JCTC Journal of Chemical Theory and Computation

# Association of Aminoglycosidic Antibiotics with the Ribosomal A-Site Studied with Brownian Dynamics

Maciej Długosz,*,† Jan M. Antosiewicz,‡ and Joanna Trylska†

*Interdisciplinary Centre for Mathematical and Computational Modelling and Department of Biophysics, University of Warsaw, Zwirki i Wigury 93, Warsaw 02-089, Poland*

Received September 4, 2007

**Abstract:** Brownian dynamics methodology was applied to simulate the encounter of aminoglycosidic antibiotics with the ribosomal A-site RNA. Studied antibiotics included neamine, neomycin, ribostamycin, and paromomycin which differ in chemical structure, the number of pseudosugar rings, and the net charge. The influence of structural, electrostatic, and hydrodynamic properties of antibiotics on the kinetics of their association with the ribosomal A-site was analyzed. The computed diffusion limited rates of association are of the order of $10^{10}$ 1/M·s, and they weakly depend on ionic strength. Prior to binding, antibiotics often slide along the RNA groove with the time scale of approximately 10 ns per base pair in the case of neamine. We observed that upon forming the encounter complex aminoglycosides displace magnesium ions from the binding pocket.

## Introduction

Aminoglycosidic antibiotics are a family of antibacterial drugs which have been widely used in medical therapy for over 60 years. Most aminoglycosides interfere with translation by binding to the prokaryotic tRNA decoding A-site of the 16S RNA in the 30S ribosomal subunit.[1] They interfere with the decoding process by decreasing the accuracy of translation and by blocking proper peptide synthesis.[2,3] Upon binding, aminoglycosides displace two universally conserved adenine residues (A1492 and A1493) which are involved in contacts with the mRNA−tRNA hybrid. This leads to reduced discrimination against noncognate tRNAs and decreases translational fidelity.[4–8] Unfortunately, aminoglycosides suffer from moderate affinity and inadequate specificity and are toxic to mammalian ear and kidney cells. Moreover, bacterial resistance limits their effectivenes in medical therapy. Therefore, there is a widely recognized need to understand their binding mechanism in order to improve their selectivity and efficiency.

Aminoglycosides are sugar derivatives with various numbers of amine and hydroxyl groups. Antibiotics considered in this work, i.e., neamine, ribostamycin, paromomycin, and neomycin (see Figure 1), belong to the 4,5-disubstituted 2-deoxystreptamine neomycin class. They consist of a 2-deoxystreptamine ring with amine sugar ring substitutions at positions 4 and 5. Position 5 is the attachment point for auxiliary rings. Aminoglycosides are all positively charged at physiological pH due to the number of their amine groups[9] and, therefore, possess high affinity for the negatively charged RNA. The nature of the RNA−aminoglycoside interactions was found to be predominantly electrostatic.[10–12]

It was found that aminoglycosides bind to the A-site-containing RNA oligonucleotides in a manner similar to that of the full ribosome.[13–16] In recent years, crystal structures of several aminoglycosides bound to the A-site RNA constructs became available.[17–19] Crystal structures of aminoglycosides bound to the entire 30S ribosomal subunit were solved as well.[20–22] Because the structures of aminoglycosides complexed with the A-site RNA motifs reproduce the way of binding to the 30S subunit, they provide good models to study aminoglycosidic recognition which was also confirmed by fluorescence experiments.[23,24]

Detailed knowledge of how aminoglycosides interact with their binding sites on the ribosome may help in understanding

* To whom correspondence should be addressed. E-mail: mdlugosz@icm.edu.pl.

† Interdisciplinary Centre for Mathematical and Computational Modelling.

‡ Department of Biophysics.

10.1021/ct700210n CCC: $40.75 
Published on Web 03/06/2008

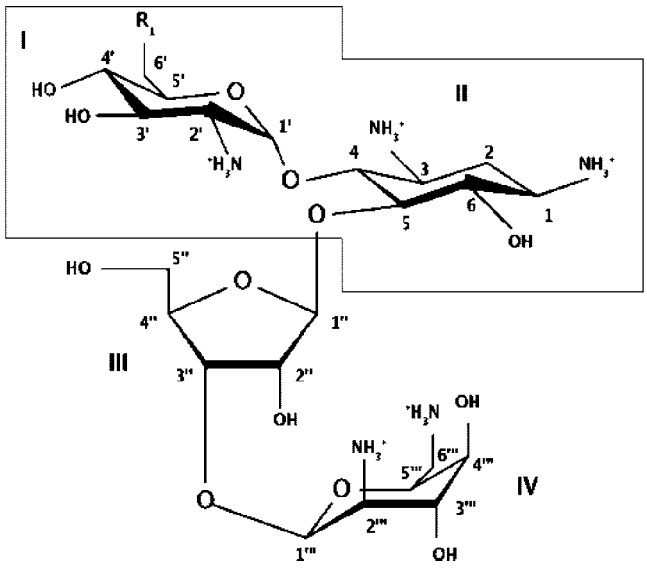


***Figure 1.*** Studied aminoglycosides. (neamine) rings I and II (inside the contour) with $R_1 = NH_3^+$; (ribostamycin) rings I−III with $R_1 = NH_3^+$; (neomycin) rings I−IV with $R_1 = NH_3^+$; (paromomycin) rings I−IV with $R_1 = OH$.

why antibiotics block certain stages of translation. Previous studies, both experimental[12,25–30] and theoretical[10–12,31] considered both structural properties of antibiotic/RNA complexes and thermodynamics of binding. For aminoglycosides' inhibitory role, it is not only important how strong are the bound complexes but how fast they can be formed. Therefore, in our work, we focus on the first stages of the binding process, namely the diffusion toward the RNA and association to form the encounter complex. We inspect how electrostatics influences these processes. There is no available experimental data which considers this problem in case of aminoglycosidic antibiotics and RNA, neither such studies have been conducted so far with theoretical methods.

To study the kinetics of association of aminoglycosides with the RNA, we apply Brownian dynamics (BD) methodology, which is a powerful technique to simulate the diffusional motion between interacting solutes.[32–34] BD is commonly used to provide theoretical estimates of the association rates of diffusion controlled reactions. This method was proven to be particularly useful for studying protein–ligand,[35,36] protein–protein,[37,38] and DNA−ligand interactions.[39] The interacting molecules move in a continuum solvent that exerts stochastic forces which lead to random collisions and induce a Brownian motion. BD trajectories are generated by solving the Ermak−McCammon equation[32] in a series of time steps to reach microsecond time scales with interparticle electrostatic interactions described with the Poisson–Boltzmann model.[40,41] On the basis of the number of association events, the probability of encounter and the rate of association are estimated.[42,43]

From the BD simulations, we determine the association rates of aminoglycosides to the A-site RNA and investigate their diffusion toward the binding site together with the influence of mobile counterions. On the basis of the reactive trajectories, we analyze the mechanism of the encounter complex formation with the RNA and its dependence on the net charge, chemical composition, and hydrodynamic properties of antibiotics. The simulations enabled us to describe the interactions governing the diffusion and binding of antibiotics to the ribosomal A-site.

## Methods

**Simulation of Diffusional Motion with Brownian Dynamics.** The theory of Brownian motion can be applied to particles immersed in solvent provided that their mass and size are larger than the mass and size of the solvent molecules. Such particles undergo a continuous irregular Brownian motion due to collisions with solvent molecules and their displacement $\Delta\vec{r}$ in time $\Delta t$ is given by[44,45]

$$\langle \Delta\vec{r}^2 \rangle = 6D\Delta t \tag{1}$$

where $D$ is the translational diffusion coefficient of the particle (for nonspherical particles $D$ has a meaning of an average diffusion coefficient) and

$$D = \frac{k_B T}{6\pi\eta a} \tag{2}$$

where $k_B$ is the Boltzmann constant, $T$ is absolute temperature, $\eta$ is solvent viscosity, and $a$ is the hydrodynamic radius of the particle. The dynamics of diffusional motion is described by the Langevin equation. One possible way of solving this equation is a propagation scheme presented by Ermak and McCammon.[32] The motion of a ligand, composed of spherically symmetric subunits, diffusing to a fixed receptor and subjected to both intersubunit and external forces can be derived from the following equation

$$r_i^{n+1} = r_i^n + \sum_j \frac{\Delta t}{k_B T} D_{ij} F_j^n + R_i(\Delta t) \tag{3}$$

where indices $i$ and $j$ run over the particle coordinates ($1 \le i,j \le 3N$), $r_i$ is the position vector component, $F_i$ is the sum of intersubunit and external forces acting in direction $i$, integer $n$ represents discrete times $t = n\Delta t$ at intervals (time steps) $\Delta t$, $D_{ij}$ is the diffusion tensor which is configuration dependent, and $R_i(\Delta t)$ is a random displacement whose average value is zero and $\langle R_i(\Delta t) R_j(\Delta t) \rangle = 2D_{ij}^n \Delta t$. Typically a relative motion of a ligand to a fixed receptor is simulated and the hydrodynamic interactions between them are neglected. Therefore, the diffusion coefficient of the receptor is summed into diagonal parts of the diffusion tensor.

**Modeling of Forces.** Because for a number of biomolecules, the electrostatic steering is the main force driving the encounter, it is important to have an accurate description of interparticle electrostatic interactions. In a BD simulation, intermolecular forces, $F_i$, are given as a sum of exclusion and electrostatic terms.[46] Other interactions, such as hydrogen bonding and van der Waals forces are neglected as they are too computationally demanding with regard to microsecond time scales achieved by this method.

Treatment of electrostatic interactions is based on the Poisson–Boltzmann model.[40,41] A molecule, immersed in a continuum solvent characterized with a high dielectric constant of $\epsilon \sim 80$, is represented as a set of beads with centrally assigned partial charges and with $\epsilon$ typically in the range 2–12.[47–50] Effects arising from dielectric heterogeneity

and ionic strength are also included. Electrostatic properties of such system can be derived from the nonlinear Poisson–Boltzmann equation[40,41]

$$\nabla\epsilon(\vec{r})\nabla\psi(\vec{r}) + 4\pi\left[\rho(\vec{r}) + \sum_i \mathrm{e}z_i c_i^{\mathrm{b}} \mathrm{e}^{-\frac{z_i \mathrm{e}\psi(\vec{r})}{k_{\mathrm{B}}T}}\right] = 0 \qquad (4)$$

where $\epsilon(\vec{r})$ is a function of the position, $(\vec{r})$ is the electrostatic potential, and $\rho(\vec{r})$ is the fixed molecular charge density. The sum represents the mean concentration of $z_i$-valent ions given by the Boltzmann distribution where $c_i^{\mathrm{b}}$ is the ion concentration in the bulk and e is the proton charge. For biomolecules of arbitrary shape, eq 4 needs to be solved numerically, e.g. with the finite-difference method[46,51,52] which gives as an output the electrostatic potential of a molecule on a 3D grid. In BD, a ligand moves in the potential generated by a receptor obtained from the solution of the Poisson–Boltzmann equation. Intermolecular forces are computed considering the ligand as a set of point charges immersed in the continuum solvent.[53]

**Estimation of Association Rates.** To compute bimolecular association rate constant, a solution of the diffusion equation is required and it can only be provided for systems with simple geometry. However, BD simulations allow one to estimate association rates and include the effects of molecular shape, charge distribution, internal motion, and hydrodynamic interactions. The association rate $k$ computed based on a BD simulation is given by[43]

$$k = k_{\mathrm{D}}(b)\beta^{\infty} \qquad (5)$$

where $k_{\mathrm{D}}(b)$ is the steady state rate constant for two particles separated with the distance $b$ (see Figure 2) and $\beta^{\infty}$ is the probability that having reached that distance particles will form an encounter complex. The value of $k_{\mathrm{D}}(b)$ can be computed analytically,[42,54] but $\beta^{\infty}$ must be estimated based on the BD simulation in which a large number of trajectories is generated.[43]

**Preparation of systems for Brownian Dynamics Simulations.** The coordinates of an oligonucleotide A-site duplex complexed with two paromomycin molecules (Figure 3) were taken from the Protein Data Bank (entry code 1J7T[17]). For computations of electrostatic potential and generation of BD trajectories, based on this structure, three variants of the oligonucleotide A-site model were derived: a structure of the bare A-site duplex (total charge of −40 e), of the A-site duplex with one bound paromomycin (total charge of −35 e), and of the bare A-site but with six explicitly modeled $Mg^{2+}$ ions (total charge of −28 e). The magnesium ions were not present in the original crystal structure of the A-site RNA, and their initial positions near the oligonucleotide surface (Figure 3) were obtained based on the coordinates of the *Thermus thermophilus* 30S subunit complexed with paromomycin (PDB entry code 1FJG[7,22]). Such an approach seemed reasonable because the root-mean-square deviation between the single A-site fragment of the studied duplex and the corresponding fragment of the 30S subunit with regard to backbone phosphate groups is 1.18 Å; therefore, the conformation of the RNA fragments containing the A-site is very similar in both structures. We decided to perform test simulations with explicit $Mg^{2+}$ ions because the RNA

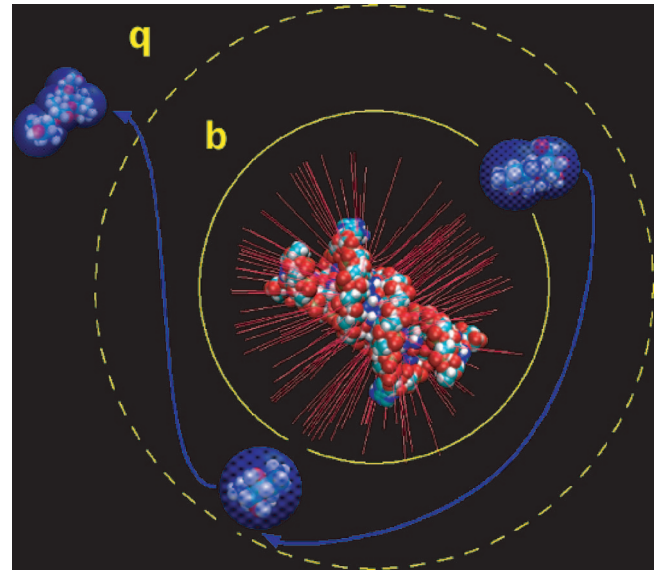


***Figure 2.*** Schematic diagram illustrating the BD method. Antibiotic (depicted with blue beads) moves through the electric field generated by the RNA (shown as van der Waals spheres). Lines of the electrostatic field are shown in red. Simulations are performed in coordinates defined relative to the position of the central oligonucleotide. At the beginning of a trajectory, the ligand is placed with a randomly chosen orientation at a randomly chosen point on the surface of the sphere with radius *b*. A BD trajectory is then generated. During a BD simulation, the ligand either diffuses outside a sphere of radius *q* and the trajectory is truncated or satisfies the predefined reaction criteria and the formation of an encounter complex takes place. The *b* radius is chosen such that outside this sphere forces acting on the ligand are centrosymmetric. The ratio of the number of reactive trajectories to the total number of trajectories allows one to compute the association rate constant.

is known to bind both monovalent and divalent ions which neutralize the backbone phosphate charges and are required for proper folding. Moreover, $Mg^{2+}$ ions are thought to be required for the formation of specific tertiary contacts.[55–57] Also, aminoglycosides are believed to displace ions from their RNA binding site upon complexation.[11,58] We aimed to test whether the presence of positive divalent ions in the proximity of aminoglycosidic binding site can influence the kinetics of binding to the RNA. We performed two kinds of simulations involving $Mg^{2+}$ ions, either fixing their positions near the RNA surface or allowing them to diffuse freely in the solution.

Partial charges and radii were assigned according to Amber force field parameters;[59] hydrogens were added to heavy atoms, and their positions were energy-minimized with the SANDER module of the AMBER 8 package using 10 000 steps of the steepest descent algorithm. Structures of neamine, ribostamycin, and neomycin were constructed based on the coordinates of the A-site RNA complexed with two paromomycins by simple removal or replacement of atoms with the Insight II 2000 software.[60] Partial charges for antibiotics were computed with the bcc option of the ANTECHAMBER application.[61] The aminoglycosides were determined to be fully protonated upon binding to RNA,[62] and accordingly,

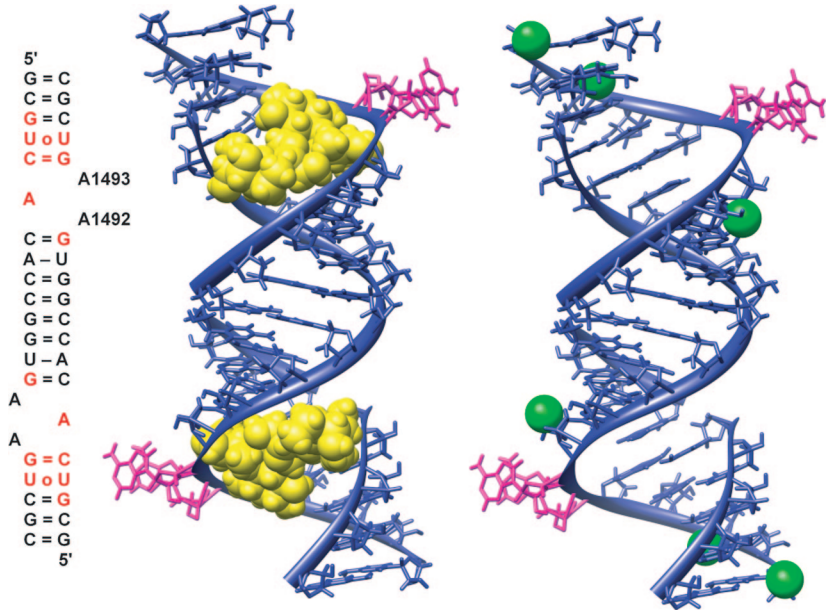


***Figure 3.*** Overall view of paromomycin/A-site oligonucleotide complex. (left) Secondary structure of the crystallized RNA duplex (two similar A-site models). A1492 and A1493 are labeled according to *E. coli* numbering. Base pairs are represented with = and − (corresponding to three or two hydrogen bonds, respectively) for Watson–Crick pairs or o for non-Watson–Crick pairs. Red denotes nucleotides in direct contact with paromomycin. (middle) Three-dimensional structure of the RNA duplex (blue) with paromomycins shown in yellow as van der Waals spheres and A1492 and A1493 denoted in magenta. (right) Positions of explicit $Mg^{2+}$ ions (green spheres) docked to the oligonucleotide A-site duplex based on their coordinates in the 30S ribosomal structure.

neamine and ribostamycin were assigned a total net charge of +4 e, paromomycin of +5 e, and neomycin that of +6e (Figure 1).

For the BD simulation, hydrodynamic parameters (such as translational diffusion coefficients and Stokes translational radii) are required. One also needs to construct hydrodynamic models of ligands, i.e., the representation of each ligand as a set of beads with hydrodynamic properties corresponding to its all-atom structure (see Figure 4). Because there is no experimental data regarding diffusion of those molecules, to compute the hydrodynamic properties of all antibiotics, we used the HYDROPRO software of de la Torre.[63,64] To validate our results, we parametrized the used software based on the ATP molecule for which experimental results are known.[65] We represented each antibiotic with a different number of beads corresponding to the number of its pseudosugar rings (Figure 4). Beads were centered on the geometric centers of rings and were assigned a total charge according to the number of amine groups connected to the given ring. Hydrodynamic radii of beads were chosen such that the resulting bead models reproduced the values of diffusional coefficients of antibiotics modeled with the all-atom representation (Table 1). To construct appropriate bead models, in-house software developed by one of us was used.[66] Magnesium ions were represented as spheres with a central +2 e charge and a size sufficient to enclose a fully solvated ion, i.e. an ion surrounded by six water molecules (Table 1). Translational diffusion coefficient and hydrodynamic radius of the oligonucleotide fragment required for BD simulations were also determined with the HYDROPRO software.

**Electrostatic Calculations.** All electrostatic calculations were carried out with the University of Houston Brownian Dynamics (UHBD) package.[53] The electrostatic potential and forces were calculated by solving the nonlinear Poisson–Boltzmann equation. Cubic 3D grids centered on the RNA with dimensions of 365 × 365 × 365 and 1.0 and 0.5 Å spacings (focusing feature of the UHBD program) were used. For the largest grid, the multiple Debye–Hückel boundary conditions were applied.[53] The ionic strength of the solvent was varied from 50 to 300 mM in equal steps of 50 mM at

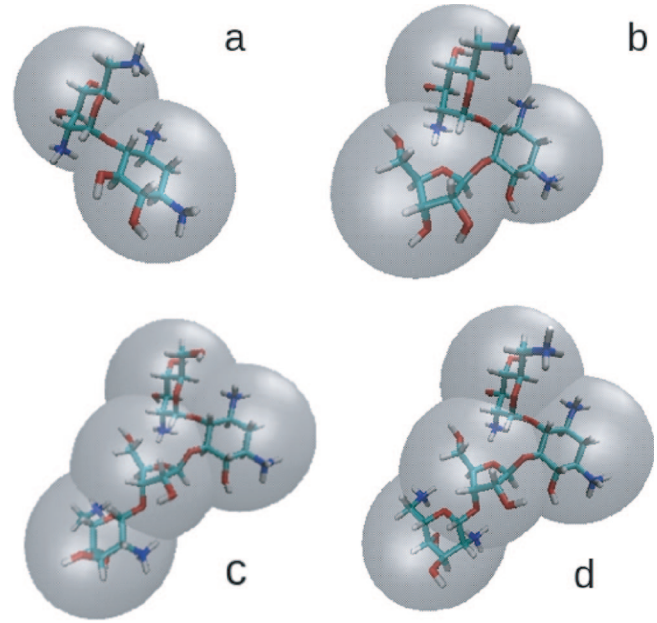


***Figure 4.*** Atomic structures and bead models used in BD simulations of antibiotics: (a) neamine, (b) ribostamycin, (c) paromomycin, and (d) neomycin. Each antibiotic is represented as a set of beads which are centered on the appropriate rings.

***Table 1.*** Parameters Used in Brownian Dynamics Simulations[a]

| molecule | $D_{tr}$ (cm/s) | $N_b$ | $R_{HB}$ [Å] | $Q$ |
|---|---|---|---|---|
| neamine | $4.15 \times 10^{-6}$ | 2 | 4.30 | 2.0, 2.0 |
| ribostamycin | $3.78 \times 10^{-6}$ | 3 | 4.70 | 2.0, 2.0, 0.0 |
| paromomycin | $3.21 \times 10^{-6}$ | 4 | 4.40 | 2.0, 2.0, 0.0, 1.0 |
| neomycin | $3.18 \times 10^{-6}$ | 4 | 4.50 | 2.0, 2.0, 0.0, 2.0 |
| $Mg^{2+}$ | $5.71 \times 10^{-6}$ | 1 | 3.75 | 2.0 |
| A-site duplex | $1.02 \times 10^{-6}$ | | | |

[a] The following parameters are shown: translational diffusion coefficients ($D_{tr}$), number of beads modeling ligands ($N_b$), hydrodynamic radii ($R_{HB}$) of beads, and charge assigned to each bead $Q$.

a constant temperature of 293 K. The dielectric constant of the RNA interior was set to 4 and, that of the solvent, to 78. We also calculated the electrostatic potentials with lower (2) and higher (12) $\epsilon$ values inside the RNA. Test BD simulations with these $\epsilon$ were performed, but no significant influence on the results was observed. The Richards probe-accessible surface[67] of the molecule was used for the definition of the dielectric boundary; the value of 1.4 Å was used for the solvent probe radius and an initial set of 280 surface dots per atom;[68] the Stern ion exclusion layer was defined with a 2 Å radius.

**BD Simulations.** BD simulations were performed with the UHBD package.[53] To compute the association rates, for each antibiotic, a total number of 5000 trajectories at six values of ionic strengths was simulated. All trajectories began with the antibiotic and the RNA fragment at a center-to-center distance of 90 Å ($b$-sphere). The radius of the $q$-sphere was set to 300 Å. The ligand model implemented in UHBD takes into account internal flexibility of molecules as each bead is able to move, interacting hydrodynamically and electrostatically with others. To maintain the overall shape of the diffusing molecule, holonomic constraints are applied (SHAKE algorithm[53]) which ensure that a molecule rotates and translates as a whole according to its diffusional properties. To maintain the lengths of the pseudobonds connecting the beads, the SHAKE algorithm was applied with a 0.2 Å tolerance. A variable time step was used; the value of 0.1 ps in the region within 50 Å around the RNA, 0.5 ps in the region within 50–75 Å, 1.0 ps within 75–120 Å, and 1.5 ps within 120–300 Å. Apart from the hydrodynamic radii, each bead was assigned an exclusion radius of 2 Å in order to account for the steric exclusion of the ligand by the RNA fragment. This did not permit any of the antibiotic bead (i.e., its center) to come closer than this radius to the van der Waals surface of any RNA atom. The exclusion radius chosen for each bead is smaller than its hydrodynamic radius in order to account, at least partially, for the flexibility of the RNA fragment; larger exclusion radii would prohibit closer contacts between the antibiotic and RNA.

BD simulations including $Mg^{2+}$ ions (either fixed or mobile) and a mobile paromomycin were performed at 150 mM ionic strength. A constant time step of 0.1 ps was used in this case. During simulations with mobile $Mg^{2+}$ ions, all BD trajectories were initiated with paromomycin placed randomly on the $b$-sphere. The starting positions of $Mg^{2+}$ were chosen as in Figure 3. In each case, 5000 BD trajectories were simulated. In both cases, interactions between the antibiotic and ions are taken into account; however, they are treated differently. Fixed ions are treated as subunits of the receptor because their presence influences the shape and electrostatic potential around the RNA fragment and therefore the movement of the antibiotic. When $Mg^{2+}$ ions and the paromomycin diffuse in the potential of the bare RNA fragment, their mutual influence is modeled with Coulombic and excluded volume interactions.

**Reaction Criteria.** As a measure of a successful formation of the encounter complex, for each antibiotic, we defined three reaction criteria (distances between the centers of beads mimicking aminoglycosidic rings and RNA atoms, see Figure 5). These criteria were chosen based on intermolecular distances observed in the crystal structure of the paromomycin/RNA complex. We assumed that the formation of the complex occurred when all of the observed distances differed from those of the crystal structure by less than 5 Å. Such a definition is rather tight, but we aimed to avoid ambiguities arising from the cylindrical shape of the RNA, as well as those trajectories which finished successfully but with the ligand bound on the other side of the A-site. Because the structure of the oligonucleotide is a fully symmetric duplex containing two antibiotic binding sites (Figure 3), during BD simulations the reaction criteria were checked for both sites.

## Results

**Diffusion of Antibiotics near the Surface of the RNA.** Computing of reaction rates based on BD simulations requires gathering a large number of trajectories. According to the theory described in the Methods section, these trajectories can terminate either as successful ones, i.e., satisfying the reaction criteria, or as nonreactive, in the case when the ligand escapes beyond the $q$-sphere. The probability of the reaction is determined based on the ratio of successful and the total number of trajectories.[43] However, BD trajectories are of finite length because in the calculations their duration is restricted by a predefined maximal time and maximal number of steps. Hence, it is possible that after exceeding of a predefined allowed number of steps, the trajectory terminates and the diffusing molecule remains inside the $q$-sphere but without satisfying the reaction criteria. This can pose a problem if one deals with a molecular system consisting of strongly interacting species. In the presented case, where the backbone phosphate groups of the RNA are a source of highly negative potential attracting positively charged antibiotics, the number of such trajectories turned out to be statistically important.

We observed three cases which led to termination of trajectories with the ligand remaining inside the $q$-sphere. One group included those where the antibiotic visited both binding pockets, i.e., two reactive events were recorded, in between which the ligand moved through a groove that connects the binding sites. The second group consisted of trajectories during which the ligand got trapped in one of the binding pockets and did not leave until the maximal number of allowed steps ($10^7$) was reached; in that case, only one reaction event was recorded. The third group included trajectories in which no reaction events were observed because the ligand was trapped at the molecular

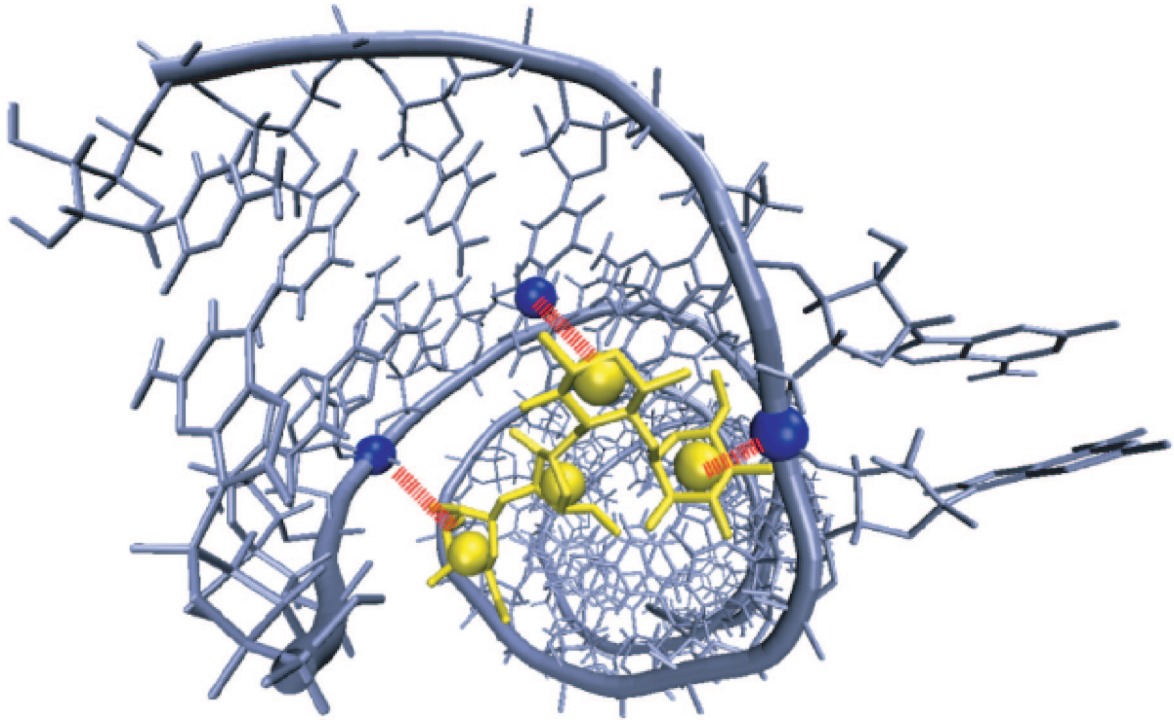

***Figure 5.*** Exemplary reaction criteria used in BD simulations of paromomycin. The distances between the centers of three beads of the antibiotic (shown in yellow), oxygen (residue U39), and phosphorus (residue U4 and A37) atoms (shown in blue) were monitored. RNA chains are shown with ribbons.

surface of the RNA and was not able to get to the binding site in the finite trajectory time. The latter cases were not observed for the smallest antibiotic i.e., neamine. The largest number of events ending with the ligand trapped on the RNA surface was observed for ribostamycin and was of the order of 10% of the total number of generated trajectories. For paromomycin and neomycin, which are the biggest and most highly charged from the studied set, the number of such trajectories was not statistically significant.

Taking into account the overall shape of molecules, their size, the total charge, and diffusion coefficients, the observed trappings of ligands inside the RNA groove can be explained as a result of not only purely electrostatic origin but also of steric interactions. The smallest neamine modeled with only two beads diffused most efficiently through the RNA groove. For ribostamycin, of similar to neamine total charge (+4 e) but composed of three beads, diffusion was less efficient and the strongest capture was observed. For paromomycin (+5 e) and neomycin (+6 e) composed of four beads, trapping inside the RNA groove was also observed but their diffusion near the RNA surface was more efficient than that of ribostamycin. Because we were interested not only in the mechanism of encounter but also in the association rate constants, to avoid any "truncated" trajectories due to the time or steps limit, we enlarged the maximal number of steps for generating BD trajectories hard-coded in the UHBD program.

For trajectories during which a ligand visited both A-sites, we determined the time it takes for the antibiotic to move between both binding sites. For neamine, the smallest one from the set, we present the times of its residence near the surface of the RNA fragment. Figure 6 shows the distribution of time intervals between subsequent binding events of

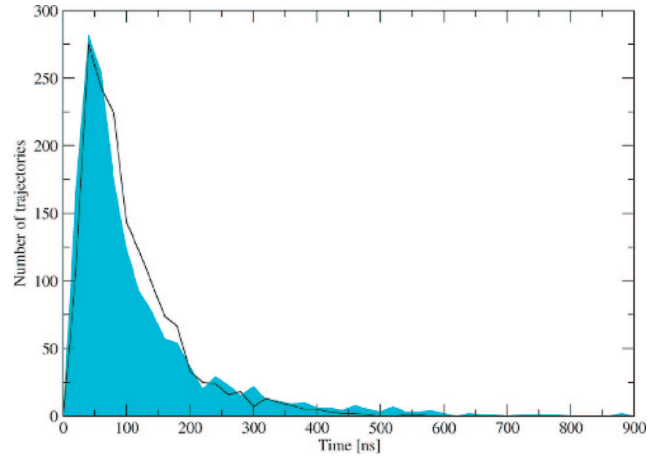


***Figure 6.*** (black line) Distribution of time intervals between subsequent antibiotic binding events in the two A-sites, i.e, ligand diffusion time from one A-site to another. (cyan polygon) Distribution of time intervals between the beginning of the BD trajectory (molecule positioned on the *b*-sphere) and the first binding event.

neamine to the RNA A-sites and distribution of time intervals between the beginning of the BD trajectory and the first act of binding. The latter distribution is shown only for comparison—those intervals depend on the radius of the *b*-sphere, thus, the fact that both shown distributions have similar width and maxima is a coincidence and has no physical meaning. The average time needed to diffuse between the binding pockets is in case of neamine of the order of 100 ns which leads to an average of 10 ns per base pair.

***Table 2.*** Association Rate Constants and Their Dependence on Ionic Strength Derived from BD Simulations

| | Rate $10^{10} \pm$ Error $10^8$ 1/M·s | | | |
|---|---|---|---|---|
| *I* [mM] | neamine | ribostamycin | paromomycin | neomycin |
| 50.0 | $3.42 \pm 5.61$ | $3.36 \pm 4.65$ | $3.33 \pm 4.78$ | $3.43 \pm 4.58$ |
| 100.0 | $2.89 \pm 5.74$ | $2.96 \pm 4.77$ | $2.73 \pm 4.81$ | $2.90 \pm 4.75$ |
| 150.0 | $2.59 \pm 5.60$ | $2.63 \pm 4.84$ | $2.43 \pm 4.82$ | $2.61 \pm 4.80$ |
| 200.0 | $2.41 \pm 5.57$ | $2.45 \pm 4.73$ | $2.27 \pm 4.80$ | $2.46 \pm 4.80$ |
| 250.0 | $2.33 \pm 5.53$ | $2.30 \pm 4.72$ | $2.20 \pm 4.79$ | $2.29 \pm 4.79$ |
| 300.0 | $2.25 \pm 5.52$ | $2.17 \pm 4.69$ | $2.18 \pm 4.77$ | $2.26 \pm 4.79$ |

The behavior of ligands observed in BD trajectories suggests that binding of antibiotics to the A-site can be accomplished in two ways: the ligand either finds the binding site directly or, due to the highly negative potential of the RNA fragment, associates to any place on the oligonucleotide surface and then slides along the groove between backbone phosphate groups, in a manner of one-dimensional diffusion, until it finds the binding pocket (see Supporting Information). Such a model of association kinetics and its applications to a regulatory protein which finds its specific site on the DNA chain were previously described.[69,70] However, in our case, to properly test this hypothesis, BD simulations with a longer fragment of the RNA oligonucleotide are required.

**Rates of Association and Their Dependence on Ionic Strength.** Rates of association and their dependence on ionic strength for the studied antibiotics are shown in Table 2 and Figure 7. All rate constants are of the same magnitude at all ionic strengths. Slower association (especially at ionic strengths above 50 mM) is observed for paromomycin. Paromomycin can be compared to neomycin because their size and overall shape are similar what results in similar diffusion coefficients (Table 1). However, neomycin associates slightly faster due to its higher net charge (+6 versus +5 e for paromomycin) and stronger interaction with the RNA.

Association rates for other ligand−protein and protein–protein systems have been measured in the range of $10^3-10^9$ 1/M·s where in the upper limit association is enhanced by strong electrostatic interactions.[71] For example, experimentally determined rate constants for positively charged ligands binding to acetylocholinesterase range up to $4 \times 10^9$ 1/M·s.[72] Therefore, our calculated rates seem reasonable even though, due to lack of experimental values, we do not focus our study on their absolute values but on relative ones and comparisons among various aminoglycosides.

For studied antibiotics, the decrease in computed association rates upon change of ionic strength from 50 to 300 mM is about 30%. This is a rather weak dependence in comparison with a 10-fold decrease observed experimentally for protein–protein association.[73–77] Such weak dependence on ionic strength observed in our case results from the fact that a significant positive antibiotic net charge is accumulated and distributed in a rather small volume and all partial charges assigned to beads are positive. Proteins, on the other hand, are characterized by nonuniform charge distributions and often a dipolar character, hence provide stronger shielding of protein–protein electrostatic interactions by the solvated ionic charges and stronger dependence of protein–protein association on the ionic strength of the solution.

Because the studied RNA fragment possesses two antibiotic binding sites, we checked if occupying one of the A-sites influences the mechanism and rate of encounter of the other antibiotic. Therefore, for paromomycin, we also computed its rate of association with the RNA while one binding pocket was already occupied by another paromomycin. The second paromomycin diffused in the electrostatic potential generated by the RNA complexed with all-atom model of the other bound paromomycin. From the simulations conducted at ionic strength of 150 mM, we obtained the rate constant equal to $1.16 \times 10^{10} \pm 3.42 \times 10^8$ 1/M·s. This number is nearly two times smaller than $2.43 \times 10^{10} \pm 4.82 \times 10^8$ 1/M·s presented in Table 2 for a naked oligonucleotide. This suggests that binding of one antibiotic does not influence the association rate constant of another one, and there is no cooperativity in subsequent acts of binding of antibiotics to the oligonucleotide. Such a lack of cooperativity was also seen in the relative binding free energy calculations conducted in our earlier studies.[12]

To explain this result, we visualized and compared 1000 successful trajectories of paromomycin diffusing in the potential generated by the naked RNA fragment and of paromomycin diffusing in the potential generated by the RNA fragment with one of the A-sites permanently occupied by another paromomycin; potentials were generated at ionic strength of 150 mM. On the basis of those trajectories, we prepared density maps representing the preferred positions of the antibiotic near the RNA surface. These maps are presented in Figure 8. As expected, in case of the bare RNA, the shape of the constant density surface is identical near both A-sites, showing that negatively charged backbone phosphate groups attract positively charged ligand and strongly influence its diffusion. In the second case, placing the paromomycin in one of the binding pockets modifies the shape of the density map moving it away from the RNA surface close to the part already occupied by the bound paromomycin. However, the presence of one positively charged aminoglycoside in the binding site does not introduce substantial changes in the electrostatic potential generated by the RNA fragment which would be sufficient to expel another antibiotic from this region. Therefore, the diffusing molecule can be still initially directed toward the already occupied region. Previously published experimental work[78] revealed by means of crystallography that binding of two aminoglycosides to one A-site of an RNA duplex is possible (with the second antibiotic molecule bound in various ways in the RNA groove at the edge of the A-site). Our results along with the above-described effect of trapping of antibiotics inside the RNA groove seem to be in agreement with this observation.

**Role of Magnesium Ions.** Figure 9 shows density maps representing the preferred positions of magnesium ions and antibiotics near the surface of the oligonucleotide. Maps were constructued based on BD trajectories of paromomycin and six $Mg^{2+}$ ions diffusing simultaneously in the electrostatic potential generated by the empty RNA fragment. Figure 9

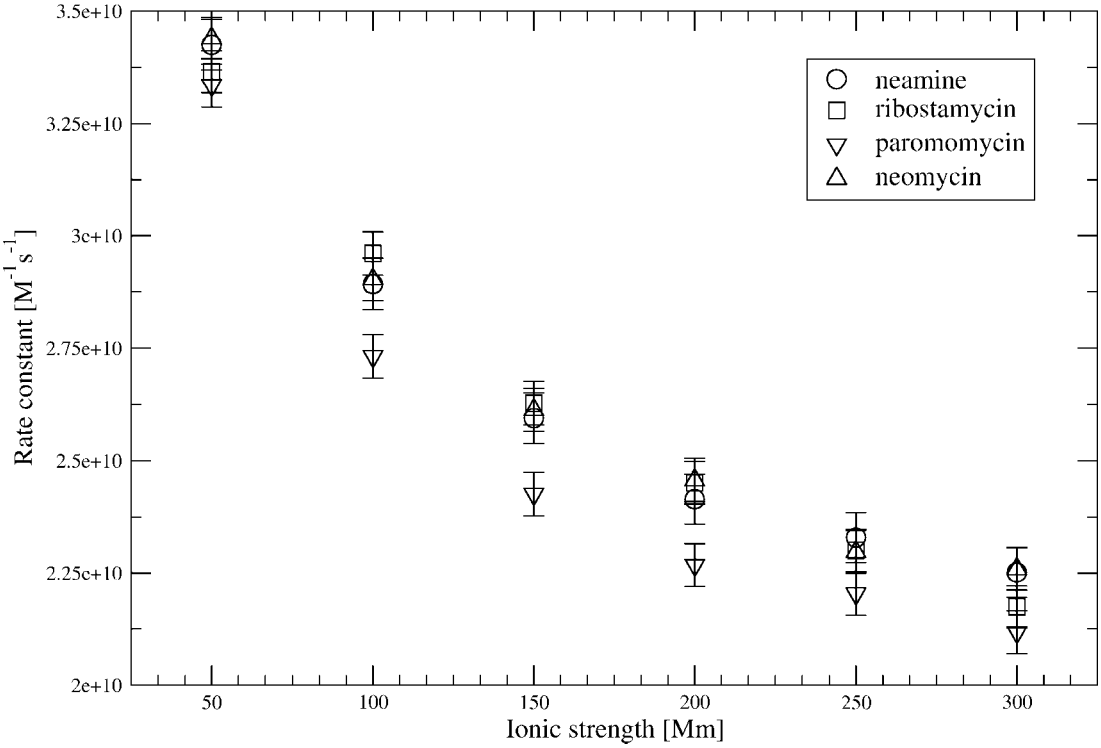


***Figure 7.*** Ionic strength dependence of the association rates computed from BD simulations. Error values are estimated at the 90% confidence level.

indicates that ions and antibiotics favor the same regions of space, i.e. both A-sites and the major groove of the RNA. Together with Coulombic interactions between positively charged antibiotics and positively charged $Mg^{2+}$ ions this leads to competition of both species upon binding.

Figure 10 presents the dependence of paromomycin association rates on the reaction criteria for three types of simulations: without $Mg^{2+}$ ions, with fixed positions of $Mg^{2+}$ ions, and with mobile $Mg^{2+}$ ions. When magnesium ions are not present, no dependence of association rates on the used reaction criteria is observed. A similar result is obtained when the positions of $Mg^{2+}$ ions are fixed during simulations but the computed association rate constant is lower ($1.80 \times 10^{10}$ 1/M·s) than the one obtained from the simulation without ions ($2.43 \times 10^{10}$ 1/M·s). Decrease in association probability is caused by a lower total charge of the RNA fragment (−28 e with $Mg^{2+}$ vs −40 e without $Mg^{2+}$) and less attractive electrostatic potential influencing the movement of antibiotic.

A different scheme is observed when magnesium ions are allowed to diffuse freely in solution. Association rate constants obtained for larger distances (less strict reaction criteria) are lower ($2.13-2.25\ 10^{10}$ 1/M·s) than the ones computed based on the simulations without ions but higher than those computed with fixed positions of ions. This is a result of electrostatic repulsion between ions and the antibiotic which are competing for the position inside the same A-site. When shorter reaction distances are taken into account there is an abrupt change in the computed rates which decrease to $6.90 \times 10^{9}$ 1/M·s. This arises from both an excluded volume effect (when the $Mg^{2+}$ ions are bound inside the A-site there is not enough space to accommodate the antibiotic) and electrostatic repulsion between the antibiotic and ions (as the antibiotic needs to expel ions from the binding site in order to satisfy reaction criteria; see Supporting Information).

We observe that magnesium ions are able to diffuse away from the oligonucleotide. From the total of 6 $Mg^{2+}$ initially positioned near the oligonucleotide, the maximal number of $Mg^{2+}$ ions present near the RNA surface and competing with antibiotic was 4. Electrostatic interactions between ions and RNA are not strong enough to prevent ions' escape. However, one should bear in mind that forces between ions and RNA as implemented within the BD method lack the terms describing specific short-range interactions and during our simulations mobility of magnesium ions is probably overestimated.

## Conclusions

We investigated the mechanism and kinetics of association of four aminoglycosides to the ribosomal A-site model using Brownian dynamics method. For the studied antibiotics, we established the order of association rates with the RNA fragment and calculated how these rates depend on ionic strength. The mechanism of one-dimensional diffusion of antibiotics near the RNA surface was also analyzed.

Association of antibiotics studied in this work is strongly electrostatically driven which results in large values of the computed association rates which are of the order of $10^{10}$ 1/M·s. Comparable rates are observed for neamine, which is composed of two pseudosugar rings and bears a total charge of +4 e, ribostamycin, which carries an equal total charge and whose structure differs from that of neamine by an additional ring, and composed of four rings neomycin (+6 e). Smaller rates are obtained for paromomycin, composed of four rings and possessing a similar diffusion

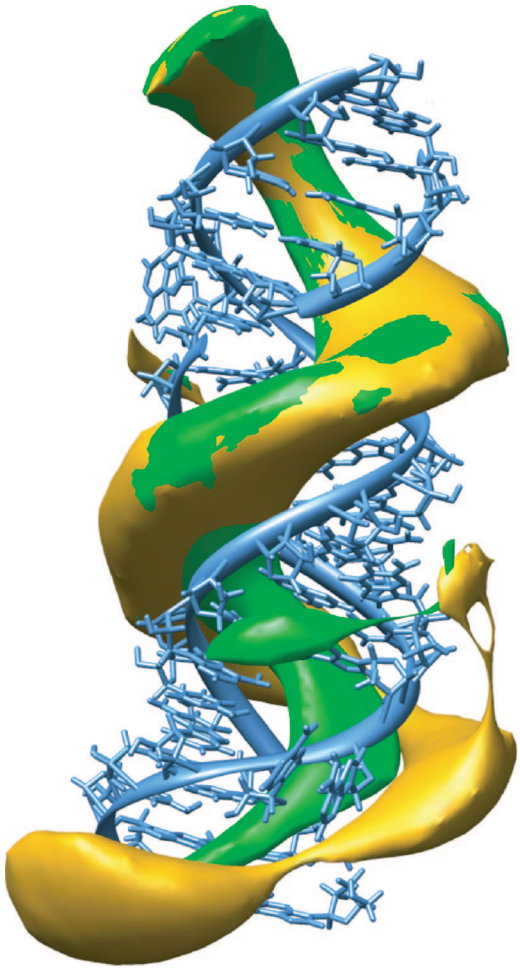

***Figure 8.*** BD derived density maps of paromomycin molecules around the RNA fragment. (green) Constant density surface generated when both binding pockets were empty. (yellow) Constant density surface when one binding pocket (lower half of the picture) was permanently occupied by the second antibiotic. Surfaces are constructed from points describing the positions of the geometric center of the ligand.

