}^{\text{MOPAC}}|^2}{MN}} \quad (3)$$

Here, LocalSCF shows more accurate geometry gradients than MOZYME for heavy atoms for each Hamiltonian, except PM3. For hydrogen atoms, MOZYME gradients are systematically more accurate than LocalSCF for all Hamiltonians. In both of the cases, the LocalSCF and MOZYME gradient errors are below 2% of the corresponding target values, which is a good indication of reliability of both linear scaling methods.

The last factor considered in this comparison is CPU time. According to Table 2, the LocalSCF program using its default settings is slower than MOZYME by a factor of 1.2–2.0. This is because LocalSCF is optimized for large systems and is, in part, due to difficulties of making two computer programs run at the same level of accuracy, with the LocalSCF program settings being tuned up for a better agreement with matrix diagonalization and for a maximal savings of computer memory. Moreover, the performance comparison conducted on a small system may not be representative of the real performance when applied to a larger system. Because calculation time is critical for the calculation of large molecules, the faster program modes and larger systems will be considered further in this study.

To complete our analysis of the data presented in Table 2, we consider how well LocalSCF and MOZYME work with particular Hamiltonians. In MOZYME calculations, the MNDO, AM1, and PM3 Hamiltonians score equally with slightly higher errors obtained for PM5. The difference between the Hamiltonians is sharper in the case of LocalSCF. Here, the least errors are observed for the MNDO Hamiltonian, followed by AM1, then PM3, and at last, PM5. Because the AM1 Hamiltonian is considered to be of primary interest among other semiempirical Hamiltonians for biological applications,<sup>22–26</sup> it is especially important that our results confirm that AM1 does not introduce any particular complications for the linear scaling regimen. Regarding the PM5 Hamiltonian, both LocalSCF and MOZYME results indicate that PM5 makes more difficult in comparison to other Hamiltonians the task of achieving the linear scaling regimen.

**3.2. Fast Mode Calculation of Small Proteins.** In the above tests, LocalSCF calculations were performed with Coulomb integrals treated explicitly. However, the straightforward calculation time of Coulomb integrals scales quadratically with the number of atoms. This becomes unacceptable for larger systems. This problem is addressed in the LocalSCF program with the help of the fast multipole method. The next series of LocalSCF and MOZYME calculations on the insulin snapshots consisting of 1247 atoms are performed in the fast program settings. LocalSCF calculations were performed with the FMM option turned on. The comparable faster mode of MOZYME is obtained by setting the NDDO cutoff to 6 Å. FMM is not available in MOPAC2002 program. Because the studied system is still the same as the one used to collect data for Tables 1 and 2, the existing matrix diagonalization data were used as

reference data here as well. The obtained results are summarized in Table 3.

As in the previous tests, the total energy is underestimated by LocalSCF; however, all of the relative properties are well-reproduced. MOZYME energy differences in the fast mode also grew larger. Again, both linear scaling methods show the largest differences with the matrix diagonalization when the PM5 Hamiltonian is employed. Here, the difference is larger for MOZYME, which is unusual. Similar to the previous test, the LocalSCF conformational energies are in better agreement with matrix diagonalization than the ones calculated by MOZYME for all of the Hamiltonians, including PM5. LocalSCF dipole moments are more accurate for all Hamiltonians except PM5, where MOZYME is a little more accurate. Partial atomic charge differences are smaller for LocalSCF. Bond orders are equally good for both methods. A comparison of the data shown in Tables 2 and 3 indicates that, with the exception of the PM5 Hamiltonian, LocalSCF in its fast mode (Table 3) still provides more accurate conformational energies, dipole moments, and partial charges than MOZYME in its regular mode (Table 2). For the programs in their fast mode, the gradients are systematically more accurate in MOZYME with LocalSCF errors twice as large on average. The largest RMS differences in LocalSCF gradients are observed for the PM5 Hamiltonian.