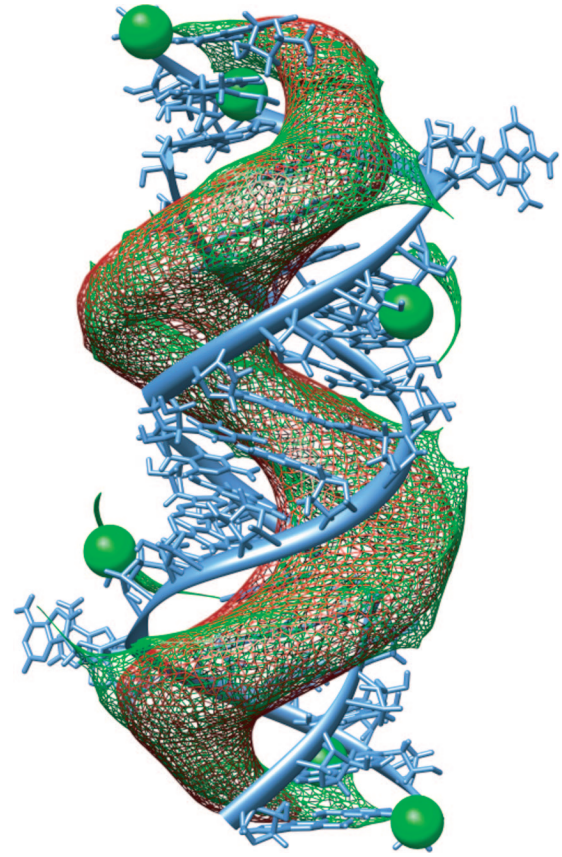

***Figure 9.*** Density map representing the preferred positions of $Mg^{2+}$ ions and paromomycin near the RNA fragment constructed based on BD simulations with mobile $Mg^{2+}$ ions. (green) Positions of $Mg^{2+}$ ions (for comparison, initial postions of ions are shown as spheres). (red) Positions of antibiotic.

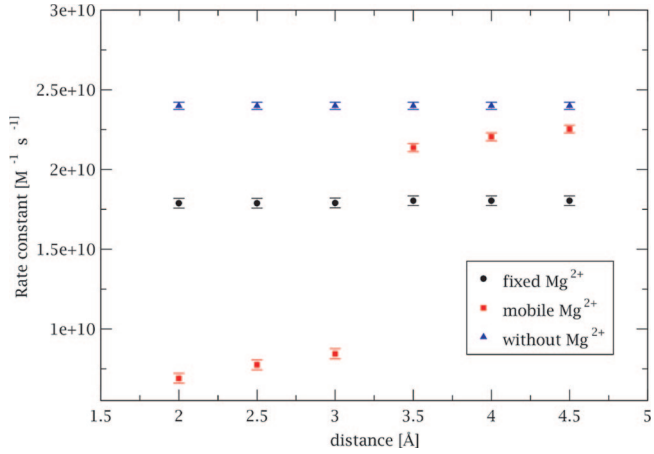


***Figure 10.*** Dependence of association rate constant on the definition of the reaction distance based on the simulation of paromomycin at 150 mM ionic strength and with different treatment of $Mg^{2+}$ ions.

coefficient to neomycin, but bearing a total charge lower than that of neomycin (+5 e). For all antibiotics, a weak dependence of rates on the ionic strength is observed.

The applied BD methodology is not free of limitations and some of them were addressed in this study. First of all, the oligonucleotide and encounter complexes are represented with a single conformation. Also, the internal mobility of ligands is not explicitly taken into account. However, we chose the RNA configuration with two adenines flipped out of the binding bulge and "ready" for the incoming antibiotic to avoid steric restrictions for binding in the A-site. Treating flexibility of associating molecules explicitly would allow the description of the formation of the encounter complex in a more realistic way but may not necessarily influence the rates of association. Antibiotics are represented as sets of beads with partial charges assigned to their geometric centers—this is a simplified model which is not able to reproduce the effects arising from the distribution of partial charges and their possible influence on the observed kinetics of binding. We, however, distribute the net charge of each bead in a way which mimics the charge distribution of each aminoglycosidic ring. Effects of charge desolvation and their influence on the computed rates are also not taken into account in our simulations; these effects were studied for the case of protein–protein association with the application of effective charges model.[79,80] However, validity of this model is questionable in the case of nonlinear Poisson–Bolt-

zmann equation and it should be carefully examined before application to highly charged systems, such as the studied RNA fragment. On the other hand, a proper model to be used to generate the electrostatic potential in case of highly charged RNA should be the nonlinear Poisson–Boltzmann equation. Charge desolvation for RNA/DNA systems is an interesting problem itself, but it was out of the scope of the present study. We were mostly interested in the relative order of association rates and the mechanism of association itself; therefore, we believe that despite the limitations the applied model serves well for this purpose and was able to give estimates of the association rates and explain the mechanism of aminoglycoside diffusion toward the RNA together with a possible role of $Mg^{2+}$ ions.

The kinetics of the RNA oligonucleotide/antibiotic complex formation has not been studied previously neither with experimental nor computational methods and we were not able to validate or compare our results with those obtained with another method. However, we have shown the details of the association mechanism, e.g. sliding of antibiotics along the RNA, the role of cations upon association, which may become helpful for designing or narrowing a group of compounds targeting the bacterial ribosome.

**Acknowledgment.** The authors acknowledge support from University of Warsaw (115/30/E-343/S/2007/ICM BST 1255), Polish Ministry of Science and Higher Education (3 T11F 005 30, 2006–2008), Fogarty International Center (NIH Research Grant no R03 TW07318), and Foundation for Polish Science. ICM University of Warsaw (grant no. G-31-4). M.D. wishes to thank Dr. Razif Gabdoulline for helpful disscusion.

**Supporting Information Available:** Animations from BD trajectories showing two kinds of behavior of aminoglycosides near the A-site oligonucleotide, i.e., sliding along the RNA groove between backbone phosphate groups and trapping inside the RNA groove and an animation showing interactions between mobile magnesium ions and paromomycin together with the displacement of $Mg^{2+}$ ions upon binding of antibiotic in the A-site RNA. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Walter, F.; Vincens, Q.; Westhof, E. *Curr. Opin. Chem. Biol.* **1999**, *3*, 694–704.

(2) Davies, J.; Gorini, L.; Davis, B. D. *Mol. Pharmacol.* **1965**, *1*, 93–106.

(3) Moazed, D.; Noller, H. F. *Nature* **1987**, *327*, 389–394.

(4) Karimi, R.; Ehrenberg, M. *Eur. J. Biochem.* **1994**, *226*, 355–360.

(5) Yoshizawa, S.; Fourmy, D.; Puglisi, J. D. *Science* **1999**, *285*, 1722–1725.

(6) Pape, T.; Wintermeyer, W.; Rodnina, M. V. *Nat. Struct. Biol.* **2000**, *7*, 104–107.

(7) Carter, A. P.; Clemons, M. W.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. *Nature* **2000**, *407*, 340–348.

(8) Vicens, Q.; Westhof, E. *Chembiochem.* **2003**, *4*, 1018–1023.

(9) Kaul, M.; Barbieri, C. M.; Kerrigan, J. E.; Pilch, D. S. J. *Mol. Biol.* **2003**, *326*, 1373–1387.

(10) Wang, H.; Tor, Y. *J. Am. Chem. Soc.* **1997**, *119*, 8734–8735.

(11) Hermann, T.; Westhof, E. *J. Med. Chem.* **1999**, *42*, 1250–1261.

(12) Yang, G.; Trylska, J.; Tor, Y.; McCammon, J. A. *J. Med. Chem.* **2006**, *49*, 5478–5490.

(13) Purohit, P.; Stern, S. *Nature* **1994**, *370*, 659–662.

(14) Miyaguchi, H.; Narita, H.; Sakamoto, K.; Yokoyama, S. *Nucleic Acids Res.* **1996**, *24*, 3700–3706.

(15) Recht, M. I.; Fourmy, D.; Blanchard, S. C.; Dahlquist, K. D.; Puglisi, J. D. *J. Mol. Biol.* **1996**, *262*, 421–436.

(16) Blanchard, S. C.; Fourmy, D.; Eason, R. G.; Puglisi, J. D. *Biochemistry* **1998**, *37*, 7716–7724.

(17) Vicens, Q.; Westhof, E. *Structure* **2001**, *9*, 647–658.

(18) Vicens, Q.; Westhof, E. *Chem. Biol.* **2002**, *9*, 747–755.

(19) Vicens, Q.; Westhof, E. J. *Mol. Biol.* **2003**, *326*, 1175–1188.

(20) Ogle, J. M.; Murphy, F. V.; Tarry, M. J.; Ramakrishnan, V. *Cell* **2002**, *111*, 721–732.

(21) Ogle, J. M.; Carter, A. P.; Ramakrishnan, V. *Trends Biochem. Sci.* **2003**, *28*, 259–266.

(22) Ogle, J. M.; Brodersen, D. E.; Clemons, W. M., Jr.; Tarry, M. J.; Carter, A. P.; Ramakrishnan, V. *Science* **2001**, *292*, 897–902.

(23) Kaul, M. *J. Am. Chem. Soc.* **2004**, *126*, 3447–3453.

(24) Shandrick, S.; Zhao, Q.; Han, Q.; Ayida, B. K.; Takahashi, M.; Winters, G. C.; Simonsen, K. B.; Vourloumis, D.; Hermann, T. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 3177–3182.

(25) Griffey, R. H.; Hofstadler, S. A.; Sannes-Lowery, K. A.; Ecker, D. J.; Crooke, S. T. *Proc. Natl. Sci. U.S.A* **1999**, *96*, 10129–10133.

(26) Sucheck, S. J.; Wong, A. L.; Koeller, K. M.; Boehr, D. D.; Draker, K.; Sears, P.; Wright, G. D.; Wong, C.-H. *J. Am. Chem. Soc.* **2000**, *122*, 5230–5231.

(27) Ryu, D.-H.; Rando, R. R. *Bioorg. Med. Chem.* **2001**, *9*, 2601–2608.

(28) Wong, C.-H.; Hendrix, M.; Priestley, E. S.; Greenberg, W. A. *Chem. Biol.* **1998**, *5*, 397–406.

(29) Pilch, D. S.; Kaul, M.; Barbieri, C. M.; Kerrigan, J. E. *Biopolymers* **2003**, *70*, 58–79.

(30) Pfister, P.; Hobbie, S.; Brull, C.; Corti, N.; Vassela, A.; Westhof, E.; Bottger, E. C. *J. Mol. Biol.* **2005**, *346*, 467–475.

(31) Vaiana, A. C.; Westhof, E.; Auffinger, P. *Biochimie* **2006**, *88*, 1061–1073.

(32) Ermak, D. L.; McCammon, J. A. *J. Chem. Phys.* **1978**, *69*, 1352–1360.

(33) Wade, R. C.; Gabdoulline, R. R. *Methods* **1998**, *3*, 329 341.

(34) Beard, D.; Schlick, T. *J. Chem. Phys.* **2000**, *112*, 7313–7322.

(35) Tan, R. C.; Truong, T. T.; McCammon, J. A.; Sussman, J. L. *Biochemistry* **1993**, *32*, 401–403.

(36) Wade, R. C.; Gabdoulline, R. R.; Lüdemann, S. K.; Lounnas, V. *Proc. Natl. Acad. Sci. U.S.A* **2007**, *95*, 5942–5949.

(37) Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **1997**, *72*, 1917–1929.

(38) Speer, A.; Dammer, Ch.; Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **2006**, *90*, 1913–1924.

(39) Das, A.; Jayaram, B. J. *Mol. Liquids* **1998**, *77*, 157–163.

(40) Gilson, M. K.; Honig, B. *Proteins: Struct., Func., Genet.* **1988**, *4*, 7–18.

(41) Zhou, H. *J. Chem. Phys.* **1994**, *100*, 3152–3162.

(42) Northrup, S. H.; Hynes, J. T. *J. Chem. Phys.* **1979**, *71*, 871–883.

(43) Northrup, S. H.; Allison, S. A.; McCammon, J. A. *J. Chem. Phys.* **1984**, *80*, 1517–1524.

(44) Einstein, A. *Ann. Phys. (Leipzig)* **1905**, *17*, 549–560.

(45) Smoluchowski, M. V. *Ann. Phys. (Leipzig)* **1906**, *21*, 756–760.

(46) Madura, J. D.; Davis, M. E.; Gilson, M. K.; Wade, R. C.; Luty, B. A.; McCammon, J. A. Biological applications of electrostatic calculations and Brownian dynamics simulations. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1994; Vol. 5, pp 229–267..

(47) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochemistry* **1996**, *35*, 7819–7833.

(48) van Vlijmen, H. H. T.; Schaefer, M.; Karplus, M. *Proteins: Struct., Func., Genet.* **1998**, *33*, 145–148.

(49) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415–436.

(50) You, T. J.; Bashford, D. *Biophys. J.* **1995**, *69*, 1721–1733.

(51) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435–445.

(52) Holst, M.; Baker, N.; Wang, E. *J. Comput. Chem.* **2000**, *21*, 1319–1342.

(53) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187–197.

(54) Smoluchowski, M. V. *Phys. Z.* **1916**, *17*, 557–571.

(55) Pyle, A. M. *J. Biol. Inorg. Chem.* **2002**, *7*, 679–690.

(56) DeRose, V. J. *Curr. Opin. Struct. Biol.* **2003**, *13*, 317–324.

(57) Draper, D. E. *RNA* **2004**, *10*, 335–343.

(58) Hermann, T.; Westhof, E. *J. Mol. Biol.* **1998**, *276*, 903–912.

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

(60) *Insight II*, version 2000; Accelrys: San Diego, CA, 2000.

(61) Jakalian, A.; Busch, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146.

(62) Barbieri, C. M.; Srinivasan, A. R.; Pilch, D. S. *J. Am. Chem. Soc.* **2004**, *126*, 14380–14388.

(63) Carrasco, B.; de la Torre, J. G. *Biophys. J.* **1999**, *76*, 3044–3057.

(64) de la Torre, J. G. *Biophys. Chem.* **2001**, *94*, 265–274.

(65) de Graaf, R. A.; van Kranenburg, A.; Nicolay, K. *Biophys. J.* **2000**, *78*, 1657–1664.

(66) Antosiewicz, J. *Biophys. J.* **1995**, *69*, 1344–1354.

(67) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176.

(68) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1988**, *9*, 327–335.

(69) Berg, O. G.; Blomberg, C. *Biophys. Chem.* **1977**, *4*, 367–381.

(70) Berg, O. G.; Blomberg, C. *Biophys. Chem.* **1977**, *7*, 33–39.

(71) Sheinerman, S. H.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153–159.

(72) Quinn, P. M. *Chem. Rev.* **1987**, *87*, 955–979.

(73) Schreiber, G.; Fersht, A. R. *Biochemistry* **1993**, *32*, 5145–5150.

(74) Escobar, L.; Root, M. J.; MacKinnon, R. *Biochemstry* **1993**, *32*, 6982–6987.

(75) Wallis, R.; Moore, G. R.; James, R.; Kleanthous, C. *Biochemistry* **1995**, *34*, 13743–13750.

(76) Wendt, H.; Leder, L.; Harma, H.; Jelesarov, I.; Baici, A.; Bosshard, H. R. *Biochemistry* **1997**, *36*, 204–213.

(77) Radic, Z.; Kirchhoff, P. D.; Quinn, D. M.; McCammon, J. A.; Taylor, P. *J. Biol. Chem.* **1997**, *272*, 23265–23277.

(78) François, B.; Russel, R. J. M.; Murray, J. B.; Aboul-ela, F.; Masquida, B.; Vicens, Q.; Westhof, E. *Nucl. Acid Res.* **2005**, *33*, 5677–5690.

(79) Gabdoulline, R. R.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 3868–3878.

(80) Gabdoulline, R. R.; Wade, R. C. *J. Mol. Biol.* **2001**, *306*, 1139–1155.

CT700210N

JCTC Journal of Chemical Theory and Computation

# Accelerated Superposition State Molecular Dynamics for Condensed Phase Systems

Michele Ceotto

*Dipartimento di Chimica Fisica ed Elettrochimica, Università degli Studi di Milano, via Golgi 19, 20133 Milano, Italy*

Gary S. Ayton and Gregory A. Voth*

*Department of Chemistry and Center for Biophysical Modeling and Simulation, University of Utah, Salt Lake City, Utah 84112*

Received December 1, 2007

**Abstract:** An extension of superposition state molecular dynamics (SSMD) [Venkatnathan and Voth *J. Chem. Theory Comput.* **2005**, *1*, 36] is presented with the goal to accelerate timescales and enable the study of "long-time" phenomena for condensed phase systems. It does not require any a priori knowledge about final and transition state configurations, or specific topologies. The system is induced to explore new configurations by virtue of a fictitious (free-particle-like) accelerating potential. The acceleration method can be applied to all degrees of freedom in the system and can be applied to condensed phases and fluids.

## I. Introduction

Molecular dynamics (MD) simulation models the time evolution of a given atomistic-level system by integrating Newton's equations of motion.[1–3] Extensions to MD employ dynamics described by a Langevin equation[4–6] or variations of the previous differential equations with additional parameters (e.g., friction coefficients) or degrees of freedom (e.g., Nosé−Hoover thermostats[1,7–9]). The integration of the underlying equations of motion are limited in time; in fact, most MD simulations are far too limited in duration to examine many important biomolecular processes which can occur on timescales longer than tens of nanoseconds. For example, lateral diffusion in lipid bilayers occurs on the second scale,[10] while protein folding occurs on the millisecond scale.[11–15] On the other hand, relying on the rapid growth in computer technology employing a "brute force" MD approach is also not feasible as a speedup of 6 orders of magnitude will be required in order to access the relevant timescales.

In some complex systems, the multiple time- and length-scales can lead to the so-called effect of "broken ergodicity", where mechanical observables (e.g., internal energy, pressure) calculated via accumulated time averages differ (sometimes greatly) from the ensemble average (where all points in phase space are considered). In other words, the timescale used for the measurement is much shorter than the actual relaxation time for the system. As a result, in this particular case, the MD simulation performs a sum of subaverages of isolated phase space subsets and misses some others. As an example of an approach to overcome these issues, Andricioaei and Straub[16] have introduced new generalized Monte Carlo (MC) and MD algorithms inspired by Tsallis statistics ($q$-jumping) and later, "smart walking" MC.[17] Other approaches include the multicanonical algorithm,[18,19] where extensive macrovariables are added, simulated tempering,[20,21] and replica exchange (REX),[22–25] where intensive thermodynamic state quantities (pressure, temperature, chemical potentials, etc.) are varied. All the above methods have been implemented both in MC and MD simulation algorithms. A main limitation involves the number of replicas; these can grow unmanageably large when many macrostates are required for the simulation to satisfy ergodicity. Fenwick et al.[26] have combined MC replica methods with biased force-field parameters in order to directly modulate the specific molecular interaction responsible for the kinetic traps and, in the same spirit, others[27,28] have modified force-field

* Corresponding author. E-mail: voth@chem.utah.edu.

10.1021/ct7003275 CCC: $40.75 
Published on Web 03/04/2008

parameters (except those related to solvent−solvent interactions) in an MD protocol.

Another approach for timescale acceleration is the parallel replica method (PRD),[29–31] where the system is replicated in parallel and independent MD trajectories are generated via different initial velocity distributions. Whenever a successful trajectory is obtained, all processors are stopped. This state is then replicated over all processors, and the whole process is restarted. A further implementation of this method is the parallel sequential synchronization (PSS)[32] which shows how PRD is easy to combine with other techniques.

To accelerate timescales and enable the study of "long-time" phenomena within the MD framework, one can perform simulations at higher temperatures. Voter[33] has developed, and later improved,[29] a method called temperature accelerated dynamics (TAD) which raises the temperature and corrects for this bias by filtering out transitions that would not occur at the original temperature. A completely different approach for "rare events" dynamics (especially for passages over high barriers) employs accelerating potentials.[34–45] Here, the potential energy surface is modified for a small set of degrees of freedom and the original state is recovered on-the-fly by means of non-Boltzmann weights. Conformational flooding[46] first selects a subconformational space and then destabilizes the initial conformation and consequently lowers the free energy barrier of structural transitions. Similarly, the hyperdynamics method[29,35,36] focuses on infrequent transition events from one potential energy basin to another and then constructs a bias potential such that the original potential energy surface changes are done without affecting the transition state regions where the rate is calculated using the harmonic limit of transition state theory. Significant boosts have been found for surface diffusion dynamics;[29,34–36,42] however, for more complex systems, the construction of such a bias potential is not trivial and may not solve the low barrier problem when the system is trapped by a set of states connected by low barriers. A simpler recipe is one offered by Tully and co-workers,[37,41] where the bias potential is chosen so that the system evolves in a flat "puddle region" instead of sinking into a local minimum. This method has been applied to dihedral degrees of freedom in small peptide dynamics[37–39,41,47] and has been implemented in MC simulation with a bias in momentum space.[37,48,49] Recent work has employed a "boost potential" to modify the original potentials that govern the system.[44,50–52] The scheme raises the wells depths in a continuous manner and has been successfully applied, for example, to alanine dipeptide in an explicit solvent. In conjunction with a quasiharmonic analysis, this approach has been used to calculate the entropy for an eight residue peptide in explicit water.[44]

A different solution to the direct kinetic dynamics of rare events is transition path sampling.[53–55] In this method reactants and products are known a priori, and path ensembles between these states are generated by constructing a random walk in path space with a MC algorithm.

In solid-state simulations, a popular choice to accelerate MD is feature activated molecular dynamics (FAMD),[56] which creates localized regions around a defect atom or cluster (active region) sites. A full MD simulation is subsequently performed while the rest of the solid lattice is kept static. This is done by the use of a thermal activity function, which decays to zero in a sigmoid fashion once out of the active region.

Although the above accelerating potential procedures partially removes the problem of broken ergodicity and accelerates the crossing of barriers, it is not suitable for calculating equilibrium thermodynamic properties as it undersamples low-energy states. Self-guided molecular dynamics approaches[57–61] can avoid this problem and can be implemented without any a priori knowledge of the system; rather they are based on the cumulative history of a system's trajectory. In particular, when this idea is combined with replica exchange,[61] the copies are self-regulating and compete during the simulation to overcome to any undersampled region. Along the same line of an adaptive algorithm, Laio and Parrinello[62–65] introduced the so-called "metadynamics" method in which a history-dependent potential, given as the sum of Gaussians centered on the trajectory in a reduced "collective variable" space, fills the free energy surfaces minima and drives the systems to explore new wells.

In condensed phases, the complexity of the system is such that it is not possible to isolate a priori a subset of degrees of freedom responsible for the long-time properties. For example, a simple picture of a double well coupled to a bath with a starting and ending configuration is, most of the time, an inadequate description of the system. Consequently, a state-to-state transition model with a transition state configuration, which follows a first order kinetic picture, is often out of the question. An example problem that enters into this category is the bilateral diffusion of phospholipids in membranes; the possible configurations are so numerous that trying to isolate any subsets of degrees of freedom will mostly turn out to be impossible, if not counterproductive. In fact, the transition state that separates two multidimensional basins in a double well picture is misleading for such systems since no specific event and reaction subset of coordinates can be identified. To deal with these issues, a more generalized version of enhancing sampling must be adopted, which is not restricted to first order kinetics. Instead, the system should enchance its ergodic properties without any specific instructions or a priori topological constraints, and should be left to explore new configurations in an unrestricted fashion.

In light of the above considerations, the goal of the present work is to develop an approach within the MD framework that can deal with all the degrees of freedom of complex condensed phase systems, for example fluids, by not requiring any a priori knowledge of the system (such as the final state), and by guaranteeing a full recovery of the unbiased potential statistics. The starting point of the present work is the superposition state molecular dynamics (SSMD) method.[45] The main idea underlying SSMD is to create fictitious potentials whose dynamics are accelerated with respect to the original (physical) one. These potentials are built up as a superposition of the upper and lower states with respect the physical one and they are coupled in such a way that the overall dynamics smoothly and continuously switch

from one potential surface to another to achieve acceleration. The coupling follows the superposition state rules of quantum mechanics, and the construction of the fictitious potential and the location of the coupling term are crucial aspects of this approach.

In previous work,[45] SSMD was developed and applied to a one-dimensional rough potential energy landscape and it was proven to enhance ergodic behavior. Here, it is extended to a multidimensional potential energy landscape and is applied to the case of an isotropic fluid. An isotropic fluid is an example of the kind of complex system described previously as all degrees of freedom equally contribute to the ergodic behavior of the system.

The remaining sections of this paper are organized as follows: section II reviews the sampling techniques utilized in this work, while in section III the extension of superposition state molecular dynamics (SSMD) to condensed phase simulations is introduced. In section IV, SSMD is implemented for Coulombic interactions. Some closing comments are then given in section VI.

## II. Non-Boltzmann Sampling

The aim of this section is to review non-Boltzmann sampling and later show its limitation to condensed phase simulations. A system of $N$ particles with phase space coordinates $\mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_N)$ and $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$, where $\mathbf{p}_i$ is the momenta and $\mathbf{r}_i$ is the position, of particle $i$ is considered. The canonical phase space probability density is given by

$$P(\mathbf{r}, \mathbf{p}) = \frac{\mathrm{e}^{-\beta H(\mathbf{r},\mathbf{p})}}{Q} \tag{1}$$

where

$$Q = \iint \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{p}\, \mathrm{e}^{-\beta H(\mathbf{p},\mathbf{r})} \tag{2}$$

is proportional to the canonical partition function. Here,

$$H(\mathbf{r}, \mathbf{p}) = \sum_{i=1}^{N} \frac{|\mathbf{p}_i|^2}{2m} + V(\mathbf{r}) \tag{3}$$

is the classical Hamiltonian and $\beta = 1/k_B T$, where $k_B$ is Boltzmann's constant, $T$ is the temperature, and $V(\mathbf{r})$ is the total potential energy function. In terms of just the coordinate space, eq 1 can be expressed in the usual way

$$P(\mathbf{r}) = \frac{\mathrm{e}^{-\beta V(\mathbf{r})}}{\int \mathrm{d}\mathbf{r}\, \mathrm{e}^{-\beta V(\mathbf{r})}} \tag{4}$$

The physical meaning of the probability density $P(\mathbf{r})$ is that the canonical ensemble average of a position-dependent observable $O(\mathbf{r})$ is given by

$$\langle O \rangle = \int \mathrm{d}\mathbf{r}\, P(\mathbf{r}) O(\mathbf{r}) \tag{5}$$

An MD estimate of this integral is given when eq 5 is replaced by the following sum:

$$\langle O \rangle \cong \frac{1}{N_T} \sum_k O_k(\mathbf{r}) \tag{6}$$

where the $k$ index stands for the MD averaging step, and $N_T$ is the total number of averaging steps. If a different "accelerated" MD potential $V'(\mathbf{r})$ is used, then the observable average in the original canonical ensemble can still be obtained by rewriting eq 5 in the following way,

$$\langle O \rangle = \frac{\int \mathrm{d}\mathbf{r}\, \mathrm{e}^{-\beta V(\mathbf{r})} O(\mathbf{r})}{\int \mathrm{d}\mathbf{r}\, \mathrm{e}^{-\beta V(\mathbf{r})}} = \frac{\int \mathrm{d}\mathbf{r}\, \mathrm{e}^{-\beta V'(\mathbf{r})} O(\mathbf{r}) \mathrm{e}^{-\beta (V(\mathbf{r}) - V'(\mathbf{r}))}}{\int \mathrm{d}\mathbf{r} \mathrm{e}^{-\beta V'(\mathbf{r})} \mathrm{e}^{-\beta (V(\mathbf{r}) - V'(\mathbf{r}))}} =$$

$$\frac{\langle O(\mathbf{r}) \mathrm{e}^{-\beta \Delta V(\mathbf{r})} \rangle}{\langle \mathrm{e}^{-\beta \Delta V(\mathbf{r})} \rangle} \tag{7}$$

or by the equivalent sum from the MD simulation as the potential $V'(\mathbf{r})$:

$$\langle O \rangle \cong \frac{\sum_k O_k(\mathbf{r}) \mathrm{e}^{-\beta \Delta V_k(\mathbf{r})}}{\sum_k \mathrm{e}^{-\beta \Delta V_k(\mathbf{r})}} \tag{8}$$

where $\Delta V_k(\mathbf{r}) = V_k(\mathbf{r}) - V_k'(\mathbf{r})$ is the difference between the original and the accelerated potential at the average step $k$ and the dynamics is carried out on the new potential $V'(\mathbf{r})$. This approach has been in used for many decades now, both in MC and MD calculations. In essence, most of the methods described in the Introduction can be reduced to this essential idea. It has the significant advantage that all the terms, both in the numerator and denominator, of eq 8, are reweighted on-the-fly in the MD simulation.

Recall that the overall aim of the present work is to develop an acceleration scheme that is capable of treating condensed phase systems. As such, a first test employs a benchmark system and consisted of a Lennard-Jones fluid of $N = 1372$ particles characterized by the pair-interaction potential

$$V(r_{ij}) = 4\varepsilon \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{9}$$

where $r_{ij} = |r_i - r_j|$ is the interparticle separation, $\sigma$ is the fundamental unit of distance, and $\varepsilon$ is the fundamental unit of energy. The fundamental unit of energy $\varepsilon$, can be expressed as $\varepsilon = \gamma \varepsilon_c$, where $\gamma$ is a dimensionless scaling parameter and $\varepsilon_c$ is a constant. The same is valid for the fundamental unit of distance $\sigma$, i.e., $\sigma = \gamma_\sigma \sigma_c$. In all simulations $N\sigma_c^3/L^3 = 0.8$, the mass, $m$, is unity, and $k_B T/\varepsilon_c = 0.8$.

In a straightforward application of eq 8, an accelerated dynamics can be directly employed by using a new potential $V'(\mathbf{r})$ with $\gamma = 0.5$ and is shown by the dashed line in panel a of Figure 1. All other parameters were kept constant. This dynamics was performed for *all degrees of freedom*, and it is equivalent to the dynamics of the original reduced potential ($\gamma = 1$) but at double the temperature. Simulations employed 500 000 time steps in the constant *NVT* ensemble.[7,8] The reduced time step was set to $\Delta t_c = 0.0025$, which is related to the unscaled time step by the relation $\Delta t = \sqrt{\varepsilon_c/m\sigma_c}\Delta t_c$. An examination of the radial distributions functions, as shown in panel b of the same figure (with the same line convention as in a), shows that the original radial distribution

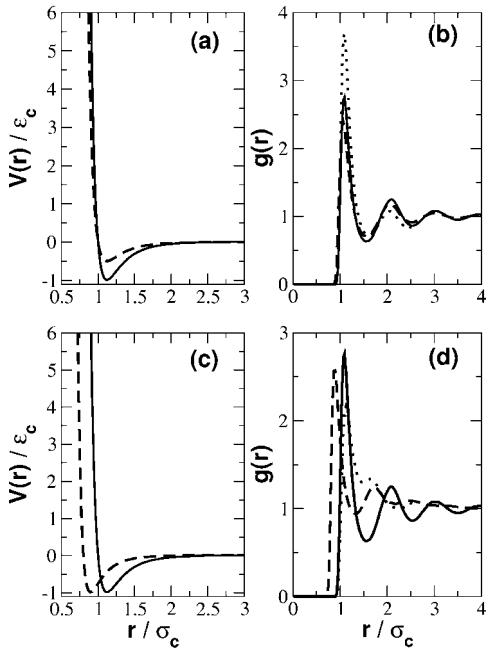


***Figure 1.*** Potentials and radial distribution functions. (a) Standard Lennard-Jones potential ($\gamma = 1$) in the solid line and for $\varepsilon = 0.5\varepsilon_c$ and $\sigma = \sigma_c$ in the dashed line. (b) Radial distribution function at a reduced temperature equal to $k_BT/\varepsilon_c = 0.8$ and reduced density $N\sigma_c^3/L^3 = 0.8$ after 500 000 time steps for the standard Lennard-Jones potential in the solid line, for the half-deep well in the dashed line, and the reweighted one in a dotted line as an estimate of the standard distribution function as obtained by eq 8. (c) Same as a but with $\varepsilon = 1$ and $\sigma = 0.8\sigma_c$ in the dashed line. (d) Same as b where the dashed line is used for the corresponding potential in panel c.

function (continuous line) is not recovered by rescaling on-the-fly according to eq 8. In this and in subsequent instances, any rescaled quantities found from a direct application of eq 8 will be denoted as "re-weighted". In fact, the reweighted radial distribution function (dotted line) largely overestimates the first peak of $g(\mathbf{r})$. The dynamics was also examined, where here the core radius was scaled to $\gamma_\sigma = 0.8$ (dashed line in panel c of Figure 1) and the reweighted radial distribution function was obtained (dotted line in panel d of the same figure). Here, the correct location of the first shell–core is observed, but the magnitude is off and the longer-ranged correlations are not accounted for correctly. Since non-Boltzmann sampling is exact regardless of the number of degrees of freedom, a much longer simulation will fully recover the original radial distribution function, but at the same time, any advantage due to the new accelerating potential is lost. In fact, if the same simulation of the dashed potential in panel a of Figure 1 is performed at a much higher temperature ($k_BT/\varepsilon_c = 4.0$) and lower density ($N\sigma_c^3/L^3 = 0.2$) (i.e., a gas phase), the original radial distribution function is recovered as shown in Figure 2 by the dotted lines, even if it slightly misses the small second solvation shell at $r_{ij} = 2.5\sigma_c$.

The above results indicate that the sources of errors are both topological and numerical in nature. They are topological because the potential landscape is multidimensional, and with a higher dimensionality, a larger number of sampling

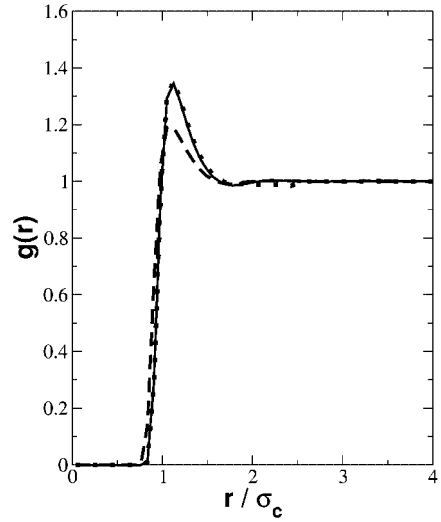


***Figure 2.*** Same plot as panel b of Figure 1 but with the simulation performed at a reduced temperature of $k_BT/\varepsilon_c = 4.0$ and a reduced density $N\sigma_c^3/L^3 = 0.2$.

events are necessary to yield converged statistics. They are numerical because at high temperatures $\beta$ is smaller, and the errors made in the exponential sum and its propagation in the exponentiation are accumulated.[43] Instead, at lower temperatures, a complex potential energy landscape may not be recovered by the reweighting expression in eq 8 in standard run lengths. This shows that an indiscriminate use of non-Boltzmann sampling techniques is not fruitful, even if these are in principle exact. This was also reported recently by de Oliveira et al.[51] for the case of a rigid water molecule condensed phase model, where only the oxygen−oxygen radial distribution function was considered. For this reason in section III, a still exact but at the same time practical and more efficient approach is presented.

## III. Superposition State Molecular Dynamics (SSMD) for Condensed Phase Simulations

In superposition states molecular dynamics (SSMD),[45] an accelerating potential $V'(\mathbf{r})$ is obtained by a superposition of "states", which include the original (physical) one characterized by the potential $V(\mathbf{r})$, along with fictitious states characterized by an additional potential, $V_f(\mathbf{r})$. In the case of a two-state combination with one fictitious and one real potential, the effective SSMD potentials can be extended to multidimensional systems by superimposing each pairwise interaction at a time in the way described above. The resulting effective potential is derived by (analytically) diagonalizing a two-state matrix as

$$V'(r_{ij}) = \frac{V(r_{ij}) + V_f(r_{ij})}{2} \pm \frac{1}{2}\sqrt{4V_{12}^{\,2}(r_{ij}) + (V(r_{ij}) - V_f(r_{ij}))^2} \tag{10}$$

where $V_f(r_{ij})$ is the fictitious (not physical and in this case) pairwise potential, $V_{12}(r_{ij})$ is the coupling, and $V'(r_{ij})$ can be given both by the sum (the highest eigenenergy) or the difference (the lowest eigenenergy) in eq 10. In this paper, the lowest eigenenergy will always be used to perform the dynamics and the fictitious and coupling potentials will be varied.

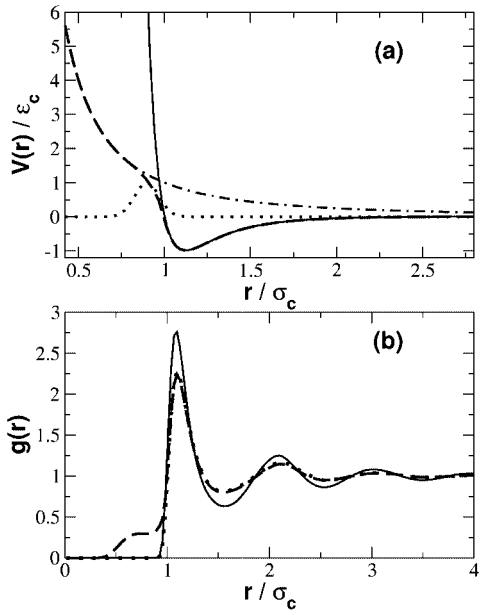


***Figure 3.*** SSMD potential and radial distribution function at the same temperature and density as in Figure 1. (a) Standard Lennard-Jones potential in the solid line; upper fictitious state potential in the dashed-dot line; coupling potential in the dotted line; and the resulting SSMD potential in the dashed line. (b) Standard Lennard-Jones fluid radial distribution function in the solid line; SSMD potential radial distribution function in the dashed line; and the reweighted one in the dotted line. Note that the reweighting is exact on the left side of the first solvation shell peak.

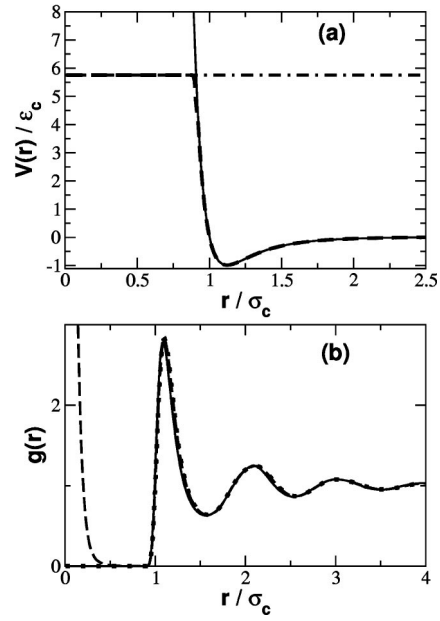


***Figure 4.*** Potential and radial distribution functions at the same temperature and density as in Figure 1. (a) Original Lennard-Jones potential in the solid line; the upper fictitious SSMD state $V'(r_{ij}) = 5.75\varepsilon_c$ in the dashed-dot line; and the resulting SSMD potential in the dashed line. (b) Original radial distribution function in the solid line; SSMD potential radial distribution function in the dashed line; and reweighted radial distribution line in the dotted line. The reweighting fully recovers the original distribution function.

However, one also has the flexibility to use the upper eigenenergy or to employ more fictitious states in an *N*-state SSMD scheme.

As an example, the original LJ potential $V(r_{ij})$ was plotted in panel a of Figure 3 with a continuous line, whereas the fictitious potential (shown as the dot-dashed line) is given by

$$V_f(r_{ij}) = \varepsilon\left(\left(\frac{\sigma}{r_{ij}}\right)^2 + \delta\right) \tag{11}$$

where the arbitrary constant is $\delta = 4$, and with the dotted line, the coupling potential is given by

$$V_{12}(r_{ij}) = V_{12}(r_0)e^{-\alpha(r_{ij} - r_0)^2} \tag{12}$$

where $V_{12}(r_0) = \varepsilon$, $\alpha = (10/\sigma)^2$, and $r_0$ is the pairwise length where the switching between the real and the fictitious potentials occurs. The resulting SSMD potential $V'(r_{ij})$ is plotted in Figure 3 with a dashed line and it clearly follows the original potential up to the beginning of the repulsive wall. The corresponding radial distribution function (dashed line), the reweighted (dotted line) and the exact one (continuous line) are shown on panel b of Figure 3. It can be seen how the original one is completely recovered on the repulsive wall side, but it does not accurately describe the long-range correlation part (rdf structure).

In light of this result, the avoided crossing between the fictitious and the real potential located at $r_0$ was pushed deeper into the repulsive region than the value shown in Figure 3. Then, in order to achieve an eventual acceleration for the entire fluid (i.e., for all degrees of freedom) at the same time, the fictitious potential was chosen to be a free particlelike one as shown in panel a of Figure 4; in other words, the interaction is such that if two particles come closer than the critical distance $r_0$, then their pairwise potential gradually flattened to a constant, while other interactions with the rest of the fluid remain unchanged. This choice exactly recovers the original radial distribution function as shown by the dotted line in panel b of Figure 4, and it was verified to be independent of system size, total simulation time, and choice of time step. When the distance between the particles is $\mathbf{r} < r_0$, the accelerated fluid contains an additional amount of kinetic energy as compared with the ordinary LJ fluid. This amount is the difference between the potential cut off and the original potential at $\mathbf{r}$, and it is the origin of the resulting acceleration. However, this kinetic surplus is quickly redistributed between all degrees of freedom via the thermostat, bath interactions, or particle collisions. For the choice of potential shown in dashed line on panel a of Figure 4, an increase in the diffusion constant of the order of 25–30% was achieved.

A rough evaluation of the area under the radial distribution function for $\mathbf{r} < r_0$ in Figure 4 gives the number of particles in this region after being correctly normalized for spherical symmetry. This number is never larger than 10% of the total number of particles, and typical values for $\langle e^{-\beta\Delta V}\rangle$ (the denominator of eq 8) were never being observed to be smaller than 0.98. Interestingly, the velocity autocorrelation function (not reported here) reaches the asymptotic zero value plateau in a smoother way and in half of the time with respect to the original one, suggesting that the dynamics is more

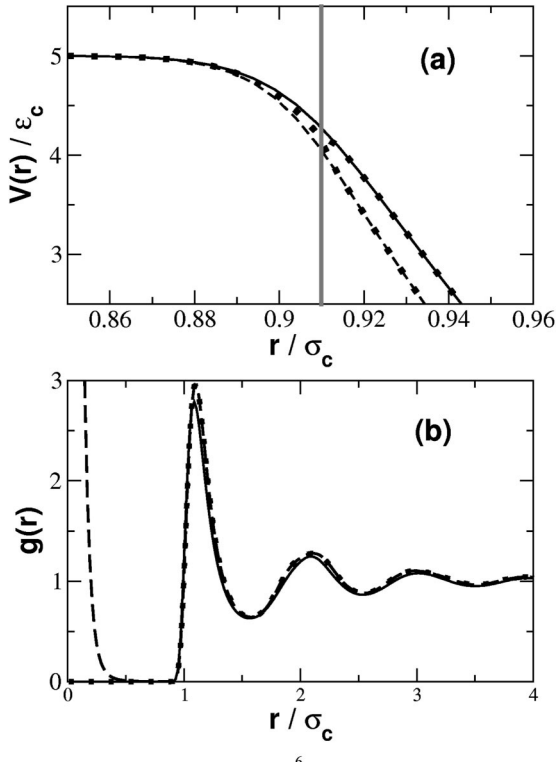


***Figure 5.*** Potential and radial distribution function of a charged electroneutral Lennard-Jones fluid. (a) Gray vertical line indicating the cutoff radius and delimiting the accelerating region to the left; SSMD potential for attractive and repulsive interactions in the dashed and solid line, respectively; the same potential as before but switching off the charge interactions for $r_{ij} < r_0$. Note the overlap between the repulsive and attractive curves in this case once $r_{ij}$ is smaller that $r_0$. (b) Same plot as panel b of Figure 4.

diffusive and less inertial. This is shown also in panel a of Figure 7, where an example of the mean square displacement is reported. The advantage of the present method is that the key parameter involves varying the value of the fictitious potential cutoff height. The disadvantages of this approach are related to the fact that the particles will remain in the $\mathbf{r} < r_0$ region if the potential cutoff is too low, which could alter the radial distribution function and decrease the diffusion constant.

## IV. Implementation for Coulombic Interactions

The next level of complexity explored in this work was to include long-range electrostatics into the underlying model. Point charges were added to the previously described Lennard-Jones fluid such that the whole system was electroneutral. The total potential acting on each pair of particles was then given by the sum of a Coulombic and a Lennard-Jones interaction. A standard switching function (which modulates the transition from a given asymptotic value to zero) was used for cutting off the long-range Coulomb interaction. A direct application of eq 10 gives the attractive (dashed) and repulsive (solid) SSMD potentials. The details of these at the crossing region are shown in panel a of Figure 5. The dynamics for this resulted in pairs of particles

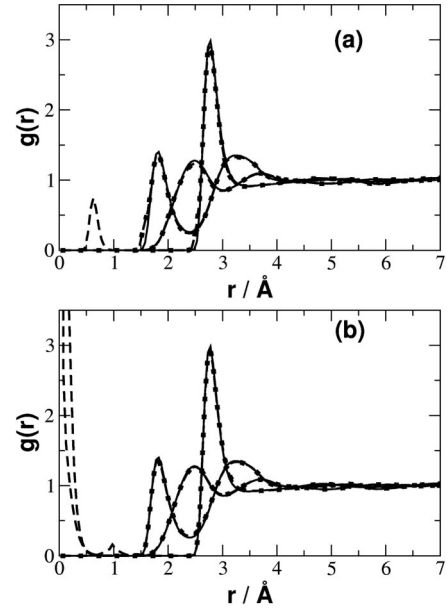


***Figure 6.*** Radial distribution functions for the oxygen−oxygen interaction (curve with the highest peak), oxygen−hydrogen interaction (curve closer in), and hydrogen−hydrogen interaction at $T = 298$ K for a simulation of 256 water molecules and 250 ps time length. (a and b) Original distributions in solid lines; SSMD distributions in dashed lines; and reweighted distributions in the dotted line. (a) Pairwise SSMD procedure for each atoms at a time (see main text). (b) Fictitious potential switched on for the entire molecular pairwise interactions whenever a pair of atoms of these molecules are in a distance less than $r_0$.

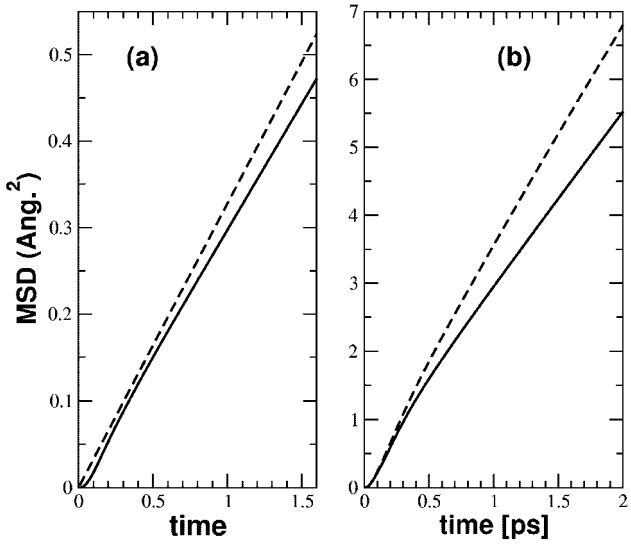


***Figure 7.*** Mean square displacements for a Lennard-Jones fluid (a) and for liquid water (b). Accelerated results are shown in dashed lines. Notice that for the LJ fluid the inertial transition is reduced upon acceleration. In this plot, the self-diffusion coefficient shows a percentage increment of 20% and 25%, respectively, for the LJ fluid and for water.

remaining bound within the $\mathbf{r} < r_0$ region. Oppositely charged particles were strongly attracted to each other. Thus, by switching off their charges once their distance is $\mathbf{r} < r_0$, their pair interaction is the same as in the standard LJ fluid. This last choice is represented by the dotted potential in the

same figure where two dotted lines for repulsive and attractive interactions are seen for $\mathbf{r} > r_0$, and one dotted line for $\mathbf{r}$ smaller than the critical distance. The corresponding radial distribution functions, displayed on the bottom panel, are once again fully recovered and with the result that the fluid dynamics is accelerated.

The next system studied selected involved 256 water molecules at $T = 298$ K in a cubic box with periodic boundary conditions and density equal to 1000 kg/m$^3$. The potentials were modeled according to the flexible TIP3P force field.[66] A typical simulation length ranged from 150 to 250 ps. SSMD was implemented in DL_POLY 2.14[67] and the time-step was equal to 0.5 fs. As also noted for the LJ fluid, the acceleration increment does not change with time-step size and total simulation time (as long as the simulation results are converged). The long-range Coulombic interactions are calculated employing both the reaction field method (with a dielectric constant equal to 78) and the Ewald summation method. The explicit implementation of this last method for SSMD is reported in the Appendix.

Two acceleration approaches were employed: an atomic and a molecular approach. In the atomic approach, each atom within the water molecule was subjected to the acceleration cutoff. In the molecular approach, the entire water molecule was treated as a single site (analogous to the previous LJ simulations).

In the first approach, each intermolecular pairwise potential was cut off in the same way as was done for the Lennard-Jones fluid example both for attraction and repulsion. In this case, hydrogen−hydrogen interactions take most advantage of the potential cut off, simply because of the smaller van der Waals radius of hydrogen−hydrogen interaction. Nevertheless, once the hydrogen atoms were closer than $\mathbf{r} < r_0$, the oxygen atoms were still out of the $\mathbf{r} < r_0$ region and thus their strong repulsive interaction pulled the water molecules back. The effect of this acceleration scheme is shown by the artificial shells in $g(\mathbf{r})$ (dashed line of Figure 6a). The reweighted SSMD curve correctly removes it in order to recover the original radial distribution function, while all other pair interactions were left unchanged as expected by this kind of acceleration scheme. Even if the system acquired new phase space regions represented by the artificial shell in Figure 6a, the overall acceleration improvement was found to be only a few percent (i.e., the diffusion constant increased by about 5%).

However, it is possible to achieve an improvement with the second type of acceleration where a molecular-based cutoff scheme is employed. Whenever atoms of two different molecules come closer than a certain cut off radius $r_0$, then all pair interactions between each component of the two molecules can be thought to smooth out to a free particle potential; the result is shown by the radial distribution function in panel b of Figure 6. The correct equilibrium distribution function is recovered and the diffusion constant is increased by 33%. An example of the mean square displacement for the accelerated dynamics is shown in panel b of Figure 7 (dashed line). This plot suggests that the reduced number of ballistic collisions is mainly responsible for the acceleration. This result is promising if it is taken into consideration that the method has been applied to all degrees of freedom without distinction between them of any sort and that no specific rare events, such that ones that arise in a double well picture, are targeted. In fact, the fluids described above are in thermodynamic equilibrium, as opposed to a multidimensional double well where a major boost in probability transitions between the two basins can be obtained only after selecting a subset of coordinates.

Between the two schemes, the first atom-based approach is easier to implement, and this could play a role when more complex (e.g., biomolecular) systems are considered. The second scheme embodies the original idea of creating a fictitious phase space region which allows entire molecules to funnel through new configurations and, as a consequence, gain an entropic boost.

## V. Concluding Remarks

Standard accelerated molecular dynamics via non-Boltzmann sampling has been shown to be problematic for condensed phase fluid systems, which do not adhere to a transition state theory (TST) picture along one or a few chosen coordinates. When applied indiscriminately to all fluid degrees of freedom, standard non-Boltzmann sampling results are poor, as shown on Figure 1. Here, an alternative approach is proposed based on the insertion of extended states into the interactions with the goal to enhance phase space sampling. The present methodology is the natural extension to complex systems of SSMD (which was previously presented for a one-dimensional example[45]). This approach, which in the present case couples the original potential to a fictitious free particle one, has been proven to recover the exact statistics via the non-Boltzmann sampling. It can be applied to selected types of pairwise interactions if one thinks these to be more relevant than others or equally to all degrees of freedom. In this case, the SSMD approach has be shown to achieve acceleration for both a standard LJ fluid and one with Coulombic interactions added. These systems are considered as good benchmarks for a condensed phase acceleration technique because their physical properties are well-known and any anomalies in the SSMD implementation can be easily recognized.

A resulting acceleration of 33% is also achieved for water after applying the SSMD implementation to all of the degrees of freedom. This can be considered a promising result if one takes into consideration that, in hyperdynamics,[29,36] a boost factor of 1400 for one degree of freedom reduces already to a factor of 2 if applied to a 20-dimensional system[68] and that the present SSMD implementation can recover the original statistical properties via non-Boltzmann sampling.

**Acknowledgment.** This research was supported by the Office of Naval Research (N00014-05-1-0457) and the National Science Foundation (CHE-0317132 and CHE-0719522). We thank Drs. Yeshitila Gebremichael, Vinod Krishna, and Satoru Iuchi for helpful discussions and comments.

## Appendix: SSMD for Ewald Summations

The Ewald method[1,69] estimates Coulombic interactions for an electrically neutral system of charged point particles under periodic boundary conditions in a given volume $L^3$, where $L$ is the length of the cubic simulation cell. These interactions are given by the sum of three contributions: the first is the electrostatic potential generated by the periodic sum of charges with density distributed as Gaussians

$$V_g(r_{ij}) = \frac{2\pi}{L^3} \sum_{(k\neq 0)} \frac{e^{-k^2/4\alpha}}{k^2} \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_d} \cos(\mathbf{k}\cdot r_{ij}) \quad \text{(A1)}$$

where $\alpha$ is the Gaussian exponential coefficient of the charges distributions, $\varepsilon_d$ the dieletric constant of the medium, $k$ is the reciprocal space vector, $q_i$ and $q_j$ are the point charges at respective locations $r_i$ and $r_j$, and $r_{ij} = |r_i - r_j|$. The second term is given by the correction of the spurious self-interactions included into the Fourier sum

$$V_{sl}(r_{ij}) = -\sum_j \frac{q_j}{4\pi\varepsilon_d} \sqrt{\frac{\alpha}{\pi}} \quad \text{(A2)}$$

and finally, the third one is the short-range interaction due to the point charges screened by oppositely charged Gaussians,

$$V_{sr}(r_{ij}) = \frac{1}{2} \sum_{(i\neq j)} \frac{q_i q_j}{4\pi\varepsilon_d} \frac{1 - \mathrm{erf}(\sqrt{\alpha} r_{ij})}{r_{ij}} \quad \text{(A3)}$$

where erf stands for "error function" and the cell summation has been taken out from all the above potentials.

In order to smooth the hard core potential, both the repulsive and attractive ones, the Fourier contribution has to be compensated by summing the exact opposite amount via real space. For this reason, eq A3 is rewritten as follows

$$V_{sr}(r_{ij}) = \frac{1}{2} \sum_{(i\neq j)} \frac{q_i q_j}{4\pi\varepsilon_d} \frac{f(r_{ij}) - \mathrm{erf}(\sqrt{\alpha} r_{ij})}{r_{ij}} \quad \text{(A4)}$$

where, given a cut off radius $r_0$, $f(r_{ij})$ interpolates between 1, when $r_{ij} > r_0$, and $r_{ij}/r_0$ when $r_{ij} < r_0$, in a SSMD fashion. On doing so, the resulting short-range potential in the forbidden region is

$$V_{sr}(r_{ij}) = \frac{1}{2} \sum_{(i\neq j)} \frac{q_i q_j}{4\pi\varepsilon_d} \left[ \left( \frac{1}{r_0} - \frac{1}{r_{ij}} \right) + \frac{\mathrm{erfc}(\sqrt{\alpha} r_{ij})}{r_{ij}} \right] \quad \text{(A5)}$$

where erfc is the complementary error function. The forces used can be calculated by performing the first derivatives of each potential piece. This implementation allows one to apply SSMD to the Ewald summation approach and exactly corrects the short-range potential such that, after the summation over all components, it brings the forces to be equal to zero for $r_{ij} < r_0$.

### References


(1) Frenkel, D.; Smit, B. *Understanding Molecular simulation: From Algorithms to Applications*, 2nd ed.; Academic: Boston, 2002.

(2) Allen, M. P.; Tildsley, D. J. *Computer Simulations of Liquids*; Oxford University Press: New York, 1987.

(3) Heermann, D. W. *Computer Simulation Methods in Theoretical Physics*; Springer: Berlin, 1986.

(4) Brunger, A.; Brooks, C. L., III.; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495.

(5) Schlick, T.; Barth, E.; Mandziuk, M. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 181.

(6) Schlick, T.; Skeel, R. D.; Brunger, A. T.; Kale, L. V.; Board, J. A.; Hermans, J.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 9.

(7) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511.

(8) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117.

(9) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

(10) Gawrisch, K. *The Dynamics of membrane Lipids*; CRC Press LLC: Boca Raton, FL, 2005.

(11) Daggett, V. *Curr. Opin. Struct. Biol.* **2000**, *10*, 160.

(12) Mayor, U.; Christopher, M. J.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci.* **2000**, *97*, 13518.

(13) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740.

(14) Shea, J.-E.; Brooks, C. L., III. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499.

(15) Sheinerman, F. B.; Brooks, C. L., III. *Proc. Natl. Acad. Sci.* **1998**, *95*, 1562.

(16) Andricioaei, I.; Straub, J. E. *J. Chem. Phys.* **1997**, *107*, 9117.

(17) Andricioaei, I.; Straub, J. E. *J. Chem. Phys.* **2001**, *114*, 6994.

(18) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249.

(19) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9.

(20) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451.

(21) Lyubartsev, A. P.; M, A. A.; Sheykunov, S. V.; Vorontsov-Velyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776.

(22) Sugita, Y.; Okamoto, Y. *J. Phys. Soc. Jpn.* **1999**, *65*, 1604.

(23) Tesi, M.; E, J. J. v. R.; Orlandini, E.; Whittington, S. G. *J. Stat. Phys.* **1996**, *82*, 155.

(24) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140.

(25) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604.

(26) Fenwick, M. K.; Escobedo, F. A. *J. Chem. Phys.* **2003**, *119*, 11998.

(27) Jang, S.; S, S.; Pak, Y. *Phys. Rev. Lett.* **2003**, *91*, 058305.

(28) Affentranger, R.; I, T.; Di Iorio, E. E. *J. Chem. Theory Comput.* **2006**, *2*, 217.

(29) Montalenti, F.; Voter, A. F. *J. Chem. Phys.* **2002**, *116*, 4819.

(30) Voter, A. F. *Phys. Rev. B* **1998**, *57*, 13985.

(31) Zagrovic, B.; E, J. s.; Pande, V. *J. Mol. Biol.* **2001**, *313*, 151.

(32) Uberuaga, B. P.; Anghel, M.; Voter, A. F. *J. Chem. Phys.* **2004**, *120*, 6363.

(33) Sorensen, M. R.; Voter, A. F. *J. Chem. Phys.* **2000**, *112*, 9599.

(34) Miron, R. A.; Fichthorn, K. A. *J. Chem. Phys.* **2003**, *119*, 6210.

(35) Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908.

(36) Voter, A. F. *J. Chem. Phys.* **1997**, *106*, 4665.

(37) Corcelli, S. A.; Rahman, J. A.; Tully, J. C. *J. Chem. Phys.* **2003**, *118*, 1085.

(38) Hamelberg, D.; Mongan, J.; McCammon, J. A. *Protein Sci.* **2004**, *13*, 76.

(39) Hamelberg, D.; T, S.; McCammon, J. A. *J. Chem. Phys.* **2005**, *122*, 241103.

(40) Steiner, M. M.; Genilloud, P.-A.; Wilkins, J. W. *Phys. Rev. B* **1998**, *57*, 10236.

(41) Rahman, J. A.; Tully, J. C. *J. Chem. Phys.* **2002**, *116*, 8750.

(42) Madun, R. A.; Fichthorn, K. A. *Phys. Rev. Lett.* **2004**, *93*, 128301.

(43) Xing, C.; Andricioaei, I. *J. Chem. Phys.* **2006**, *124*, 034110.

(44) Minth, D. D. L.; Hamelberg, D.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*, 154105.

(45) Venkatnathan, A.; Voth, G. A. *J. Chem. Theory Comput.* **2004**, *1*, 36.

(46) Grubmuller, H. *Phys. Rev. E* **1995**, *52*, 2893.

(47) Hamelberg, D.; J, M.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919.

(48) MacFadyen, J.; Andricioaei, I. *J. Chem. Phys.* **2005**, *123*, 074107.

(49) Mella, M. *J. Chem. Phys.* **2005**, *122*, 204106.

(50) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*, 155102.

(51) de Oliveira, C. A. F.; Hamelberg, D.; McCammon, J. A. *J. Phys. Chem. B* **2006**, *110*, 22695.

(52) de Oliveira, C. A. F.; Hamelberg, D.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*, 175105.

(53) Dellago, C.; Bolhius, P. G.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 9236.

(54) Bolhuis, P. G.; Dellago, C.; Geissler, P. L.; Chandler, D. *J. Phys.: Condens. Matter* **2000**, *12*, A147.

(55) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291.

(56) Prasad, M.; Sinno, T. *J. Chem. Phys.* **2004**, *121*, 8699.

(57) Wu, X.; Brooks, B. R. *Chem. Phys. Lett.* **2003**, *381*, 512.

(58) Wu, X.; Wang, S. *J. Phys. Chem. B* **1998**, *102*, 7238.

(59) Wu, X.; Wang, S. *J. Chem. Phys.* **1999**, *110*, 9401.

(60) Andricioaei, I.; Dinner, A. R.; Karplus, M. *J. Chem. Phys.* **2003**, *118*, 1074.

(61) Bitetti-Putzer, R.; Dinner, A. R.; Yang, W.; Karplus, M. *J. Chem. Phys.* **2006**, *124*, 174901.

(62) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 090601.

(63) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12562.

(64) Micheletti, C.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2004**, *92*, 170601.

(65) Oganov, A. R.; Martonak, R.; Laio, A.; Raiteri, R.; Parrinello, M. *Nature* **2005**, *438*, 1142.

(66) Durell, S. P.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198.

(67) Smih, W.; Forester, T. *J. Mol. Graphics* **1996**, *14*, 136.

(68) Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **2001**, *115*, 9657.

(69) Ewald, P. P. *Ann. Phys.* **1921**, *64*, 253.

CT7003275

JCTC Journal of Chemical Theory and Computation

# An Investigation of the Accuracy of Different DFT Functionals on the Water Exchange Reaction in Hydrated Uranyl(VI) in the Ground State and the First Excited State

Pernilla Wåhlin,[†] Cécile Danilo,[†,§] Valérie Vallet,*[,§] Florent Réal,[†] Jean-Pierre Flament,[§] and Ulf Wahlgren[†,‡]

*Department of Physics, Stockholm University, AlbaNova University Centre, 106 91 Stockholm, Sweden, NORDITA, AlbaNova University Centre, 106 91 Stockholm, Sweden, and Université des Sciences et Technologies de Lille 1, Laboratoire PhLAM, CNRS UMR 8523, CERLA, CNRS FR 2416, Bât P5, 59655 Villeneuve d'Ascq Cedex, France*

Received March 14, 2007

**Abstract:** We discuss the accuracy of density functional theory (DFT) in the gas phase for the water-exchange reactions in the uranyl(VI) aqua ion taking place both in the electronic ground state and in the first excited state (the luminescent $^3\Delta_g$ state). The geometries of the reactant and intermediates have been optimized using DFT and the B3LYP functional, with a restricted closed-shell formalism for the electronic ground state and either an unrestricted open-shell formalism or the time-dependent DFT method for the $^3\Delta_g$ state. The relative energies have been computed with wave-function-based methods such as Møller–Plesset second-order perturbation theory, or a minimal multireference perturbative calculation (minimal CASPT2); coupled-cluster method (CCSD(T)); DFT with B3LYP, BLYP, and BHLYP correlation and exchange functionals; and the hybrid DFT−multireference configuration interaction method. The results obtained with second-order perturbative methods are in excellent agreement with those obtained with the CCSD(T) method. However, DFT methods overestimate the energies of low coordination numbers, yielding to too high and too low reaction energies for the associative and dissociative reactions, respectively. Part of the errors appears to be associated with the amount of Hartree–Fock exchange used in the functional; for the dissociative intermediate in the ground state, the pure DFT functionals underestimate the reaction energy by 20 kJ/mol relative to wave-function-based methods, and when the amount of HF exchange is increased to 20% (B3LYP) and to 50% (BHLYP), the error is decreased to 13 and 4 kJ/mol, respectively.

## Introduction

Quantum chemical studies of actinides differ from those of the lighter elements in several ways. Relativistic effects are obvious, but many of the related computational problems have been mastered, at least for small complexes; the spin–orbit problem can be handled either at the two- or four-component level, or in an LS-coupled framework by spin–orbit configuration interaction (CI) programs such as EPCISO[1] or by variation-perturbation methods as implemented in, for example, the complete active space interaction with spin–orbit, CASSI-SO,[2] program in the Molcas[3] package. Another consideration is the large number of electrons. The core–electrons can be included in an effective core potential, but the highly polarizable outer-core orbitals make it necessary to include all shells with a principal quantum number above 4 in the valence space, which makes the

* To whom correspondence should be addressed. Tel.: +33 3 2033 5986. Fax: +33 3 2033 7020. E-mail: valerie.vallet@univ-lille1.fr.

[†] Stockholm University.

[§] Université des Sciences et Technologies de Lille 1.

[‡] NORDITA.

10.1021/ct700062x CCC: $40.75 
Published on Web 02/29/2008

calculations more cumbersome than those for lighter atoms. The open f-shell problem can in the scalar relativistic framework, in principle, be handled by the complete active space self-consistent field method with subsequent second-order perturbation calculation of the dynamic correlation, CASSCF[4–6]/CASPT2[7,8] levels, but calculations at those two levels where all valence excited states are included in the reference can only be used on small systems; already the bare uranyl ion with 12 active orbitals (six bonding and six antibonding $U-O_{yl}$ orbitals) is approaching the limit of about 15 active orbitals for a complete CASSCF/CASPT2 treatment.[9] It would seem that those problems are easily overcome by reverting to density functional theory (DFT) based methods, but the applicability of DFT functionals, which have been developed for lighter elements, is not obvious.

Complexes with more than one unpaired f electron give rise to multireference effects, which makes conventional unrestricted methods inaccurate.[10] In some cases, such systems can be described by DFT-based methods such as time-dependent DFT (TD-DFT) or the hybrid DFT−multireference configuration interaction (DFT-MRCI).[11] In the latter, dynamic correlation is treated by DFT, while static correlation and multireference effects in the electronic states are accounted for in a MRCI framework. The majority of the published studies concerns the uranyl(VI) ion, for which the closed-shell structure essentially removes the problems related to spin–orbit effects and open f shells. Although the f-shell problem does not appear explicitly, the f orbitals participate actively in the bonding, and the applicability of DFT-based methods relies on the ability of the functionals to treat the f electrons properly.

DFT does not give reliable results for electron reduction processes in actinyl complexes,[12] but this is not particular for the actinides. Variable results have been reported for some complexes where the oxidation state is not changed but where the bonds between the actinide and the ligands are strong, such as in $UO_3$, $UF_6$,[13–17] and so forth. Batista et al.[15,16] have calculated the dissociation energy for $UF_6$ using both a small-core relativistic effective core potentials (RECP)[18] and an all-electron approximation together with the B3LYP functional. Their spin–orbit free results are in good agreement with experiments; the difference is just a few kilojoules per mole. Peralta et al.[17] have calculated the dissociation energy for $UF_n$, $n = 1-6$, at the B3LYP level using the all-electron approximation. Their result for the series has a deviation from the experimental value varying from −36 kJ/mol ($n = 4$) up to +43 kJ/mol ($n = 1$) and with the best result for $n = 3$ and 5, where the deviation is just a few kilojoules per mole. Considering that the dissociation reaction energy is more difficult to calculate with high accuracy than the reaction we have chosen for our study, their results are good but point out the inconsistency of DFT functionals.

The solvation of uranyl(VI) has been studied by several groups.[19–31] Gutowski and Dixon[28] report differences on the order of 20 kJ/mol when calculating the relative energies of hydrated uranyl with four- and five-coordinated water at the B3LYP level compared to the Møller–Plesset second-order perturbation theory (MP2) level, using a small-core RECP on the uranium atom. Cao and Balasubramanian[25] also investigated the structure and energetics of actinyl aqua ions with four-, five-, and six-coordinated water molecules. They made a comparison using the B3LYP[32–35] hybrid density functional and MP2 and coupled cluster with single and double excitation (CCSD)[36] for actinyl complexes, both in the gas phase and aqueous solution. They used a large-core RECP and a smaller basis set for the uranium than we have used in the present study. However, they also noted discrepancies on the order of 20 kJ/mol in the relative energies of hydrated uranyl with varying numbers of coordinated water if computed with DFT or wave-function-based correlated methods. In a recent study of the water exchange in uranyl(VI) by Rotzinger,[31]the same trend is observed for BLYP[37–39] and B3LYP, using large-core effective core potentials on the uranium atom. Shamov and Schreckenbach[26] also noted differences between B3LYP and PBE[32,33,40,41] functionals. All studies so far agree that the five-coordinated ion is the preferred coordination of uranyl(VI) in solution, but they disagree on the relative energies of four- and six-coordinated species. This results in different conclusions on the nature of the water-exchange mechanism. One possible explanation to the variable results may be the important contribution from the f orbitals in the binding in, for example, actinide(VI) complexes such as the actinyl ions $AnO_2^{2+}$. Since the quality of DFT applied to actinide complexes is unclear, it is reasonable to investigate the behavior of DFT methods before launching an investigation of a new reaction.

We have previously studied ligand-exchange reactions in uranyl(VI) and uranyl(V),[22,42–44] and this line of research has now been extended to photoexcited uranyl(VI). The first reaction we have considered is the water exchange mechanism in the first excited triplet state, which is luminescent. Before entering into the chemical properties of photoexcited uranyl, we have deemed it desirable to investigate the accuracy of different DFT functionals and hybrids for this relatively simple complex that do not exhibit any evident multireference properties. The water-exchange reaction can proceed via three pathways,[22] a dissociative pathway with a four-coordinated intermediate, an associative pathway with a six-coordinated intermediate, and finally an interchange pathway. In the present investigation, we have compared the relative energies obtained with different methods between a five-coordinated reactant and the four- and six-coordinated dissociative and associative intermediates for the electronic ground state and the first excited state, the latter complexes being depicted in Figure 1. The geometries for the two states bear an overall strong resemblance to each other, naturally, with variations in bond lengths. There are experimental results for hydrated uranyl in the ground state,[45] but modeling solvent effects is outside the scope of the present study; our aim is to compare various electronic structure methods in the gas phase, and not to study the chemical problem of the solvation of the uranyl(VI) ion. Consequently, we do not have an ultimate answer to which result is "correct" for the fictitious reaction under study. However, results, which differ significantly between different methods, clearly indicate potential problems. In our experience, wave-function-based methods such as CCSD(T) or CASPT2 with a reasonably

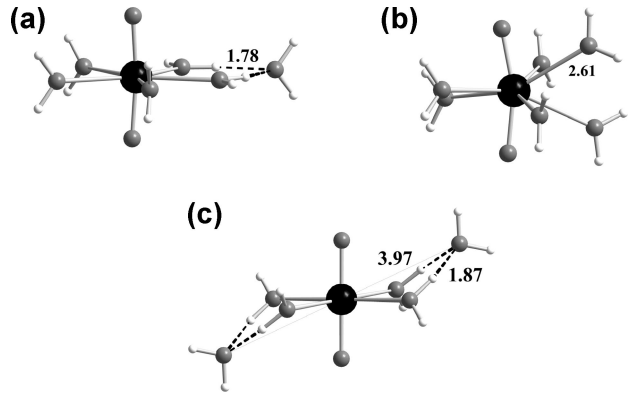


***Figure 1.*** The geometries of the (a) reactant, (b) A-intermediate, and (c) D-intermediate water-coordinated uranyl(VI) complexes in the luminescent state. Distances in the figures are given in ångströms.

large active space are reliable, and minimal CASPT2 (MP2) methods have been shown to give reliable results when compared to those of both experiments and CCSD(T) in a number of cases. Therefore, we view the wave-function-based results as more trustworthy than DFT, although we cannot ascertain this assumption; for example, the size of the systems restricts the size of the basis sets to about triple-$\zeta$ plus polarization functions. The DFT functionals considered in the present study are the hybrid B3LYP and BHLYP and the pure DFT functionals PBE, TPSS, BP86, and BLYP used with the restricted closed-shell formalism for the ground state and the unrestricted open-shell formalism or time-dependent DFT for the first excited state. The wave-function-based methods used were CCSD(T) and MP2/minimal CASPT2.

**Theory.** *The Model.* In the present study, we have used the water-exchange reaction in uranyl(VI) in the ground state and in the luminescent state as a probe to compare results obtained with different pure DFT functionals and hybrids and with wave-function-based MP2/minimal CASPT2 and CCSD(T).[46,47] The comparison is based on the energy of reaction starting with the five-coordinated precursor complex with one water in the outer hydration sphere, $[UO_2(H_2O)_5)]^{2+}$,$(H_2O)$ (Figure 1a), to the associative and dissociative intermediates, $[UO_2(H_2O)_6]^{2+}$ and $[UO_2(H_2O)_4]^{2+}$, $(H_2O)_2$, Figure 1b and c, respectively.

In the model, we consider the exchange between inner- and outer-sphere water molecules. The dissociative and associative reactions read:

$$[UO_2(H_2O)_5]^{2+},(H_2O) \leftrightarrow [UO_2(H_2O)_6]^{2+} \quad \text{A-reaction} \tag{1}$$

$$[UO_2(H_2O)_5]^{2+},(H_2O) \leftrightarrow [UO_2(H_2O)_4]^{2+},(H_2O)_2 \quad \text{D-reaction} \tag{2}$$

An alternative model refers to the complete dissociation reactions:

$$[UO_2(H_2O)_5]^{2+} + (H_2O) \leftrightarrow [UO_2(H_2O)_6]^{2+} \quad \text{A-reaction} \tag{3}$$

$$[UO_2(H_2O)_5]^{2+} \leftrightarrow [UO_2(H_2O)_4]^{2+} + H_2O \quad \text{D-reaction} \tag{4}$$

As discussed in ref 48, it is preferable to use intramolecular reactions 1 and 2, rather than reactions 3 and 4, because systematic errors of the solute–solvent interactions compensate in a more efficient way.

Our aim has been to establish the accuracy of different DFT functionals and to pinpoint the problem of the description of electron correlation for the uranyl−water bonding.

*Technical Details.* For the ground state, we examined four generalized gradient approximation exchange-correlation functionals, PBE, TPSS,[32,33,38,49] BP86,[32,33,37,50,51] and BLYP; two hybrid functionals, B3LYP and BHLYP;[32–34,52] the SAOP functional specially designed for response properties;[53] two wave-function-based methods, MP2 and CCSD(T); and the DFT-MRCI method proposed by Grimme and Waletzke.[11] The DFT-MRCI method is a combination of Kohn–Sham density functional theory and multireference CI methods and can be used for calculating excitation energies for singlet and triplet states, starting from a closed-shell reference state. In this respect, it is similar to the TD-DFT method.[54–56] The wave function is expressed in terms of Kohn–Sham orbitals, obtained with the BHLYP hybrid functional (with 50% HF exchange). A selection criterion for single and double excitations, based on the energy gap between a configuration state function and the corresponding parent configuration, reduces the number of configuration state functions. Some parameters, determined from atomic calculations, are used in the method to scale the Coulomb and exchange integrals. Currently, the optimized parameter sets for the effective DFT-MRCI Hamiltonian are available with the BHLYP functional. We refer the reader to the original publication[11] of Grimme and Waletzke for further details on the method. For the luminescent state, we examined the same functionals and wave function methods as for the ground state. For the hybrid functionals B3LYP and BHLYP, we used both the unrestricted open-shell formalism and the time-dependent method. The exchange-correlation functionals, PBE, TPSS, BP86, BLYP, and SAOP, have been studied within the time-dependent approach. For minimal CASPT2, we used two different CASs, a minimal active space that includes the orbitals necessary for the description of the electronic states of interest, namely, the $\sigma_u$ together with the two $\delta_u$ orbitals, and the slightly larger CAS where the two $\varphi_u$'s had been added to the previous CAS. The reason for including calculations with the larger CAS was to verify that our minimal CASPT2 result is not CAS-dependent. Unfortunately, it is not possible to use a nonperturbative wave-function-based method, such as for example a multireference configuration interaction, for the luminescent state due to the large number of electrons involved. For MP2, minimal CASPT2, CCSD(T), and DFT-MRCI calculations, only the valence shells, 6s, 6p, 5f of uranium and 2s and 2p of oxygen atoms, were correlated.

Calculations were performed using either RECP or an all-electron basis set with the relativistic Douglas−Kroll−Hess Hamiltonian.[57,58] In the RECP calculations, we used a small-core RECP of the Stuttgart type[18] for uranium with the corresponding basis set.[59] In the all-electron calculations, we used the (26s23p17d13f5g3h)/9s8p6d4f2g ANO-RCC basis set suggested by Roos et al.[60] with triple-$\zeta$ quality. Oxygen

and hydrogen were treated at the all-electron level in all calculations, using the TZVP basis set suggested by Schäfer et al.[61,62] The effect of the g functions was investigated in the all-electron calculations; the effect was small, between 0.2 and 1.4 kJ/mol at the DFT level and at most 1.6 kJ/mol with the wave-function-based methods. This is in agreement with previous results obtained by Vallet et al.[63] No g functions on uranium were used in the RECP calculations. The DFT−SAOP calculations were performed with the ADF2006.01 code, incorporating scalar relativistic effects by means of the zeroth-order regular approximation (ZORA) method[64–67] and with a triple-ζ Slater-type basis set enlarged by adding two polarization functions (TZ2P). This basis set is large enough to yield very similar accuracy to the Gaussian basis set used in all-electron calculations.

Effects on the ground state and the luminescent state, arising from the spin–orbit splitting of uranium, have been calculated with the EPCISO program[1] using a RECP for uranium.[18] The reference space used in the calculations of the luminescent state included five orbitals, the $\sigma_u$ and the two $\delta_u$'s and two $\varphi_u$'s, while the other electrons were kept uncorrelated, whereas for the ground state, only the $\sigma_u$ electrons were correlated. The resulting spin–orbit effect, including the single excitations from the active orbitals, was then added to the multiconfigurational CASPT2 energies. Spin–orbit coupling lowers the energies of all three complexes by 28.4–32.9 kJ/mol for the excited state and by 9.9–13.1 kJ/mol for the ground state. This affects the reaction energies with a lowering of 2.9 kJ/mol for the associated reaction and an increase of 2.5 kJ/mol for the dissociative reaction, for the excited state. For the ground state, it yields an increase of 0.5 kJ/mol for the associated reaction and a lowering of 2.7 kJ/mol for dissociated reaction.

It is interesting to note that, at the spin–orbit level, the luminescent state in the reactant is dominated by $\varphi$ character, whereas in the A-intermediate, there is a mixture of $\delta$ and $\varphi$ contributions, and in the D-intermediate, it has $\delta$ character. The mixing between $\delta$ and $\varphi$ contributions is determined by the spacing between the $^3\Delta$ and $^3\Phi$ states, as discussed in detail by Réal et al. in a recent study.[68]

There are no significant basis set superposition error (BSSE) effects on the computed reaction energies; the largest calculated effect on the ground-state model was 3.0 kJ/mol at the MP2 level and 1.6 kJ/mol for B3LYP using Gaussian 03.[69] Rotzinger[31] has in a recent paper calculated the BSSE contribution to the reaction energies for the ground state and also found them insignificant, at most 4 kJ/mol. We also calculated the BSSE effect on the dissociation of the outer-sphere water in the ground-state reactant, $[UO_2(H_2O)_5]^{2+}$, $(H_2O)$; the effect at the B3LYP level is 5 kJ/mol, while for the MP2, level it is close to 11 kJ/mol. The BSSE-corrected reaction energies were found to be respectively −102 and −110 kJ/mol, showing the tendency of B3LYP to underestimate reaction energies.

The optimizations of the geometries for the ground state were done at the B3LYP level with symmetry constraints using Gaussian 03.[69] For the reactant, we used $C_s$ point group symmetry, for the A-intermediate, $C_2$ symmetry, and for the D-intermediate, $C_{2h}$ symmetry. The geometries in the luminescent state were optimized both with TD-B3LYP using TURBOMOLE 5.7[70,71] and with U-B3LYP using Gaussian 03, without symmetry constraints. All reaction energies have been calculated using the TD-B3LYP optimized geometries with the Gaussian 03,[69] TURBOMOLE 5.7[70,71] the ADF2006.01[72–74] packages.

The program packages Molcas 6.4[3] and Molpro 2006.1[75] were used for the MP2/minimal CASPT2 and CCSD(T) calculations.

***Table 1.*** Optimized Geometries for the Reactant, A-Intermediate, and D-Intermediate in Their Electronic Ground States[a]

| | $d(U-O_{yl})$ | | $d(U-OH_2)$ first coord. shell | |
|---|---|---|---|---|
| | B3LYP | MP2 | B3LYP | MP2 |
| ref | this work | 63 | this work | 63 |
| reactant | 1.749 | 1.776 | 2.496 | 2.47(1) |
| A-intermediate | 1.755 | 1.785 | 2.518 × 4/2.645 × 2 | 2.47 × 2, 2.49 × 2, 2.69 × 2 |
| D-intermediate | 1.748 | 1.774 | 2.420 | 2.41 |

[a] Distances are in ångströms.

## Results and Discussion

**Water Exchange in the Electronic Ground State of Uranyl(VI) Penta Aqua Ion.** The optimal ground-state geometries are shown in Table 1. The agreement between MP2 and B3LYP optimized geometries is good. It is in fact a general experience that B3LYP gives geometries in good agreement with experimental results.[63] As an example, the bond distances in the bare uranyl ion are 1.701 Å at the B3LYP level and 1.706 Å at the CCSD(T) level and 1.709 Å at the average quadratic coupled-cluster level.[12] Our geometries for the bare uranyl ion, the reactant, and the D-intermediate are in good agreement with those obtained by Gutowski and Dixon.[28] In the following, we will base our discussions on geometries optimized at the B3LYP or, for the luminescent state, TD-B3LYP level.

As expected, the bonds to the coordinated water contract or expand relative to the reactant in the D-intermediate and A-intermediate. The shortening of the uranium−water bond in the D-intermediate is sizable, 0.07 Å. In the A-intermediate, four waters are located in the equatorial plane of the uranyl unit with distances about 0.01 Å longer than in the five-coordinated reactant, while two waters lie above and below the equatorial plane at longer distances, 2.65 Å. The uranium−water bond is thus considerably stronger in the D-intermediate than in the reactant and the A-intermediate. The variation in the $U-O_{yl}$ distance is smaller.

For the A-intermediate, pure DFT methods yield higher reaction energies than MP2 and CCSD(T), but the differences are small; the mean value for the DFT-based RECP calculations, including DFT-MRCI, is 41.3 kJ/mol, with a largest deviation from the mean value of 4.3 kJ/mol. The largest absolute difference (between B3LYP and TPSS) is 8 kJ/mol. The MP2 and CCSD(T) results, respectively 35.6 and 36.7 kJ/mol, are about 5 kJ/mol below the DFT mean result. It is noteworthy that the largest difference between the DFT and the wave function results, 10 kJ/mol, occurs for B3LYP and for the SAOP functional, and this difference is not negligible.

***Table 2.*** Reaction Energies in kJ/mol for A-Reaction (1) and D-Reaction (2) Computed for Uranyl(VI) in Its Electronic Ground State[a]

| | | A-reaction | | | | D-reaction | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | ECP | | AE | | ECP | | AE | |
| method | HF exchange in the DFT functional | no g | BSSE | no g | with g | no g | BSSE | no g | with g |
| B3LYP | 20% | 44.9 | 44.8 | 41.6 | 42.1 | 20.1 | 17.1 | 21.8 | 21.6 |
| BHLYP[b] | 50% | 42.7 | | 41.3 | 41.7 | 29.3 | | 31.0 | 30.7 |
| BP86 | 0% | 42.7 | | | | 13.5 | | | |
| PBE | 0% | 40.3 | | | | 13.1 | | | |
| TPSS | 0% | 37.0 | | | | 15.7 | | | |
| BLYP | 0% | 41.9 | | | | 13.5 | | | |
| SAOP[b] | 0% | | | | 43.1 | | | | 13.1 |
| MP2 | | 35.6 | 37.1 | 34.1 | 35.5 | 34.4 | 32.8 | 35.2 | 33.6 |
| MP2 SO | | 36.1 | | | | 31.7 | | | |
| CCSD(T)[c] | | 36.7 | | 31.9 | 33.2 | 34.6 | | 37.4 | 36.0 |
| DFT-MRCI | 50%[b] | 39.7 | | | | 25.4 | | | |

[a] Both an effective core potential (ECP) and an all-electron (AE) basis set were used for the uranium atom. [b] SAOP results have been obtained with the all-electron ZORA and all-electron Slater triple-$\zeta$ basis sets (see Theory section). [c] The $T_1$ diagnostic is between 0.0225 and 0.0239.

The all-electron and the RECP results are similar, differing by at most 5 kJ/mol, a satisfactory result. The effect of the g functions is negligible at the DFT level and below 2 kJ/mol in the wave-function-based calculations. It should be pointed out that g functions can be very important in strongly bound U(VI) complexes and in redox reactions.[12,14] Basis set effects for actinides are discussed in a review article by Vallet et al.[63]

The D-intermediate behaves differently. DFT consistently gives reaction energies that are lower than those obtained with MP2 and CCSD(T) (cf. Table 2), while the opposite was true for the A-intermediate. The reaction energies obtained with the pure DFT functionals are insensitive to the choice of functional, but the effect of including HF exchange is much more pronounced for the D-intermediate; the effect is depicted in the energy diagram in Figure 2. B3LYP gives a higher reaction energy by 6 kJ/mol, than the pure functionals; with BHLYP, the increase amounts to 17 kJ/mol; for MP2 and CCSD(T), the reaction energy is more than 20 kJ/mol higher than that obtained with the pure functionals. DFT-MRCI, which is based on BHLYP Kohn–Sham orbitals, also yields a value which is larger, by 12 kJ/mol, than with pure functionals and lies within the energy obtained with B3LYP and BHLYP functionals. The basis set effects and the quality of the RECP is very similar for the two intermediates.

Rotzinger[76] noted that DFT tends to overestimate the energy of transition metal complexes with a high coordination number, thus favoring associative reactions over dissociative reactions (cf. Table 2). The low reaction energy obtained for the D-intermediate at the DFT level compared to MP2 and CCSD(T) is due to the strong bonds between the four equatorial waters and the uranyl ion. It appears that the amount of HF exchange has a larger impact on the relative energies of the D-intermediate than in the reactant and in the A-intermediate. This is probably the result of the stronger bonds between the metal and the coordinated water in the D-intermediate (cf. Table 1). However, the close similarity between the reaction energies obtained with MP2 and CCSD(T) makes us confident of the reliability of wave-function-based methods as compared to DFT ones.

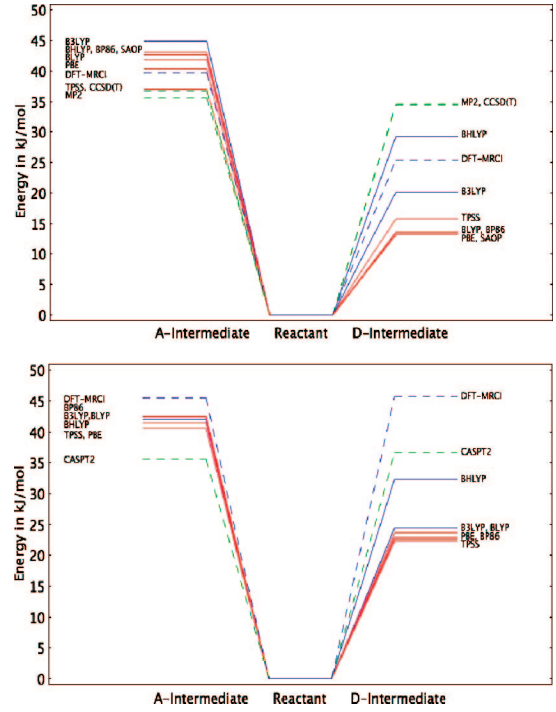


***Figure 2.*** The reaction energies relative to the reactant for the A- and D-intermediates computed for uranyl(VI) in its electronic ground state (a) and in its first electronic excited state (b). The pure functionals are plotted in red, the hybrids in blue, and the wave-function-based methods in dashed green. The DFT-MRCI result is plotted in dashed blue.

The difference between DFT and wave-function-based methods could be caused by an inaccurate description in the DFT framework of either the hydrogen bonds and dispersion effects or the solvated uranyl. These options were investigated through calculations where the uranyl charge distribution was computed for the bare ion and kept frozen in the molecule. The differences in the MP2 and DFT reaction energies with a frozen charge distribution on uranyl were at most 1.0 kJ/mol for the A-intermediate, to be compared to 9.3 kJ/mol at the B3LYP level and 6.3 kJ/mol at the BLYP level in the fully relaxed calculation. For the A-intermediate,

we can thus conclude that the differences between the MP2 and the DFT results are due primarily to the description of uranyl. However, the differences are small. For the D-intermediate, the effects are larger both in the frozen uranyl and in the fully relaxed calculations. The differences obtained in the frozen uranyl calculations were 3.7 kJ/mol at the B3LYP level and 9.0 kJ/mol at the BLYP level. In the fully relaxed calculations, the differences were 14.3 kJ/mol at the B3LYP level and 20.9 kJ/mol at the BLYP level, which is more than twice the difference found for the frozen uranyl. Taking into consideration that not only the description of the hydrogen bonds and dispersion effects contributes to the differences between wave-function-based and the DFT-based methods for the water cage, we conclude that the differences obtained with the fully relaxed model originate primarily from the solvated uranyl ion. There are also shortcomings of DFT in the description of dispersion interactions.[77,78]

Gutowski and Dixon[28] have published a study of the free energy of solvation of the uranyl(VI) ion. One of the studied reactions was reaction 4, for which the authors obtained a difference in the reaction energy of 20 kJ/mol between MP2 and B3LYP, in agreement with our results. Gutowski and Dixon also made CCSD calculations on $[UO_2(H_2O)_4]^{2+}$ and $[UO_2(H_2O)_5]^{2+}$ and obtained $T_1$ diagnostics of 0.023 and 0.022, also in agreement with our results. This indicates, as pointed out by Gutowski and Dixon[28] and discussed in details by Vallet et al.,[79] that multireference effects caused by near-degeneracies, so-called static correlation effects, are minor in these complexes. Shamov and Schreckenbach[26] have in a density functional study of actinide complexes in solution, using reactions 3 and 4, made a comparison between B3LYP and PBE. They found a similar shift, 15 and 9.2 kJ/mol between the two functionals, to what we have found, for reactions 1 and 2. Bühl et al.[27,29,30] have carried out Car–Parrinello calculations on the dissociation of one water molecule from a hydrated uranyl(VI) ion with Car–Parrinello molecular dynamics using the BLYP functional, large-core RECP, and plane-wave basis sets. In their smallest model, with only five waters and no further solvent effects, they obtain an electronic energy of reaction of −16.3 kJ/mol with a D-intermediate, with the second sphere water hydrogen bound to one water molecule (single hydrogen bond model) in the first coordination sphere.[29] We obtain a reaction energy at the BLYP level of 13.5 kJ/mol with a six-water model, with the water in the second sphere hydrogen bound to two waters in the first coordination sphere (double hydrogen bond model) for both the reactant and the D-intermediate (cf. Figure 1). Using the same model as Bühl et al., we obtain a reaction energy of −14.3 kJ/mol at the BLYP level and −7 kJ/mol at the B3LYP level, in good agreement with Bühl et al.[29] We have also investigated a six-water model, where the outer-sphere water is singly hydrogen bound to the inner-sphere water, for the D-intermediate, and found a similar reaction energy, −4 kJ/mol at the B3LYP level. Our calculated reaction energy with the double hydrogen bond model for the D-intermediate is considerably higher, 20 kJ/mol (in the optimal structure for the reactant, the outer-sphere water keeps forming two hydrogen bonds with inner-sphere water molecules). However, since the scope of this paper is a method study, the present investigation was done using a double hydrogen bond model for the D-intermediate.

***Table 3.*** Optimized Geometries for the Reactant, A-Intermediate, and D-Intermediate in the Luminescent Electronic State[a]

| | $d(U-O_{yl})$ | | $d(U-OH_2)$ first coord. shell | |
|---|---|---|---|---|
| | U-B3LYP | TD-B3LYP | U-B3LYP | TD-B3LYP |
| reactant | 1.805 | 1.795 | 2.495 | 2.495 |
| A-intermediate | 1.812 | 1.804 | 2.520/2.613 | 2.525/2.619 |
| D-intermediate | 1.803 | 1.794 | 2.43 | 2.43 |

[a] Distances are in ångströms.

***Table 4.*** Reaction Energies in kJ/mol for the A-Reaction (1) and D-Reaction(2) Computed for the Luminescent State of Uranyl(VI)

| method | HF exchange in DFT functional | A-reaction | D-reaction |
|---|---|---|---|
| B3LYP | 20% | 45.6 | 24.5 |
| BHLYP | 50% | 42.0 | 32.3 |
| TD-B3LYP | 20% | 42.4 | 24.6 |
| TD-BHLYP | 50% | 42.0 | 32.3 |
| TD-BP86 | 0% | 42.6 | 22.7 |
| TD-PBE | 0% | 40.6 | 23.0 |
| TD-TPSS | 0% | 41.5 | 22.3 |
| TD-BLYP | 0% | 42.4 | 23.7 |
| TD-SAOP | 0% | 47.9 | 13.9 |
| DFT-MRCI | 50% | 45.5 | 45.8 |
| Min-CASPT2,[a] CAS = 3 | | 35.6 | 36.6 |
| Min-CASPT2 SO | | 32.7 | 39.1 |
| Min-CASPT2,[a] CAS = 5 | | 34.9 | 36.9 |

[a] CAS = 3 and CAS = 5 refer to active spaces involving three active ($\sigma_u$ and two $5f_\varphi$) or five active orbitals ($\sigma_u$, two $5f_\varphi$, and two $5f_\delta$), respectively.

**Water Exchange in the Luminescent State of Uranyl-(VI).** Table 3 shows geometries for the reactant, the A-intermediate, and the D-intermediate optimized with U-B3LYP and TD-B3LYP. TD-B3LYP gives shorter bond distances than U-B3LYP, but the differences are small, particularly for the uranium−water distances. Our results show that geometries in (good) agreement with U-B3LYP can be obtained with TD-B3LYP, a result that opens the possibility of geometry optimizations in excited states of actinide complexes. The reaction energies are shown in Table 4 and Figure 2.