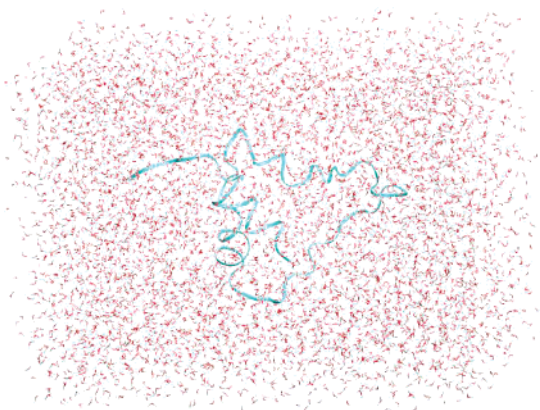
In contrast to LocalSCF, where regular and fast modes are different by turning off and on the FMM mode, the MOZYME program fast mode employs a reduced criterion for NDDO approximation (6 vs 12 Å), and therefore, the increase of MOZYME gradient errors comes entirely from the SCF part of the calculation. This allows us to compare LocalSCF gradient errors from Table 2 to MOZYME gradient errors from Table 3, because both are coming from the corresponding linear scaling replacement of the matrix diagonalization. As the data show, the gradient errors due to the MOZYME alternative to diagonalization in the fast mode (Table 3) are larger than the LocalSCF errors (Table 2). The errors are also subject to computer implementation of the algorithms, the factor which complicates the analysis of the differences. Overall, the errors due to LocalSCF and MOZYME linear scaling solutions are relatively small. Our work to reduce gradient errors is in progress and will be reported in a subsequent publication. As in the case of the default program settings (Table 2), LocalSCF in fast program settings is slower than the corresponding fast mode of MOZYME (Table 3); however, the LocalSCF program in the fast mode is in better agreement with matrix diagonalization for conformational energies, dipole moments, and partial atomic charges than MOZYME in default (accurate) settings for all Hamiltonians except PM5.

**3.3. Fast Mode Calculation of Protein in a Water Box.** The hydrated insulin samples calculated above consisted of 1247 atoms each. These are quite small systems in comparison to the number of atoms normally treated in biomolecular simulations by modern macromolecular force fields. It might be expected that, for such a relatively small system of 1000 atoms, many linear scaling methods may perform equally well. Indeed, LocalSCF and MOZYME in the above

**Table 3.** LocalSCF (LSCFF)<sup>a</sup> and MOZYME (MOZ6)<sup>b</sup> RMS Differences for 20 Insulin Conformations Containing 1247 Atoms in Comparison with Matrix Diagonalization in Their Fast Mode Settings

	MNDO		AM1		PM3		PM5	
	LSCFF	MOZ6	LSCFF	MOZ6	LSCFF	MOZ6	LSCFF	MOZ6
$E_{\text{tot}}$	4.39	4.15	7.73	2.37	8.25	1.14	18.35	23.71
$E_{\text{conf}}$	0.37	1.39	0.55	1.38	0.37	1.23	1.40	1.74
$m$	0.03	0.17	0.06	0.16	0.08	0.15	0.12	0.11
$Q_{\text{non-hydrogen}}$	0.0005	0.0014	0.0005	0.0016	0.0006	0.0020	0.0019	0.0021
$Q_{\text{hydrogen}}$	0.0002	0.0008	0.0002	0.0009	0.0002	0.0009	0.0007	0.0009
BO <sub>AB</sub>	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
$\nabla_{\text{non-hydrogen, \%}}$	3.49	1.56	3.73	1.64	4.03	1.59	7.73	3.09
$\nabla_{\text{hydrogen, \%}}$	4.81	2.51	6.87	2.59	4.44	2.64	9.03	3.36
CPU time, sec	102	49	104	60	111	75	111	68

<sup>a</sup> FMM turned on. <sup>b</sup> Cutoff for NDDO approximation is 6 Å.