For the A-intermediate, pure and hybrid TD-DFT calculations give higher reaction energies, on average 42.4 kJ/mol, which is higher by about 7 kJ/mol than the two minimal CASPT2 calculations. The largest deviation, 10 kJ/mol, occurs for TD-SAOP and U-B3LYP, as in the ground state, but also for DFT-MRCI, the same difference is noticed. The spread within the reaction energies is even less than in the ground state; the standard deviation is 0.8 compared with 2.7, and the largest absolute difference, between U-B3LYP and TD-PBE, is only 5 kJ/mol.

Like in the ground state, the situation is different for the D-intermediate. Here, we can notice that the reaction energy varies with the chosen DFT functional, the largest difference with the CASPT2 value being obtained with the TD-SAOP functional, although the latter has been specially designed to describe the excited state. The effect of including HF

***Table 5.*** The adiabatic transitions, in $cm^{-1}$, computed for uranyl(VI) using an effective core potential (ECP)[a]

| | B3LYP | BHLYP | BP86 | PBE | TPSS | BLYP | SAOP[b] | MP2/min CASPT2 |
|---|---|---|---|---|---|---|---|---|
| reactant | 18704 | 21157 | 15903 | 15997 | 17179 | 15566 | 25008 | 30115 |
| A-reaction | 18843 | 20950 | 15847 | 15972 | 17493 | 15574 | 25009 | 30115 |
| D-reaction | 19021 | 21284 | 16622 | 16766 | 17700 | 16387 | 25409 | 30304 |

[b] SAOP results have been obtained with the all-electron ZORA and all-electron Slater triple-$\zeta$ basis sets (See Theory section). [a] For all functionals, except the unrestricted (U-) B3LYP, the TD-DFT results have been used in the table.

exchange is not as pronounced as for the ground state. U- and TD-B3LYP give reaction energies very close to those of the pure functionals, with an average of 23 kJ/mol. As in the ground state, the use of the BHLYP functional increases the energy by 9 kJ/mol. The minimal CASPT2 reaction energy is here 14 kJ/mol higher than that obtained with the pure functionals. The DFT-MRCI, which is based on BHLYP Kohn–Sham orbitals, gives a reaction energy 10 kJ/mol higher than the one received with minimal CASPT2; for the ground state, the picture was different; DFT-MRCI was sharing the behavior of the BHLYP functional and underestimated the reaction energy at 9 kJ/mol. If we look at both the A- and D-reaction, we can see that DFT-MRCI has the same feature as the minimal CASPT2 but shifted up to 10 kJ/mol. The basis set effects and the quality of the RECP are, as expected from the ground-state result, very similar for the two intermediates.

We have reported in Table 5 the adiabatic energies of the luminescent state with respect to the ground state in the reactant and A- and D-intermediates for the DFT functionals and MP2/min CASPT2. There are differences in the absolute transition energy values, as predicted by the various methods. Our group has discussed the reasons for such discrepancies in detail in another publication.[68] In the present context, it is interesting to note that all methods predict that the adiabatic transition energies to the luminescent state are very similar, within a few hundred wave numbers, that is, within 10 kJ/mol. This implies that the reaction potential energy surfaces that connect the reactant to the D- and A-intermediates in the excited-state are essentially parallel to those of electronic ground state.

## Conclusions

In this study, we have made a detailed investigation of the effect of different DFT functionals and the wave-function-based methods MP2 or minimal CASPT2 and CCSD(T) on the associative and dissociative pathways in the water exchange reaction for hydrated uranyl(VI) in the ground state and in the luminescent state. CCSD(T) is the most accurate method used in the present investigation. In our calculations, MP2 and CCSD(T) give very similar results, which provides support for the accuracy of MP2. For the luminescent state, we have, so far, no other accurate method with which to compare our minimal CASPT2 results, but the consistency between the two different CASs (three or five active orbitals) proves that our result does not depend on a wrongly chosen CAS, an argument frequently used for CASPT2 calculations deviating from other accurate methods.

DFT consistently overestimates the reaction energy compared to MP2/minimal CASPT2 and CCSD(T) for the associative pathway, both in the ground state and the luminescent state, but the differences are minor, less than 10 kJ/mol in all cases. For the dissociative pathway, DFT consistently underestimates the reaction energy, by up to 22 kJ/mol for the pure functionals without HF exchange. This is probably an effect of the stronger uranium−water bonds in the D-intermediate. HF exchange is important in the ground-state dissociative reaction, where the difference between MP2 or CCSD(T) and the DFT reaction energies decreases monotonically with an increasing amount of HF exchange in the functionals. The same trend is noted for the luminescent state but not as "clear" as in the ground state.

Looking at the overall picture of the behavior of DFT results compared with wave function methods, see the reaction energy in Figure 2, we can notice that the consistent overestimation and underestimation of the reaction energy for, respectively, the associative mechanism and the dissociative mechanism compared to the wave function result strongly points out that, at this moment, DFT has to be used with caution when reaction mechanisms within actinide chemistry are investigated. The dissociative mechanism is always favored, with 14–29 kJ/mol for the ground state, and for the luminescent state, the corresponding numbers are 10–20 kJ/mol. For the wave function methods, this difference is a few kilojoules per mole.

All DFT functionals have been developed by comparison with lighter elements, and our results show that these functionals are not ideally suited to describe reactions involving actinides where the 5f contribution in the bonds is substantial and the charge distribution on the actinide is significantly perturbed during the reaction. Examples of this are the gas-phase reactions discussed in refs 13 and 14, but also the coordination of uranyl complexes with sulfate ligands.[80] In actinide complexes, with ligands composed of light elements, DFT should behave well as long as the charge distribution on the actinides is not significantly perturbed.

Water interacts fairly strongly with the uranyl ion. That alone has an influence on the uranyl bond distance when the number of coordinated waters is changed. This is enough to generate errors, which in some cases may be significant. The error obtained for B3LYP compared to CCSD(T) or CASPT2 is 10–15 kJ/mol. This is certainly an acceptable result in many cases, but the error can be expected to increase when a metal–ligand bond is strengthened. Moreover, the error has opposite signs for the associative and dissociative mechanisms, and DFT should thus be used with great caution when discussing reaction energetics of actinide complexes.

**Acknowledgment.** We thank Prof. Ingmar Grenthe for fruitful discussions. This study is made within a joint project (JRP 01-12) within the EC supported ACTINET network of excellence. It has been supported by a generous grant from SKB, Swedish Research Council, Laboratoire

de Physique des Lasers, Atomes et Molecules is Unité Mixte de Recherche Mixte du CNRS. The Centre d'Etudes et de Recherches Lasers et Applications is supported by the Ministère chargé de la Recherche, the Région Nord-Pas de Calais, and the Fonds Européen de Développement Economique des Régions. C.D. acknowledges support by a mobility fellowship financed by the Région Nord Pas de Calais. Computational resources have been provided by the National Supercomputer Center in Linköping Sweden (Project 007-05-36); the Institut de Développement et de Ressources en Informatique Scientifique du Centre National de la Recherche Scientifique, IDRIS France (Project 61859); and the Centre Informatique National de l'Enseignement Supérieur, CINES France (Project phl2531).

**Note Added after ASAP Publication.** This article was released ASAP on February 29, 2008 with minor text errors. The correct version was posted on March 12, 2008.

## References

(1) Vallet, V.; Maron, L.; Teichteil, C.; Flament, J.-P. *J. Chem. Phys.* **2001**, *113*, 1391–1402.

(2) Malmqvist, P.-Å.; Roos, B. O.; Schimmelpfennig, B. *Chem. Phys. Lett.* **2002**, *35*, 230–240.

(3) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222–239.

(4) Roos, B. O. In *Advances in Chemical Physics*; Lawley, K. P., Ed.; John Wiley & Sons Ltd.: Chichester, England, 1987; Ab Initio Methods in Quantum Chemistry - II, Chapter 69, page 399.

(5) Malmqvist, P.-Å.; Rendell, A.; Roos, B. O. *J. Phys. Chem.* **1990**, *94*, 5477–5482.

(6) Roos, B. O. *Int. J. Quantum Chem.* **1980**, *17*, 175–189.

(7) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(8) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218–1226.

(9) Gagliardi, L.; Roos, B. O. *Chem. Soc. Rev.* **2007**, *36*, 893–903.

(10) Wahlgren, U.; Tsushima, S.; Grenthe, I. *J. Phys. Chem. A.* **2006**, *110*, 9025–9027.

(11) Grimme, S.; Waletzke, M. *J. Chem. Phys.* **1999**, *111*, 5645–5655.

(12) Vallet, V.; Schimmelpfennig, B.; Maron, L.; Teichteil, C.; Leininger, T.; Gropen, O.; Grenthe, I.; Wahlgren, U. *Chem. Phys.* **1999**, *244*, 185–193.

(13) Privalov, T.; Schimmelpfennig, B.; Wahlgren, U.; Grenthe, I. *J. Phys. Chem. A* **2002**, *106*, 11277–11282.

(14) Schimmelpfennig, B.; Toraishi, T.; Wahlgren, U.; Grenthe, I. *J. Phys. Chem. A* **2003**, *107*, 9456–9462.

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

(16) Batista, E. R.; Martin, R. L.; Hay, P. J. *J. Chem. Phys,* **2004**, *121*, 11104–11111.

(17) Peralta, J. E.; Batista, E. R.; Scuseria, G. E.; Martin, R. L. *J. Chem. Theory. Comput.* **2005**, *1*, 612–616.

(18) Küchle, W.; Dolg, M.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1994**, *100*, 7535–7542.

(19) Spencer, S.; Gagliardi, L.; Handy, N. C.; Ioannou, A. G.; Skylaris, C.-K.; Willetts, A. *J. Phys. Chem. A* **1999**, *103*, 1831–1837.

(20) Hay, P. J.; Martin, R. L.; Schreckenbach, G. *J. Phys. Chem. A* **2000**, *104*, 6259–6270.

(21) Tsushima, S.; Suzuki, A. *J. Mol. Struct.* **2000**, *529*, 21–25.

(22) Vallet, V.; Wahlgren, U.; Schimmelpfennig, B.; Szabó, Z.; Grenthe, I. *J. Am. Chem. Soc.* **2001**, *123*, 11999–12008.

(23) Clavaguéra-Sarrio, C.; Brenner, V.; Hoyau, S.; Marsden, C. J.; Millié, P.; Dognon, J.-P. *J. Phys. Chem. B* **2003**, *107*, 3051–3060.

(24) Moskaleva, L. V.; Krüger, S.; Spörl, A.; Rösch, N. *Inorg. Chem.* **2004**, *43*, 4080–4090.

(25) Cao, Z.; Balasubramanian, K. *J. Chem. Phys.* **2005**, *123*, 114309–114321.

(26) Shamov, G. A.; Schreckenbach, G. *J. Phys. Chem. A* **2005**, *109*, 10961–10974.

(27) Bühl, M.; Diss, R.; Wipff, G. *J. Am. Chem. Soc.* **2005**, *127*, 13506–13507.

(28) Gutowski, K. E.; Dixon, D. A. *J. Phys. Chem. A* **2006**, *110*, 8840–8856.

(29) Bühl, M.; Kabrede, H.; Diss, R.; Wipff, G. *J. Am. Chem. Soc.* **2006**, *128*, 6357–6368.

(30) Bühl, M.; Kabrede, H. *Inorg. Chem.* **2006**, *45*, 3834–3836.

(31) Rotzinger, F. P. *Chem.—Eur. J.* **2007**, *13*, 800–811.

(32) Dirac, P. A. M. *Proc. R. Soc. London, Ser. A* **1929**, *123*, 714–733.

(33) Slater, J. C. *Phys. Rev.* **1951**, *81*, 385–390.

(34) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.

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

(36) Scheiner, A. C.; Scuseria, G. E.; Rice, J. E.; Lee, T. J.; Schaefer, H. F., III *J. Chem. Phys.* **1987**, *87*, 5361–5373.

(37) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098–3100.

(38) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.

(39) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200–206.

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

(41) Perdew, J. P.; Wang, Y. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *45*, 13244–13249.

(42) Vallet, V.; Wahlgren, U.; Szabó, Z.; Grenthe, I. *Inorg. Chem.* **2002**, *41*, 5626–5633.

(43) Vallet, V.; Moll, H.; Wahlgren, U.; Szabó, Z.; Grenthe, I. *Inorg. Chem.* **2003**, *42*, 1982–1993.

(44) Vallet, V.; Privalov, T.; Wahlgren, U.; Grenthe, I. *J. Am. Chem. Soc.* **2004**, *126*, 7766–7767.

(45) Farkas, I.; Bányai, I.; Szabó, Z.; Wahlgren, U.; Grenthe, I. *Inorg. Chem.* **2000**, *39*, 799–805.

(46) Watts, J. D.; Gauss, J.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 8718–8733.

(47) Knowles, P. J.; Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1993**, *99*, 5219–5227. See also erratum: *J. Chem. Phys.* **2000**, *112*, 3106–3107.

(48) Vallet, V.; Wahlgren, U.; Grenthe, I. *J. Am. Chem. Soc.* **2003**, *125*, 14941–14950.

(49) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(50) Vosko, S.; Wilk, L.; Nussair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(51) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822–8824.

(52) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(53) Schipper, P. R. T.; Gritsenko, O. V.; van Gisbergen, S. J. A.; Baerends, E. J. *J. Chem. Phys.* **2000**, *112*, 1344–1352.

(54) Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2002**, *117*, 7433–7447.

(55) Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2004**, *121*, 12772–12773.

(56) Bauernschmitt, R.; Ahlrichs, R. *Chem. Phys. Lett.* **1996**, *256*, 454–464.

(57) Douglas, D.; Kroll, N. M. *Ann. Phys.* **1974**, *82*, 89–155.

(58) Hess, B. *Phys. Rev. A: At., Mol., Opt. Phys.* **1986**, *33*, 3742–3748.

(59) Cao, X.; Dolg, M. *J. Mol. Struct.* **2004**, *673*, 203–209.

(60) Roos, B. O.; Lindh, R.; Malmqvist, P.-Å.; Veryazov, V.; Widmark, P. O. *Chem. Phys. Lett.* **2005**, *409*, 295–299.

(61) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

(62) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(63) Vallet, V.; Macak, P.; Wahlgren, U.; Grenthe, I. *Theor. Chem. Acc.* **2006**, *115*, 145–160.

(64) van Lenthe, E.; Baerends, E. J.; Snidjers, J. G. *J. Chem. Phys.* **1993**, *99*, 4597–4610.

(65) van Lenthe, E.; Baerends, E. J.; Snidjers, J. G. *J. Chem. Phys.* **1994**, *101*, 9783–9792.

(66) van Lenthe, E.; Ehlers, A. E.; Baerends, E. J. *J. Chem. Phys.* **1999**, *110*, 8943–8953.

(67) van Lenthe, E.; Snijders, J. G.; Baerends, E. J. *J. Chem. Phys.* **1996**, *105*, 6505–6516.

(68) Réal, F.; Vallet, V.; Marian, C. M.; Wahlgren, U. *J. Chem. Phys.* **2007**, *127*, 214302.

(69) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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*, revision D.02; Gaussian, Inc.: Wallingford, CT, 2006.

(70) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(71) *TURBOMOLE*, v.5-7-1. http://www.turbomole.com (accessed Dec 12, 2007).

(72) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931–967.

(73) van Gisbergen, S. J. A.; Snijders, J. G.; Baerends, E. J. *Comput. Phys. Commun.* **1999**, *118*, 119–138.

(74) *Amsterdam Density Functional (ADF)*; SCM, Theoretical Chemistry, Vrije Universiteit: Amsterdam, The Netherlands. http://www.scm.com (accessed Dec 12, 2007).

(75) Werner, H. J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. O. J.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R. M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO*, version 2006.1, a package of ab initio programs. See http://www.molpro.net (accessed Feb 2008).

(76) Rotzinger, F. P. *J. Phys. Chem. B* **2005**, *109*, 1510–1527.

(77) Wesolowski, T. A.; Parisel, O.; Ellinger, Y.; Weber, J. *J. Phys. Chem. A* **1997**, *101*, 7818–7825.

(78) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I. *J. Phys. Chem. A* **2005**, *109*, 11015–11021.

(79) Vallet, V.; Wahlgren, U.; Grenthe, I. *Chem.—Eur. J.* **2007**, *13*, 10294–10297.

(80) Vallet, V.; Grenthe, I. *C. R. Chim.* **2007**, *10*, 905–915.

CT700062X

JCTC Journal of Chemical Theory and Computation

# Response of Scalar Fields and Hydrogen Bonding to Excited-State Molecular Solvation of Carbonyl Compounds

Anant D. Kulkarni,* Benedetta Mennucci,* and Jacopo Tomasi

*Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Risorgimento 35, 56126 Pisa, Italy*

Received September 21, 2007

**Abstract:** An attempt has been made to understand the mechanism of excited-state molecular solvation and its effect on hydrogen bonding in carbonyl compounds in aqueous solution. The correlation between solvation and electronic transitions has been investigated by comparing results obtained either with a supermolecular description in terms of hydrogen-bonded clusters or with a combined method embedding such clusters with a polarizable continuum dielectric mimicking the bulk water. Popular scalar fields such as molecular electrostatic potential and molecular electron density have been used as useful tools to probe the changes in the hydrogen bonding passing from ground to excited states in the gas as well as solvent phase.

## 1. Introduction

Water is the omnipresent solvent in almost all important chemical and biological processes. Thus, it is no wonder that the study of molecular hydration, dealing with the interaction of water with other molecules, has become an area of prime importance in chemistry and biology.[1,2] The hydration of molecules containing a carbonyl group has been the focus of several studies.[3–15] In particular, the carbonyl double bond shows spectroscopically important transitions such as $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$, which are well-known examples of environment-sensitive processes. Two prototype examples of molecules containing carbonyl groups, namely, formaldehyde ($H_2CO$) and urea ($CO(NH_2)_2$), are considered in the present study.

The earlier studies on formaldehyde and water (in the ground state) complexes at the HF level were reported by Tomasi and co-workers[5a] with reference to the environmental effects on biomolecules, followed by yet another *ab initio* study[5b] discussing the solvent shift of electron absorption spectra. Later on, 1:1 complexes of formaldehyde···water systems were studied at the MPn[6a,b] and CCSD level.[6c] Wolfe et al.[7] performed *ab initio* calculations on the hydration of formaldehyde with up to four water molecules and reported the thermochemical parameters and vibrational frequencies. More recently, combined quantum mechanics/molecular mechanics approaches have been used to calculate molecular response properties and excitation energies of formaldehyde in water.[8]

The water-soluble nature of urea is well-known. The effect of the molecular environment on the optical susceptibility of urea in solution has been studied by Ledoux and Zyss.[9] According to the earlier studies[10] based on MD simulations, the urea molecule can enter the cluster of water molecules without appreciable distortion in the hydrogen-bonded network of water molecules. A clear conclusion about the hydrophobic or hydrophilic nature of urea could not be drawn from this work. A detailed study addressing the solvent and the vibrational effect on static and dynamic polarizabilty as well as hyperpolarizability within the quantum chemical framework has been reported by our group.[11] Lee et al.[12] performed density functional theory (DFT) computations on the complexes between urea, water, and urea dimers, using VWN and BLYP functionals and DZ94 and DZ94P basis sets. According to this study, the urea dimer will be unstable due to the interaction between the urea and water molecules when surrounded by a large number of water molecules. Åstrand and co-workers[13] employed the empirical-potential-based as well as the quantum-chemical approach to study urea···water complexes. The most favorable structure for the urea···water complex has two hydrogen bonds and is cyclic in nature.

* Corresponding author. E-mail: anantkul@dcci.unipi.it and anantkul@chem.unipune.ernet.in (A.D.K.), bene@dcci.unipi.it.

10.1021/ct7002429 CCC: $40.75 
Published on Web 02/29/2008

Following this historical summary, a brief preamble to the scalar-field-based studies is useful since this study involves the extensive use of popular scalar fields, namely, molecular electrostatic potential (MESP, also denoted by $V_{es}$) and molecular electron density (MED also denoted by $\rho(\mathbf{r})$).

Bader and co-workers pioneered the topographical studies of MED.[16] Also, Gatti and co-workers[17] performed extensive studies involving molecular electron-density-based approaches. The use of MESP as a useful tool to predict the molecular properties was first advocated by Scrocco and Tomasi[18a] more than 3 decades ago. Tomasi and co-workers[18b] have summarized the applications of MESP for studying the intermolecular interactions in great detail by highlighting the important aspects of molecular structure and solvation effects. Politzer and Thruhlar[19] have also advocated the utility of a MESP-based approach through various studies. Extensive topographical investigations of MED and MESP have been carried out by Gadre and co-workers[20–22] to understand hydrogen-bonded interactions over the past decade or so.

In view of this, it is worthwhile to study the excited-state molecular solvation process and the corresponding spectroscopic properties and to probe the correlation between this process and hydrogen-bonding interactions using MESP and MED, the popular scalar fields. In particular, the following questions will be addressed:

What is the effect of an electronic excitation on the hydrogen bonding of explicitly bound water molecules in the ground state?

Can one use the information from scalar fields such as MESP and MED to understand this process?

What are the effects on this process due to the solvent, treated as an external, continuous medium, thereby taking into account the bulk solvation effect?

The present paper is organized as follows: section 2 deals with methodological aspects employed, and the results and discussion are presented in section 3, where numerical results are presented and discussed. Section 4 concludes the article and also presents future extensions of the present study.

## 2. Methodology

The structures of M···$H_2O$ clusters (where M = formaldehyde and urea) were generated from the knowledge of earlier studies[6,22] (both in the gas and solution phases) and the cooperativity offered by MESP.[20–22] The geometry optimizations for the ground and excited states, in the gaseous state as well as in the solution (water) state, were achieved with the Gaussian program package.[23] The ground states were first obtained at the HF/6-31G(d,p) level, whereas the excited-state geometries were computed employing the simple *ab initio* configuration interaction, including only single excited configurations (CIS).

In order to enhance the level of sophistication, we have also employed DFT. Application of DFT-based methods in studying hydrogen-bonded complexes could be a question of debate. However, the success of using a DFT approach for studying hydrogen-bonded interactions has been highlighted by several researchers.[24] The popularity of these methods may be attributed to the fact that they model the exchange correlation effects, at a considerably lower computational cost than other correlated methods.[24] As highlighted by Dreuw and Head-Gordon,[25] time-dependent DFT (TD-DFT) has emerged as a reliable and prominent method to study the excited-state properties of medium- and large-sized molecular systems. Also, the earlier studies,[26,27] by other researchers and by one of the present authors, have demonstrated the reliability of TD-DFT-based methods for studying the excited states. Taking a cue from these studies, we have used the B3LYP functional, which is a combination of Becke's three-parameter functional[28] with the correlation functional of Lee, Yang, and Parr.[29] In order to increase the level of sophistication, we have employed a split valence triple-$\zeta$ basis set augmented with a diffuse and polarization function, 6-311++G(d,p), along with the time-independent and time-dependent versions of density functional theory using B3LYP and TD B3LYP frameworks[30,31] for ground- and excited-state properties, respectively. The extended basis set employed in this work, that is, 6-311++G(d,p), has been used in the (6d) version as implemented in Gaussian.[23] The molecular orbitals were visualized using GVIEW,[32] to follow the electronic transitions.

**2.1. Solvation Study.** The present study employs the integral equation formalism−polarizable continuum model (*IEF-PCM*) solvation model,[33] an accurate reformulation of the PCM model[34] due to Tomasi and co-workers. The IEF-PCM model involves mimicking the solvent in a dielectric continuum fashion with a dielectric constant ($\epsilon$) which surrounds the molecular cavity, with the shape and dimension adjusted on the real geometric structure of the solute molecule. The latter polarizes the solvent and induces an electric field (the "reaction field") that interacts with the solute. In the IEF-PCM model, the solute–solvent electrostatic interaction is represented in terms of an apparent charge density spreading on the cavity surface, giving rise to specific operators which contribute to the Hamiltonian of the isolated system (to obtain the final effective Hamiltonian). Such solvent terms depend on the solute wave function they contribute to modify, and thus, the problem requires the solution of a modified self-consistent field scheme.

In general, the solute electronic and nuclear charge distribution and solvent reaction field are allowed to mutually equilibrate. However, when the solute undergoes an abrupt change of its electronic state through a vertical transition, the relaxation of the reaction field in the direction of the new solute state may be incomplete (nonequilibrium solvation). Considering the typical time scales characterizing electronic and nuclear (or molecular) motions, we assume that only the part of the solvent reaction which is induced by the polarization of its electrons can immediately be modified according to the new electronic state, which is reached by the solute in the transition process. The other part in the system remains frozen under the previous equilibrium conditions as determined by the initial state. In a reasonable approximation, the fast component can be taken proportional to the dielectric constant at infinite frequency $\epsilon_\infty$ where $\epsilon_\infty \approx n^2$ and $n$ is the refractive index of the solvent. In IEF-PCM framework, this scheme is implemented by introducing two sets of apparent charges representing the

electronic (or fast) and the slow contributions of the solvent reaction. IEF-PCM has been generalized to CIS and TD-DFT calculations which also include solvent perturbation operators in the coupled perturbed scheme; for more details on the formalism, see refs 35 and 36. The values used in the following calculations for the $\epsilon$ and $\epsilon_\infty$ of water are 78.39 and 1.776, respectively.

**2.2. MESP and MED Analysis and Visualization.** The present work involves extensive use of MESP and MED and their topographical analysis in terms of critical points (CPs). The CPs provide valuable information about the structure, bonding features, and environment of the molecule. The details of electrostatic potential, electron density, topography, and related concepts may be found elsewhere.[14,16,20,21] MESP and MED over the regular grids enclosing the molecular system under study were computed using the Gaussian package.[23] The topographical analysis for various electronic states was performed by following the MOs for each specific transition and the density matrices computed at respective levels employing the UNIPROP[37] and UNIVIS-2000[38] packages. In the present study, we have employed the information provided by (3, −1) MESP and MED CPs computed in the bonding region between the two molecules. The Laplacian of the electronic density ($\nabla^2\rho(\mathbf{r})$) has also been used to understand the bonding features in this study.[16] The negative value of $\nabla^2\rho(\mathbf{r})$ implies a preponderance of electron density and hence is an indicator of a covalent bond, whereas its positive value indicates nonbonding or closed-shell interaction such as hydrogen bonds between the two atoms.[16–20] The geometry generation and visualization of scalar fields were carried out with the help of the versatile visualization package, UNIVIS-2000.[38]

## 3. Results and Discussion

The results of the quantum chemical investigation of the effects of H bonds on excited states in terms of a combined molecular solvation approach (i.e., a supermolecular description involving hydrogen-bonding effects and the bulk effects as represented by a continuum solvation) are discussed as follows.

The analysis is focused on $n-\pi^*$ excited states of the gas phase and solvated M···$H_2O$ clusters studied at the HF/6−31G(d,p) and CIS/6–31G(d,p) levels (referred to as level I); B3LYP/6−311++G(d,p) and TDB3LYP/6−311++G(d,p) (referred to as level II). The additional results for $\pi \rightarrow \pi^*$ transitions are also reported as a comparison.

In order to probe the geometrical and energetic changes in the excited-state formation/relaxation, we consider the excitation process to occur via a stepwise mechanism. In the gas phase, such a mechanism is clear: a relaxed ground state is first vertically excited into a Franck–Condon (FC) (i.e., geometrically unrelaxed) excited state, which then relaxes into a completely equilibrated excited state.

As discussed in section 2.1., in solution, the picture is more difficult as two different solvation regimes have to be considered when fast modifications in the solute occur. Here, these two regimes give rise to the following process: initially an equilibrated ground state is vertically excited into a FC state embedded in a nonequilibrated solvent (step 1); successively, the solvent relaxes toward an equilibrium regime with respect to the FC state (step 2), and finally both the solute and solvent reach a completely relaxed state (step 3). This schematization assumes that solvent dielectric relaxation proceeds faster than the solute geometrical relaxation. An alternative approach would be that of deleting step 2 and assuming that solute and solvent relaxations proceed contemporaneously. Both of these approaches are clearly approximations of the real process, but they are useful in the present context as they provide insights about solvation effects on excited-state formation and relaxation.

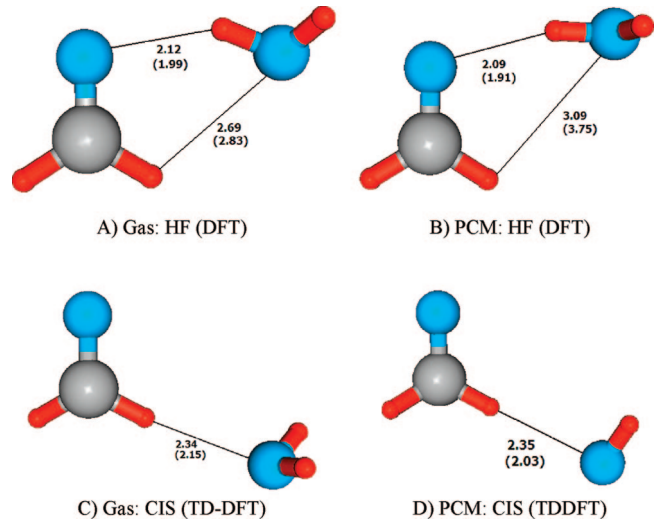


***Figure 1.*** Schematic representation of gas- and solvent-phase structures of ground (A and B) and $n-\pi^*$ excited (C and D) states of *formaldehyde···water* complexes at HF-CIS and DFT−TD-DFT levels. All of the distances are in angstroms, and the values in parentheses indicate the distances computed at the B3LYP/6−311++G(d,p) level of theory. Refer to the text for details.

Figures 1 and 2 present the schematic representation of the most stable structures obtained for the ground and relaxed $n-\pi^*$ excited states at level I and level II for the gas and solvent phases, whereas in Table 1, we report the corresponding vertical excitation energies (for level I $\pi-\pi^*$, data are also reported).

**3.1. Gas-Phase H-Bonded Clusters.** For both systems, a cyclic structure with two hydrogen bonds is found in the ground state, while the geometrical relaxation of the excited state results in overall modification of the hydrogen-bonding picture by pushing the water molecule away from the carbonyl oxygen (toward the hydrogen of formaldehyde and urea). Going into a more detailed comparison between the two systems, we find that H-bond distances are shorter in urea both for the CO···HOH in the ground state and the H···$OH_2$ in the relaxed excited state. Such a picture does not change passing from one to the other level of calculations. However, DFT computations predict a shortening of the hydrogen-bond lengths for both the ground and the relaxed excited states.

Moving to the absorption process, it may be seen from Table 1 that the vertical excitation energies for $n \rightarrow \pi^*$ transitions are always lower than the corresponding $\pi \rightarrow \pi^*$ transitions. Also, TD-DFT results are comparatively lower than their

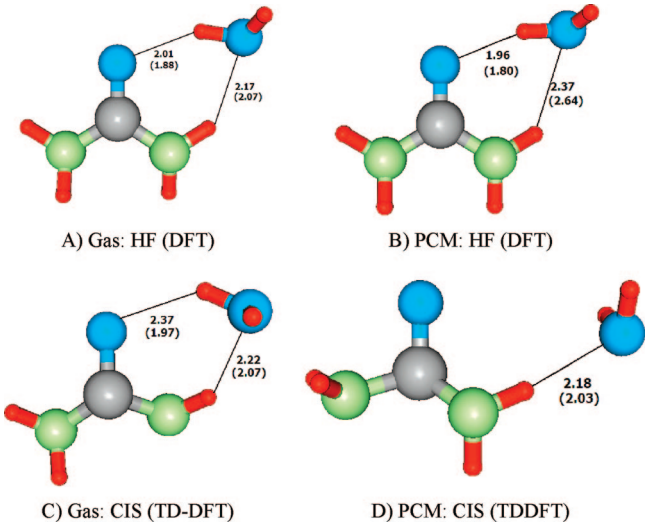


***Figure 2.*** Schematic representation of gas- and solvent-phase structures of ground (A and B) and $n-\pi^*$ excited (C and D) states of *urea···water* complexes at HF-CIS and DFT−TD-DFT levels. All of the distances are in angstroms, and the values in parentheses indicate the distances computed at the B3LYP/6−311++G(d,p) level of theory. Refer to the text for details.

***Table 1.*** Excitation Energies ($n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ Transitions) for M···$H_2O$ Clusters (M = Formaldehyde and Urea) in the Gas Phase and in Water[a]

| | | level I[b] | | level II[c] |
|---|---|---|---|---|
| molecular system | gas/solvent | $n \rightarrow \pi^*$ | $\pi \rightarrow \pi^*$ | $n \rightarrow \pi^*$ |
| formaldehyde···$H_2O$ | gas | 4.91 | 10.57 | 4.09 |
| | water | 5.02 | 10.50 | 4.17 |
| urea···$H_2O$ | gas | 8.36 | 9.97 | 7.17 |
| | water | 8.77 | 10.35 | 7.27 |

[a] Excitation energies are in electrovolts. [b] CIS/6-31G(d,p) level. [c] TD B3LYP/6−311++G(2d,2p) level. (Refer to the text for further details.)

corresponding values at the CIS level toward a better agreement with experiments (see below). Here, however, it is more interesting to quantify the effect of the H bonds as described by the two levels of calculations.

In $n \rightarrow \pi^*$ transitions, the electronic density on the heteroatom (here, the carbonyl oxygen) decreases upon excitation. As a result, the tendency of the heteroatom to form hydrogen bonds is reduced with respect to the ground state. Introducing the effect of the H-bonded water should thus lead to higher absorption energy (i.e., a blue-shift), and the stronger the hydrogen bond is, the larger is the shift.

To verify this picture, we have optimized the structures of isolated (i.e., in the absence of H-bonded water) formaldehyde and urea and calculated the $n-\pi^*$ excitation energies (this analysis has been limited to level I). The results obtained are 4.73 and 7.79 eV for formaldehyde and urea, respectively; by combining these results with the data reported in Table 1 for the gas-phase M···$H_2O$ clusters, we obtain blue-shifts of 0.18 and 0.57 eV for formaldehyde and urea, respectively.

Correlating these shifts with the H-bond distances reported in Figures 2 and 3, it may be seen that our computations perfectly reproduce the criterion which is generally used in

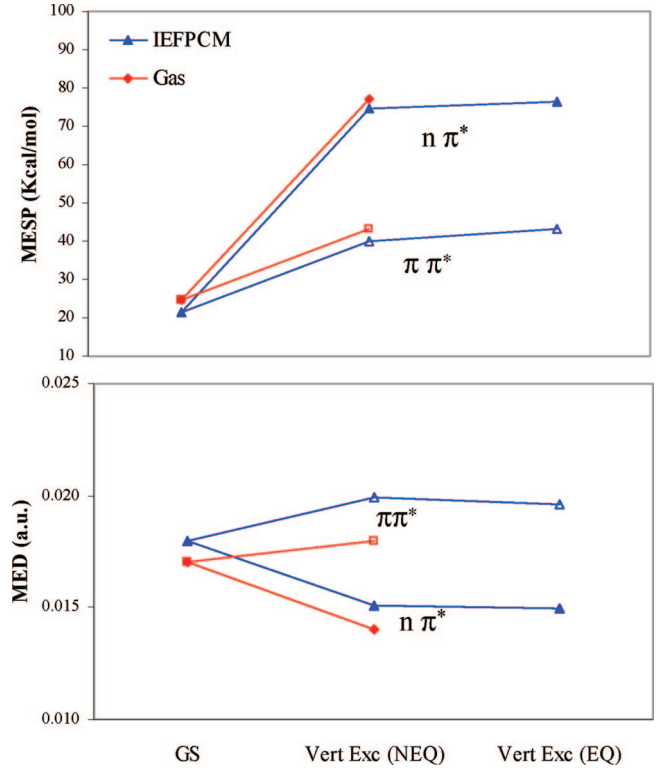


***Figure 3.*** MESP (upper panel) and MED (lower panel) values for formaldehyde···$H_2O$ clusters at the RHF-CIS/6−31G(d,p) level in the ground and vertical excited states ($n\pi^*$ and $\pi\pi^*$) in the gas phase and in water. For the solvated systems, two vertical excited states are reported corresponding to nonequilibrium and equilibrium solvation, respectively.

the analysis of the shift in the $n \rightarrow \pi^*$ band for determining the energy of the hydrogen bond.

The $\pi \rightarrow \pi^*$ transitions show an altogether different behavior as compared to that exhibited by $n \rightarrow \pi^*$. In case of these transitions, it is generally observed that the heteroatom is more basic in the excited state than in the ground state. The resulting excited molecule thus involves a stronger hydrogen bond as compared to the ground state, and a red-shift is expected for the excitation energies.

As a result of this, we expect the clusters with stronger H bonds to give smaller transition energies (i.e., larger red-shifts). By comparing $\pi \rightarrow \pi^*$ transition energies of the isolated formaldehyde and urea (10.58 and 10.09 eV) with those reported in the table for the M···$H_2O$ systems, we see that, once again, the predicted behavior is confirmed by the calculations in which urea shows the largest red-shift.

**3.2. Solvated H-Bond Clusters.** A look at the Figures 1 and 2 reveals that the solvent effect as incorporated by IEF-PCM leads to a shortening of (C)O···H(OH) hydrogen bonds for the ground states of both systems at both levels of calculations. This shortening indicates strengthening of the H bond due to the effects of the additional water molecules (here, represented by the polarizable continuum dielectric).

In the case of the completely relaxed excited states, we observe structures with a nonplanar solute and only one hydrogen bond, similarly to what was found in the gas phase. The larger urea···$H_2O$ system presents a larger nonplanarity due to rotation around the C−N bond. This change may be

regarded as a signature of solvation in terms of structural deformation.

The differences in the geometrical structures of ground and excited states passing from the gas phase to solvated clusters are reflected in the shifts in the excitation energies. The introduction of the continuum solvent leads to a significant blue-shift for $n \rightarrow \pi^*$ transitions and a smaller red-shift for $\pi \rightarrow \pi^*$ transitions. This result can be better explained considering in more details all of the possible effects that the solvent has on the absorption process.

In the previous section, we quantified the effect of H bonds on the $n \rightarrow \pi^*$ transition in terms of the blue-shifts with respect to the isolated molecule. Here, we can extend such an analysis by adding the effect of all the other solvent molecules (i.e., the bulk of the solvent). Such a bulk effect acts in a 2-fold way: it modifies the structure of the H-bonded clusters, as commented above, and it directly affects the ground and excited states by differentially stabilizing them. As a result of these two effects, the excitation energies will be changed with respect to the gas-phase clusters. From Table 1, we see that the net bulk effect leads to further blue-shifts of 0.11 and 0.41 eV for formaldehyde and urea at level I (this picture remains consistent at level II but with smaller shifts).

At this stage it is felt worthwhile to dissect these shifts in terms of their (structural and direct) components. This is easily obtained by recalculating the excitation energies of the solvated clusters but, this time, keeping their structure fixed in the gas-phase geometry. The results obtained are 5.08 and 8.69 eV for formaldehyde and urea, respectively; these data when compared with the corresponding ones reported in Table 1 (namely, 5.08 and 8.77 eV) show that the effects the bulk solvent has on the geometry of the H-bonded systems (and indirectly on the excitation energies) are negligible for formaldehyde, whereas they lead to a further blue-shift (of 0.08 eV) for urea.

The $\pi \rightarrow \pi^*$ transitions show an altogether different behavior as compared to that exhibited by $n \rightarrow \pi^*$. In case of these transitions, we have previously observed that the heteroatom is more basic in the excited state than in the ground state. The resulting excited molecule thus involves a stronger hydrogen bond as compared to the ground state, and as a result, a red-shift on the absorption of the cluster is found with respect to the isolated molecule. Within this picture, the addition of the bulk effects should induce a further red-shift: this is confirmed by our calculations in which IEF-PCM $\pi \rightarrow \pi^*$ energies are always smaller than the corresponding ones calculated in the gas-phase clusters.

Even if the present work is not primarily aimed at a comparison with experiments, it is useful to conclude this analysis with some information on the observed spectra of these compounds. Although the pure electronic $n-\pi^*$ transition of formaldehyde is forbidden, a band due to the coupling with vibrational modes mainly involving the out-of-plane bending motion is observed in the experiments.[38] The experimental value[38] for the gas phase is 3.5–4.0 eV. An experimental value of the blue-shift for monomeric formaldehyde in water is not available due to

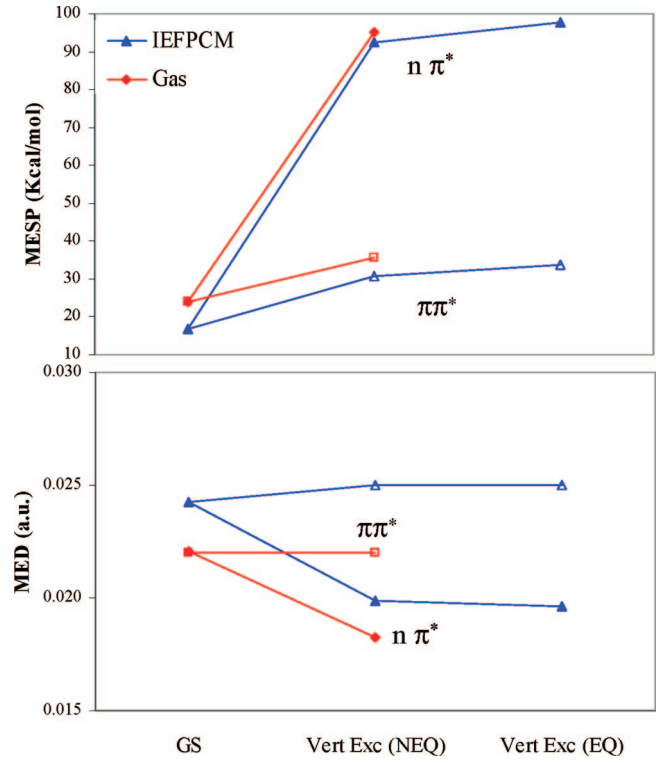


***Figure 4.*** MESP (upper panel) and MED (lower panel) values for urea···$H_2O$ clusters at the RHF-CIS/6−31G(d,p) level in the ground and vertical excited states ($n\pi^*$ and $\pi\pi^*$) in the gas phase and in water. For the solvated systems, two vertical excited states are reported corresponding to nonequlibrium and equilibrium solvation, respectively.

the formation of oligomers. However, it is likely to be around 0.23 eV, the value observed for acetone. The results reported in Table 1 when compared to the values of the free molecule in the gas phase (4.73 eV at level I and 3.96 eV at level II) lead to shifts on the order of 0.16 eV (level I) and 0.13 eV (level II) for the isolated supermolecule and 0.26 eV (level I) and 0.21 eV (level II) for the solvated (IEF-PCM) one.

To the best of our knowledge, UV absorption spectra of urea have not been investigated in depth, and the electronic transitions of the urea chromophore have not been determined either experimentally or theoretically.

**3.3. MESP and MED Analysis.** Figures 3 and 4 depict the graphical comparison of the MESP and MED values for $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ (the electrostatic potential values in the graphs are converted into kcal/mol counterparts) in the gas phase and in solution. The details of the $n \rightarrow \pi^*$ transition for the gas phase and solvent phase are summarized in Tables 2 and 3, respectively.

The variation of MESP plotted in a plane containing the molecular system and hydrogen bonds is finally depicted in Figure 5 for solvated clusters at the ground, vertical excitation, and completely relaxed $n-\pi^*$ excited states. The positive-valued electrostatic potential is indicated by the red color and the negative valued potential by blue. The white region between the molecules stands for the small or zero-valued potential value or bonding region of the cluster.

The trends in the MESP and MED for formaldehyde···$H_2O$ and urea···$H_2O$ complexes are qualitatively similar. The $-NH_2$

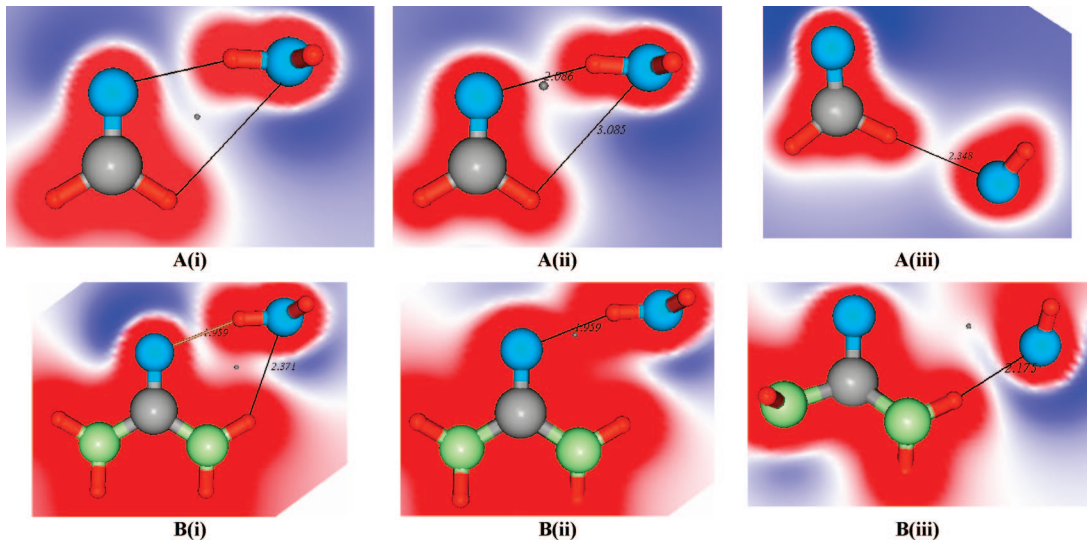


***Figure 5.*** Schematic representation of structures and MESP mapped on the planes containing molecules and hydrogen bonds computed for (A) formaldehyde···$H_2O$ and (B) urea···$H_2O$ complexes at the RHF-CIS/6−31G(d,p) level in the (i) ground state (solvent), (ii) vertical excitation state (solvent), and (iii) fully relaxed excited states (solvent).

***Table 2.*** Details of MESP and MED Critical Points (3, −1) for $n \rightarrow \pi^*$ Transition of M···$H_2O$ clusters (M = Formaldehyde and Urea) in the Gas Phase[a]

| | level I | | | level II | | |
|---|---|---|---|---|---|---|
| | ground state | vertical exc. state | relaxed exc. state | ground state | vertical exc. state | relaxed exc. state |
| | | | (A) Formaldehyde···Water Complex | | | |
| MESP | 0.039 | 0.123 | 0.039 | 0.069 | 0.157 | 0.007 |
| $\rho(\mathbf{r})$ | 0.017 | 0.014 | 0.012 | 0.023 | 0.019 | 0.016 |
| $\nabla^2\rho(\mathbf{r})$ | 0.057 | 0.053 | 0.037 | 0.082 | 0.077 | 0.057 |
| | | | (B) Urea···Water Complex | | | |
| MESP | 0.038 | 0.151 | 0.029 | 0.063 | 0.225 | 0.069 |
| $\rho(\mathbf{r})$ | 0.016 | 0.017 | 0.007 | 0.026 | 0.027 | 0.024 |
| | 0.022 | 0.018 | 0.015 | | | |
| $\nabla^2\rho(\mathbf{r})$ | 0.071 | 0.065 | 0.049 | 0.103 | 0.107 | 0.085 |

[a] MESP and MED values at the CPs are in atomic units (au). Level I: CIS/6−31G(d,p). Level II: TD B3LYP/6−311++G(2d,2p). (Refer text for further details.)

***Table 3.*** Details of MESP and MED Critical Points (3, −1) for $n \rightarrow \pi^*$ Transition of M···$H_2O$ Clusters (M = Formaldehyde and Urea) in Water[a]

| | level I | | | | level II | | | |
|---|---|---|---|---|---|---|---|---|
| | ground state | vertical exc. state (NEQ) | vertical exc. state (EQ) | relaxed exc. state | ground state | vertical exc. state (NEQ) | vertical exc. state (EQ) | relaxed exc. state |
| | | | | (A) Formaldehyde···Water Complex | | | | |
| MESP | 0.034 | 0.119 | 0.122 | 0.025 | 0.069 | 0.150 | 0.155 | 0.035 |
| $\rho(\mathbf{r})$ | 0.018 | 0.015 | 0.015 | 0.015 | 0.026 | 0.023 | 0.023 | 0.023 |
| $\nabla^2\rho(\mathbf{r})$ | 0.056 | 0.052 | 0.052 | 0.035 | 0.095 | 0.094 | 0.095 | 0.074 |
| | | | | (B) Urea···Water Complex | | | | |
| MESP | 0.027 | 0.148 | 0.156 | −0.009 | 0.059 | 0.106 | 0.115 | 0.009 |
| $\rho(\mathbf{r})$ | 0.024 | 0.019 | 0.019 | 0.017 | 0.035 | 0.035 | 0.035 | 0.019 |
| $\nabla^2\rho(\mathbf{r})$ | 0.077 | 0.070 | 0.070 | 0.048 | 0.117 | 0.122 | 0.120 | 0.069 |

[a] MESP and MED values at the CPs are in atomic units (au). Level I: CIS/6−31G(d,p). Level II: TD B3LYP/6−311++G(2d,2p). (Refer text for further details.)

groups in the case of urea have marginal effects on the hydrogen-bond-forming ability. The results in Figures 2 and 3 indicate that the MESP CP becomes more positive on $n \rightarrow \pi^*$ excitation, whereas for $\pi \rightarrow \pi^*$ vertical transition, this change is rather small due to marginal perturbation in the H-bonding environment. In the case of electron density, the $n \rightarrow \pi^*$ excitation show a decrease in the transition from the ground to the vertical excited state but only a very small increase for the $\pi \rightarrow \pi^*$ transition.

In general, it may be noticed that the CP values computed within the DFT framework (level II) are higher than their counterparts within the HF framework (level I). The reason for the differences in the absolute values may be attributed to the difference in the molecular geometries at the HF and DFT/B3LYP levels. In the case of the DFT level, the proton of $H_2O$ was pulled more toward the carbonyl oxygen, and the angle of approach has been found to be larger at the HF level (see Figures 1 and 2).

The inclusion of the external IEF-PCM continuum does not change the qualitative behavior of MESP and MED values observed in the gas phase, but it affects the absolute values.

In greater detail, the first step of the excitation process (namely, the vertical excitation with a nonequilibrium solvent) presents the same features observed in the gas phase: MESP values at a CP increase of ~3–4 times with respect to the ground-state value for level I (and ~2 times for level II). These CP values show a further small increment when solvent is allowed to become equilibrated. Once again, the final relaxation of the solute geometry (in equilibrium with the solvent) is accompanied by a significant reduction of the MESP to a value nearly close to the ground-state one. The sudden modification of electrostatic potential values while going from the ground state to different excited states was reported previously by Shukla and Leszczynski,[27b] whereas preliminary analyses on the effects of excitations on MESP were given by Tomasi and co-workers more than 3 decades ago.[41]

As a final analysis, let us consider $\nabla^2\rho(\mathbf{r})$ values evaluated at the MED CP (see Tables 2 and 3). We recall that negative values of $\nabla^2\rho(\mathbf{r})$ at the MED CP have been used as an indication of a covalent bond, whereas positives value are an indicator of noncovalent or hydrogen-bonded character. From the results reported in Tables 1 and 2, we have confirmation of this indication: $\nabla^2\rho(\mathbf{r})$ values evaluated at the MED CP are all positive for both systems either in the gas phase or in water. However, for this property, the effect of excitation is rather small.

## 4. Summary

This article presents an investigation on the correlation between hydration effects and $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ excitation processes in carbonyl compounds.

In such an analysis, two different models have been compared, one using gas-phase H-bonded clusters and the other using H-bonded clusters plus an external continuum solvation (IEF-PCM). To describe the formation/relaxation of the excited state, a different stepwise mechanism is introduced in both models. For the gas phase, the standard picture in terms of a two-step process passing through the Franck–Condon state is used, while for IEF-PCM solvated clusters, a further intermediate step corresponding to solvent dielectric relaxation in succession to the vertical excitation is added to include bulk water effects. The changes in geometries and energetics of the ground and excited states have been analyzed using two popular scalar fields, that is, the molecular electrostatic potential and the molecular electron density.

The MESP CPs show a drastic increase (about 2–3 fold with respect to the ground state) upon first excitation, and this remains fairly similar when solvent relaxation is considered. However, solvent equilibrium along with the complete geometrical relaxation leads to a stabilization of the MESP value close to the ground-state value. A different behavior is found for MED, for which a significant decrease is observed for $n \rightarrow \pi^*$ transitions and a minor increase for $\pi \rightarrow \pi^*$ transitions. These trends in the scalar fields may be attributed to structural distortions in the planarity of the molecule and in the modification of the hydrogen-bond network. By contrast, the Laplacian of electron density, $\nabla^2\rho(\mathbf{r})$, values show very subtle changes after excitation. Hence, the information from $\nabla^2\rho(\mathbf{r})$ seems not to be usefully exploited to draw any specific conclusion about possible correlations between hydrogen bonds, bulk effects, and electronic excitations.

It is expected that more studies based on a similar approach for different carbonyl compounds are necessary to generalize the conclusions. Further studies on a similar theme are underway in our laboratory.

**Acknowledgment.** A.D.K. is thankful to Professors Shridhar R. Gadre, Rajeev K. Pathak (University of Pune, Pune, India) and Carlo Gatti (CNR-ISTM, Milan, Italy) for useful discussions and suggestions. The authors acknowledge the financial support of Gaussian Inc.

## References

(1) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford: New York, 1997.

(2) Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond*; Oxford: New York, 1999.

(3) Lewell, X. Q.; Hillier, I. H.; Field, M. J.; Morris, J. J.; Taylor, P. J. *J. Chem. Soc., Faraday Trans.* **1988**, *84*, 893.

(4) Cai, Z.; Reimers, J. R. *J. Phys. Chem. A* **2007**, *111*, 954.

(5) (a) Alagona, G.; Pullman, A.; Scrocco, E.; Tomasi, J. *Int. J. Peptide Protein Res.* **1973**, *5*, 251. (b) Cimiraglia, R.; Miertus, S.; Tomasi, J. *Chem. Phys. Lett.* **1981**, *80*, 286.

(6) (a) Kumpf, R. A.; Damewood, J. R., Jr. *J. Phys. Chem.* **1989**, *93*, 4478. (b) Dimitrova, R. A.; Peyerimhoff, G. D. *J. Phys. Chem.* **1993**, *97*, 12731. (c) Ramelot, T. A.; Hu, C.-H.; Fowler, J. E.; Deleeuw, B. J.; Schaefer, H. F. *J. Chem. Phys.* **1994**, *100*, 4347.

(7) Wolfe, S.; Kim, C.; Yang, K.; Weinberg, N.; Shi, Z. *J. Am. Chem. Soc.* **1995**, *117*, 4240.

(8) (a) Canuto, S.; Coutinho, K. *Int. J. Quantum Chem.* **2000**, *77*, 192. (b) Kawashima, Y.; Dupuis, M.; Hirao, K. *J. Chem. Phys.* **2002**, *117*, 248. (c) Kongsted, J.; Osted, A.; Pedersen, T. B.; Mikkelsen, K. V.; Christiansen, O. *J. Phys. Chem. A* **2004**, *108*, 8624.

(9) Ledoux, I.; Zyss, J. *Chem. Phys.* **1982**, *73*, 203.

(10) Tanaka, H.; Touhara, H.; Nakanishi, K.; Watanabe, N. *J. Chem. Phys.* **1984**, *80*, 5170.

(11) Mennucci, B; Cammi, R.; Cossi, M.; Tomasi, J. *THEOCHEM* **1989**, *426*, 191.

(12) Lee, C.; Stahlberg, E. A.; Fitzgerald, G. *J. Phys. Chem.* **1995**, *99*, 17737.

(13) Åstrand, P.-O.; Wallqvist, A.; Karlström, G.; Linse, P. *J. Chem. Phys.* **1991**, *95*, 8419.

(14) (a) Wolfe, S.; Kim, C.; Yang, K.; Weinberg, N.; Shi, Z. *J. Am. Chem. Soc.* **1995**, *117*, 4240. (b) Sivanesan, D.; Subramaniam, V.; Babu, K.; Gadre, S. R. *J. Phys. Chem. A* **2000**, *104*, 10887. and the references therein. (c) Shishkin, O. V.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. B* **2000**, *104*, 5357. (d) Kim, N. J.; Kang, H.; Jeong, G.; Kim, Y. S.; Lee, K. T.; Kim, S. K. *J. Phys. Chem. A.* **2000**, *104*, 6552.

(15) (a) Kim, S.; Wheeler, S. E.; Schaefer, H. *J. Chem. Phys.* **2006**, *124*, 204310. (b) Kim, N. J. *Bull. Korean Chem. Soc.* **2006**,

*27*, 1009. (c) Schiedt, J.; Weinkauf, R.; Neumark, D. M.; Schlag, E. W. *Chem. Phys.* **1998**, *239*, 511. (d) Li, X.; Cai, Z.; Sevilla, M. D. *J. Phys. Chem. A* **2002**, *106*, 1596. (e) Pal, S. K.; Peon, J.; Zewail, A. H. *Chem. Phys. Lett.* **2002**, *363*, 57.

(16) (a) Bader, R. F. W. *Atoms in Molecules - A Quantum Theory*; Oxford University Press: Oxford, 1990. (b) Bader, R. F. W.; Essen, H. *J. Chem. Phys.* **1984**, *80*, 1943. (c) Bone, R. G. A.; Bader, R. F. W. *J. Phys. Chem.* **1996**, *100*, 10892.

(17) Gatti, C; May, E.; Destro, R.; Cargnoni, F. *J. Phys. Chem. A* **2002**, *106*, 2607. and the references therein.

(18) (a) Scrocco, E.; Tomasi, J. *Top. Curr. Chem.* **1973**, *42*, 95. (b) Tomasi, J.; Mennucci, B.; Cammi, M. *Molecular Electrostatic Potentials: Concepts and Applications*; Murray, J. S., Sen, K. D., Eds.; Elsevier: Amsterdam, 1996.

(19) (a) Politzer, P.; Thruhlar, D. G. *Chemical Applications of Atomic and Molecular Electrostatic Potential*; Plenum: New York, 1981. (b) Murray, J. S.; Sen, K. D. *Molecular Electrostatic Potentials: Concepts and Applications*, Murray, J. S., Sen, K. D., Eds.; Elsevier: Amsterdam, 1996.

(20) (a) Gadre, S. R.; Shirsat, R. N. *Electrostatics of Atoms and Molecules*; Universities Press: Hyderabad, 2000. (b) Gadre, S. R. *Computational Chemistry: Reviews of Current Trends* Lesczynski, J. Ed.; World Scientific: Singapore, 2000; Vol. 4, and the references therein.

(21) (a) Pathak, R. K.; Gadre, S. R. *Proc. Indian Acad. Sci., Chem. Sci.* **1990**, *102*, 189. (b) Gadre, S. R.; Pathak, R. K. *J. Chem. Phys.* **1990**, *93*, 1170.

(22) (a) Pingale, S. S.; Gadre, S. R.; Bartolotti, L. J. *J. Phys. Chem. A* **1998**, *102*, 9987. (b) Gadre, S. R.; Bhadane, P. K. **1999**, *103*, 3512. (c) Gadre, S. R.; Babu, K.; Rendell, A. P. *J. Phys. Chem A* **2000**, *104*, 8976. (d) Suresh, C. H.; Koga, N.; Gadre, S. R. *J. Org. Chem.* **2001**, *66*, 6883. (e) Kulkarni, A. D.; Babu, K.; Gadre, S. R.; Bartolotti, L. J. *J. Phys. Chem. A.* **2004**, *108*, 2492. (f) Gadre, S. R.; Kulkarni, A. D. *Indian J. Chem.* **2001**, *39A*, 50.

(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.; 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 Devlopment Version*; Gaussian, Inc.: Wallingford, CT, 2004.

(24) (a) Cramer, C. J. *Essentials of Computational Chemistry*; Wiley: Chichester, 2002. (b) Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. *J. Phys. Chem. A* **1998**, *102*, 3782. (c) Estrin, D. A.; Paglieri, L.; Corongiu, G.; Clementi, E. *J. Phys. Chem.* **1996**, *100*, 8701. (d) Szczepaniak, K.; Szczesniak, M.; Person, W. B. *J. Phys. Chem. A* **2000**, *104*, 3852. (e) Thakkar, A. J.; Kassimi, N. E.; Hu, S. *Chem. Phys. Lett.* **2004**, *387*, 142.

(25) Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009.

(26) Mennucci, B. *J. Am. Chem. Soc.* **2001**, *124*, 1506.

(27) (a) Shukla, M. K.; Lesczynski, J. *J. Phys. Chem. A* **2002**, *106*, 11338. (b) Shukla, M. K.; Lesczynski, J. *J. Phys. Chem. A* **2003**, *107*, 5538. (c) Shukla, M. K.; Lesczynski, J. *J. Phys. Chem. B* **2005**, *109*, 17333. and the references therein. (d) Shukla, M. K.; Lesczynski, J. *J. Phys. Chem. A.* **2003**, *107*, 5538.

(28) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1998**, *38*, 3098.

(29) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(30) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1998**, *109*, 8218.

(31) (a) Bauernschmtt, R.; Alhrichs, R. *Chem. Phys. Lett.* **1996**, *256*, 454. (b) Casida, M. E.; Jamorski, C.; Casida, K. C.; Salahub, D. R. *J. Chem. Phys.* **1998**, *108*, 4439.

(32) Gview: A graphical user interface for Gaussian 03. More information may be found at http://www.gaussian.com/gv_plat.htm (accessed Nov. 15, 2007).

(33) (a) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032. (b) Cancès, E.; Mennucci, B. *J. Math. Chem.* **1998**, *23*, 309. (c) Mennucci, B.; Cancès, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506.

(34) (a) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117. (b) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239. (d) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327. (e) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210. (f) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. *Chem. Phys. Lett.* **1998**, *286*, 253. (g) Barone, V.; Cossi, M.; Tomasi, J. *J. Comput. Chem.* **1998**, *109*, 404. (h) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151.

(35) Cammi, R.; Mennucci, B.; Tomasi, J. *J. Phys. Chem. A* **2000**, *104*, 5631.

(36) Scalmani, G.; Frisch, M. J.; Mennucci, B.; Tomasi, J.; Cammi, R.; Barone, V. *J. Chem. Phys.* **2006**, *124*, 094107.

(37) *UNIPROP*, a molecular property calculation package; Theoretical Chemistry Group, Department of Chemistry, University of Pune: Pune, India. Bapat, S. V.; Shirsat, R. N.; Gadre, S. R. *Chem. Phys. Lett.* **1992**, *200*, 373.

(38) *Univis-2000*, see: Limaye, A. C.; Gadre, S. R. *Curr. Sci. (India)* **2001**, *80*, 1296.

(39) (a) Clouthier, D. J.; Ramsey, D. A. *Annu. Rev. Phys. Chem.* **1983**, *34*, 31. (b) Robin, M. B. *Higher Excited States of Polyatomic Molecules*; Academic Press: New York, 1985; *Vol. III*. (c) Walzl, K. N.; Koerting, C. F.; Kuppermann, A. *J. Chem. Phys.* **1987**, *87*, 3796.

(40) (a) Hosoya, H.; Tanaka, J.; Nagakura, S. *Bull. Chem. Soc. Jpn.* **1960**, *33*, 850. (b) Janssen, M. J. *Recl. Trav. Chim. Pays-Bas* **1960**, *79*, 454. (c) Rang, K.; Sandstrom, J.; Svensson, C. *Can. J. Chem.* **1998**, *76*, 811.

(41) (a) Bonaccorsi, R.; Scrocco, E.; Tomasi, J. *Jerusalem Symposia vol 6: Chemical and Biochemical Reactivity*; Pullman, B., Bergman, E., Eds.; Israeli Academy of Science: Jerusalem, 1974; p 387. (b) Cimiraglia, R.; Tomasi, J. *J. Am. Chem. Soc.* **1977**, *99*, 1135. (c) Daudel, R.; Le Rouzo, H.; Cimiraglia, R.; Tomasi, J. *Int. J. Quantum Chem.* **1978**, *13*, 537.

CT7002429

JCTC Journal of Chemical Theory and Computation

# Embedding Fragment ab Initio Model Potentials in CASSCF/CASPT2 Calculations of Doped Solids: Implementation and Applications

Ben Swerts,[†] Liviu F. Chibotaru,*[,†] Roland Lindh,[‡] Luis Seijo,[§] Zoila Barandiaran,[§] Sergiu Clima,[†] Kristin Pierloot,[†] and Marc F. A. Hendrickx[†]

*Division of Quantum and Physical Chemistry, Katholieke Universiteit Leuven, Celestijnenlaan 200F, B-3001 Heverlee, Belgium, Department of Theoretical Chemistry, Chemical Center, Lund University, P.O. Box 124, 221 00 Lund, Sweden, and Departamento de Química, C-14, and Instituto Universitario de Ciencia de Materiales Nicolás Cabrera, Universidad Autónoma de Madrid, 28049 Madrid, Spain*

Received November 17, 2007

**Abstract:** In this article, we present a fragment model potential approach for the description of the crystalline environment as an extension of the use of embedding ab initio model potentials (AIMPs). The biggest limitation of the embedding AIMP method is the spherical nature of its model potentials. This poses problems as soon as the method is applied to crystals containing strongly covalently bonded structures with highly nonspherical electron densities. The newly proposed method addresses this problem by keeping the full electron density as its model potential, thus allowing one to group sets of covalently bonded atoms into fragments. The implementation in the MOLCAS 7.0 quantum chemistry package of the new method, which we call the embedding fragment ab inito model potential method (embedding FAIMP), is reported here, together with results of CASSCF/CASPT2 calculations. The developed methodology is applied for two test problems: (i) the investigation of the lowest ligand field states $^2A_1$ and $^2B_1$ of the Cr(V) defect in the $YVO_4$ crystal and (ii) the investigation of the lowest ligand field and ligand–metal charge transfer (LMCT) states at the Mn(II) substitutional impurity doped into $CaCO_3$. Comparison with similar calculations involving AIMPs for all environmental atoms, including those from covalently bounded units, shows that the FAIMP treatment of the $YVO_4$ units surrounding the $CrO_4^{3-}$ cluster increases the excitation energy $^2B_1 \rightarrow {}^2A_1$ by ca. 1000 $cm^{-1}$ at the CASSCF level of calculation. In the case of the $Mn(CO_3)_6^{10-}$ cluster, the FAIMP treatment of the $CO_3^{2-}$ units of the environment give smaller corrections, of ca. 100 $cm^{-1}$, for the ligand-field excitation energies, which is explained by the larger ligands of this cluster. However, the correction for the energy of the lowest LMCT transition is found to be ca. 600 $cm^{-1}$ for the CASSCF and ca. 1300 $cm^{-1}$ for the CASPT2 calculation.

## I. Introduction

When using the molecular orbital (MO) approach for the description of local properties of crystals, one has to take special care because of their periodic nature. The MO method of choice is normally applied to a representative part of the structure, usually a part of the unit cell, but the effects of the infinite environment cannot be ignored. All proposed solutions are based on the principles of localization and the separability of a many-electron system into subsystems. One of the simplest solutions was pioneered by Sugano and Shulman.[1] They surrounded the structure with point charges to reproduce the electrostatic potential. In the field of organic

* Corresponding author phone: +32-16-327424; fax: +32-16-327992; e-mail: Liviu.Chibotaru@chem.kuleuven.be.

[†] Katholieke Universiteit Leuven.

[‡] Lund University.

[§] Universidad Autónoma de Madrid.

10.1021/ct7003148 CCC: $40.75 
Published on Web 03/05/2008

crystal research, this principle was extended to the super molecule[2] (SM) and quantum mechanics/molecular mechanics (QM/MM)-based[3] models, which are of value in geometrical analysis. For the study of inorganic crystals, however, accurate SM-based models quickly become computationally too expensive because they include the nearest neighbors into the wave function. Crystals containing metals need to be described using larger basis sets and more accurate multiconfigurational expansions of the wave function. It is also beneficial to be able to limit any electron correlation treatment to the central part of the system.[4]

The embedding ab initio model potential (AIMP) method,[5] which proved to be successful in many applications,[6] addresses this by replacing the nearest neighbors by a set of frozen electron densities. These densities are represented by spherical model potentials centered on the atomic positions. Herein lies its major limitation: the environment cannot always easily be divided into spherical ions. When strongly covalently bonded structures are present, the resulting electron density is anisotropic and not accurately representable either by a set of spherical densities or a single large spherical density. A natural following step would be the generalization of the embedding AIMP method to covalently bounded groups of atoms (which we call fragments).

The fragment approach has a long history in quantum chemistry, and many versions of this approach have been proposed in the past. For instance, in the integrated ab initio plus molecular mechanics geometry optimization (IMMOM) method,[7] the chemical groups linked to the active site through a single bond are replaced by the hydrogen atom, while nonbonded interactions of the active site with other atoms in the molecule are described by the MM force field. More rigorous approaches are based on the theory of separability of many-electron systems consisting of weakly interacting parts[8,9] for which effective group potentials (EGPs) can be rigorously introduced. Thus, Katsuki[10] and Mejias Romero and Sanz[11] have developed EGPs for chemical groups linked to the active site by intermolecular interactions, without taking into account charge transfer effects. These effects are incorporated in the effective fragment potential (EFP) approach, which includes a small basis set on the fragment, simulating the covalent interactions of some fragment electrons with the active site, while the interactions with other electrons of the fragment are described by a model potential. Ohta et al.[12] have proposed an EFP for the $NH_3$ groups which included only the lone pair orbital of the nitrogen in the basis set. von Arnim and Peyerimhoff[13,14] have developed an EFP version for small chemical groups where the short-range part of the potential is stored in an intermediate atomic orbital basis set and the long-range part of the potential is simulated by multipole expansions. Another version of this approach, proposed by Colle et al.,[15,16] uses the nonlocal representation for the short-range part of the fragment potential, including the short-range part of the Coulomb interaction, expressed via molecular orbitals of the fragment. An alternative approach is the EGP method introduced by Durand and Malrieu,[17] which is a shape-consistent potential aimed at the reproduction of the active valence orbitals of the fragment, rather than its entire effect on the active site, as was the goal of the EFP. The EGP method was developed by the Toulouse group[18] and proved to be often a reliable tool of fragment calculations of the molecules.[19]

In this article, we propose the embedding fragment ab inito model potential (embedding FAIMP) method, which is basically an extension of the conventional embedding AIMP over polyatomic groups. It uses exact potentials in the sense that a multiatom fragment can be treated as a single entity and is represented by its full electron density. When used with single atom fragments, the method is functionally identical to the embedding AIMP method. The details of the method are presented in the next section, and the details of its implementation into the MOLCAS-7.0 quantum chemistry software are given in section III. Then, in section IV, we apply this method for two substitutional impurity problems.

## II. Method Description

The FAIMP method assumes some of the approximations of the AIMP method and improves other ones. In particular, FAIMP assumes the frozen environment approach (typical of embedding techniques), which makes it applicable only to the calculation of local properties, namely, those which depend strongly on the local geometry and electronic structure of a reference cluster and depend only secondarily on the electronic structure of the environment. The frozen environment approach is a basic assumption in the AIMP embedded cluster method, and although improvements including lattice relaxation and polarization have been explored,[6] it has been found that it is very accurate when applied to very ionic hosts where monatomic ions are easily distinguished. It is reasonable to expect that the frozen environment approach should equally apply to more complex hosts where ionic interactions also occur among fragments (which can be monatomic but also polyatomic ions), whereas covalent interactions may occur within the polyatomic fragments. In these cases, the existing covalent interactions within the polyatomic fragments should be adequately treated at the stage of generating the effective embedding potential, so that the effective potential corresponds to the electronic structure of a polyatomic density instead of corresponding to a set of monatomic electronic densities (examples of hosts of this type are $YVO_4$ and $CaCO_3$, treated in section IV). Otherwise, the interactions between the reference cluster and the external fragments are subject to the same approximations and, presumably, to the same accuracy, as in previous applications of the AIMP embedded cluster method. Consistently, the frozen fragment electronic structure would generate polyatomic Coulomb, exchange, and projection operators which can either be calculated explicitly, this being the alternative in the present implementation, or be subject to further approximation along the usual AIMP recipes for representing local and nonlocal operators, this being the target of future implementations. This latter step should result in significant savings in the evaluation of the FAIMP one-electron integrals in the cluster basis set.

**II.1. The Energy Expression.** For the derivation of the embedding FAIMP Hamiltonian, we consider a central cluster surrounded by a frozen environment consisting of

multiatom fragments. The many-electron nonrelativistic Hamiltonian of this system with $N_{clus} + N_{env}$ electrons reads

$$H_{tot} = \sum_{i}^{N_{clus}+N_{env}} \left\{ -\frac{1}{2}\nabla_i^2 - \sum_K \frac{Z_K}{|r_i - R_K|} \right\} + \sum_{i>j}^{N_{clus}+N_{env}} \frac{1}{r_{ij}} + \sum_{K>L} \frac{Z_K Z_L}{|R_K - R_L|} \quad (1)$$

Within the theory of separability of many-electron systems (group-function theory),[8,9] the total wave function for the system is written as a generalized antisymmetric product of group wave functions. Each group wave function can be a single or multiconfigurational expansion with the added limitation that the number of electrons in each group is constant. This means any electron correlation or electron transfer between groups is ruled out. If the group wave functions fulfill the strong orthogonality condition,[20] the effective electronic Hamiltonian for a single group $G$ can be written as

$$H_{eff}^G = H^G + \sum_i \sum_{L \notin G} \frac{Z_L}{|r_i - R_L|} + \sum_i \sum_{j \notin G} \frac{1}{r_{ij}} \quad (2)$$

It includes the interactions of the electrons of the group with the nuclei (core-attraction) and electrons (Coulomb repulsion and exchange) of all other groups. In practice, this equation cannot be used as-is, however, as its rigorous application would lead to variational collapse of the active electron orbitals onto the frozen orbital space of the fragments because the orthogonality conditions are not imposed. They can be applied following the procedure by Huzinaga and Cantu.[6,8]

As most solutions of the many-electron Hamiltonian are based on orbital expansions, we assume for simplicity that we are dealing with a closed-shell Hartree–Fock (HF) calculation. In this case, the orbitals are solutions of the following Fock equation:

$$F_{tot}|\varphi_i\rangle = \left\{ -\frac{1}{2}\nabla^2 - \sum_K \frac{Z_K}{|r - R_K|} + \sum_j (2J_j - K_j) \right\} |\varphi_i\rangle = \epsilon_i |\varphi_i\rangle \quad (3)$$

Now, when we split the system into cluster and environment electrons, subject the orbitals to the following orthogonality conditions:[6,8]

$$\langle \varphi_i^{env} | \varphi_j^{env} \rangle = \delta_{ij} \qquad \langle \varphi_i^{clus} | \varphi_j^{env} \rangle = 0 \qquad \langle \varphi_i^{clus} | \varphi_j^{clus} \rangle = \delta_{ij} \quad (4)$$

and minimize the total energy under the variational restriction that $\varphi^{env}$ remain frozen, we obtain

$$\left\{ F^{tot} - \left[ \sum_{env} |\varphi^{env}\rangle\langle\varphi^{env}| F^{tot} + F^{tot} \sum_{env} |\varphi^{env}\rangle\langle\varphi^{env}| \right] \right\} |\varphi^{clus}\rangle = \epsilon^{clus} |\varphi^{clus}\rangle \quad (5)$$

Then, if we choose the frozen environment orbitals to be eigenfunctions of $F^{tot}$, we obtain a Huzinaga−Cantu-like equation: [6,8]

$$\left\{ F^{tot} + \sum_{env} (-2\epsilon^{env}) |\varphi^{env}\rangle\langle\varphi^{env}| \right\} |\varphi^{clus}\rangle = \epsilon^{clus} |\varphi^{clus}\rangle \quad (6)$$

Combining this with group-function theory, we obtain the following Hamiltonian for the central cluster:

$$H_{eff}^{clus} = H^{clus} + \sum_i^{N_{clus}} \sum_{L \in env} \frac{Z_L}{|r_i - R_L|} + \sum_i^{N_{clus}} \sum_j^{N_{env}} \frac{1}{r_{ij}} + \sum_F^{N_{frag}} \sum_{o \in F}^{occ} (-2\epsilon_o) |\varphi_o^F\rangle\langle\varphi_o^F| \quad (7)$$

with $\varphi_0{}^F$ being an occupied orbital of fragment $F$ and $N_{frag}$ the number of fragments in the system.

The first two correction terms are trivial to implement, but the last term (the projection operator) needs to be rewritten in linear combination of atomic orbitals (LCAO) form, based on the expansion $\varphi_o^F = 2\sum_{\lambda \in F} c_{o\lambda} \lambda$ $\varphi_o = \sum_{\lambda \in F} c_{o\lambda} \lambda$ and the expression for the energy-weighted density matrix $W_{\lambda\sigma} = 2\sum_{o \in F}^{occ} \epsilon_o c_{o\lambda} c_{o\sigma}$ :

$$\begin{aligned} H_{\mu\nu}^{proj} &= \sum_F \sum_{o \in F}^{occ} (-2\epsilon_o) \langle \mu | \varphi_o^F \rangle \langle \varphi_o^F | \nu \rangle \\ &= \sum_F \sum_{o \in F}^{occ} \sum_{\lambda\sigma \in F} (-2\epsilon_o) c_{o\lambda} c_{o\sigma} \langle \mu | \lambda \rangle \langle \sigma | \nu \rangle \\ &= -\sum_F \sum_{\lambda\sigma \in F} W_{\lambda\sigma}^F \langle \mu | \lambda \rangle \langle \sigma | \nu \rangle \end{aligned} \quad (8)$$

Noting that, in this and the following expressions, the indices $\mu$ and $\nu$ loop over the basis functions of the cluster and the indices $\lambda$ and $\sigma$ loop over the basis functions of the fragments $F$ (which means both indices should always point to basis functions of the same fragment),[21] the complete electronic energy can be written in LCAO form:

$$E^{clus,eff} = E^{clus} + \sum_{\mu\nu} D_{\mu\nu} \left\{ \sum_{L \in env} \left\langle \mu \left| \frac{Z_L}{|r - R_L|} \right| \nu \right\rangle + \sum_{\lambda\sigma} D_{\lambda\sigma}^F (\mu\nu || \lambda\sigma) - \sum_{\lambda\sigma} W_{\lambda\sigma}^F \langle \mu | \lambda \rangle \langle \sigma | \nu \rangle \right\} \quad (9)$$

where the density matrices of the cluster and the fragments are defined as

$$D_{\mu\nu} = 2\sum_{o \in S}^{occ} c_{o\mu} c_{o\nu} \quad (10a)$$

and

$$D_{\lambda\sigma}^F = 2\sum_{o \in F}^{occ} c_{o\lambda} c_{o\sigma} \quad (10b)$$

respectively. When expression 9 is applied to fragments consisting of single atoms, the resulting energies are comparable to those obtained using the electronic embedding AIMP Hamiltonian. The energy of the AIMP embedded cluster, however, also contains an effective nuclear repulsion term between the nuclei in the cluster and in the environment:

$$E_{nuc}^{AIMP} = \sum_{K \in clus} \sum_{L \in env} \frac{Z_K Z_L}{R_{KL}} \quad (11)$$

**II.2. First Derivatives of the Energy.** In order to determine the first derivative of the full FAIMP Hamiltonian,

one first has to determine all constant terms: the fragments' orbital energies and coefficients and the fragment atoms' Mulliken charges. This leads to the following expression for taking the derivative with respect to the positions of the atoms of the central cluster $\{R\}$:

$$\frac{\partial E^{\text{clus,eff}}}{\partial R} = \frac{\partial E^{\text{clus}}}{\partial R} + \sum_{K\in\text{clus}}\sum_{L\in\text{env}} Z_K Z_L \frac{\partial R_{KL}^{-1}}{\partial R} + \sum_{\mu\nu} D_{\mu\nu}\left\{\sum_{K\in\text{env}} \frac{\partial}{\partial R}\left\langle\mu\left|\frac{Z_K}{|r-R_K|}\right|\nu\right\rangle + \sum_{\lambda\sigma} D^F_{\lambda\sigma}\frac{\partial}{\partial R}(\mu\nu\|\lambda\sigma) - \sum_{\lambda\sigma} W^F_{\lambda\sigma}\frac{\partial}{\partial R}\langle\mu|\lambda\rangle\langle\sigma|\nu\rangle\right\}\left[+\sum_{\mu\nu}\frac{\partial D_{\mu\nu}}{\partial R}\left\{\sum_{L\in\text{env}}\left\langle\mu\left|\frac{Z_L}{|r-R_L|}\right|\nu\right\rangle + \sum_{\lambda\sigma} D^F_{\lambda\sigma}(\mu\nu\|\lambda\sigma) - \sum_{\lambda\sigma} W^F_{\lambda\sigma}\langle\mu|\lambda\rangle\langle\sigma|\nu\rangle\right\}\right] \quad (12)$$

The last set of terms contains the derivative of the cluster's density matrix. Because the fragment orbitals are also eigenfunctions of $F^{\text{clus}}$ (see eq 5), these terms should be added to similar terms occurring when determining the derivative of $E^{\text{clus}}$ in the Hartree–Fock case.[23] This means they are also included in the term

$$-\sum_{\mu\nu} W_{\mu\nu}\frac{\partial S_{\mu\nu}}{\partial R} \quad (13)$$

already calculated in $\partial E^{\text{clus}}/\partial R$. This means that eq 12 is in principle only valid when the set of terms in square brackets is removed.

## III. Implementation Details

The Hamiltonian of the considered system in the environment of fragments, represented by their full molecular density, can be expressed as follows:

$$H^{\text{FAIMP}}_{\mu\nu} = H_{\mu\nu} - \sum_E\left\langle\mu\left|\frac{Z_E}{|r-R_E|}\right|\nu\right\rangle + \sum_M\sum_{ab\in M} D^M_{ab}\left[(\mu\nu|ab) - \frac{1}{2}(\mu a|\nu b)\right] + \sum_M\sum_{A\in M}(-2\epsilon_A)\langle\mu|A\rangle\langle A|\nu\rangle$$

where $\mu\nu$ are the basis functions of the central cluster and $ab$ are the basis functions on atoms $E$ of each fragment $M$ used to compute the fragment orbitals $A$ and the density matrices $D_{ab}{}^M$. In the derivation of the energy expression, an all-electron description of the cluster is assumed. The energy expression is equally valid for usage with effective core potentials[6] if the core potential Hamiltonian is used instead of $H^{\text{clus}}$. The same argument can be used for the fragments. They can also be constructed using ECP-type basis sets. As in the regular case, only interaction integrals are calculated, so constant one-center contributions are omitted. Finally, the relativistic effects can be included in the same fashion as in the AIMP approach.[6]

The FAIMP energy and first derivatives are implemented in the MOLCAS 7.0 package.[24] After the geometry of the system is read, where the fragments are specified just as one center (which is normally taken to be an obvious location like the symmetry center or the center of mass), fragments

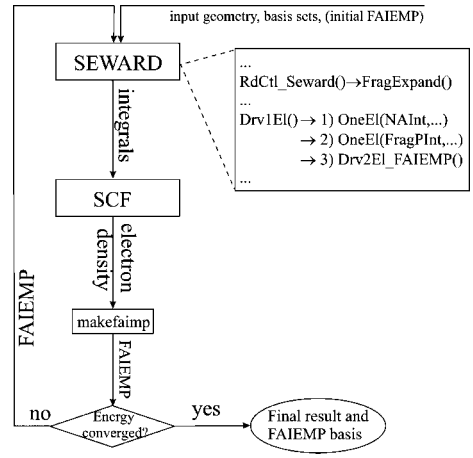


***Figure 1.*** Flowchart of the FAIMP procedure implemented in the MOLCAS-7.0 package.

are expanded (new atoms created from the fragment's atoms) according to the specifications in the fragment density library, and all other data for the fragments are read. Several routines are modified/added to the SEWARD module from MOLCAS suite of programs to evaluate the fragment-related integrals (Figure 1), beside the regular integrals $H_{\mu\nu}$:

1.Nuclear attraction integrals between the cluster's electrons and the fragment nuclei of the expanded fragment atoms:

$$\sum_E \frac{Z_E}{r_\mu - R_E}$$

2.Projection integrals, which are assembled from energy-weighted density matrix $W_{\lambda\sigma}{}^M$ and the two overlap integrals $\langle\mu|\lambda\rangle$ and $\langle\sigma|\nu\rangle$ and contracted afterward. The results are added to the one electron Hamiltonian:

$$P^{\text{FAIMP}}_{\mu\nu} = \sum_M\sum_{A\in M}(-2\epsilon_A)\langle\mu|A\rangle\langle A|\nu\rangle = \sum_M\sum_{A\in M}(-2\epsilon_A)\langle\mu|\sum_{\lambda\in M} c_{A\lambda}\lambda\rangle\langle\sum_{\lambda\in M} c_{A\lambda}\lambda|\nu\rangle = -\sum_M\sum_{\lambda\in M}\sum_{A\in M} 2\epsilon_A c_{A\lambda}c_{A\sigma}\langle\mu|\lambda\rangle\langle\sigma|\nu\rangle = -\sum_M\sum_{\lambda\sigma\in M} W^M_{\lambda\sigma}\langle\mu|\lambda\rangle\langle\sigma|\nu\rangle$$

3.Two-electron interaction integrals. A relative efficiency is obtained with proper prescreening at this stage, by eliminating the intracluster and intra- and interfragment integrals and calculating only the cluster-fragment integrals. These are added locally to the one-electron Hamiltonian:

$$\sum_M\sum_{ab\in M} D^M_{ab}\left[(\mu\nu|ab) - \frac{1}{2}(\mu a|\nu b)\right]$$

With all one- and two-electron integrals computed, the SCF module computes a HF electronic density, from which a small utility (MAKEFAIMP) generates the fragment AIMP basis set. The resulting so-called FAIMP basis set can be included in the Fragment library, but it is not a regular basis

set. It consists of a standard name (X.FRAGMENT.author.0s. 0s.0e-FAIMP-compoundName-etc), the name of the standard basis sets of the participating atoms, relative (to the positioning center that is specified in the input) coordinates of the fragment atoms, orbital energies and coefficients of the occupied fragment orbitals, and Mulliken charges on each atom of the fragment from the calculated SCF wave function. The fragment AIMP generated in this way for a particular crystalline environment can be used in other crystals as well, but when generated for the specific environment, it will give better results. It is, though, a good idea to use FAIMP from other crystals as a starting point for the considered environment. The possibility to perform geometry optimization is implemented in the ALASKA module, where the first derivatives are calculated in the way described above.

The fragment AIMP method was designed to be a generalization of the AIMP method,[5,6] and the same iterative procedure is used to obtain the fragments' orbital energies and coefficients so that (energy-weighted) density matrices are obtained that correspond to fragments in a perfect crystalline environment. The iterative procedure uses restricted Hartree–Fock calculations to obtain the basis sets. Methods incorporating electron correlation can be used too, though the equations are not formally valid for them and have not been tested. The starting point (when no FAIMPs are available to start with) is a single-point SCF calculation of each multiatom molecular fragment that can be considered as a single entity in the crystalline structure. The resulting total density of the fragment from the first single-point calculation is taken as the embedding fragment's basis set for the subsequent run. By alternating the different fragments on the position of a central cluster and employing densities from the previous steps as input for the embedding fragments, a new and improved electron density is generated. The central cluster is usually embedded in a few shells of FAIMP and eventually a few shells of point charges. The self-consistent iterative procedure continues until convergence (usually ~25–30 steps) and is implemented as a shell script, which can be concisely summarized as follows:

*while* [SCF Energy not converged]
*do*
*for* EachFragment *in* AllFragmentTypes
*do*
Molcas (SEWARD) compute integrals
Molcas (SCF) calculate SCF wavefunction
MAKEFAIMP generate FAIMP basis set out of the SCF wavefunction
*done*
*done*

The described implementation of the FAIMP method in the MOLCAS 7.0 package still lacks two essential features. First, it is not yet in the AIMP representation[6] but is still represented by a collection of bielectronic Coulomb and exchange integrals between cluster and fragment orbitals. In order to achieve the AIMP representation, the short-range Coulomb and exchange interaction should be represented via nonlocal operators, as it was proposed, for instance, for EFP by von Arnim and Peyerimhoff.[13,14] Second, the symmetry is not yet implemented for the FAIMP procedure.

The FAIMP is particularly suitable for ionic hosts formed by polyatomic ions or charged fragments, as commented upon above. Consequently, the fragment group functions are expected to be naturally localized within the fragment volume (the same is true for the reference cluster, as commented upon above). Thus, the basis set used to obtain the fragment molecular orbitals can be restricted to include only the bases of the atoms forming the fragment. This natural localization allows for the use of smaller fragment basis sets than the ones that would be presumably needed if standard Hartree–Fock calculations with (partly frozen) localized orbitals would be performed. The latter would be superior, however, in cases where the environment is not naturally localized, as it has been demonstrated in the study of defects and chemisorption in metallic surfaces.[25]

## IV. Illustrative Calculations

In order to assess the importance of the FAIMP approach for the treatment of the effects of covalently bonded groups on the electronic structure of transition metal clusters, we made test calculations for two substitutional impurity systems: (i) $YVO_4$:$Cr^{5+}$ and (ii) $CaCO_3$:$Mn^{2+}$. In both of these cases, no geometry optimization has been done. The main goal of these calculations was the comparison of FAIMP and AIMP approaches.

**IV.1. Cr(V) Impurity in $YVO_4$ Crystal.** Cr(V)-doped yttrium vanadate ($YVO_4$) is a member of a class of compounds with a potential use as tunable solid-state lasers. In this system, chromium has a high oxidation state, which has only been found to be stable in a tetraoxo coordination. If the $CrO_4^{3-}$ structure was in a pure tetrahedral environment, it would have an $^2E$ ground state and a $^2T_2$ excited state several thousand wavenumbers higher in energy. The $YVO_4$ crystal, however, exhibits a distortion with an elongation along the binary axis[26] (in contrast with a more common compression along this axis),[27] lifting the degeneracy of the $^2E$ state. For the case of $CrO_4^{3-}$, crystal field theory (CFT) predicts the $^2B_1$ state ($d_{x^2-y^2}$) to be the ground state. EPR[28] and optical absorption[29] experiments, however, predict an $^2A_1$ ($d_{z^2}$) ground state. This is surprising, even more so considering the fact that the splitting of the $^2T_2$ state does occur as predicted by CFT.[29]

A number of explanations for this phenomenon have been proposed. It was suggested that it is due to strong covalency in the Cr−O bonding[28] or strong interactions with $Y^{3+}$ ions in the second coordination sphere of chromium as revealed by DFT calculations.[29] A recent study by Pascual et al.[30] used CASSCF calculations on the $CrO_4^{3-}$ cluster in combination with the AIMP method for the description of the crystalline environment. Their findings are in agreement with experiments regarding structure and ordering of the states. They concluded that the ordering is 76% due to direct and indirect embedding effects and 24% due to strong covalency. In order to do this type of calculation, the $VO_4^{3-}$ ions had to be modeled as $V^{5+}$ and $O^{2-}$ ions, imposing spherical electron densities. A FAIMP description of the crystal is more in line with the nature of the crystal, as the entire fragment can be described as a single entity.

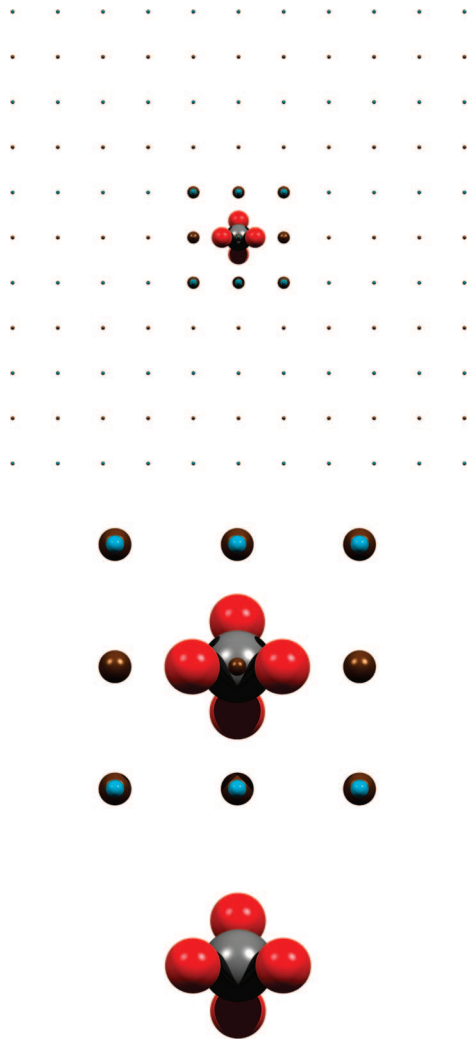

***Figure 2.*** Schematic representation of the cluster $CrO_4^{3-}$ (bottom) embedded into one layer of $VO_4^{3-}$ FAIMP (small balls) and one layer of $Y^{3+}$ AIMP (large balls) (middle) and eight layers of point charges (top). The view is along the 4-fold symmetry axis.

In the present study, the Cr(V) impurity in the $YVO_4$ crystal has been modeled by the $CrO_4^{3-}$ central cluster surrounded by one layer of $Y^{3+}$ and $VO_4^{3-}$ fragments and eight layers of point charges in a $I4_1/amd$ crystalline structure reoriented to conform to a $D_{2d}$ site symmetry (Figure 2). The geometry of the $CrO_4^{3-}$ cluster was taken from ref 30 for the ground state $^2A_1$, where it was optimized in the embedded AIMP CASSCF calculation. The fragment densities were constructed and optimized for three entities: one for yttrium and two for two orientations of the vanadate ion, as the current MOLCAS implementation does not provide automatic rotation of fragments and their density matrices. The point charges have the values of the net charge of the fragments located at their fragment centers. The frontier charges were scaled according to Evjen's method[31] in order to attain a zero-charged environment.