**Figure 2.** Snapshot #1 of insulin in a water box (20 058 atoms) with total charge 0 neutralized by chloride counterions.

tests showed similar performances and comparable levels of agreement with matrix diagonalization. To see a difference between these two methods by CPU time and memory consumption, one should consider a bigger system for the tests. One of the practical challenges driving the development of linear scaling QM methods is to enable realistic calculations of large biological systems in their physiological environment, for example, quantum-mechanical simulations of biological macromolecules in the condensed phase. Although this goal is not yet quite achieved, it would be interesting to compare the performance of the LocalSCF and MOZYME methods on the insulin placed in a large water box, which we used in a classical MD simulation for the purpose of the generation of insulin conformations. A total of 20 snapshots containing 20 058 atoms (Figure 2) were considered in the test. The LocalSCF and MOZYME

programs were used in their fast modes as described above. These results are compared with conformational energies, dipole moments, and partial atomic charges calculated by LocalSCF in default settings (LSCFR) because the latter shows the best agreement with matrix diagonalization according to Tables 2 and 3. The reference LSCFR calculations required 198 MB of memory and 34 826 s of CPU time per box for the AM1 Hamiltonian, whereas MOZ12 calculation required 1500 MB of memory, which did not fit into 1 GB of the available memory.

The analysis of the performed calculations is presented in Table 4. Use of the FMM in a LocalSCF calculation resulted in significant deviations of LSCFF results from the reference data for the PM5 Hamiltonian, whereas existing FMM implementation worked well for all other Hamiltonians. LSCFF conformational energies are in much better agreement with reference data than the MOZ6 results. Similarly, LSCFF dipole moments are closer to the reference data than the MOZ6 ones. Deviations of partial atomic charges are all below 0.002 electron units for both methods, with MNDO and AM1 Hamiltonians favoring the MOZYME method and PM3 favoring the LocalSCF one. LSCFF calculation is about 3-fold faster than MOZ6 and 3-fold more economical in memory requirement.

The computational advantage of the VFL approximation, LocalSCF linear scaling solution, and short LMOs extends beyond the test cases discussed in this work. For example, performing LocalSCF calculations with disabled LMO expansion on the insulin in a water box provides an additional 3-fold speedup for the AM1 Hamiltonian with an average LMO size of 23 atomic centers. However, it is not yet clear whether the existing semiempirical Hamiltonians will remain

**Table 4.** LocalSCF (LSCFF) and MOZYME (MOZ6) in Their Fast Mode Settings; RMS Differences for 20 Insulin–Water Boxes Containing 20 058 Atoms in Comparison with LocalSCF in Accurate Settings (LSCFR)

	MNDO		AM1		PM3		PM5	
	LSCFF	MOZ6	LSCFF	MOZ6	LSCFF	MOZ6	LSCFF	MOZ6
$E_{\text{conf}}$ , kcal/mol	4.07	32.75	5.42	25.56	1.94	23.82	54.06	16.33
$\mu$ , Debye	0.86	3.71	1.13	3.50	0.22	3.40	10.15	2.00
$Q_{\text{non-hydrogen, } e}$	0.0013	0.0010	0.0016	0.0010	0.0003	0.0012	0.0157	0.0013
$Q_{\text{hydrogen, } e}$	0.0009	0.0008	0.0011	0.0008	0.0002	0.0008	0.0109	0.0009
CPU time, sec	3707	10039	4018	11682	5824	19035	5560	14388
memory, MB	241	816	256	816	283	816	270	816

valid for such extremely short LMOs. New semiempirical parameters are necessary in order to utilize in full the performance advantages of the extremely short LMOs.

## Conclusions

Linear scaling LocalSCF and MOZYME methods were studied in this work with respect to computational performance and the ability to reproduce matrix diagonalization results. Both diagonalization alternatives were found to reliably reproduce the target data for the MNDO, AM1, and PM3 Hamiltonians with the differences staying below a typical accuracy of semiempirical methods. Somewhat, larger differences between linear scaling and diagonalization results were observed for the PM5 Hamiltonian, pointing to the higher level of difficulty introduced by the PM5 Hamiltonian for the linear scaling LocalSCF and MOZYME methods. In the advent of modern linear scaling methods, future efforts on the development of new semiempirical Hamiltonians for biological applications should ideally target the task of facilitating the linear scaling. The performed tests indicate that MOZYME is more accurate in the prediction of total energy, which is systematically underestimated by LocalSCF due to the employment of relatively short LMOs. However, the LocalSCF method is shown to be more reliable in the prediction of conformational energies, dipole moments, and atomic charges, thereby supporting the conclusion about the predictive abilities of the LocalSCF approximation for the linear scaling problem. The increase of the LocalSCF geometry gradient errors in the FMM mode points to the need for the better optimization of the LocalSCF–FMM interface. For the large box of insulin in water, containing 20 058 atoms, the LocalSCF computer program is about 3-fold faster and 3-fold more economical by memory in comparison to the MOZYME program. However, the relative performances obtained are only rough estimations of the true capabilities of the LocalSCF and MOZYME theoretical methods because the details of their computer implementations cannot be entirely eliminated from influencing the test results. Also, the performed tests are limited to the studied systems, and the proposed methodology of the comparison should be applied to other molecular systems before the final conclusions can be drawn for a general case.

**Acknowledgment.** The authors thank Fujitsu Limited for granting the access to the LocalSCF and MOPAC2002 computer programs.

**Supporting Information Available:** Tables listing computational details for each individual snapshot and particular Hamiltonian. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Anikin, N. A.; Anisimov, V. M.; Bugaenko, V. L.; Bobrikov, V. V.; Andreyev, A. M. LocalSCF Method for Semiempirical Quantum-Chemical Calculation of Ultralarge Biomolecules. *J. Chem. Phys.* **2004**, *121* (3), 1266–1270.
- (2) Bugaenko, V. L.; Bobrikov, V. V.; Andreyev, A. M.; Anikin, N. A.; Anisimov, V. M. *LocalSCF2 User Manual*; Fujitsu Ltd.: Tokyo, Japan, 2005.
- (3) Stewart, J. J. P. Application of Localized Molecular Orbitals to the Solution of Semiempirical Self-Consistent Field Equations. *Int. J. Quantum Chem.* **1996**, *58* (2), 133–146.
- (4) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23* (2), 69–89.
- (5) Hall, G. G. The Molecular Orbital Theory of Chemical Valency VIII: A Method for Calculating Ionization Potentials. *Proc. R. Soc. London* **1951**, *A205*, 541–5552.
- (6) Lee, T.-S.; York, D. M.; Yang, W. Linear-Scaling Semiempirical Quantum Calculations for Macromolecules. *J. Chem. Phys.* **1996**, *105* (7), 2744–2750.
- (7) Dixon, S. L.; Merz, K. M., Jr. Semiempirical Molecular Orbital Calculations with Linear System Size Scaling. *J. Chem. Phys.* **1996**, *104* (17), 6643–6649.
- (8) Millam, J. M.; Scuseria, G. E. Linear Scaling Conjugate Gradient Density Matrix Search as an Alternative to Diagonalization for First Principles Electronic Structure Calculations. *J. Chem. Phys.* **1997**, *106* (13), 5569–5577.
- (9) Seijo, L.; Barandiaran, Z. Parallel, Linear-Scaling Building-Block and Embedding Method Based on Localized Orbitals and Orbital-Specific Basis Sets. *J. Chem. Phys.* **2004**, *121* (14), 6698–6709.
- (10) Goedecker, S.; Scuseria, G. E. Linear Scaling Electronic Structure Methods in Chemistry and Physics. *Comput. Sci. Eng.* **2003**, *5* (4), 14–21.
- (11) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (12) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616.
- (13) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103* (11), 4613–4621.
- (14) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.
- (15) York, D. M.; Darden, T. A.; Pedersen, L. G. The Effect of Long-Range Electrostatic Interactions in Simulations of Macromolecular Crystals: A Comparison of the Ewald and Truncated List Methods. *J. Chem. Phys.* **1993**, *99* (10), 8345–8388.
- (16) Greengard, L. *The Rapid Evaluation of Potential Fields in Particle Systems*; MIT Press: Cambridge, MA, 1988.
- (17) White, C. A.; Head-Gordon, M. Derivation and Efficient Implementation of the Fast Multipole Method. *J. Chem. Phys.* **1994**, *101* (8), 6593–6605.
- (18) Dewar, M. J. S.; Thiel, W. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.* **1977**, *99* (15), 4899–4907.