For the description of the central $CrO_4^{3-}$ cluster, two basis sets were used: The first was an ANO-RCC basis set,[32] contracted to [7s6p4d3f2g] for chromium and [4s3p2d1f] for oxygen and designated as "RCC" in the discussions. The second employed basis set, accompanying the core CG-AIMP by Barandiarán and Seijo,[33] was augmented with three f functions[34] and contracted as [4s4p5d1f] for chromium and [2s3p1d] for oxygen and referred to as "ECP". These are the same basis sets as used in a previous AIMP study by Pascual et al.,[30] thus allowing us to directly compare the present FAIMP results to these AIMP results. For the fragments, we constructed the FAIMP densities from three atomic basis sets, more specifically, an ANO-DK3 basis set[35] for all atoms (denoted as "DK3"), an ANO-RCC basis set, using a DZP contraction for all atoms (denoted as "RCC"), Cowan−Griffin relativistic core model potentials with a [3s3p4d] contraction of Barandiarán's AIMP[36] for yttrium, a [3s3p4d] contraction of Seijo's AIMP[37] for vanadium, and a [2s4p1d] contraction of Barandiarán's AIMP[33] for oxygen (denoted as "ECP"). The combinations of cluster basis and fragment basis sets will be denoted as RCC + DK3, RCC + RCC, and ECP + ECP. The FAIMP basis sets for the fragments were optimized to a convergence criterion of $\Delta E < 10^{-8}$ Hartree, which was achieved in 20–25 iterations, compared to an average of seven iterations for the atomic AIMP method.

The relative energies of the $^2A_1$ and $^2B_1$ states of the $CrO_4^{3-}$ cluster were determined using the aforementioned combinations of basis sets using the CASSCF/CASPT2 method.[38–40] The active space consisted of the 3d orbitals of chromium and the 2p orbitals of the four oxygens for a total of 25 electrons in 17 active orbitals. The dynamical correlations were computed at the CASPT2 stage by correlating all but 1s of oxygen and 1s, 2s, and 2p electrons of Cr and V atoms. All ab initio calculations were performed with the MOLCAS 7.0 software.

The results for the first excitation energy are shown in Table 1. The calculated energies show a stronger dependence on the basis set in the case of a cluster embedded in the crystal than in the gas phase. This is especially the case for the CASPT2 calculations. The CASSCF excitation energy for the ECP + ECP basis can be compared directly with a similar AIMP calculation in ref 30, which gave a value of 1650 $cm^{-1}$ for the direct $^2B_1 \rightarrow {}^2A_1$ gap. As we can see from Table 1, this result differs from the FAIMP calculation by ca. 1000 $cm^{-1}$. Although we cannot check the accuracy of these predictions by confronting them with experimental results, the obtained difference in the two approaches is large enough to justify the need for the FAIMP method in this case.

**IV.2. $Mn^{II}$ Impurity in Calcite.** Divalent manganese in calcite is one of the most investigated substitutional impurities in molecular crystals. Calcite is the rhombohedral form of $CaCO_3$ and belongs to the space group $D_{3d}^6$.[41] There are two nonequivalent Ca(II) sites in the calcite corresponding to the alternation of the orientations of the $CO_3^{2-}$ ions in the successive carbonate planes. The manganese(II) ions substituting the calcium(II) ions in calcite are octahedrally coordinated to six nearest-neighbor oxygen atoms of carbonate ions (Figure 3, bottom). Detailed structural investigations by X-ray standing waves and extended X-ray absorption fine structure (EXAFS) have shown[42] that the Mn−O distance is found to be the same as in the isostructural $MnCO_3$ (2.18 Å). Since this is shorter than the Ca−O distance in calcite

***Table 1.*** Relative Energies (in $cm^{-1}$) of the Lowest $^2B_1$ and $^2A_1$ Terms in the $CrO_4^{3-}$ Cluster for Different Combinations of Basis Sets Specified in the Text

| | RCC + DK3 | | RCC + RCC | | ECP + ECP | |
|---|---|---|---|---|---|---|
| | CASSCF | CASPT2 | CASSCF | CASPT2 | CASSCF | CASPT2 |
| gas phase | −1461 | −826 | −1461 | −826 | −1415 | −823 |
| crystal | 1112 | 1936 | 1346 | 2276 | 658 | 1184 |
| $\Delta_{G-C}$ | 2573 | 2762 | 2807 | 3102 | 2073 | 2007 |

by 0.18 Å, in order to match other interatomic distances revealed by EXAFS, the relaxation of neighboring $CO_3^{2-}$ ions was supposed, the main feature being the rotation of Mn−O(1)−C planes by 20°.[42] The recent ligand-field (LF) simulations of optical transitions and EPR in $CaCO_3$:$Mn^{2+}$ [43]

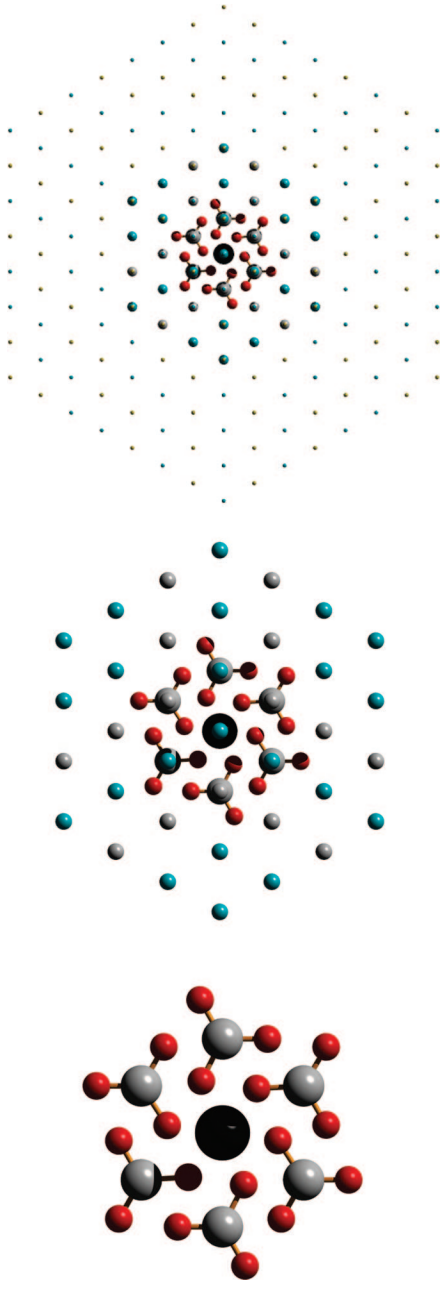

***Figure 3.*** Schematic representation of the cluster $Mn(CO_3)_6^{10-}$ (bottom) embedded into two layers of $CO_3^{2-}$ FAIMP (large balls) and two layers of $Ca^{2+}$ AIMP (small balls) (middle) and eight layers of point charges (top). The view is along the 3-fold symmetry axis.

have further refined the geometry of six oxygens surrounding the manganese ions; however, the structural changes were found to be rather small. In the following, we adopted a simplified structural model for the manganese impurity, which only included the relaxation of the Mn−O bond by 0.18 Å, while all other nuclear coordinates were left unchanged.

We performed CASSCF/CASPT2 calculations of the $Mn(CO_3)_6^{10-}$ cluster (Figure 3, bottom) with Cowan−Griffin relativistic core model potentials with a [3s3p4d] contraction for Mn,[33] a [2s3p1d] contraction for carbon, and a [2s4p1d] contraction for oxygen. The embedding into the calcite lattice was simulated by two layers of FAIMPs (or atomic AIMPs for performance comparison with the AIMP method) on $CO_3^{2-}$ ions and two layers of AIMPs on $Ca^{2+}$ ions around the central cluster (Figure 3, middle) and seven layers of point charges replacing these two types of ions (Figure 3, top). For the $CO_3$ fragments, two different FAIMPs were constructed and optimized corresponding to two orientations of the carbonates in the calcite crystal. The five unpaired electrons on the $Mn^{II}$ impurity make the $^6A_1$ ground state, relative to which the first 24 quartet LF excited states were calculated. For the LF states, a minimal active space was employed, consisting of five 3d orbitals of manganese and an additional five double-shell orbitals, that is, five electrons in 10 orbitals of active space. Dynamical correlation effects were computed at the CASPT2 stage by correlating all electrons (the core–electrons were represented by ECP). To reduce the computational effort for these calculations, the virtual space was reduced by 200 orbitals out of a total of 470 functions. Besides LF excitations, the lowest ligand-to-metal charge transfer (LMCT) state was evaluated as well with both atomic AIMPs and FAIMPs in the same environment and with an enlarged active space. The relatively large size of the ligands leads to a closely spaced manifold of molecular orbitals; therefore, in order to have converged CASSCF and CASPT2 calculations, it was necessary to use a rather extended active space of 35 electrons in 20 orbitals.

The results of the calculations are shown in in Table 2. The first column in the table shows the free ion Mn(II) parentage of the LF terms, which is meaningful given that the weak ligand-field scenario is realized in the complex $Mn(CO_3)_6^{10-}$.[43] The trigonal symmetry of the cluster and the environment require that the T terms split into nondegenerate A and double degenerate E representations of the trigonal symmetry group, which can be easily recognized in the results. Comparison with the assigned transitions of the optical absorption spectra for $Mn^{2+}$ ions in shells[43] shows differences with the calculated values in Table 2 of several thousand wavenumbers. This is probably explained by the nonoptimized geometry of the impurity center and the poor

***Table 2.*** Energies ($cm^{-1}$) of LF and LMCT Excited States of the $Mn(CO_3)_6^{10-}$ Cluster Calculated with AIMP and FAIMP Methods

| | | | AIMP | | FAIMP | |
|---|---|---|---|---|---|---|
| | | $O_h$ | CASSCF | CASPT2 | CASSCF | CASPT2 |
| ligand field | $^6S$ | $^6A_1$ | 0 | 0 | 0 | 0 |
| | $^4G$ | $^4T_1$ | 25256 | 22740 | 25325 | 22788 |
| | | | 25428 | 22947 | 25577 | 23075 |
| | | | 25428 | 22951 | 25577 | 23068 |
| | | $^4T_2$ | 28619 | 26936 | 28606 | 26905 |
| | | | 28894 | 27300 | 28996 | 27414 |
| | | | 28896 | 27308 | 28996 | 27410 |
| | | $^4E$ | 30055 | 28818 | 30059 | 28800 |
| | | | 30056 | 28809 | 30059 | 28808 |
| | | $^4A_1$ | 30167 | 28871 | 30189 | 28896 |
| | $^4D$ | $^4T_2$ | 35502 | 32240 | 35546 | 32271 |
| | | | 35502 | 32248 | 35546 | 32267 |
| | | | 35634 | 32423 | 35723 | 32501 |
| | | $^4E$ | 37407 | 34260 | 37337 | 34190 |
| | | | 37407 | 34263 | 37337 | 34196 |
| | $^4P$ | $^4T_1$ | 38466 | 35373 | 38358 | 35239 |
| | | | 38731 | 35768 | 38754 | 35771 |
| | | | 38732 | 35761 | 38754 | 35718 |
| | $^4F$ | $^4T_1$ | 49819 | 45764 | 49846 | 45760 |
| | | | 49819 | 45766 | 49846 | 45759 |
| | | | 50067 | 46060 | 50139 | 46110 |
| | | $^4A_2$ | 50578 | 45857 | 50605 | 45926 |
| | | $^4T_2$ | 52655 | 49187 | 52602 | 49104 |
| | | | 52656 | 49179 | 52603 | 49112 |
| | | | 52814 | 49356 | 52810 | 49355 |
| LMCT | | $^6A$ | 56739 | 94942 | 56157 | 93673 |

treatment at the CASPT2 level, which we were enforced to adopt. Another source of errors is the insufficient basis set on the manganese ion which, in particular, leads to the overestimation of the excitation energies to states with different spin multiplicities.[44] However, this drawback is not expected to affect much the assessment of the FAIMP method for this system.

The comparison of the results obtained by FAIMP and AIMP methods shows differences which do not exceed 100 $cm^{-1}$ for the calculated energies, which are much lower than the differences obtained for $YVO_4:Cr^{5+}$ in the previous section. This is due to the fact that the ligands in the present case are much larger and, therefore, screen efficiently the short-range potential of the fragments. This is not expected to be so in the case of LMCT excitations. Indeed, as the last line of Table 2 shows, the effect of FAIMP is much stronger, giving the difference with the AIMP method of about 600 $cm^{-1}$ for the CASSCF and 1300 $cm^{-1}$ for the CASPT2 calculation. This excitation corresponds to the transfer of one electron from a doubly occupied ligand orbital delocalized over two oxygens of the carbonate, one of them being the closest to the manganese ion (Figure 4a) and the other to the singly occupied 3d orbital of manganese (Figure 4b). As Figure 4a shows, there is a direct overlap of the ligand orbital with the nearest-neighbor $CO_3$ group from the first layer of the embedding, which makes its energy sensitive to the interaction with this group.

## V. Conclusions

The fragment AIMP method is a useful generalization of the AIMP method. It permits a more accurate description of the

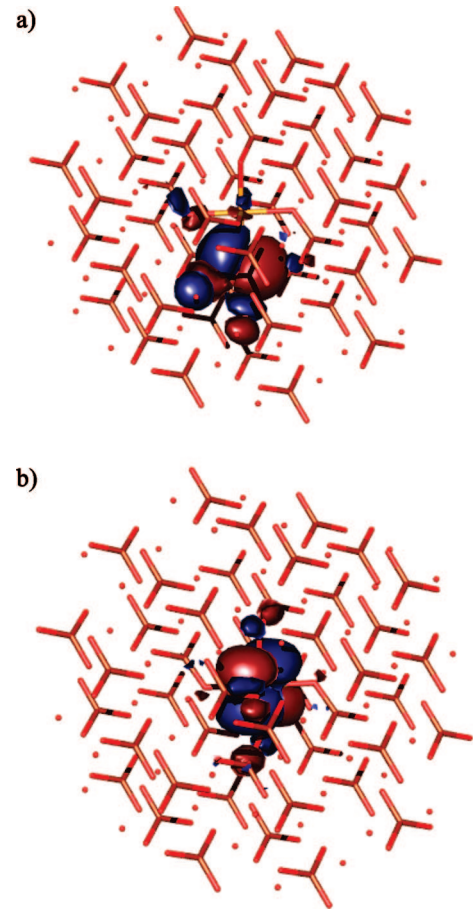


***Figure 4.*** The ligand orbital (a) and the metal orbital (b) involved in the lowest LMCT of the $Mn(CO_3)_6^{10-}$ cluster.

(crystalline) environment of a molecular system without imposing limits on the frozen densities used to represent this environment. This opens the door for a more accurate treatment of the local states and the related spectroscopy in carbonates, sulfates and many natural minerals, molecular solids, and so forth. The method is also more flexible in its choice of basis sets for the fragments' atoms. These basis sets are also easier to construct. When used with single atom fragments, the method essentially reduces to the AIMP implementation. The main downside of this method is the fact that it is computationally more expensive, albeit only in the calculation of the one-electron integrals for the cluster. In subsequent calculations, FAIMP corrections are present in the one-electron matrices and do not increase the computational time in any way. The limiting step in the calculation of the FAIMP integrals is the contraction of the fragment density matrices with the two-electron interaction integrals. Test calculations for $YVO_4:Cr^{5+}$ and $CaCO_3:Mn^{2+}$ systems show that the corrections introduced by FAIMP treatment compared to the conventional AIMP method are important.

Consistent with the frozen environment approximation, nonlocal properties of perfect or imperfect crystals should not be the target of the FAIMP method as it is presented here. Furthermore, the extent or definition of the reference cluster should be consistent with the frozen environment approximation in the calculation of local properties. For dealing with more covalent hosts or very extended defects,

other alternative methods, which can be used along a building block route or as embedding methods, should be preferred.[25,45]

As already mentioned, the described implementation of the FAIMP method in the MOLCAS 7.0 package still lacks two essential features: (i) the AIMP representation of bielectronic and projection operators and (ii) the account of symmetry of the supermolecule (cluster + fragments). These are tasks for further development. Their accomplishment would greatly facilitate the use of the FAIMP method for embedded calculations, especially for the geometry optimization of impurity systems, which is done routinely for the AIMP method.

**Acknowledgment.** Financial support by the Flemish Science Foundation and the Flemish Government under the Concerted Action Scheme are gratefully acknowledged. B.S. thanks the MOLCAS group, while L.C. and M.H. would like to thank the Madrid group for hospitality during their visits.

## References

(1) Sugano, S.; Shulman, R. G. *Phys. Rev.* **1963**, *130*, 517–530.

(2) Peeters, A.; Van Alsenoy, C.; Lenstra, A. T. H.; Geise, H. J. *Int. J. Quantum Chem.* **1993**, *46*, 73–80.

(3) Swerts, B.; Van Droogenbroeck, J.; Peeters, A.; Van Alsenoy, C. *J. Phys. Chem. A* **2002**, *106*, 4245–4250.

(4) Whitten, J. L. *Chem. Phys.* **1993**, *177*, 387–397.

(5) Barandiarán, Z.; Seijo, L. *J. Chem. Phys.* **1988**, *89*, 5739–5746.

(6) Seijo, L.; Barandiarán, Z. In *Computational Chemistry: Reviews of Current Trends*; Leszcynski, J., Ed.; World Scientific: Singapur, 1999; Vol. 4, pp 55–152.

(7) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170–1179.

(8) Huzinaga, S.; Cantu, A. A. *J. Chem. Phys.* **1971**, *55*, 5543–5549.

(9) Kleiner, M.; McWeeny, R. *Chem. Phys. Lett.* **1973**, *19*, 476–479.

(10) Katsuki, S. *J. Chem. Phys.* **1993**, *98*, 496–501.

(11) Mejias Romero, J. A.; Sanz, J. F. *J. Chem. Phys.* **1993**, *99*, 1255–1261.

(12) Ohta, K.; Yoshioka, Y.; Morokuma, K.; Kitaura, K. *Chem. Phys. Lett.* **1983**, *101*, 12–17.

(13) von Arnim, M.; Peyerimhoff, S. D. *Theor. Chim. Acta* **1993**, *87*, 41–57.

(14) von Arnim, M.; Peyerimhoff, S. D. *Chem. Phys. Lett.* **1993**, *210*, 488–494.

(15) Colle, R.; Salvetti, O. *Theor. Chim. Acta* **1991**, *80*, 63–70.

(16) Colle, R.; Curioni, A.; Salvetti, O. *Theor. Chim. Acta* **1993**, *86*, 451–465.

(17) Durand, Ph.; Malrieu, J. P. In *Advances in Chemical Physics: Ab Initio Methods in Quantum Chemistry*; Lawley, K. P., Ed.; Wiley: New York, 1987; vol. *LXVII*, part I, p 321.

(18) Poteau, R.; Ortega, I.; Alary, F.; Ramirez Solis, A.; Bartelat, J. C.; Daudey, J. P. *J. Phys. Chem. A* **2001**, *105*, 198–205.

(19) Carissan, Y.; Bessac, F.; Alary, F.; Heully, J. L.; Poteau, R. *Int. J. Quantum Chem.* **2006**, *106*, 727–733.

(20) Parr, R. G.; Elisson, F. O.; Lykos, P. G. *J. Chem. Phys.* **1956**, *24*, 1106.

(21) Note that, in general, the cluster basis set does not include basis functions of elements that belong to the environment. However, in some particular cases, the cluster basis set should be enlarged so as to include basis functions of next neighbours. This is only necessary in cases where geometry optimizations are done with respect to coordinates that confront cluster and environment components along bonding directions of the cluster (ref 6) or in cases where delocalized cluster states are calculated, such as impurity-trapped excitons.[22] None of these cases apply here.

(22) Ordejón, B.; Seijo, L.; Barandiaran, Z. *J. Chem. Phys.* **2007**, *126*, 194712.

(23) Pople, J.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. *Int. J. Quantum Chem.* **1979**, *S13*, 225.

(24) MOLCAS 7.0: Karlström, G.; Lindh, R.; Malmqvist, P.-A.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mat. Sci.* **2003**, *28*, 222–239.

(25) Whitten, J. L.; Yang, H. *Surf. Sci. Rep.* **1996**, *24*, 55.

(26) Baglio, J. A.; Gashurov, G. *Acta Crystallogr., Sect. B* **1968**, *24*, 292–293.

(27) (a) Banks, E.; Greenblatt, M.; Post, B. *Inorg. Chem.* **1970**, *9*, 2259–2264. (b) Albrecht, C.; Cohen, S.; Mayer, I.; Reinen, D. *J. Solid State Chem.* **1993**, *107*, 218–228.

(28) Greenblatt, M.; Pifer, J. H. *J. Solid State Chem.* **1979**, *29*, 1–7.

(29) Hazenkamp, M. F.; Stückl, A. C.; Cavalli, E.; Güdel, H. U. *Inorg. Chem.* **2000**, *39*, 251–254.

(30) Pascual, J. L.; Barandiarán, Z.; Seijo, L. *Int. J. Quantum Chem.* **2002**, *90*, 751–758.

(31) Evjen, H. M. *Phys. Rev.* **1932**, *39*, 675–687.

(32) Roos, B. O.; Lindh, R.; Malmqvist, P.-A.; Veryazov, V.; Widmark, P.-O. *J. Phys. Chem. A* **2004**, *108*, 2851–2858.

(33) Barandiarán, Z.; Seijo, L. *Can. J. Chem.* **1992**, *70*, 409–415.

(34) Petterson, L. G. M. Private communication.

(35) Tsuchiya, T.; Abe, M.; Nakajima, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 4463–4472.

(36) Barandiarán, Z.; Seijo, L.; Huzinaga, S. *J. Chem. Phys.* **1990**, *93*, 5843–5850.

(37) Seijo, L.; Barandiarán, Z.; Huzinaga, S. *J. Chem. Phys.* **1989**, *91*, 7011–7017.

(38) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. *Chem. Phys.* **1980**, *48*, 157–173.

(39) Andersson, K.; Mamlqvist, P.-A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(40) Andersson, K.; Mamlqvist, P.-A.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218–1226.

(41) Marshall, S. A.; Serway, R. A. *Phys. Rev.* **1968**, *171*, 345–349.

(42) Cheng, L.; Sturchio, N. C.; Bedzyk, M. J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2001**, *63*, 144104.

(43) Cheng, L.; Xiao-Yu, K.; Xiao-Ming, T.; Xiong, Y. *J. Phys. Chem. A* **2007**, *111*, 2783–2789.

(44) Pierloot, K.; Vancoillie, S. *J. Chem. Phys.* **2006**, *125*, 124303.

(45) Seijo, L.; Barandiarán, Z. *J. Chem. Phys.* **2004**, *121*, 6698.

CT7003148

JCTC Journal of Chemical Theory and Computation

# Initial Hardness Response and Hardness Profiles in the Study of Woodward–Hoffmann Rules for Electrocyclizations

F. De Proft,*,† P. K. Chattaraj,‡ P. W. Ayers,§ M. Torrent-Sucarrat,† M. Elango,|| V. Subramanian,|| S. Giri,‡ and P. Geerlings†

*Eenheid Algemene Chemie (ALGC), Vrije Universiteit Brussel (VUB), Faculteit Wetenschappen, Pleinlaan 2, 1050 Brussels, Belgium, Department of Chemistry and Center for Theoretical Studies, Indian Institute of Technology, Kharagpur 721302, India, Department of Chemistry, McMaster University, Hamilton, Ontario, Canada L8S 4M1, and Chemical Laboratory, Central Leather Research Institute, Adyar, Chennai 600020, India*

Received October 28, 2007

**Abstract:** The fundamental principles of pericyclic reactions are governed by the Woodward–Hoffmann rules, which state that these reactions can only take place if the symmetries of the reactants' molecular orbitals and the products' molecular orbitals are the same. As such, these rules rely on the nodal structure of either the wave function or the frontier molecular orbitals, so it is unclear how these rules can be recovered in the density functional reactivity theory (or "conceptual DFT"), where the basic quantity is the strictly positive electron density. A third, nonsymmetry based approach to predict the outcome of pericyclic reactions is due to Zimmerman which uses the concept of the aromatic transition states: allowed reactions possess aromatic transition states, while forbidden reactions possess antiaromatic transition states. Based on our recent work on cycloadditions, we investigate the initial response of the chemical hardness, a central DFT based reactivity index, along the reaction profiles of a series of electrocyclizations. For a number of cases, we also compute complete initial reaction coordinate (IRC) paths and hardness profiles. We find that the hardness response is always higher for the allowed modes than for the forbidden modes. This suggests that the initial hardness response along the IRC is the key for casting the Woodward–Hoffmann rules into conceptual DFT.

## 1. Introduction

The well-known Woodward–Hoffmann rules[1] provide the basis of our current understanding of pericyclic reactions—reactions involving a cyclic rearrangement of electrons. Examples of pericyclic reactions include cycloadditions, electrocyclizations, sigmatropic rearrangements, and chelotropic reactions.[2] Three different approaches to interpreting the pericyclic reactivity trends are common in the literature; two out of three use symmetry arguments based on signs of either the frontier molecular orbitals or the total wave function of the system.[1,3] A third method, introduced by Zimmerman,[4] for the prediction of the outcome of these reactions uses the concept of the (anti)aromaticity of the transition state. In this approach, reactions with aromatic transition states are asserted to be allowed, while those with an antiaromatic transition state are considered forbidden. Depending on the number of nodes in the $\pi$ system at the transition state of a given mode (e.g., conrotatory vs disrotatory, suprafacial vs antarafacial approach) in the pericyclic reaction, systems can be classified as either Hückel (zero or even number of nodes) or Möbius (odd number of

* Corresponding author. Fax: +32-2-6293317. E-mail: fdeprof@vub.ac.be.

† Vrije Universiteit Brussel (VUB).

‡ Indian Institute of Technology.

§ McMaster University.

|| Central Leather Research Institute.

10.1021/ct700289p CCC: $40.75 
Published on Web 03/14/2008

nodes) systems. Hückel and Möbius systems are aromatic for $(4n + 2)$ and $4n$ $\pi$ electrons and antiaromatic for $4n$ and $(4n + 2)$ $\pi$ electrons, respectively.[5]

It has been argued that density functional theory (DFT)[6] provides an excellent framework for the formulation of a general theory of chemical reactivity.[7] This involves the introduction of a series of chemical concepts via the so-called response functions of the energy of the system with respect to either the number of electrons, the external potential (for an isolated atomic or molecular system, this is the potential due to the nuclei), or both. Their use in connection with a series of chemical principles rooted theoretically in DFT, often within a perturbative perspective on chemical interactions, was shown to provide powerful machinery in problems of reactivity and stability; this area of research has been termed "conceptual DFT".[7] However, none of these concepts or principles are connected to the phase of the wave function and the question thus arises as to how the Woodward–Hoffmann rules can be recovered from the conceptual DFT. The DFT itself uses the electron density, a quantity which is always larger than or equal to zero, of the system as the basic source of information of all atomic and molecular properties. Conceptual DFT based reactivity descriptors (especially the hardness) have been used in the past to analyze the Woodward–Hoffmann rules.[8] Recently, we have found that the concept of initial hardness response (i.e., the change of the hardness of the reagents along the initial stage of a model reaction coordinate of the pericyclic reaction) can be used to explain the Woodward–Hoffmann rules in the case of two prototypical cycloaddition reactions, i.e. the [4 + 2] Diels–Alder reaction of 1,3-butadiene with ethylene and the [2 + 2] cycloaddition of two ethylene molecules.[9] This finding connects with the proposed linear relationship between hardness and aromaticity.[10,11] It was concluded that the initial hardness response could be considered to be a predictor of the activation hardness, which encapsulates information about the aromaticity of the transition state, which in turn provides a rationale for the Woodward–Hoffmann rules.[9] In addition, we have used a related quantity, the "dual descriptor"[12] in the rationalization of these rules for cycloadditions, electrocyclizations, and sigmatropic shift reactions.[13]

In this contribution, we will investigate in detail the use of the initial hardness response descriptor in the study of electrocyclization reactions, for which we will also construct energy and hardness profiles. In these pericyclic reactions, the $\pi$ system will undergo an intramolecular rearrangement, leading to the formation of a $\sigma$ bond and a ring. Depending on the number of $\pi$ electrons involved in this process, the process will follow a conrotatory or a disrotatory pathway, resulting in a different stereochemical outcome. In the cyclization of linear polyenes, the conrotatory mode corresponds to a Möbius-like ring closure, which corresponds to an aromatic system when it involves $4n$ $\pi$ electrons. The disrotatory movement, yielding a Hückel-like ring closure, is aromatic when $(4n + 2)$ $\pi$ electrons are involved.[4]

We have focused our attention on five examples of electrocyclizations, depicted in Figure 1. This paper can be divided into two parts. In the first part, the hardness response

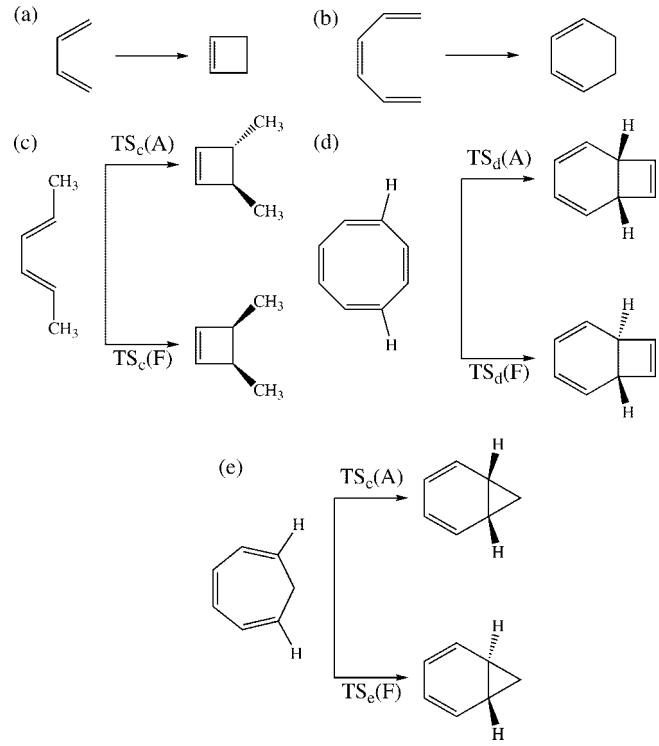


***Figure 1.*** Summary of the different electrocyclizations investigated in this work.

for the thermal and photochemical electrocyclization reactions of 1,3-butadiene and 1,3,5-hexatriene (reactions a and b in Figure 1) have been studied along the initial stages of model *conrotatory* and *disrotatory* reaction coordinates. Next, we have investigated the *allowed* and *forbidden* cyclizations for a series of larger systems, 2,4-hexadiene, cyclooctatetraene, and cycloheptatriene in their ground states to obtain more insights into the complete reaction and hardness profiles of these reactions (reactions c, d, and e in Figure 1).

## 2. Theoretical Background and Computational Details

The chemical hardness $\eta$[14,15] is a central quantity for use in the study of reactivity and stability, through the hard and soft acids and bases principle[14,16] and the principle of maximum hardness.[14c,17,18] Parr and Pearson have introduced the chemical hardness as the second derivative of the energy of the system with respect to the number of electrons (sometimes preceded by an arbitrary factor of $^1/_2$, which we have dropped in our treatment) which can be estimated by the difference of the vertical ionization energy and electron affinity of the system:[15]

$$\eta = \left(\frac{\partial^2 E}{\partial N^2}\right)_v \approx I - A \tag{1}$$

Recently, Tozer and De Proft have introduced an approximate method to compute this quantity, requiring only the calculation of the neutral and cationic systems[19]

$$\eta \approx (\varepsilon_{\mathrm{LUMO}} - \varepsilon_{\mathrm{HOMO}}) + 2(\varepsilon_{\mathrm{HOMO}} + I) \tag{2}$$

where $\varepsilon_{\mathrm{HOMO}}$ and $\varepsilon_{\mathrm{LUMO}}$ are the Kohn–Sham orbital energies of the highest occupied (HOMO) and lowest unoccupied

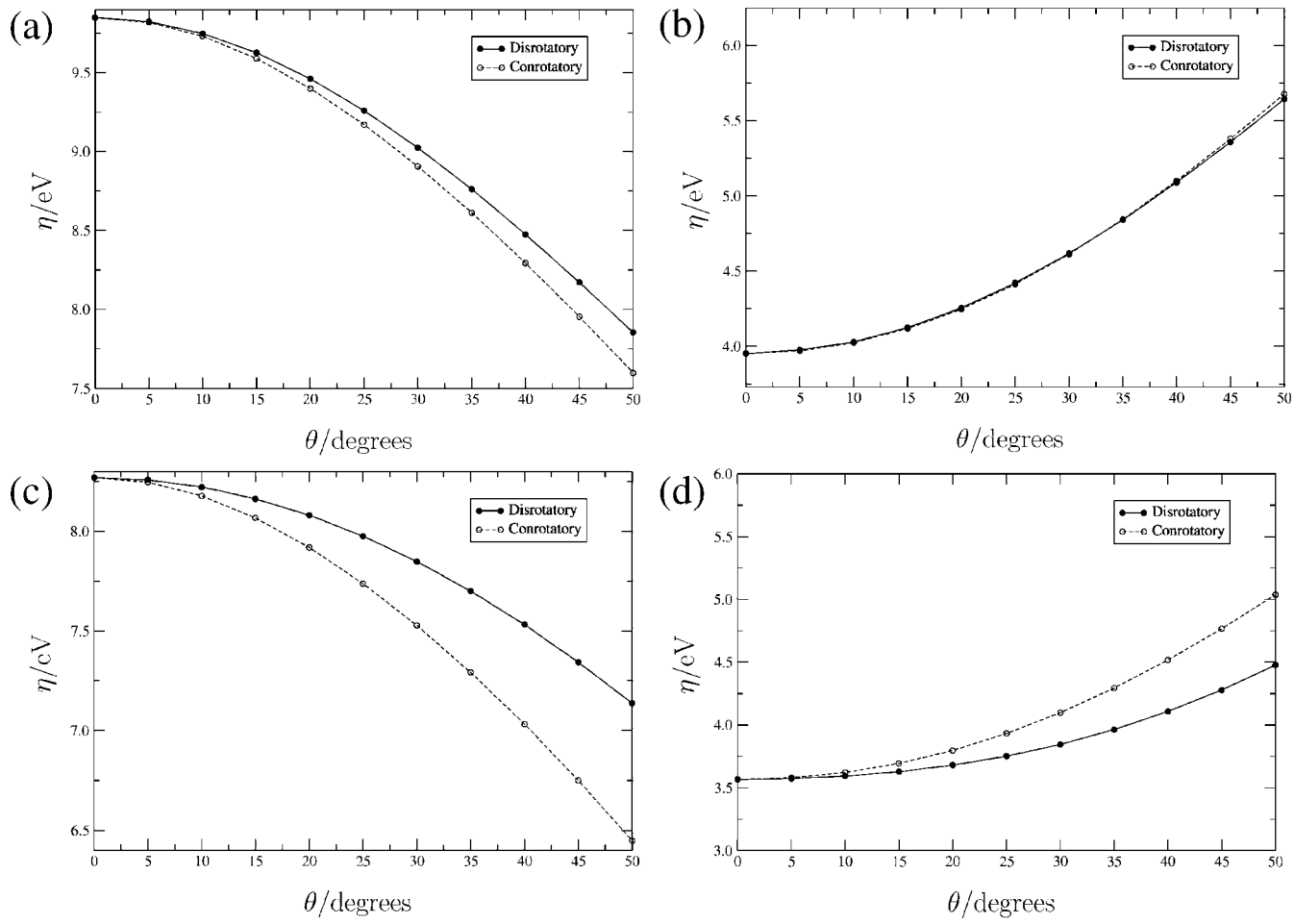


***Figure 2.*** Evolution of the chemical hardness $\eta$ along a model reaction coordinate, quantified by the dihedral angle $\theta$, for electrocyclization of singlet and triplet 1,3-butadiene (parts a and b, respectively) and 1,3,5-hexatriene (parts c and d, respectively).

(LUMO) molecular orbitals, respectively, obtained using a pure density functional. It was shown that this equation corresponds to eq 1 but using the following approximation for the computation of the electron affinity:[19,20]

$$A \approx -(\varepsilon_{\mathrm{LUMO}} + \varepsilon_{\mathrm{HOMO}}) - I \quad (3)$$

This approximate method was shown to give reasonable estimates for the electron affinities of systems possessing metastable anions.[19,20] In this work, we will need to compute the hardness not only of molecules in their (singlet) ground state but also in the lowest triplet excited state. For the triplet excited state, we need to modify eqs 2 and 3. The vertical electron affinity for a system in the first excited triplet state $A(\mathrm{T})$, using eq 3 for the estimation of the electron affinity of its singlet ground state $A(\mathrm{S})$, can be expressed as

$$A(\mathrm{T}) = A(\mathrm{S}) + \Delta E_{\mathrm{ST}} \approx -(\varepsilon_{\mathrm{LUMO}}(\mathrm{S}) + \varepsilon_{\mathrm{HOMO}}(\mathrm{S})) - I(\mathrm{S}) + \Delta E_{\mathrm{ST}} \quad (4)$$

where $I(\mathrm{S})$ is the vertical ionization energy of the singlet ground state and $\Delta E_{\mathrm{ST}}$ is the vertical singlet–triplet gap. The hardness of the first excited triplet state can thus be estimated as

$$\eta(\mathrm{T}) = I(\mathrm{T}) - A(\mathrm{T}) \approx I(\mathrm{T}) + I(\mathrm{S}) + (\varepsilon_{\mathrm{LUMO}}(\mathrm{S}) + \varepsilon_{\mathrm{HOMO}}(\mathrm{S})) - \Delta E_{\mathrm{ST}} \quad (5)$$

Note that this equation depends not only on the ionization energies and orbital energies but also on the singlet–triplet gap. The singlet–triplet gap can, in this case, be estimated reliably using DFT methods. (For a recent example, see ref 21.)

Gaussian 03 program[22] geometries, transition states, and intrinsic reaction coordinates were obtained at the B3LYP[23]/6-311+G**[24] level; single point hardness calculations using eqs 2 and 5 were performed using the Perdew, Burke, and Ernzerhof (PBE) functional[25] with the same basis set.

## 3. Results and Discussion

In the first part of this contribution, we have computed the hardness response along an initial reaction coordinate of the conrotatory and disrotatory modes of the cyclization of 1,3-butadiene and 1,3,5 hexatriene, both in the singlet and the triplet states. This model reaction coordinate involves a conrotatory or a disrotatory movement of the terminal $CH_2$ groups of these molecules, as quantified by the angle $\theta$, which is the dihedral angle between the terminal hydrogens and the completely fixed, planar carbon skeleton. It is to be mentioned that the planar $C_{2v}$ conformation of both of these compounds is not a minimum on the potential energy surfaces of these molecules, but this conformation has been used to keep the analysis as simple as possible and in

agreement with the spirit (and the molecular geometry) of the original Woodward–Hoffmann work. Figure 2 depicts the initial hardness changes for these four initial steps. In the case of the ground-state reaction of 1,3-butadiene (Figure 2a), the hardness along the initial reaction path in the forbidden disrotatory mode is higher than along the allowed, conrotatory mode, contrary to our earlier findings for cycloadditions.[9] In the excited state (Figure 2b), hardness profiles for both the model reactions are almost indistinguishable. Figure 2c depicts the hardness change along the initial model reaction coordinate of the ground-state electrocylization of 1,3,5-hexatriene. In this case, one can see that the hardness along the allowed, disrotatory mode is always higher than the hardness along the forbidden mode. In the triplet state, this ordering is reversed, in agreement with the Woodward–Hoffmann rules. Indeed, the results for the electrocyclization of 1,3,5-hexatriene are in agreement with the fact that the allowed mode of the reaction should have a higher hardness than the forbidden one. In order to gain insight into the apparently contradicting results for 1,3-butadiene, both the conrotatory (allowed) and disrotatory (forbidden) initial reaction coordinate (IRC) paths in the singlet state were computed at the CASSCF(4,4)/6-31G* level of theory. The transition structures for these electrocyclizations were taken from ref 26. Indeed, in a previous study,[8] it was found that the transition structure for the allowed process of this electrocyclization is harder than the one for the forbidden pathway. Along the initial stages of the computed IRC path, the hardness was evaluated at the PBE/6-311+G** level using eq 2. Figure 3 depicts both the energy (Figure 3a) and the hardness profiles of the allowed (Figure 3b) and forbidden (Figure 3c) cyclizations. The reaction path of the cyclization in the allowed mode clearly shows a steeper onset than the forbidden mode; after some point, however, both reaction paths cross, resulting, as could be expected, in an activation barrier for the allowed reaction that is smaller than the forbidden reaction; the difference in activation energy amounts to 13.1 kcal/mol. The hardness profiles are in agreement with the allowed or forbidden character of the reactions. The hardness along the allowed conrotatory mode is always larger than the hardness of the forbidden, disrotatory mode. The hardness shows a very steep decrease in the vicinity of the transition state for the forbidden reaction, with the HOMO–LUMO gap, part of the hardness expression in eq 2, essentially going to zero. This implies that the unexpected result for the initial hardness response for the cyclization of 1,3-butadiene is probably due to the fact that the model reaction coordinate is a rather bad approximation to the IRC for this case. However, the simplified model reaction coordinate performs appreciably better as the size of the molecule increases, because larger molecules have greater conformational flexibility.

The reaction and hardness profiles (Figure 4) for the electrocyclization of 2,4-hexadiene (reaction c) bear a close resemblance with the ones obtained for 1,3-butadiene, which might be expected since both reactions involve 4 $\pi$ electrons. Also in this case, a crossing along the IRC occurs, yielding a lower energy transition state for the allowed, conrotatory mode than for the forbidden mode.

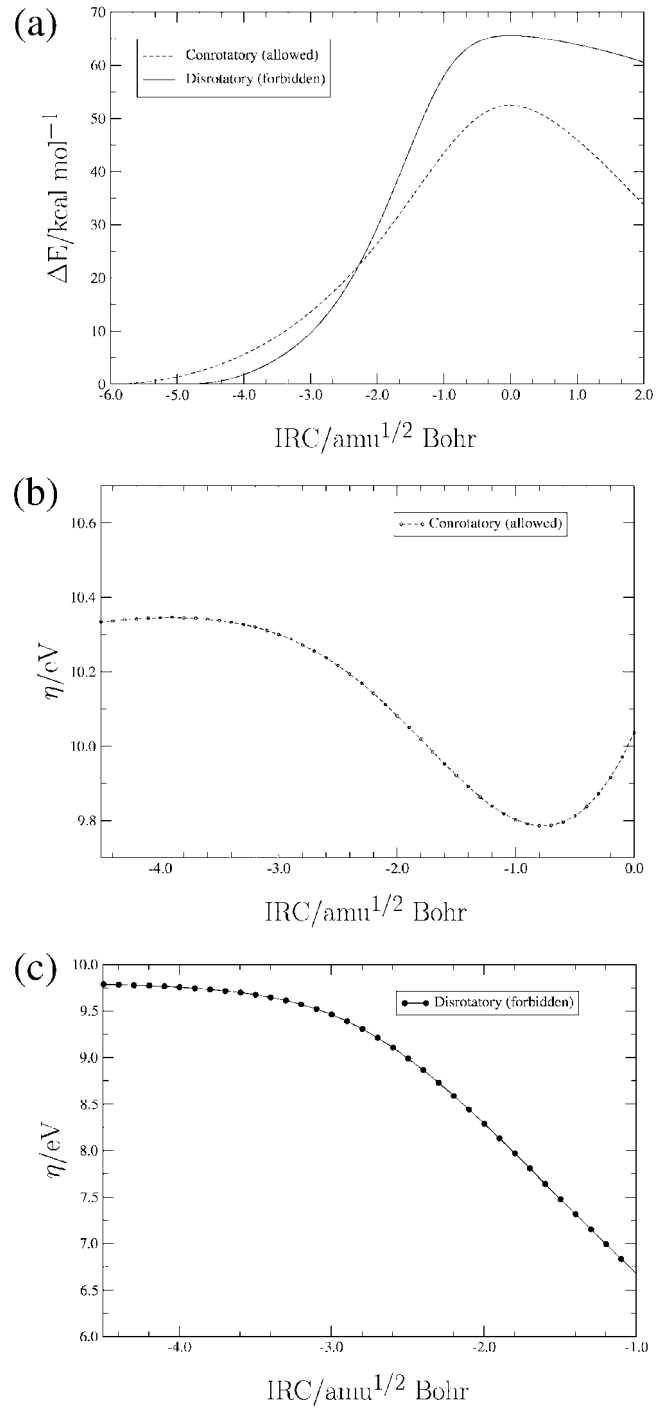


***Figure 3.*** Energetic reaction paths (a) and hardness profiles (b and c) for the conrotatory and disrotatory modes of the electrocyclization of 1,3-butadiene.

In this case, however, the difference in activation barriers is only 5.2 kcal/mol. On the other hand, this implies that although the calculation of the reaction path was performed using a single reference methodology, the same qualitative results are obtained as in the case of the multireference calculations, which is the main aim of our approach. The hardness profile confirms that, for the allowed reaction, the hardness in the neighborhood of the transition state along the IRC path is always larger for the allowed mode. However, initially, the hardness along the forbidden

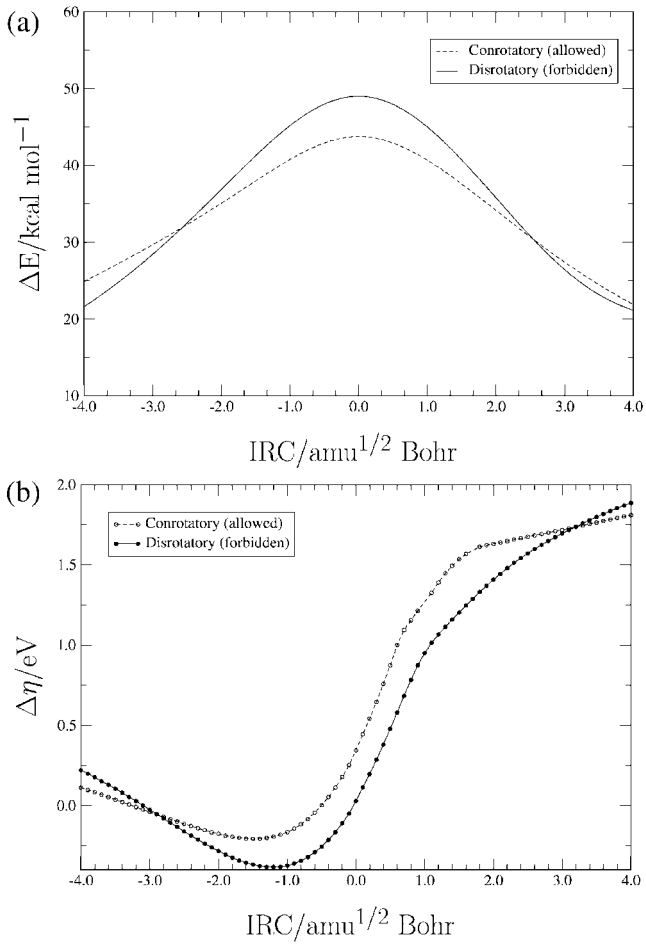


***Figure 4.*** Energetic reaction paths (a) and hardness profiles (b) for the conrotatory and disrotatory modes of the electrocyclization of 2,4-hexadiene.

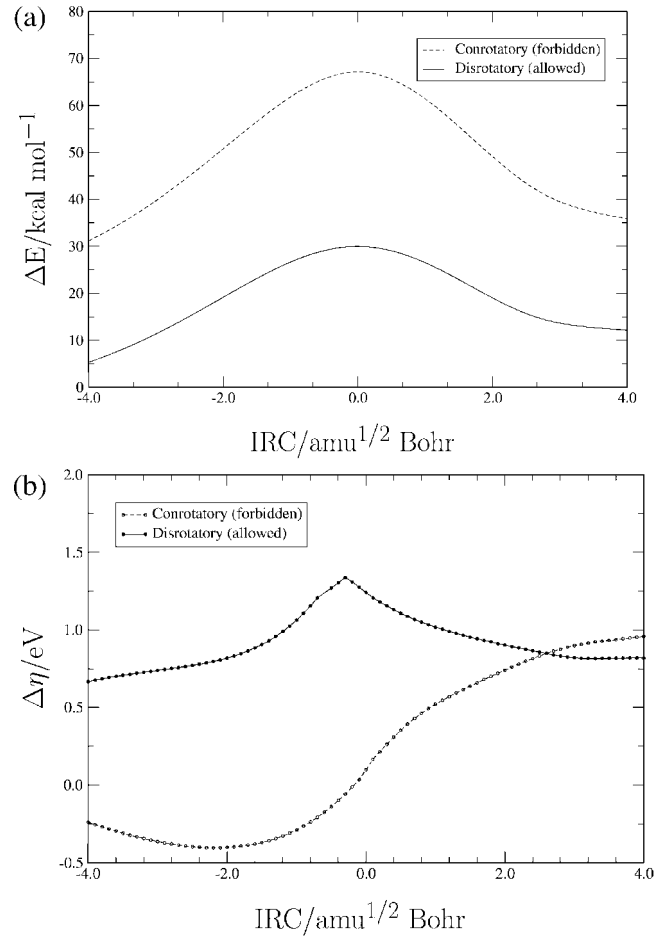


***Figure 5.*** Energetic reaction paths (a) and hardness profiles (b) for the conrotatory and disrotatory modes of the electrocyclization of cyclooctatetraene.

mode pathway is larger, which explains the failure of the initial hardness response observed for the 1,3-butadiene.

Next, we have analyzed the profiles in the case of two reactions involving 6 $\pi$ electrons, i.e. the cyclizations of cyclooctatetraene (reaction d) and cycloheptatriene (reaction e). Respective plots are provided in Figures 5 and 6, respectively. As can be seen, in both cases, there is no crossing of the energy profiles and the energy of the disrotatory, allowed mode is always below the energy of the conrotatory, forbidden mode. Also, the hardness along the allowed mode is almost always higher than along the forbidden mode, in agreement with the result presented earlier for the cyclization of the 4 $\pi$ electron systems. A difference between this case and the former can however be noticed. In the 6 $\pi$ electron cases, where the allowedness of the reaction can be explained by the Hückel aromaticity of the transition state, the profile for the allowed mode shows an increase in hardness when going from the reactant to the transition state. This sort of "violation" of the maximum hardness principle is not especially uncommon because the requirements for the maximum hardness principle to hold are very stringent.[18b,c] (refs 27–29 elucidate the reasons for violations of the maximum hardness principle in more detail.)

In the case of the cyclization of the 4 $\pi$ electron systems, however, involving a Möbius aromatic transition state, the profiles show the inverse trend, i.e. the more "classically" observed behavior of decreasing hardness along the reaction path yielding the maximum softness in the vicinity of or at the transition state.

In Table 1, we have listed the activation energies and activation hardnesses[30] for reactions c, d, and e. Note that, following ref 30, the activation hardnesses have been computed as the difference of the hardness of the reactants and the transition state. As expected, the allowed reactions always have lower activation barriers and smaller activation hardnesses. The latter finding is consistent with observations in ref 30, where it was stated that the smaller this difference, the faster the reaction.

In a final part, we quantitatively assess the initial hardness responses $(\partial\eta/\partial Q)_N$ for all the reactions considered in this work. In the case of the ground- and excited-state cyclizations of 1,3-butadiene and 1,3,5-hexatriene, $Q$ corresponds to the dihedral angle $\theta$, described above, which was used to construct a simple reaction coordinate for these cases. For the other cases, $Q$ is the mass-weighted internal coordinate corresponding to the reactive mode. Table 2 lists the finite difference approximations to the initial hardness responses. As can be seen, in all but one of the cases, the initial hardness response for the allowed reaction is larger than that of the forbidden one. The only exception, the electrocyclization of 1,3-butadiene in the singlet state, is most probably due to the fact that the simplified reaction coordinate used in this

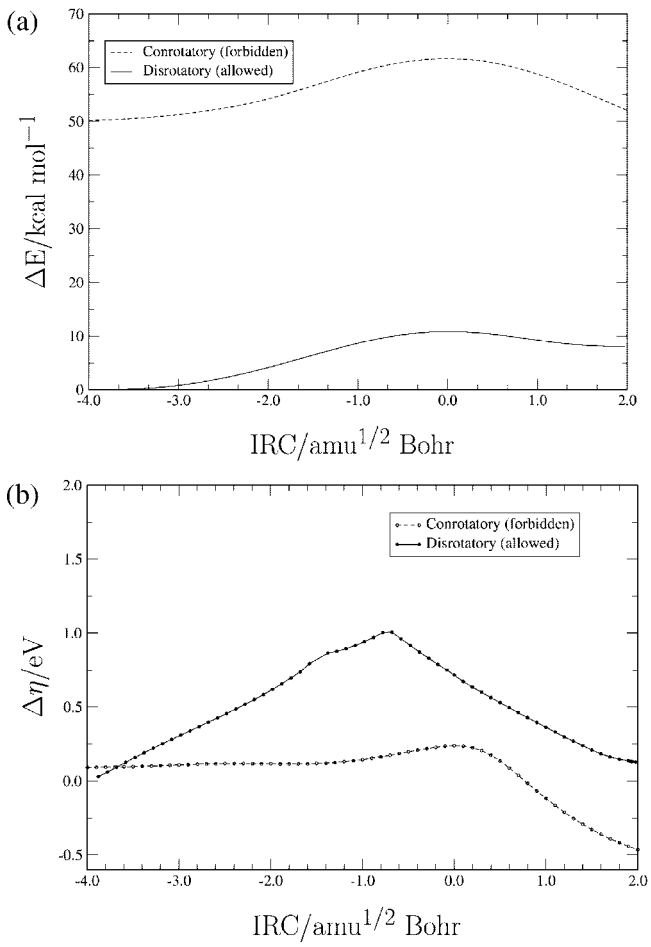


***Figure 6.*** Energetic reaction paths (a) and hardness profiles (b) for the conrotatory and disrotatory modes of the electrocyclization of cycloheptatriene.

***Table 1.*** Activation Energies $\Delta E^{\ddagger}$ and Hardnesses $\Delta\eta^{\ddagger}$ of the Allowed (A) and Forbidden (F) Modes of the Electrocyclizations of 2,4-Hexadiene, Cyclooctatetraene, and Cycloheptatriene[a]

| reactant | mode | $\Delta E^{\ddagger}$ | $\Delta\eta^{\ddagger}$ |
|---|---|---|---|
| 2,4-hexadiene[b] | conrotatory (A) | 43.8 | −0.343 |
| | disrotatory (F) | 49.0 | −0.028 |
| cyclooctatetraene[c] | disrotatory (A) | 30.0 | −1.212 |
| | conrotatory (F) | 67.2 | −0.098 |
| cycloheptatriene[d] | disrotatory (A) | 10.9 | −0.715 |
| | conrotatory (F) | 61.7 | −0.238 |

[a] All values are relative to the minimum energy structure along the reaction path of the allowed mode of each reaction. Energy differences were obtained at the B3LYP/6-311+G** level and are given in kilocalories per mole; hardness differences were obtained using eq 2 at the PBE/6-311+G** level and are given in electonvolts. [b] $E$(B3LYP/6-311+G**) = −234.700014 au, $\eta$(PBE/6-311+G**) = 9.395 eV. [c] $E$(B3LYP/6-311+G**) = −309.669028 au, $\eta$(PBE/6-311+G**) = 8.012 eV. [d] $E$(B3LYP/6-311+G**) = −271.584515 au, $\eta$(PBE/6-311+G**) = 8.906 eV.

case is not a good approximation to the IRC. We finally mention that, for this case, we carefully investigated the step-size dependence of the numerical differentiation. Computing the derivative using a very small step size (a 1° change in the dihedral angle, $\theta$), did not change the trends in the magnitudes of the derivatives and the conclusions drawn remain unchanged.

***Table 2.*** Initial Hardness Responses $(\partial\eta/\partial Q)_N$ of the Allowed (A) and Forbidden (F) Modes of the Electrocyclizations of 1,3-Butadiene, 1,3,5-Hexatriene, 2,4-Hexadiene, Cyclooctatetraene, and Cycloheptatriene[a]

| reactant | multiplicity | mode | $(\partial\eta/\partial Q)_N$ |
|---|---|---|---|
| 1,3-butadiene | S | conrotatory (A) | −0.00733 |
| | S | disrotatory (F) | −0.00498 |
| | T | disrotatory (A) | 0.00496 |
| | T | conrotatory (F) | 0.00228 |
| 1,3,5-hexatriene | S | disrotatory (A) | −0.00240 |
| | S | conrotatory (F) | −0.00469 |
| | T | conrotatory (A) | 0.00273 |
| | T | disrotatory (F) | 0.00130 |
| 2,4-hexadiene | S | conrotatory (A) | −0.152 |
| | S | disrotatory (F) | −0.210 |
| cyclooctatetraene | S | disrotatory (A) | 0.0952 |
| | S | conrotatory (F) | −0.158 |
| cycloheptatriene | S | disrotatory (A) | 0.310 |
| | S | conrotatory (F) | 0.00816 |

[a] In the first two cases, this derivative is evaluated for the reactant in its $C_{2v}$ symmetry and is expressed in electronvolts per degree. For the three other cyclizations, the derivative has been evaluated at the mass-weighted coordinate of −4.0. All derivatives were obtained numerically at the PBE/6-311+G** level (in the first two cases, with a step size of 5° in $Q$ and 0.1 in the other cases).

## 4. Conclusions

In our ongoing investigation to cast the Woodward–Hoffmann rules for pericyclic reactions into density functional reactivity theory, we have investigated a number of electrocyclizations involving both 4 and 6 $\pi$ electrons.

First, we have investigated the hardness response along the initial stages of the conrotatory and disrotatory modes of the electrocyclization of 1,3-butadiene and 1,3,5-hexatriene, both in the singlet and the triplet states; previously, this quantity was found, through its connection with the maximum hardness principle and the aromatic transition state concepts, to predict the allowedness of a cycloaddition reaction. In the 6 $\pi$ electron case, the allowed mode of the electrocyclization was found to correspond to the highest initial hardness, whereas in the 4 $\pi$ electron case, this result was not recovered. Upon consideration of the fully relaxed IRC path of this reaction, however, the expected result, i.e. higher hardness for the allowed mode, was recovered.

In the second part, we have studied the full IRC reaction and hardness profiles for the electrocyclization of 2,4-hexadiene, a 4 $\pi$ electron case, and the cyclizations of cyclooctatetraene and cycloheptatriene, two 6 $\pi$ electron cases. In all of these cases, the hardness of the allowed mode is higher than the hardness of the forbidden mode; also, the activation hardnesses for the former processes are always smaller than those of the latter ones.

Upon consideration of the magnitudes of the initial hardness response along the IRC, it can be concluded that for the electrocyclizations considered in this work, this quantity can be invoked to recover the Woodward–Hoffmann rules within the framework of the DFT-based perturbative perspective on chemical reactivity. In general, allowed electrocyclization reactions always have a higher initial hardness response than forbidden ones. This result, with the results in our earlier papers, supports the identification of the initial hardness response (or the closely related dual

descriptor) as the right key DFT-based reactivity indicator for describing pericyclic reactions.

**Acknowledgment.** P.G. and F.D.P. wish to thank the Fund for Scientific Research—Flanders (FWO) and the Free University of Brussels (VUB) for financial support. They also want to thank S. Janssens for his collaboration along the initial stages of this work. P.W.A. thanks NSERC and the Canada Research Chairs for research funding. M.T. thanks the European Community for financial help through the postdoctoral grants MEIF-CT-2006-025362. P.K.C. and S.G. thank CSIR, New Dehli, and BNRS, Mumbai, for financial assistance. M.E. and V.S. thank CSIR for financial assistance.

## References

(1) (a) Woodward, R. B.; Hoffmann, R. Stereochemistry of Electrocyclic Reactions. *J. Am. Chem. Soc.* **1965**, *87*, 395. (b) Hoffmann, R.; Woodward, R. B. Conservation of Orbital Symmetry. *Acc. Chem. Res.* **1968**, *1*, 17. (c) Woodward, R. B.; Hoffmann, R. The Conservation of Orbital Symmetry. *Angew. Chem.* **1969**, *81*, 797.; *Angew. Chem. Int. Ed. Engl.* **1969**, *8*, 781. (d) Woodward, R. B.; Hoffmann, R. *The Conservation of Orbital Symmetry*; Academic: New York, 1970.

(2) Gilchrist, T. L.; Storr, R. C. *Organic reactions and Orbital Symmetry*, second ed.; Cambridge University Press: Cambridge, 1978.

(3) (a) Goddard, W. A. Orbital Phase Continuity Principle and Selection Rules for Concerted Reactions. *J. Am. Chem. Soc.* **1970**, *92*, 7520. (b) Goddard, W. A.; Ladner, R. C. Generalized Orbital Description of the Reactions of Small Molecules. *J. Am. Chem. Soc.* **1971**, *93*, 6750. (c) Goddard, W. A. Selection Rules for Chemical Reactions Using the Orbital Phase Continuity Principle. *J. Am. Chem. Soc.* **1972**, *94*, 793.

(4) Zimmerman, H. Möbius-Hueckel Concept in Organic Chemistry. Application of Organic Molecules and Reactions. *Acc. Chem. Res.* **1971**, *33*, 272.

(5) (a) Hückel, E. Quantum-theoretical Contributions to the Benzene Problem. I. The Electron Configuration of Benzene and Related Compounds. *Z. Phys.* **1931**, *70*, 204.; Quantum Theoretical Contributions to the Problem of Aromatic and Non-Saturated Compounds. III. *Z. Phys.* **1932**, *76*, 628; Quantum Theoretical Contributions to the Problem of Aromatic and Unsaturated Compounds. IV. Free Radicals in Organic Chemistry. *Z. Phys.* **1933**, *83*, 632. (b) Heilbronner, E. Hückel Molecular Orbitals of Möbius-type Conformations of Annulenes. *Tetrahedron Lett.* **1964**, *5*, 1923.

(6) (a) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864. (b) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989. (c) Dreizler, R. M.; Gross, E. K. U. *Density Functional Theory*; Springer-Verlag: Berlin, Heidelberg, and New York, 1990. (d) Parr, R. G.; Yang, W. Density-Functional Theory of the Electronic Structure of Molecules. *Annu. Rev. Phys. Chem.* **1995**, *46*, 701. (e) Kohn, W.; Becke, A. D.; Parr, R. G. Density Functional Theory of Electronic Structure. *J. Phys. Chem.* **1996**, *100*, 12974. (f) Koch, W.; Holthausen, M. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, 2001.

(7) (a) Geerlings, P.; De Proft, F.; Langenaeker, W. Density Functional Theory: A Source of Chemical Concepts and a Cost-Effective Methodology for Their Calculation. *Adv. Quantum Chem.* **1999**, *33*, 303. (b) Chermette, H. Chemical Reactivity Indexes in Density Functional Theory. *J. Comput. Chem.* **1999**, *20*, 129. (c) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, *103*, 1793. (d) Ayers, P. W.; Anderson, J. S. M.; Bartolotti, L. J. Perturbative Perspectives on the Chemical Reaction Prediction Problem. *Int. J. Quantum Chem.* **2005**, *101*, 520. (e) Chattaraj, P. K. *J. Chem. Sci.* **2005**, 117.

(8) Chattaraj, P. K.; Fuentealba, P.; Gómez, B.; Contreras, R. Woodward—Hoffmann Rule in the Light of the Principles of Maximum Hardness and Minimum Polarizability: DFT and Ab Initio SCF Studies. *J. Am. Chem. Soc.* **2000**, *122*, 348.

(9) De Proft, F.; Ayers, P. W.; Fias, S.; Geerlings, P. Woodward-Hoffmann Rules in Density Functional Theory: Initial Hardness Response. *J. Chem. Phys.* **2006**, *125*, 214101.

(10) (a) Zhou, Z.; Parr, R. G.; Garst, J. F. Absolute Hardness as a Measure of Aromaticity. *Tetrahedron Lett.* **1988**, *29*, 4843. (b) Zhou, Z.; Parr, R. G. New Measures of Aromaticity: Absolute Hardness and Relative Hardness. *J. Am. Chem. Soc.* **1989**, *111*, 7371.

(11) (a) De Proft, F.; Geerlings, P. Conceptual and Computational DFT in the Study of Aromaticity. *Chem. Rev.* **2001**, *101*, 1451. (b) De Proft, F.; Geerlings, P. Relative Hardness as a Measure of Aromaticity. *Phys. Chem. Chem. Phys.* **2004**, *6*, 242.

(12) (a) Morell, C.; Grand, A.; Toro-Labbé, A. New Dual Descriptor for Chemical Reactivity. *J. Phys. Chem. A.* **2005**, *109*, 205. (b) Morell, C.; Grand, A.; Toro-Labbé, A. Theoretical Support for Using the $\Delta f(r)$ Descriptor. *Chem. Phys. Lett.* **2006**, *425*, 342. (c) Padmanabhan, J.; Parthasarathi, R.; Elango, M.; Subramanian, V.; Krishnamoorthy, B. S.; Gutierrez-Oliva, S.; Toro-Labbé, A.; Roy, D. R.; Chattaraj, P. K. A Multiphilic Descriptor for Chemical Reactivity and Selectivity. *J. Phys. Chem. A* **2007**, *111*, 9130.

(13) Ayers, P. W.; Morell, C.; De Proft, F.; Geerlings, P. Understanding the Woodward-Hoffmann rules by using changes in electron density. *Chem.—Eur. J.* **2007**, *13*, 8240.

(14) (a) Pearson, R. G. Hard and Soft Acids and Bases. *J. Am. Chem. Soc.* **1963**, *85*, 3533. (b) Pearson, R. G. *Hard and Soft Acids and Bases*; Dowen, Hutchinson and Ross: Stroudsberg, 1973. (c) Pearson, R. G. *Chemical Hardness*; Wiley-VCH; New York, 1997.

(15) Parr, R. G.; Pearson, R. G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512.

(16) (a) Chattaraj, P. K.; Lee, H.; Parr, R. G. HSAB Principle. *J. Am. Chem. Soc.* **1991**, *113*, 1855. (b) Chattaraj, P. K.; Schleyer, P. v. R. An Ab initio Study resulting in a Greater Understanding of the HSAB Principle. *J. Am. Chem. Soc.* **1994**, *116*, 1067. (c) Chattaraj, P. K.; Maiti, B. HSAB Principle Applied to the Time Evolution of Chemical Reactions. *J. Am. Chem. Soc.* **2003**, *125*, 2705. (d) Cedillo, A.; Chattaraj, P. K.; Parr, R. G. An Atoms- in- Molecules Partitioning of the Density. *Int. J. Quantum Chem.* **2000**, *77*, 403. (e) Ayers, P. W. The Maximum Hardness Principle Implies the Hard/Soft Acid/Base Rule. *J. Chem. Phys.* **2005**, *122*, 141102. (f) Chattaraj, P. K.; Ayers, P. W. The Maximum Hardness Principle Implies the Hard/Soft Acid/Base Rule. *J. Chem. Phys.* **2005**, *123*, 086101. (g) Ayers, P. W.; Parr, R. G.; Pearson, R. G. Elucidating the Hard/Soft Acid/Base Principle: A Perspective based on Half-Reactions. *J. Chem. Phys.* **2006**, *124*, 194107/1. (k) Chattaraj, P. K.; Ayers, P. W.; Melin, J. Further Links Between the Maximum Hardness

Principle and the Hard/Soft Acid/Base Principle: Insights from Hard/Soft Exchange reactions. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3853.

(17) (a) Pearson, R. G. Recent Advances in the Concept of Hard and Soft Acids and Bases. *J. Chem. Educ.* **1987**, *64*, 561. (b) Pearson, R. G. The Principle of Maximum Hardness. *Acc. Chem. Res.* **1993**, *26*, 250. (c) Pearson, R. G. Maximum Chemical and Physical Hardness. *J. Chem. Educ.* **1999**, *76*, 267.

(18) (a) Chattaraj, P. K.; Liu, G. H.; Parr, R. G. The Maximum Hardness Principle in the Gyftopoulos- Hatsopoulos Three-Level Model for An Atomic or Molecular Species and its Positive and Negative Ions. *Chem. Phys. Lett.* **1995**, *237*, 171. (b) Parr, R. G.; Chattaraj, P. K. Principle of Maximum Hardness. *J. Am. Chem. Soc.* **1991**, *113*, 1854. (c) Ayers, P. W.; Parr, R. G. Variational Principles for Describing Chemical Reactions: The Fukui Function and Chemical Hardness Revisited. *J. Am. Chem. Soc.* **2000**, *122*, 2010. (d) Torrent-Sucarrat, M.; Luis, J. M.; Duran, M.; Solà, M. On the Validity of the Maximum Hardness and Minimum Polarizability Principles for Nontotally Symmetric Vibrations. *J. Am. Chem. Soc.* **2001**, *123*, 7951.

(19) Tozer, D. J.; De Proft, F. Computation of the Hardness and the Problem of Negative Electron Affinities in Density Functional Theory. *J. Phys. Chem. A* **2005**, *109*, 8923.

(20) De Proft, F.; Sablon, N.; Tozer, D. J.; Geerlings, P. Calculation of Negative Electron Affinity and Aqueous Anion Hardness Using Kohn-Sham HOMO and LUMO Energies. *Faraday Discuss.* **2007**, *135*, 151.

(21) Olah, J.; De Proft, F.; Veszpremi, T.; Geerlings, P. Spinphilicity and spin-donicity of Substituted Carbenes, Silylenes, Germylenes, and Stannylenes. *J. Phys. Chem. A* **2004**, *108*, 490.

(22) 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 B03, Gaussian, Inc.: Wallingford, CT, 2004.

(23) (a) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648. (b) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785. (c) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623.

(24) For a detailed account on these types of basis sets, see, e.g., Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(25) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(26) Sakai, S. Theoretical Study on the Aromaticity of Transition States in Pericyclic Reactions. *J. Phys. Chem. A* **2006**, *110*, 6339.

(27) Torrent-Sucarrat, M.; Luis, J. M.; Duran, M.; Solà, M. The Hardness Profile as A Tool to Detect Spurious Stationary Points in the Potential Energy Surface. *J. Chem. Phys.* **2004**, *120*, 10914.

(28) Torrent-Sucarrat, M.; Duran, M.; Luis, J. M.; Solà, M. Generalizing the breakdown of the maximum hardness and minimum polarizabilities principles for nontotally symmetric vibrations to non-pi-conjugated organic molecules. *J. Phys. Chem. A* **2005**, *109*, 615.

(29) Torrent-Sucarrat, M.; Solà, M.; Toro-Labbé, A. Gas-phase structures, rotational barriers, and conformational properties of hydroxyl and mercapto derivatives of cyclohexa-2,5-dienone and cyclohexa-2,5-dienthione. *J. Phys. Chem. A* **2006**, *110*, 8901.

(30) Zhou, Z.; Parr, R. G. Activation Hardness: New Index for Describing the Orientation of Electrophilic Aromatic Substitution. *J. Am. Chem. Soc.* **1990**, *112*, 5720.

CT700289P

JCTC Journal of Chemical Theory and Computation

# Theory of High-Spin $d^4$ Complexes: An Angular-Overlap Model Parametrization of the Ligand Field in Vibronic-Coupling Calculations

Graham Carver,[†] Markus Thut,[‡] Christopher Noble,[§] and Philip L. W. Tregenna-Piggott*,[†]

*Laboratory for Neutron Scattering, PSI, CH-5232 Villigen, Switzerland, Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, CH-3012 Bern, Switzerland, and Centre for Magnetic Resonance, The University of Queensland, Brisbane, Australia*

Received December 22, 2007

**Abstract:** A new theoretical approach for the calculation of the electronic and molecular structures of octahedrally-coordinated high-spin $d^4$ complexes is described. A prescription for the construction of an effective $^3T_1 + {}^5E$ ($O$) Hamiltonian from the ligand-field matrices of a complex with general trigonal symmetry is given, where the ligand field is parametrized in terms of the angular-overlap model (AOM). The Jahn–Teller matrices for the $^3T_1 + ({}^5E\otimes e)$ vibronic Hamiltonian are constructed and the lowest eigenvalues are calculated by a numerical method. The model obviates the need to assume a temperature dependence of bonding parameters, inherent to the conventional ligand-field-theory approach and is applicable over the whole range of vibronic-coupling strengths, as demonstrated by example calculations on the $[Mn(OD_2)_6]^{3+}$ cation and $MgO:Cr^{2+}$.

## 1. Introduction

The resurgence of interest in the theoretical description of non-Kramers ions is undoubtedly due to the advent of the high-field EPR (HFEPR) technique. Oxidative stability, topicality, and the sheer fact that they invariably yield good spectra have all contributed to compounds of Mn(III) emerging as the deliciae of HFEPR spectroscopists. The strong Jahn–Teller effect, inherent to complexes of Mn(III), is usually incorporated into the analyses, only insofar as it leads to a tetragonally-distorted geometry. Dynamical Jahn–Teller coupling, which affects both the electronic and molecular structures of high-spin $d^4$ complexes, is usually ignored. This work describes a theoretical approach to calculating the electronic and molecular structures of octahedral high-spin $d^4$ complexes, which combines the chemically intuitive angular-overlap model with the rigor of vibronic-coupling calculations.

This work follows on from a recent article, in which crystallographic and spectroscopic data were presented for the $[Cr(OD_2)_6]^{2+}$ cation in the Cr(II) Tutton's salt and modeled using a $^5E\otimes e$ vibronic-coupling Hamiltonian.[1] The calculations provided a good account of the available spectroscopic and structural data and aptly demonstrated the need for a theoretical approach that goes beyond the conventional ligand-field model to include the vibrational coordinates of the molecule. The model is limited in its application, however, as the matrix elements of the spin–orbit interaction in the effective $^5E$ ($O$) electronic matrix were derived by perturbation theory, assuming perfect cubic symmetry in the strong-ligand-field limit. Even in cases of exact cubic symmetry, this model can serve only as a guide as to the magnitude of the spin–orbit splitting of the $^5E$ term. For this reason, the discussion focused on an interpretation of the ratio of the axial and rhombic zero-field-splitting parameters, $D$ and $E$, as a function of temperature, rather than their absolute values.

For octahedral Mn(III) complexes, this approach is wholly inappropriate for two reasons. First, the ratios of the

* E-mail: philip.tregenna@psi.ch.
[†] Laboratory for Neutron Scattering.
[‡] University of Bern.
[§] The University of Queensland.

10.1021/ct7003484 CCC: $40.75 
Published on Web 03/18/2008

octahedral-splitting parameter, 10Dq to the Racah parameters, $B$ and $C$, for a Mn(III) complex are larger than for the corresponding Cr(II) complex. Hence, the difference in the energies of the $^5E$ and excited $^3T_1$ terms will be much reduced, being comparable in magnitude to the spin–orbit-coupling parameter, $\zeta$. In this case the effect of the spin–orbit-coupling matrix elements between these terms, which largely governs the ground-state spin-Hamiltonian parameters, can no longer be treated as a perturbation. Second, for octahedrally coordinated Mn(III) complexes the ground-state zero-field splitting is far more susceptible to the trigonal component of the ligand field.[2]

In order for vibronic-coupling calculations of Mn(III) complexes to be meaningful, the effective electronic Hamiltonian must therefore be generated in the point group of the undistorted complex and off-diagonal spin–orbit-coupling matrix elements between the $^5E$ and excited $^3T_1$ terms must be included explicitly. This article presents a procedure for calculating the effective electronic Hamiltonian spanning the basis functions of the $^3T_1$ and $^5E$ terms, using ligand-field matrices generated by the angular-overlap model (AOM). Calculations of the ground-state electronic and low-temperature molecular structures of $MgO:Cr^{2+}$ and the $[Mn(OD_2)_6]^{3+}$ cation in the cesium manganese alum, using the $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonian serve to illustrate the model's wide applicability. The dependence of the electronic and molecular structures on the parameters of the vibronic Hamiltonian is discussed in the light of existing theoretical models.

## 2. Review of Previous Models

In the following sections, we summarize the various approaches that have been applied to model experimental data obtained for high-spin $d^4$ complexes.

**2.1. Ligand-Field Theory.** The ground-state energies and geometries of high-spin $d^4$ transition-metal complexes are frequently rationalized in terms of a ligand-field or crystal-field model.[3,4] The $^5D$ free-ion ground term is split through a cubic field, giving rise to a $^5E_g$ ($O_h$) ground term and a higher-lying $^5T_{2g}$ term. In the ligand-field description the ground term is split further by a tetragonal component of the ligand field, resulting in either a $^5A_{1g}$ or $^5B_{1g}$ ($D_{4h}$) term lower lying. Off-diagonal spin–orbit-coupling matrix elements between the ground singlet and higher-lying terms result in a zero-field splitting (zfs) of the orbital singlet. The resulting ground-state energy scheme is usually well described by the spin Hamiltonian,

$$H = D\left[S_z^2 - \frac{1}{3}S(S+1)\right] + E[S_x^2 - S_y^2] \qquad (1)$$

correct to second order, operating in the basis of the $S = 2$ spin functions, where $D$ and $E$ are the axial and rhombic zfs parameters, respectively.[5]

In symmetries lower than $O_h$ there are matrix elements of spin–orbit coupling within the $^5E$ ground term. However, the dominant contributions to the zfs of the orbital-singlet ground term are the off-diagonal spin–orbit-coupling matrix elements between the $^5E_g$ and excited $^3T_{1g}$ and $^5T_{2g}$ ($O_h$) electronic terms. The following approximate expression for the axial zfs parameter, $D$ has been derived by perturbation theory,[6]

$$D = \pm\left(\frac{\zeta^2}{6B + 5C - 10\mathrm{Dq}}\right) \pm \frac{3}{16}\left(\frac{\zeta^2}{10\mathrm{Dq}}\right) \qquad (2)$$

where $B$ and $C$ are the Racah parameters and 10Dq and $\zeta$ are the octahedral-splitting and spin–orbit-coupling parameters, respectively. The two contributions to $D$ given here are those originating from the $^3T_{1g}$ and $^5T_{2g}$ terms, respectively; the upper sign holds for the $^5A_{1g}$ ($D_{4h}$) term lower lying, and the lower, for $^5B_{1g}$ lower lying. If the $^5B_{1g}$ term is lower lying, $D$ is negative and the molecule has an elongated $D_{4h}$ geometry. On the other hand, if the sign of the tetragonal field is reversed and the $^5A_{1g}$ term is lower lying, $D$ is positive and the molecule has a compressed $D_{4h}$ geometry.

Slight deviations of the symmetry of the $ML_6$ framework away from perfect $D_{4h}$ symmetry correspond to an additional rhombic component in the crystal field and give rise to a nonzero rhombic zfs parameter. In the case of the $[Mn(OH_2)_6]^{3+}$ cation in the cesium manganese alum, where the rhombic distortion of the $\sigma$-bonding framework is barely significant, it has been demonstrated that the magnitude of the rhombic zfs parameter is governed by the strength and nature of the $\pi$-anisotropy of the surrounding ligands,[2] imposing a large trigonal ligand field.

The low-level energy schemes and associated molecular geometries of the majority of octahedral high-spin Mn(III) complexes can be rationalized in terms of ligand-field theory. Most hexa-coordinate homoleptic Mn(III) complexes with oxygen or nitrogen donors exhibit a tetragonally elongated geometry with a negative $D$.[2–4,7–9] In the comparatively few studies of octahedral Cr(II) complexes, a negative $D$ with tetragonal elongation is also observed.[10] Interestingly, while a negative $D$ was recently reported for the tetragonally-elongated $\gamma$-$Mn(acac)_3$ complex,[4] powder-magnetic-susceptibility studies by Gregson et al. performed on the tetragonally-compressed $\beta$-modification indicate a positive $D$ value,[11] also in accordance with ligand-field theory.

By contrast, in the case of *trans*-$[Mn(cyclam)I_2]I$, despite a tetragonal elongation a positive $D$ value is nonetheless observed.[12] The low-energy excitations in this complex are no longer of d−d character, and the observations are explained by considering interactions of the ground term with the low-energy LMCT terms; hence, in this instance, the ligand-field predictions are no longer valid.

**2.2. Jahn–Teller Coupling.** In instances of near or actual orbital degeneracy, high-spin $d^4$ complexes can exhibit a marked variation in the ligand-field parameters and bond lengths as a function of temperature, which cannot be explained within the framework of a "static" ligand-field model. A more rigorous approach allows for coupling of the electronic states to lattice vibrations via the Jahn–Teller interaction.

The Jahn–Teller effect is based on a theorem first described by Jahn and Teller in 1937,[13] following an original suggestion by Landau.[14] This theorem affirms that, in a nonlinear molecule having an electronically degenerate state (excepting

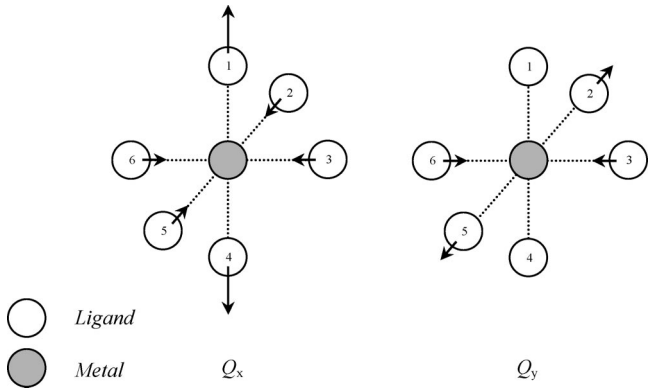


***Figure 1.*** The two components of the $\nu_2(ML_6)$ vibration. Arrows indicate the directions and relative magnitudes of the displacements.

odd-electron systems having simple Kramers degeneracy), the nuclei of the molecule will displace themselves in order to destroy the symmetry of their configuration and to remove the electronic degeneracy. The Jahn–Teller effect cannot, however, reduce the overall degeneracy of a level, since the overall Hamiltonian retains its symmetry and remains totally symmetric under the operations of the point group of the system. The consequence of the Jahn–Teller interaction is thus to replace the original electronic degeneracy by vibronic degeneracy. Accordingly, the E⊗e Jahn–Teller interaction, operating within a $^5$E electronic ground term will give rise to a vibronic-energy-level structure characterized by a $^5$E vibronic ground term.

A Hamiltonian including both ligand-field and vibronic contributions has the general form,

$$H = H_{el} + H_{ph} + H_{JT} \tag{3}$$

where $H_{ph}$ and $H_{JT}$ are the phonon- and Jahn–Teller-coupling terms, respectively. $H_{el}$ is the electronic part of the Hamiltonian given by

$$H_{el} = H_{ER} + H_{LF} + H_{SO} + H_{st} \tag{4}$$

where $H_{ER}$, $H_{LF}$, $H_{SO}$, and $H_{st}$ designate the electron-repulsion, high-symmetry ligand-field, spin–orbit coupling, and low-symmetry strain contributions to the energy respectively.

*2.2.1. E⊗e Vibronic Hamiltonian.* In the following formulation of the E⊗e Jahn–Teller Hamiltonian, we make use of the so-called "cluster model",[15] where we consider coupling of the E orbital state to a single pair of distortion coordinates, $Q_x$ and $Q_y$, representing the appropriate modes of distortion of the Jahn–Teller ion and its nearest neighbors.

The matrices of the E⊗e vibronic Hamiltonian are traditionally expressed in terms of the cubic $|\theta\rangle,|\varepsilon\rangle$ orbital basis, and the explicit form of the Jahn–Teller matrices in these bases is given elsewhere.[1,16,17] In the present article, however, we present the matrices in the complex-trigonal orbital basis. This basis is a natural choice for the construction of the effective Hamiltonian matrix in section 3. With respect to the $|u_+\rangle,|u_-\rangle$ orbital basis, the vibronic Hamiltonian, correct to second order, has the following form,

$$H_{JT} = A_1\begin{pmatrix} 0 & Q_x + iQ_y \\ Q_x - iQ_y & 0 \end{pmatrix} + A_2 \begin{pmatrix} 0 & {Q_y}^2 - {Q_x}^2 + 2iQ_xQ_y \\ {Q_y}^2 - {Q_x}^2 - 2iQ_xQ_y & 0 \end{pmatrix} \tag{5}$$

$$H_{ph} = \frac{1}{2}\begin{pmatrix} ({P_x}^2 + {P_y}^2 + {Q_x}^2 + {Q_y}^2)\hbar\omega & 0 \\ 0 & ({P_x}^2 + {P_y}^2 + {Q_x}^2 + {Q_y}^2)\hbar\omega \end{pmatrix} \tag{6}$$

where $Q_x$ and $Q_y$ are the components of the degenerate $\nu_2(ML_6)$ stretching vibration, transforming as the two components of the e irreducible representation in a trigonal point group. Displacements along these coordinates may be expressed in a basis of increments in the metal–ligand (M-L) bond lengths:

$$Q_x = \frac{1}{\sqrt{12}}(2r_1 + 2r_4 - r_2 - r_5 - r_3 - r_6)$$

$$Q_y = \frac{1}{2}(r_2 + r_5 - r_3 - r_6) \tag{7}$$

where $r_i$ is a unit displacement along the M-$L_i$ bond vector and the ligands are numbered according to the scheme used in Figure 1.

$A_1$ and $A_2$ are the first-order and second-order Jahn–Teller coupling constants and $\hbar\omega$ is the energy of the vibration. $H_{ph}$ is the Hamiltonian for an e phonon mode before coupling, and $Q_i$ and $P_i$ are dimensionless operators, related to the observables for position and momentum, $q_i$, $p_i$ by,

$$Q_i = \sqrt{\frac{\mu\omega}{\hbar}}q_i, \quad P_i = \frac{1}{\sqrt{\mu\hbar\omega}}p_i \tag{8}$$

where $\mu$ is the reduced mass of the phonon mode. The unit displacements, $r_i$ are related to $q_x$ and $q_y$ by,

$$r_1 = r_4 = \frac{1}{\sqrt{12}}2q_x; \quad r_2 = r_5 = \frac{(-q_x + \sqrt{3}q_y)}{\sqrt{12}}; \quad r_3 = r_6 = \frac{(-q_x - \sqrt{3}q_y)}{\sqrt{12}} \tag{9}$$

Finally, $H_{st}$ has the following form in the E orbital basis.

$$H_{st} = \begin{pmatrix} 0 & e_x + ie_y \\ e_x - ie_y & 0 \end{pmatrix} \tag{10}$$

where $e_x$ and $e_y$ are the two components of low-symmetry strain, having the same transformation properties as $Q_x$ and $Q_y$ above.

*2.2.2. Potential-Energy Surfaces Resulting from E⊗e Jahn–Teller Coupling.* In the adiabatic approximation, in which the kinetic contribution to the energies is neglected, diagonalization of the Jahn–Teller matrices defined in eqs 5 and 6 yields analytical expressions for the two adiabatic potential-energy sheets (APES) as functions of the $Q_x$ and $Q_y$ distortion coordinates. In Figure 2 are presented plots of the two APES, calculated using the parameters $A_1 = -900$, $A_2 = 0$, and $\hbar\omega = 254$ cm$^{-1}$.

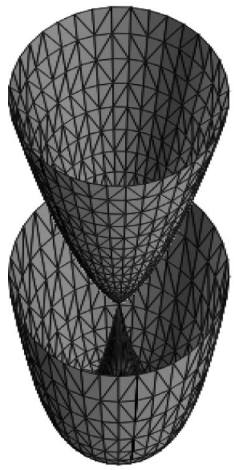

***Figure 2.*** 3D-plots of the two APES resulting from first-order Jahn–Teller coupling.

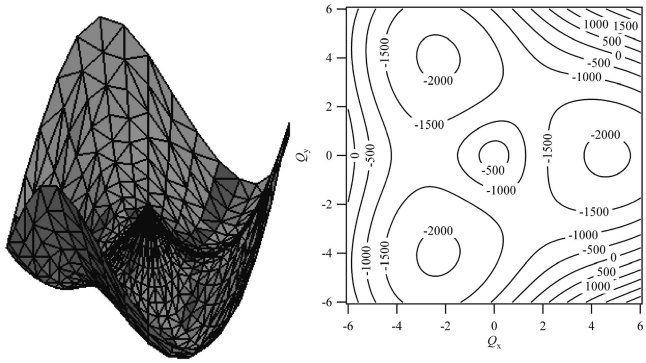


***Figure 3.*** 3D and contour plots of the lower APES resulting from first-order and second-order Jahn–Teller coupling. Energies are given in wavenumbers.

The lower potential-energy surface has the form of a "Mexican hat" and the difference in the energies between the two potential-energy surfaces at distortion coordinates at the bottom of the trough is $4E_{JT}$, where $E_{JT}$ is the Jahn–Teller stabilization energy, defined as $E_{JT} = A_1^2/2\hbar\omega$. The vibronic structure resulting from diagonalization of the full vibronic Hamiltonian exhibits a particularly large density of states at $\sim 4E_{JT}$ above the ground state. Optical transitions observed between the ground-state and these excited states are therefore helpful in obtaining estimates for $E_{JT}$ and consequently for the first-order Jahn–Teller coupling constant, $A_1$. In the framework of the ligand-field model the observed spectral band is commonly assigned to the d−d transition between the $^{2S+1}B_{1g}$ and $^{2S+1}A_{1g}$ ($D_{4h}$) terms.

When, in addition to first-order coupling, second-order coupling is also considered, warping of the lower potential-energy surface occurs, resulting in an undulating variation of the potential energy along the bottom of the trough. In Figure 3 are presented 3D and contour plots of the lower APES, calculated using the parameters $A_1 = -900$, $A_2 = 30$ cm$^{-1}$, and $\hbar\omega = 254$ cm$^{-1}$. These are the parameters employed to model the data of $[Cr(OD_2)_6]^{2+}$ in the Tutton's salts.[1] If the undistorted geometry has $O_h$ point symmetry, the resulting three potential minima and three maxima in the trough are at distortion coordinates corresponding to special points of $D_{4h}$ cokernel symmetry. All other points

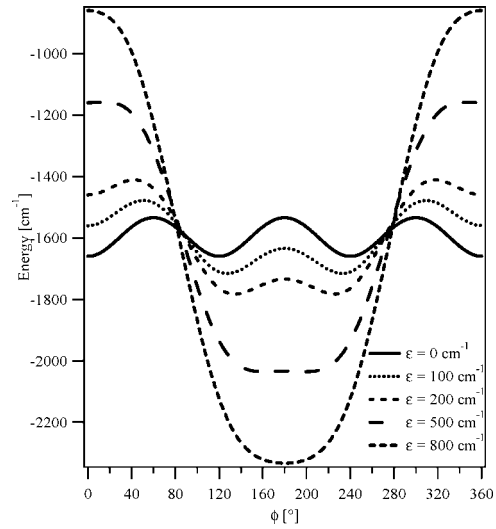


***Figure 4.*** Plots along the path of minimum potential energy of the lowest APES as a function of $\phi$, calculated for various strain magnitudes, $\varepsilon$.

on the potential-energy surface correspond to geometries of $D_{2h}$ kernel symmetry. Alternatively, if the original undistorted geometry has $S_6$ symmetry, any distortion in the $\{Q_x, Q_y\}$ coordinate space will correspond to a lowering of symmetry to $C_i$ symmetry. Defining a position in $\{Q_x, Q_y\}$ coordinate space in terms of the polar coordinates $\{\rho, \phi\}$ such that $Q_x = \rho \cos \phi$ and $Q_y = \rho \sin \phi$, the turning points of the lower potential-energy surface are

$$\rho = \frac{\pm A_1}{\hbar\omega \pm (-1)^n 2A_2}, \quad \phi = \frac{n\pi}{3},\ n = 0, 1, ..., 5 \qquad (11)$$

assuming that $\hbar\omega$ is larger than $A_2$. The upper and lower signs correspond to the cases $A_1 > 0$ and $A_1 < 0$ respectively. If $A_1/A_2 < 0$, the minima points occur for $n = 0$, 2, 4 and saddle-points for $n = 1$, 3, 5, while for the case $A_1/A_2 > 0$, the two types of points are interchanged.[18] In accordance with our definitions of $Q_x$ and $Q_y$ above, turning points occurring at $\phi = 0$, $2\pi/3$, and $4\pi/3$ correspond to tetragonally-elongated geometries, whereas those occurring at $\phi = \pi/3$, $\pi$, and $5\pi/3$ equate to tetragonally-compressed geometries.

Analytical expressions for the lowest APES, resulting from first-order and second-order Jahn–Teller coupling and low-symmetry strain, can be obtained by diagonalization of (5), (6), and (10), where the terms in $P$ are once again ignored. If we express the components of strain in terms of the strain magnitude, $\varepsilon$, and strain angle, $\theta$, such that $e_x = -\varepsilon \cos \theta$ and $e_y = -\varepsilon \sin \theta$, the effect of strain on the form of the APES is easily rationalized. Using this scheme, in the limit where $A_2 = 0$, the effect of strain is to create one potential minimum at $\{Q_x, Q_y\}$ distortion coordinates corresponding to $\phi = \theta$ with a depth of $2\varepsilon$. In Figure 4 are presented plots along the path of minimum potential energy as a function of $\phi$ for various $\varepsilon$. The lowest APES was calculated using the parameters: $A_1 = -900$, $A_2 = 5$, $\hbar\omega = 254$, $e_x = \varepsilon$ and $e_y = 0$ cm$^{-1}$. This particularly simple choice of parameters, corresponding to $\theta = \pi$, shows the effect of strain on the APES very clearly.

In the limit $A_2/\varepsilon \rightarrow 0$, the minimum of the potential-energy surface is determined by the components of strain

and for $\varepsilon = 800$ cm$^{-1}$ and $\theta = \pi$ is found at $\{Q_x, Q_y\}$ coordinates corresponding to $\phi = \pi$ and hence to a tetragonal compression. When $\varepsilon = 0$, the minima of the potential-energy surface are determined completely by $A_2$ and are at $\{Q_x, Q_y\}$ coordinates corresponding to $\phi = 0$, $2\pi/3$, and $4\pi/3$ and thus to tetragonal elongations. In the intermediate regime, the form of the potential-energy surface is dictated by the relative magnitudes of the $A_2$ and $\varepsilon$ parameters. In our example, as $\varepsilon$ is increased, the positions of the potential minima are distorted more and more toward $\{Q_x, Q_y\}$ coordinates corresponding to $\phi = \pi$.

In the limit of infinitely large nuclear masses, i.e., for infinitely small phonon energies, we expect the equilibrium configuration of the complex to correspond to the minimum of the potential-energy surface. While in reality this condition is never met, if $\varepsilon$ or $A_2$ are large compared to the thermal or phonon energies, the distortion coordinates at the minimum of the potential-energy surface will yield a reasonable first approximation of the molecular geometry. If $\varepsilon$ or $A_2$ are comparable in magnitude to the thermal energies a dependence of the observed molecular geometry on the temperature is expected. In this case, a more accurate calculation of the molecular geometry can be obtained by solving the full vibronic Hamiltonian. Once the eigenvalues and eigenvectors have been obtained, the geometry can be determined from the expectation values of the $Q_x$ and $Q_y$ operators in the thermally occupied vibronic states, using eqs 8 and 9.[1]

*2.2.3. $^5$E⊗e Vibronic Hamiltonian.* In the case of $d^1$ and $d^9$ systems, an extension of the E⊗e vibronic Hamiltonian to include spin is relatively straightforward. These systems are uncomplicated by the effects of electron repulsion, and the small basis size, only ten electronic states, means that $H_{el}$ can be explicitly included in the calculations. By contrast, the $d^4$ electronic configuration gives rise to a total of 210 electronic states. The sheer magnitude of the resulting vibronic-coupling matrix renders an explicit treatment of $H_{el}$ impractical. Instead, the effects of $H_{el}$ on the energies of the $^5$E ground term are absorbed into an effective Hamiltonian, constructed using second-order perturbation theory.