- (19) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902–3909.
- (20) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method. *J. Comput. Chem.* **1989**, *10* (2), 209–220.
- (21) Stewart, J. J. P. *MOPAC2002*, release 2.1; Fujitsu Limited: Tokyo, Japan, 2002.
- (22) Anisimov, V. M.; Anikin, N.; Bugaenko, V.; Bobrikov, V.; Andreyev, A. Accuracy Assessment of Semiempirical Molecular Electrostatic Potential of Proteins. *Theor. Chem. Acc.* **2003**, *109* (4), 213–219.
- (23) Giese, T. J.; Sherer, E. C.; Cramer, C. J.; York, D. M. A Semiempirical Quantum Model for Hydrogen-Bonded Nucleic Acid Base Pairs. *J. Chem. Theory Comput.* **2005**, *1* (6), 1275–1285.
- (24) Villar, R.; Gil, M. J.; García, J. I.; Martínez-Merino, V. Are AM1 Ligand–Protein Binding Enthalpies Good Enough for Use in the Rational Design of New Drugs? *J. Comput. Chem.* **2005**, *26* (13), 1347–1358.
- (25) Metzger, T. G.; Ferguson, D. M.; Glauser, W. A. A Computational Analysis of Interaction Energies in Methane and Neopentane Dimer Systems. *J. Comput. Chem.* **1998**, *18* (1), 70–79.
- (26) Vasilyev, V.; Bliznyuk, A. Application of Semiempirical Quantum Chemical Methods as a Scoring Function in Docking. *Theor. Chem. Acc.* **2004**, *112* (4), 313–317.

CT600222T



# JCTC

Journal of Chemical Theory and Computation

## A Second Look at Canonical Sampling of Biomolecules Using Replica Exchange Simulation.

[J. Chem. Theory Comput. 2, 1200–1202 (2006)]

Daniel M. Zuckerman and Edward Lyman

Pages 1200–1202. Our recent Letter describing several factors which could limit the efficiency of the replica exchange (RE) approach, while mostly accurate to our knowledge, contains an important error of logic. When we correct our picture of the RE approach, our previous fairly negative assessment of RE must be amended to become more positive. We believe all our specific ‘Observations,’ except **Obs. V**, are correct, although the text following **Obs. III** and at the end of the table’s caption are not correct.

Our Letter effectively equated sampling efficiency with ‘speed’ - - by which we meant the rate at which a simulation diffuses through configuration space. While our assessment of speed as such appears to have been correct, it does not properly reflect the true goal of canonical sampling: to sample all energy basins with the appropriate frequencies (probabilities). In RE, every level can sample new basins, which in turn can be ‘transmitted’ to the canonical ensemble desired for temperature  $T_0$ . Therefore, one can expect an overall *improvement* in canonical efficiency by a factor of

$$f < f_{\max} \sim (M + 1)^{-1} \sum_{i=0}^M k_a(T_i)/k_a(T_0) \quad (1)$$

where  $k_a(T)$  is the Arrhenius rate for temperature  $T$  as defined in our recent Letter. This new relation accounts for the basin-hopping rates of all levels. Of course, inefficiencies in transmitting new basins to the  $T_0$  ensemble will cause  $f < f_{\max}$ .

What is the overall outlook for biomolecular systems? The (fully correct) table entries from our Letter suggest  $1 < f_{\max} < 10$  for barriers  $\lesssim 6k_B T$  and maximum temperatures  $T_M < 500$  K. By this analysis, assuming modest barriers and maximum temperatures for biomolecular systems, RE is clearly much more promising than suggested in our original Letter. Whether it is sufficient for canonical sampling of large biomolecular systems remains an open question, however. Additional discussion of this issue can be found at our Web site, <http://www.ccbb.pitt.edu/Zuckerman/>.

We would especially like to thank Paul Maragakis who brought eq 1 to our attention.

CT600297Q