The temperature dependence of crystallographic and inelastic-neutron-scattering (INS) data for the $(ND_4)_2Cr(OD_2)_6(SO_4)_2$ salt was interpreted previously in terms of a $^5$E⊗e vibronic Hamiltonian, perturbed by low-symmetry strain.[1] The electronic part of the Hamiltonian, $H_{el}$, was accounted for by means of an effective $^5E_g$ ($O_h$) Hamiltonian,

$$H_{\text{eff}} = \frac{b}{2}\begin{pmatrix} S_\theta & -S_\varepsilon \\ -S_\varepsilon & -S_\theta \end{pmatrix} \tag{12}$$

expressed in the cubic $|\theta\rangle,|\varepsilon\rangle$ orbital basis, where,

$$S_\theta = S_z^2 - \frac{1}{3}S(S+1); \quad S_\varepsilon = \frac{1}{\sqrt{3}}(S_x^2 - S_y^2) \tag{13}$$

and $S_x$, $S_y$, and $S_z$ are the spin-angular-momentum operators operating in the basis of the $S = 2$ spin states.[1,19] The effective-Hamiltonian parameter, $b$, describes the perturbation of the $^5$E ground term through spin–orbit coupling, and an approximate expression for this parameter in terms of the spin–orbit-coupling, cubic-field, and electron-repulsion parameters was derived explicitly using second-order perturbation theory by Weihe and Mossin.[16]

A more general effective $^5$E ($C_{3v}$) Hamiltonian was presented previously by Simpson et al.[20] This Hamiltonian is also applicable to $D_{3d}$, $D_{3h}$, $S_6$, and $O_h$ symmetries and, given in the complex-trigonal $|u_+\rangle,|u_-\rangle$ orbital basis, has the following general form,

$$H_{\text{eff}} = \begin{pmatrix} \lambda S_z + dS_z^2 & \frac{\mu_1}{2}(S_+S_z + S_zS_+) + \mu_2 S_-^2 \\ \frac{\mu_1}{2}(S_-S_z + S_zS_-) + \mu_2 S_+^2 & -\lambda S_z + dS_z^2 \end{pmatrix} \tag{14}$$

where $S_-$, $S_+$, and $S_z$ are the spin-angular-momentum operators operating in the basis of the $S = 2$ spin functions, and $\lambda$, $d$, $\mu_1$, and $\mu_2$ are the effective-Hamiltonian parameters. In perfect cubic symmetry, the effective $^5$E Hamiltonian reduces to a particularly simple form. In this instance $\mu_1 = 2\sqrt{2}\mu_2$ and $d = \lambda = 0$. Making the substitution $\mu_2 = b/6$, the energies of the $^5$E spinor levels are then $-2b$, $-b$, 0, $b$, and $2b$. The result is an effective cubic $^5$E Hamiltonian, expressed in the complex-trigonal basis.

*2.2.4. Ham Theory.* The E⊗e Jahn–Teller interaction, operating within a $^{2S+1}$E electronic ground term, gives rise to a $^{2S+1}$E vibronic ground term. In the case of first-order Jahn–Teller coupling the first excited levels are accidentally degenerate vibronic $^{2S+1}A_1$ and $^{2S+1}A_2$ singlet levels. When second-order Jahn–Teller coupling is included, the pseudo-degeneracy of the excited vibronic singlets is lifted. As the magnitude of second-order coupling is increased, the $^{2S+1}$E vibronic ground term is approached asymptotically by either the $^{2S+1}A_1$ or $^{2S+1}A_2$ excited vibronic singlet term. The difference in the energies of the lower-lying doublet and the first excited singlet is referred to in the literature as the tunnelling splitting, $3\Gamma$, and depends inversely on the strength of the vibronic coupling.

Ham derived a generalized matrix form for interactions which may perturb a vibronic $^{2S+1}$E ground term and the next excited orbital-singlet level, resulting from strong first-order and second-order Jahn–Teller coupling.[21] When the lowest excited orbital-singlet state is close in energy to the ground state, we must include it in setting up the secular equation for the effect of any perturbation on the ground state. In the case where the excited singlet is of symmetry $A_1$, this matrix has the following form with respect to the basis states $|A_1\rangle$, $|g\theta\rangle$, and $|g\varepsilon\rangle$,

$$H = \begin{pmatrix} 3\Gamma & rG_\theta & rG_\varepsilon \\ rG_\theta & -qG_\theta & qG_\varepsilon \\ rG_\varepsilon & qG_\varepsilon & qG_\theta \end{pmatrix} \tag{15}$$

where $|g\theta\rangle$ and $|g\varepsilon\rangle$ are the two components of the E vibronic ground term, which transform as the $\theta$ and $\varepsilon$ components of the E representation. The corresponding matrix for $|A_2\rangle$ lower is

$$H = \begin{pmatrix} 3\Gamma & rG_\varepsilon & -rG_\theta \\ rG_\varepsilon & -qG_\theta & qG_\varepsilon \\ -rG_\theta & qG_\varepsilon & qG_\theta \end{pmatrix} \tag{16}$$

In the above matrices $q$ and $r$ are the Ham reduction factors. For zero vibronic coupling, $q = 1$ and $r = 0$; for strong first-order Jahn–Teller coupling, $q = -r = {}^1/_2$; and for strong first-order and appreciable second-order coupling, $q = -r/\sqrt{2} = {}^1/_2$. The operators $G_\theta$ and $G_\varepsilon$ are functions of the relevant perturbations which transform as the $\theta$ and $\varepsilon$ components of the E representation in cubic symmetry. The effective spin–orbit-coupling matrix is generated by making the following substitutions: $G_\theta = -b/2S_\theta$ and $G_\varepsilon = -b/2S_\varepsilon$ where $S_\theta$ and $S_\varepsilon$ are the operators defined in eq 13 above, operating in a basis of the spin functions. The corresponding substitutions required to generate the strain matrix are: $G_\theta = e_\theta$ and $G_\varepsilon = e_\varepsilon$, where $e_\theta$ and $e_\varepsilon$ are the tetragonal and rhombic strain parameters, respectively.

In the limit of weak first-order Jahn–Teller coupling, where the energy separation between the $^{2S+1}$E vibronic ground term and excited vibronic-singlet terms is large relative to the magnitude of an applied perturbation, the effect of Jahn–Teller coupling on the ground term is simply to reduce the parameter entering the description of the perturbation. The following example shall serve to illustrate.

By replacing the operators $G_\theta$ and $G_\varepsilon$ in eq 15 above with the expressions corresponding to the spin–orbit interaction, an effective spin–orbit-coupling matrix, operating in the $^5$E and excited $^5$A$_1$ vibronic bases, can be constructed. In the limit where $3\Gamma >> b$, we can neglect the off-diagonal matrix elements between the $^5$E vibronic ground term and excited singlets. Expanding the $|g\theta\rangle,|g\varepsilon\rangle$ part of the matrix in eq 15 into a basis consisting of product functions of the vibronic states and the five $S = 2$ spin states, the eigenvalues of the matrix are $-2qb$, $-qb$, 0, $qb$, and $2qb$. When $q = 1$, the matrix is identical to the $^5$E effective-Hamiltonian matrix given in eqs 12 and 13, and the eigenvalues correspond to the energies of the spinor levels of the uncoupled $^5$E electronic ground term. The effect of Jahn–Teller coupling on the spinor energies of the $^5$E ground term is then to replace the parameter $b$ by $b' = qb$ resulting in a uniform reduction in the energies.

## 3. $^3$T$_1$ + ($^5$E⊗e) Vibronic Hamiltonian

In all of the Jahn–Teller-coupling calculations carried out thus far to model the electronic and molecular structures of high-spin d$^4$ complexes a $^5$E effective Hamiltonian has been employed. The effects of spin–orbit coupling and the trigonal-splitting of higher-lying terms have been included only insofar as they are allowed to perturb the energies of the $^5$E basis functions. Such an approach takes no account of the contamination of these functions through the in-mixing of higher-lying states. In particular, when these states are close in energy compared to the magnitude of spin–orbit coupling, the zeroth-order wave functions are no longer a good approximation to the ground state, and perturbation theory breaks down. As we will later see, a contamination of the wave functions has a profound effect on the degree of quenching of the electronic operators within the $^5$E ground term.

We present here a new $^3$T$_1$ + ($^5$E⊗e) vibronic-coupling model, where the effective Hamiltonian is extended to span both the $^5$E and excited $^3$T$_1$ bases. The matrix elements of spin–orbit coupling between the $^5$E and $^3$T$_1$ terms, and the trigonal splitting of the $^3$T$_1$ term are included explicitly. Spin–orbit coupling to higher-lying terms is still treated as a second-order perturbation on the $^5$E part of the electronic basis only. The general form of the $^5$E part of the electronic Hamiltonian in the complex-trigonal basis is given in eq 14. The form of the $^3$T$_1$ part of the matrix and of the off-diagonal spin–orbit-coupling matrix between the $^5$E and $^3$T$_1$ terms has been presented elsewhere.[22,23]

The explicit form of the effective Hamiltonian for a given set of ligand-field parameters can be derived from the d$^4$ ligand-field matrices. Here, we parametrize the ligand field in terms of the AOM, where the d-orbital energies are expressed in terms of the metal–ligand $\sigma$- and $\pi$-bonding parameters, $e_\sigma$, $e_{\pi\|}$, and $e_{\pi\perp}$. This model provides an obvious link between the d-orbital energies and the structure of the complex. The pictorial simplicity of the AOM approach and transferability of the AOM parameters from one complex to another confer obvious advantages over the crystal-field approach.

We shall now outline the procedure for constructing the effective Hamiltonian from the AOM matrices. First, the ligand-field, spin–orbit coupling, spin- and orbital-angular-momentum matrices are constructed in the full d$^4$ electronic basis, using the program LIGFIELD.[24] For a given set of AOM and Racah parameters the ligand-field and electron-repulsion matrices are diagonalized together. The nineteen lowest-lying states then correspond to the $^5$E and $^3$T$_1$ terms, and the eigenfunctions are now used to transform the Hamiltonian matrices into this new "ligand field" basis.

We require that the orbital components of the $^5$E basis functions transform as $|u_+\rangle$ and $|u_-\rangle$. In trigonal symmetry, this is easily achieved by diagonalization of the matrix representation of $L_z$ within this term. However, in the special case of perfect octahedral geometry, the operator $L_z$ has no nonzero matrix elements within the $^5$E ground term. To facilitate the identification of the orbital states, we perturb the AOM input geometry away from perfect octahedral symmetry by introducing a very small trigonal distortion. With a little additional effort, $^5$E and $^3$T$_1$ basis functions that are eigenfunctions of both the $L_z$ and $S_z$ operators can be created. This is accomplished by the diagonalization of the matrix representation of the operator: $L_z + S_z$ operating only within the $^5$E and $^3$T$_1$ terms.

As the Jahn–Teller and strain matrices have been constructed separately, the relative phases of the resulting $^5$E basis functions do matter here. From the table of Clebsch−Gordan coefficients for trigonal bases,[25] we can construct the following operator equivalents, transforming as the u$_+$ and u$_-$ components of the E irreducible representation,

$$O_{Eu_+} = \frac{1}{2}(L_x^2 - L_y^2 - i(L_xL_y + L_yL_x))$$
$$O_{Eu_-} = \frac{1}{2}(L_y^2 - L_x^2 - i(L_xL_y + L_yL_x)) \qquad (17)$$

In the complex-trigonal $|u_+\rangle,|u_-\rangle$ orbital basis, the matrices of these operators must have the following form,

$$O_{Eu_+} = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}, \quad O_{Eu_-} = \begin{pmatrix} 0 & 0 \\ +1 & 0 \end{pmatrix} \qquad (18)$$

Following inspection of the form of the matrices of these operators and of the $S_+$ operator in a basis of the new $^5E$ functions, the phases of the eigenfunctions are corrected accordingly. Finally these functions are used to transform the ligand-field and spin–orbit-coupling matrices into the new basis.

Having now obtained the numerical ligand-field and spin–orbit-coupling matrices in an appropriate basis, we calculate the matrix elements of the spin–orbit perturbation, $V$ on the $^5E$ part of the basis according to the prescription,

$$\langle i|V|j\rangle = \sum_{n=20}^{210} \frac{\langle i|H_{SO}|n\rangle\langle n|H_{SO}|j\rangle}{E_i - E_n} \tag{19}$$

The indices $i,j$ refer to the states belonging to the $^5E$ term; the index $n$ runs over all 210 ligand-field states resulting from the $d^4$ electronic configuration, excepting the states belonging to the ground $^5E$ and $^3T_1$ states; and $E_i$ and $E_n$ are the energies of the ligand-field states resulting from diagonalization of the ligand-field and electron-repulsion matrices. The resulting perturbation, $V$, has the general matrix form given in eq 14 above. Finally, the complete effective Hamiltonian in the new $^3T_1 + {}^5E$ basis is constructed according to,

$$H_{eff} = V + H_{SO} + H_{LF} + H_{ER} \tag{20}$$

where $H_{SO}$ includes the matrix elements of spin–orbit coupling within and between the $^5E$ and $^3T_1$ terms, not accounted for in the perturbation. In this basis $H_{LF} + H_{ER}$ is diagonal, corresponding to the energies of the terms before spin–orbit coupling.

The matrix of the $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonian can now be constructed from eqs 5, 6, and 10 and from the effective $^3T_1 + {}^5E$ Hamiltonian, derived from the AOM matrices, in a basis of products of the electronic states and the states of the $\{Q_x, Q_y\}$ harmonic oscillator. A typical calculation with $n_v = 30$ results in a sparse matrix of dimension 9424.

## 4. Calculations and Discussion

In the following sections, we present example calculations for a variety of Jahn–Teller-coupling strengths, employing the $^3T_1 + (^5E\otimes e)$ model. The results are compared with those obtained using other models, and the dependence of the electronic and molecular structures on the parameters of the vibronic Hamiltonian is discussed.

**4.1. First-Order Jahn–Teller Coupling.** The three-state model of Ham constitutes a reasonable approximation to the Jahn–Teller Hamiltonian in the limit where spin–orbit coupling and low-symmetry-strain contributions to the Hamiltonian are small compared to Jahn–Teller coupling. Until now, however, no-one has tested the validity of this model by direct comparison of the results with those obtained by solution of the full $^5E\otimes e$ vibronic Hamiltonian. In addition, effects arising from a contamination of the $^5E$ wave functions by coupling to higher-lying electronic terms have been completely neglected.

In this section, we examine the degree of quenching of the spin–orbit coupling operator within the $^5E$ ground term for various strengths of first-order Jahn–Teller coupling. We compare and discuss the results of calculations made using both the $^5E\otimes e$ and $^3T_1 + (^5E\otimes e)$ coupling models with those predicted in accordance with the theory of Ham. In Figure 5 are presented plots of the energies of the spinor levels of the $^5E$ ground term as a function of increasing $A_1$, calculated using the three approaches.

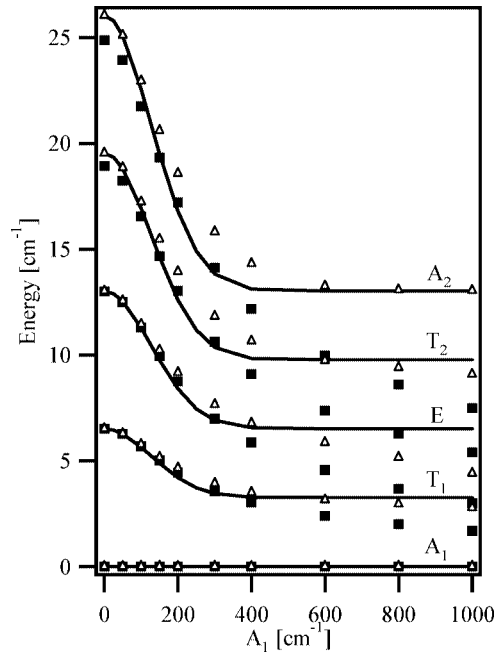


***Figure 5.*** Plots of the $^5E$ spinor energies as a function of $A_1$. Solid lines indicate the energies obtained from Ham theory. Open triangles and full squares denote the energies obtained by diagonalization of the $^5E\otimes e$ and $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonians respectively.

The full squares show the energies calculated by numerical solution of the $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonian. The effective $^3T_1 + {}^5E$ Hamiltonian was constructed from the AOM matrices, assuming octahedral geometry, with $e_\sigma = 4000$, $e_{\pi\perp} = e_{\pi\parallel} = 0$, $B = 647$, $C = 2640$, and $\zeta = 193.2$ cm$^{-1}$ in accordance with the procedure described in section 3 above. The vibronic Hamiltonian was solved for $\hbar\omega = 254$, $A_2 = 0$, $e_x = e_y = 0$ cm$^{-1}$, and $n_v = 30$. The ligand-field parameters and phonon energy used are those employed previously to model the EPR and neutron-scattering data of the deuteriated Cr(II) Tutton's salt.[1]

The open triangles show the energies calculated by numerical solution of the $^5E\otimes e$ vibronic Hamiltonian, assuming an effective-Hamiltonian parameter, $b = 6.51$ cm$^{-1}$. The latter is simply the energy of the first spinor level of the $^3T_1 + {}^5E$ Hamiltonian. Superimposed on this graph are plots of the energies of the spinor levels given by the expressions: $E = -2qb$, $-qb$, 0, $qb$, and $2qb$. Ham's reduction factor, $q$, was calculated using the relation,[19]

$$q = \frac{1}{2}[1 + e^{-4E_{JT}/\hbar\omega}] \tag{21}$$

The Mulliken symbols $A_1$, $A_2$, E, $T_1$, and $T_2$ denote the symmetry and degeneracy of the spinor levels.

For low values of $A_1$, up to ~100 cm$^{-1}$, the energies obtained from diagonalization of the full $^5E\otimes e$ vibronic

Hamiltonian agree very well with the predictions of Ham. In the range $A_1 = 200–400\ cm^{-1}$ the reduction factor $q$ is too small. As $A_1$ is increased, the relative energies of the spinor levels begin to deviate from the pattern, 0, $b'$, $2b'$, $3b'$, $4b'$, deviating most strongly at $A_1 = 1000\ cm^{-1}$ where the excited $^5A_1$ and $^5A_2$ levels are only $\sim 20\ cm^{-1}$ above the ground state. The excited $^5A_1$ and $^5A_2$ terms give rise to spinor levels of E and $T_2$, and E and $T_1$ symmetry, respectively. If the excited vibronic-singlet terms are close in energy to the ground vibronic doublet, mixing of spinor levels of the same symmetry occurs, resulting in deviations of the energies of the E, $T_1$, and $T_2$ levels from the expected energy scheme. Ham's expression nevertheless gives a surprisingly good reproduction of the energies over the entire range.

Similarly, for low values of $A_1$, the energies obtained from diagonalization of the $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonian are well approximated by Ham theory. A reasonable reproduction of the energies is achieved for $A_1 < 300\ cm^{-1}$; however, for large values of $A_1$, the energies calculated from Ham theory are considerably higher than those obtained using the vibronic Hamiltonian. The dominant contribution to the zfs of the $^5E$ term arises from off-diagonal matrix elements linking the $^3T_1$ and $^5E$ electronic terms. Though the form of the off-diagonal spin–orbit-coupling matrix between the $^5E$ vibronic ground term and the electronic $^3T_1$ term must from symmetry be the same as that of the matrix connecting the original $^5E$ and $^3T_1$ electronic terms, the factor reducing these off-diagonal matrices does not need to be identical to the factor $q$ reducing the operators within the $^5E$ term. Indeed, in the strong-coupling limit, the effective reduction factor is far smaller than the factor $q = {}^1/_2$, predicted by Ham theory.

*4.1.2. Reduction Factors and Ligand-Field Theory.* In the limit of weak first-order Jahn–Teller coupling, the low-level energy schemes, calculated above, can also be reproduced using a crystal-field or ligand-field model, with a reduced spin–orbit-coupling parameter. From Ham's reduction factor, $q = 0.5$ for strong first-order Jahn–Teller coupling, the effective spin–orbit-coupling parameter, $\zeta'$, required to model the low-level energy scheme with a ligand-field model is calculated to be $\zeta' = \sqrt{0.5}\zeta \simeq 0.707\zeta$. In contrast, the energies obtained using the $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonian, with $A_1 = 1000\ cm^{-1}$, correspond to $q = 0.3$ and an effective spin–orbit-coupling parameter, $\zeta' = \sqrt{0.3}\zeta \simeq 0.55\zeta$. This reduction of the spin–orbit-coupling parameter is much greater than the usual reductions applied to the electronic-repulsion and spin–orbit-coupling parameters, attributed to the effects of covalency.[26]

In the case of strong first-order Jahn–Teller coupling, the higher-lying singlet terms are close in energy to the ground orbital doublet and the spinor energy-level scheme differs significantly from the regular spacing expected from ligand-field theory. In addition, the observation of an optical transition of energy $\sim 4E_{JT}$, typical for first-order Jahn–Teller coupling, cannot be reconciled with a perturbed $^5E$ ground term within the ligand-field model.

**4.2. Strong First-Order and Weak Second-Order Jahn–Teller Coupling.** Acoustic paramagnetic resonance (APR), far-infrared (FIR), phonon-scattering, and thermal-conductivity experiments have been carried out on MgO:$Cr^{2+}$, $KMgF_3$:$Cr^{2+}$, and CaO:$Cr^{2+}$.[27–34] In these hosts, the chromous ion is embedded substitutionally for $Mg^{2+}$ or $Ca^{2+}$ at a site of approximately cubic symmetry and is very strongly coupled to the lattice vibrations. The energy splittings observed in these systems can be interpreted with the aid of Ham's three-state model, assuming strong first-order and weak second-order Jahn–Teller coupling with very small lattice strain.

The magnitude of the tunnelling splitting in the MgO:$Cr^{2+}$ system is of the order of $\sim 10–30\ cm^{-1}$.[35] The degeneracies of the $^5E$ ground term and the excited $^5A_1$ and $^5A_2$ terms, arising from Jahn–Teller coupling, are then lifted through the spin–orbit interaction. Although a considerable amount of data has been amassed for the MgO:$Cr^{2+}$ system, a number of different parameter sets have been proposed.[32,36,37] In the following, we do not attempt to model the data for this system. Instead these calculations should serve to illustrate, qualitatively, the energy scheme resulting from strong first-order and weak second-order Jahn–Teller coupling, as observed in such systems.

In Figure 6 are presented plots of the energies of the lowest fifteen spinor levels of a representative high-spin $d^4$ cation as a function of $\zeta$, calculated by diagonalization of (a) the $^5E\otimes e$ vibronic Hamiltonian and (b) the $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonian. The effective $^3T_1 + {}^5E$ Hamiltonian was created from the AOM matrices, assuming octahedral coordination. The ligand-field parameters were those used to construct the matrices in section 4.1. The effective-Hamiltonian parameter, $b$, is directly proportional to $\zeta^2$, and an expression for this parameter was derived from the spinor energies of the $^3T_1 + {}^5E$ Hamiltonian. In both calculations, the parameters $\hbar\omega = 254$, $A_1 = -900$, $A_2 = 1$, $e_x = e_y = 0\ cm^{-1}$, and $n_v = 30$ were used.

Superimposed on these plots are calculations based on Ham's three-state model. Once again, we have constructed the effective spin–orbit-coupling matrix by substituting the relevant functions for the $G_\theta$ and $G_\varepsilon$ operators. The matrix given in eq 16 above, with $|A_2\rangle$ lower lying was expanded into a new basis consisting of product functions of the three vibronic states and the five $S = 2$ spin states. $3\Gamma = 16.66\ cm^{-1}$ was obtained by numerical solution of the $^5E\otimes e$ vibronic Hamiltonian, without spin–orbit coupling. Finally, the energies were obtained by diagonalization of the resulting matrix, with the parameters $3\Gamma = 16.66\ cm^{-1}$, $q = -r = {}^1/_2$ and $b$ as determined above.

In Figure 6a, the energies calculated using Ham's three-state model are in very good agreement with those obtained from diagonalization of the $^5E\otimes e$ vibronic Hamiltonian. In Figure 6b, however, the spinor energies calculated using Ham's model give poor agreement with the solutions of the $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonian. This is a consequence of an overestimate in the factor $q$, as already mentioned. This result emphasizes the need for caution when using second-order perturbation theory. While the effective $^5E$ Hamiltonian may yield reasonable approximations to the energies of the spinor levels when the $^3T_1$ term is relatively high in energy, the Hamiltonian does not allow for any contamination of the ground-state wave functions, which will most certainly

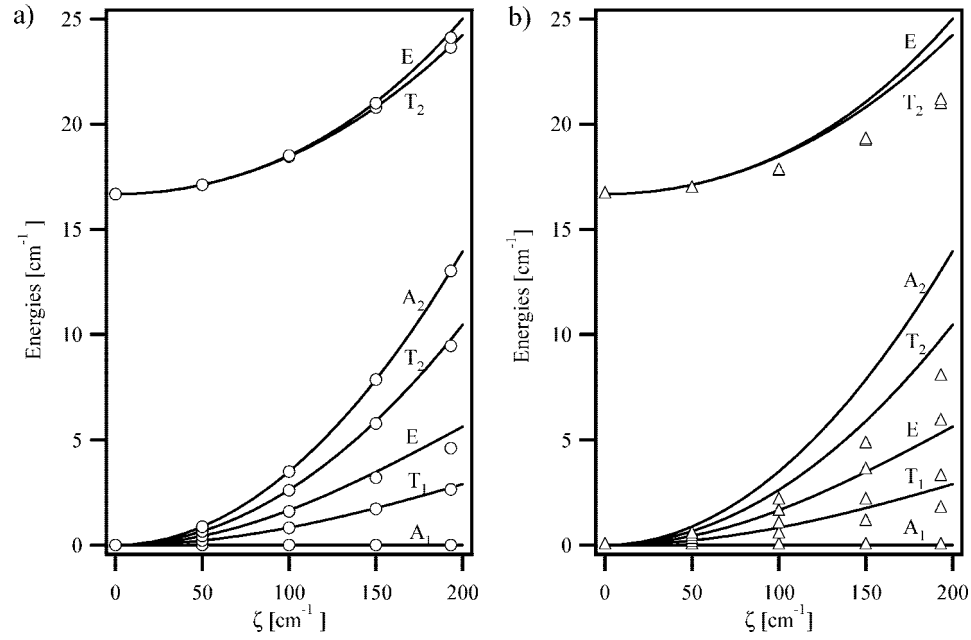


***Figure 6.*** Plots of the lowest fifteen spinor levels as a function of $\zeta$. Lines indicate the energies obtained using Ham theory. Open circles and open triangles denote the energies calculated by diagonalization of the $^5E\otimes e$ vibronic and $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonians, respectively. The parameters used are given in the text.

have an influence on the value of $q$. In this respect, the effective $^3T_1 + {}^5E$ Hamiltonian must be considered to be more exact.

**4.3. Strong First-Order and Appreciable Second-Order Jahn–Teller Coupling.** Of the high-spin $d^4$ centers for which experimental data are known, the instance of strong first-order coupling and very weak second-order coupling, considered in the previous section, is found only for the chromous ion doped into MgO and CaO. The remainder are primarily Mn(III) coordination complexes for which the large ratio of the strain to the tunnelling splitting renders a static Jahn–Teller distortion at liquid-helium temperatures. The ground-state multiplet may then be adequately described by an $S = 2$ spin Hamiltonian.

Correlating the parameter values of this Hamiltonian to the molecular structure traditionally proceeds by ligand-field theory, from which information regarding the bonding within the complex is obtained. Before proceeding further, it is vital to outline the approach and emphasize its limitations. The ligand-field Hamiltonian is often constructed using the AOM. Some dependence of the parameters upon the distance, $r$, then has to be assumed. In the spirit of crystal-field theory, $e_\sigma$ and $e_\pi$ would be assigned a $1/r^5$ and $1/r^6$ dependence, respectively.[2] The values of the bonding parameters so derived then have chemical significance only if the complex is strongly localized in the temperature range where the experimental data are acquired, for dynamic effects cannot be modeled by a Hamiltonian in which the vibrational coordinates of the complex are neglected. For example, room-temperature crystallographic data are often interpreted in conjunction with low-temperature spectroscopic data. This strategy is potentially dangerous since the room-temperature crystal structure often does not reflect the structure of the molecule at base temperature.

If the crystallographic and spectroscopic data are collected at the same temperature but dynamic effects are still prevalent, then the results of the analysis will still be misleading, as the values of the spin-Hamiltonian parameters derived will depend on the time scale of the technique relative to the rate at which the molecule undergoes intramolecular reorientations between the potential-energy minima. By contrast, in the theoretical method which we have outlined, the angular-overlap matrices are constructed in the undistorted geometry and are always valid regardless of the coupling strength. The calculated molecular and electronic coordinates appear naturally from the vibronic-coupling calculations.

As an illustration, we consider here the $[Mn(OD_2)_6]^{3+}$ cation in the deuteriated cesium manganese alum, $CsMn(SO_4)_2\cdot 12D_2O$. Single-crystal neutron-diffraction studies reveal that the site symmetry of the $[Mn(OD_2)_6]^{3+}$ cation in the alum at 170 K is $S_6$. The water molecules are rotated about the Mn−O bond vectors by the angle $\varphi = -19$ degrees with respect to the $MnO_6$ framework.[35] At ∼156 K, however, the alum undergoes a cubic to orthorhombic phase transition, and at 5 K, the aqua ion is tetragonally elongated, with Mn−O bond distances of 2.129(2), 1.929(1), and 1.924(1) Å and all O−Mn−O bond angles within 1.4 of 90°.[38]

Despite the $MnO_6$ framework being close to tetragonal, an analysis of the low-temperature HFEPR data indicates a large rhombic anisotropy, with spin-Hamiltonian parameters $D = -4.491(7)$ cm$^{-1}$ and $E = 0.248(5)$ cm$^{-1}$.[2] Similar parameters were obtained from 1.5 K INS spectra.[3] The AOM analysis of the low-temperature data has been reported in detail and may be summarized as follows. The optical spectra exhibit intra-$^5D$ transitions, which energies may be written down in terms of AOM bonding parameters. As-

suming parameters corresponding to the $[V(OH_2)_6]^{3+}$ cation, and allowing the $e_\pi$ parameters to vary as a function of $1/r^6$, it was concluded that the rhombic anisotropy results primarily from the anisotropic nature of the Mn(III)−water $\pi$-bonding.

Recently, we have collected data at elevated temperatures showing that the spin-Hamiltonian parameters are temperature dependent. In the spirit of the AOM, this result can be modeled only by allowing the parameters to be temperature dependent, implying that the nature of the bonding interaction is itself temperature dependent. This is not necessarily so. We now show how the low-temperature bond lengths and spin-Hamiltonian parameters can be reproduced by our model and how these experimental observables can become temperature dependent without implying a change in parameters describing the chemical bonding.

*4.3.1. Calculation of the Low-Level Energy Structure.* Using the $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonian, the low-level energy structure of the Mn(III) aqua ion in the cesium manganese alum can be reproduced. AOM matrices were first constructed using the bond angles of the high-symmetry ($S_6$) structure and the parameters, $e_\sigma = 6950$, $e_{\pi\perp} = 930$, $e_{\pi\parallel} = 0$, $B = 848$, and $C = 3341$ cm$^{-1}$, from which an effective $^3T_1 + {}^5E$ Hamiltonian was derived. The Racah parameters are ca. 85% of the free-ion values, and the $e_{\pi\perp}$ and $e_\sigma$ parameters are those used to model spectroscopic data for the $[V(OH_2)_6]^{3+}$ cation in the cesium vanadium alum.[39]

A band observed at ∼10 500 cm$^{-1}$ in the absorption spectra of the cesium manganese alum corresponds to the intra-$^5$E transition of energy $\sim 4E_{JT}$.[2] An estimate of the energy of the $\upsilon_2(MnO_6)$ stretching vibration, $\hbar\omega \sim 428$ cm$^{-1}$, was made based on Raman data available for related systems[40,41] and, using the relation $E_{JT} = A_1^2/2\hbar\omega$, the first-order Jahn–Teller-coupling parameter $A_1 = -1462$ cm$^{-1}$ was obtained. Values of the second-order-coupling and strain parameters were chosen such as to give rise to a sufficiently isolated $S = 2$ ground manifold. Diagonalization of the $^3T_1 + (^5E\otimes e)$ vibronic Hamiltonian using the parameters, $A_1 = -1462$, $A_2 = 30$, $e_x = -100$, $e_y = 0$, $\hbar\omega = 428$ cm$^{-1}$, $\zeta = 317$ cm$^{-1}$ (ca. 90% of the free-ion value) and $n_\upsilon = 35$, yielded a low-level energy scheme characterized by $D = -4.6421$ and $E = 0.2277$ cm$^{-1}$, in reasonable agreement with the experiment.

*4.3.2. Variation in Bond Lengths with Temperature.* In addition to calculating the low-level energy structure, it is also possible to calculate the Mn−O bond lengths of the Mn(III) aqua ion. Using the eigenvectors obtained in the calculation above, the expectation values of the $Q_x$ and $Q_y$ operators within the ground multiplet can be determined. From eqs 8 and 9, $\mu = 20$ (the reduced mass of the $D_2O$ ligand) and $\hbar\omega = 428$ cm$^{-1}$, the distortions $r_i$ of the Mn−$O_i$ bond distances were calculated. Using the average (5 K) Mn−O bond length of 1.994 Å, the resulting Mn−O bond distances are 2.14, 1.93, and 1.93 Å, in very good agreement with the 5 K experimental data.[35]

The solution of the vibronic Hamiltonian described in section 4.3.1 yields two excited spin multiplets at ∼150 cm$^{-1}$ $^3/_2e_x$ above the ground spin multiplet. At room temperature (293 K), the Boltzmann factors for the ground and excited spin multiplets are ∼0.54, ∼0.23, and ∼0.23, respectively. The Mn−O bond distances for the two excited multiplets, calculated from the expectation values of $Q_x$ and $Q_y$ in accordance with eqs 8 and 9, are 1.92, 2.03, and 2.03 Å, corresponding to tetragonally-compressed geometries. An average room temperature structure can be obtained by weighting the geometries of the ground and excited spin multiplets in accordance with their Boltzmann factors. The Mn−O bond distances at 293 K are thus calculated to be 2.03, 1.97, and 1.97 Å. A natural dependence of the bond distances on the temperature is seen to arise without the need to assume a temperature dependence of the parameters of the Hamiltonian.

## 5. Summary and Conclusion

The theoretical description of high-spin $d^4$ complexes has traditionally proceeded by way of a conventional ligand-field analysis, correlating spectroscopic data with crystallographic measurements. The approach necessitates assumptions as to the radial dependence of the ligand-field parameters, takes no account of dynamical Jahn–Teller effects, and is of dubious significance when the low-temperature structure of the complex is inferred from room-temperature X-ray diffraction data. In this work, we have presented a new theoretical approach which circumvents these difficulties by calculating the experimental quantities directly from the eigenfunctions and eigenvalues of a vibronic Hamiltonian, yet retains the chemically intuitive AOM in the construction of the electronic basis set.

Whereas the $^5E\otimes e$ vibronic-coupling model is applicable only to certain octahedral Cr(II) complexes, the $^3T_1 + (^5E\otimes e)$ model presented here can in principle be applied to any high-spin $d^4$ complex, where the ligand-field parameters in the undistorted trigonal or octahedral configuration can be reliably estimated. The utility of the method is demonstrated by the application of the Hamiltonian to $MgO:Cr^{2+}$ and the $[Mn(OD_2)_6]^{3+}$ cation in the deuteriated cesium manganese alum. A temperature dependence of the bond lengths arises quite naturally from the model.

The explicit inclusion of the spin–orbit interaction between the $^5E$ and excited $^3T_1$ terms has a marked effect on the degree of quenching of spin–orbit coupling within the ground term. For Cr(II) complexes with strong first-order coupling, it is shown that the effective reduction factor is considerably lower than that expected from Ham's theory. This result underlines the need for caution in the interpretation of the ground energy-level scheme observed for systems such as $MgO:Cr^{2+}$.

**Acknowledgment.** This work was funded by the Swiss National Science Foundation.

### References

(1) Dobe, C.; Noble, C.; Carver, G.; Tregenna-Piggott, P. L. W.; McIntyre, G.; Barra, A.-L.; Neels, A.; Janssen, S.; Juranyi, F *J. Am. Chem. Soc.* **2005**, *127* (10), 3642.

(2) Tregenna-Piggott, P. L. W.; Weihe, H.; Barra, A.-L. *Inorg. Chem.* **2003**, *42*, 8504.

(3) Basler, R.; Tregenna-Piggott, P. L. W.; Andres, H.; Dobe, C.; Güdel, H.-U.; Janssen, S.; McIntyre, G. J. *J. Am. Chem. Soc.* **2001**, *123* (14), 3377.

(4) Krzystek, J.; Yeagle, G. J.; Park, J.-H.; Britt, R. D.; Meisel, M. W.; Brunel, L.-C.; Telser, J. *Inorg. Chem.* **2003**, *42*, 4610.

(5) In some instances, fourth order terms may also be required to model the low-level energy scheme. However, these terms are generally very small compared to the terms in *D* and *E*.

(6) Griffiths, J. S. In *The Theory of Transition-Metal Ions*; Cambridge University Press: New York, 1961.

(7) Barra, A.-L.; Gatteschi, D.; Sessoli, R.; Abbati, G. L.; Cornia, A.; Fabretti, A. C.; Uytterhoeven, M. G. *Angew. Chem. Intl. Ed. Engl.* **1997**, *36*, 2329.

(8) Limburg, J.; Vrettos, J. S.; Crabtree, R. H.; Brudvig, G. W.; de Paula, J. C.; Hassan, A.; Barra, A.-L.; Duboc-Toia, C.; Collomb, M.-N *Inorg. Chem.* **2001**, *40*, 1968.

(9) Gerritsen, H. J.; Sabisky, E. S. *Phys. Rev.* **1963**, *132*, 1507.

(10) Telser, J.; Pardi, L. A.; Krzystek, J.; Brunel, L.-C. *Inorg. Chem.* **1998**, *37*, 5769.

(11) Gregson, A. K.; Doddrell, D. M.; Healy, P. C. *Inorg. Chem.* **1978**, *17*, 1216.

(12) Mossin, S.; Weihe, H.; Barra, A.-L. *J. Am. Chem. Soc.* **2002**, *124*, 8764.

(13) Jahn, H. A.; Teller, E. *Proc. Roy. Soc.* **1937**, *A161*, 220.

(14) Englman, R. In *The Jahn-Teller Effect in Molecules and Crystals*; Wiley-Interscience: New York, 1972.

(15) Van Vleck, J. H. *J. Chem. Phys.* **1939**, *7*, 72.

(16) Ham, F. S. *Phys. Rev.* **1968**, *166*, 307.

(17) Dobe, C.; Andres, H.-P.; Tregenna-Piggott, P. L. W.; Mossin, S.; Weihe, H.; Janssen, S. *Chem. Phys. Lett.* **2002**, *362*, 387.

(18) (a) These expressions are similar to those given by Isaac Bersuker in his review of the Jahn−Teller effect; see *Coord. Chem. Rev.* **1975** *14*, 357–412. (b) Note that our formulation of the JT matrices means that the sign in the upper part of the expression for $\rho$ is different, and the relationship between the relative signs of $A_1$ and $A_2$ and the positions of the minima are reversed.

(19) Ham, F. S. *J. Phys. Colloq.* **1971**, *32*, 952.

(20) Simpson, J. A. L.; Bates, C. A.; Barrau, J.; Brousseau, M.; Thomas, V. *Semicond. Sci. Technol.* **1988**, *3*, 178.

(21) Ham, F. S. In *Electron Paramagnetic Resonance*; Geschwind, S., Ed.; Plenum Press: New York, 1972; p 45.

(22) Brugel, D.; Bates, C. A. *Semicond. Sci. Technol.* **1987**, *2*, 501.

(23) The matrix elements given in ref 22 relate strictly to $Cr^{2+}$ in tetrahedral coordination. The matrices have, however, the same form for $Cr^{2+}$ in octahedral coordination.

(24) Bendix, J. *Compr. Coord. Chem. II* **2004**, *2*, 673.

(25) Sugano, S.; Tanabe, Y.; Kamimura, H. In *Multiplets of Transition-Metal Ions In Crystals*; Academic Press: New York, 1970.

(26) Figgis, B. N.; Hitchman, M. In *Ligand-Field Theory and Its applications*; Wiley-VCH: New York, 2000.

(27) Marshall, F. G.; Rampton, V. W. *J. Phys. C: Solid State Phys.* **1968**, *1*, 594–598.

(28) Rampton, V. W.; Bates, C. A.; Fletcher, J. R.; Jones, S. C.; Jaussaud, P. C. *Proceedings of the 1st International Conference on Phonon Scattering in Solids*, Paris, 1972; pp 243−246.

(29) Lange, J. N. *Phys. Rev. B* **1973**, *8*, 5999.

(30) Anderson, B. R.; Challis, L. J.; Stoelinga, J. H. M.; Wyder, P. *J. Phys. C: Solid State Phys.* **1974**, *7*, 2234.

(31) King, P. J.; Oates, S. G.; Rampton, V. W.; Shellard, I. J. *Proceedings of the 2nd International Conference on Phonon Scattering in Solids*, Nottingham, 1976; pp 181−183.

(32) Lange, J. N.; Guha, S. *Proceedings of the 2nd International Conference on Phonon Scattering in Solids*, Nottingham, 1976; pp 175−177.

(33) Rampton, V. W.; Shellard, I. J. *Proceedings of the 2nd International Conference on Phonon Scattering in Solids*, Nottingham, 1976; pp 178–180.

(34) Rivallin, J.; Salce, B. *Proceedings of the 2nd International Conference on Phonon Scattering in Solids*, Nottingham, 1976; pp 184–186.

(35) Patel, J. L.; Wigmore, J. K. *J. Phys. C: Solid State Phys.* **1977**, *10*, 1829.

(36) Fletcher, J. R.; Stevens, K. W. H. *J. Phys. C: Solid State Phys.* **1969**, *2*, 444–456.

(37) Ham, F. S. *Phys. Rev.* **1971**, *B4*, 3854.

(38) Tregenna-Piggott, P. L. W.; Andres, H. P.; McIntyre, G. J.; Best, S. P.; Wilson, C. C.; Cowan, J. A. *Inorg. Chem.* **2003**, *42*, 1350.

(39) Tregenna-Piggott, P. L. W.; Spichiger, D.; Carver, G.; Frey, B.; Meier, R.; Weihe, H.; Cowan, J. A.; McIntyre, G. J.; Zahn, G.; Barra, A.-L. *Inorg. Chem.* **2004**, *43* (25), 8049.

(40) Best, S. P.; Beattie, J. K.; Armstrong, R. S. *J. Chem. Soc., Dalton Trans.* **1984**, 2611.

(41) Tregenna-Piggott, P. L. W.; Best, S. P. *Inorg. Chem.* **1996**, *35*, 5730.

CT7003484

JCTC Journal of Chemical Theory and Computation

# Calculating Reaction Rates with Partial Hessians: Validation of the Mobile Block Hessian Approach

A. Ghysels, V. Van Speybroeck, T. Verstraelen, D. Van Neck, and M. Waroquier*

*Center for Molecular Modeling, Ghent University, Proeftuinstraat 86, B-9000 Gent, Belgium*

Received October 24, 2007

**Abstract:** In an earlier paper, the authors have developed a new method, the mobile block Hessian (MBH), to accurately calculate vibrational modes for partially optimized molecular structures [*J. Chem. Phys.* **2007,** *126* (22), 224102]. The proposed procedure remedies the artifact of imaginary frequencies, occurring in standard frequency calculations, when parts of the molecular system are optimized at different levels of theory. Frequencies are an essential ingredient in predicting reaction rate coefficients due to their input in the vibrational partition functions. The question arises whether the MBH method is able to describe the chemical reaction kinetics in an accurate way in large molecular systems where a full quantum chemical treatment at a reasonably high level of theory is unfeasible due to computational constraints. In this work, such a validation is tested in depth. The MBH method opens a lot of perspectives in predicting accurate kinetic parameters in chemical reactions where the standard full Hessian procedure fails.

## 1. Introduction

Ab initio prediction of reaction rate constants of chemical reactions has a high computational cost, especially when large (bio)molecular systems are involved. An accurate description of chemical kinetics of reactions in gas phase is nowadays perfectly practicable for moderate-sized molecules, but once the molecular environment comes into play, one has to adapt the level of theory in such way to make the computation feasible.[1] This puts a heavy burden on the accuracy of the numerical results. Chemical kinetics in static approaches is still widely based on transition state theory (TST).[2–5] Key parameters are the reaction energy barrier between the reactants and activated complex (the transition state) and the vibrational frequencies, which serve as an input in the partition functions, and their accurate computation is essential. In the molecular-statistical formulation of TST, they completely determine the equilibrium constant by the use of partition functions.

In the harmonic oscillator approximation, the molecular partition function is factorized in a translational, rotational, and vibrational contribution, where the latter is completely determined by the eigenfrequencies. Frequencies are usually computed by a normal-mode analysis (NMA). This is the main bottleneck in ab initio predictions of chemical kinetics in large molecular systems, since frequency calculations are computationally very demanding even if analytical second derivatives are employed, rather than numerical ones. If a molecular mechanics (MM) force field is used instead of a quantum mechanics (QM) or hybrid (QM/MM)[6–9] description, the frequency calculation becomes less problematic, though even at the full MM level other issues, such as the storage and manipulation of the huge Hessian matrices associated with very large systems, can become prohibitive in real applications. Anyway, chemical reactions inherently involve bond breaking and charge transfer; so, it is essential to provide a QM description for (at least) the reactive region and a full MM description is usually no option.

In addition, there are computational limitations in the geometry optimization of extended systems at a high level of theory (LOT). Very often one goes over to a partial optimization: the interesting region containing the active site is optimized at a high LOT, while the environment is kept fixed at a low LOT geometry. This approach permits one to obtain an ab initio description of the chemically active site

* Corresponding author. E-mail: michel.waroquier@UGent.be.

10.1021/ct7002836 CCC: $40.75 
Published on Web 02/27/2008

in large molecular systems, but at the same time, it creates several new problems. One of them is the extraction of accurate frequencies for the relevant vibrational modes. All partially optimized systems are nonequilibrium structures, and as a consequence of the residual gradients on the potential energy surface (PES), the standard full Hessian normal-mode analysis may show some unphysical results, e.g., spurious imaginary frequencies may appear. A frequency analysis in terms of a subset of coordinates that are optimized, i.e. a partial Hessian method, can avoid these problems.

The authors have succeeded recently in deriving a method that is able to calculate physical frequencies. The main idea is to group the atoms that were kept fixed during the partial optimization into one or more blocks that are able to move as rigid bodies with respect to the relaxed molecular part in the vibrational analysis.[10] This mobile block Hessian (MBH) method has shown to be very efficient for an accurate evaluation of relevant frequencies of vibrational modes. The proposed procedure remedies the artifact of imaginary frequencies occurring in standard frequency calculations for partially optimized systems. In addition, only a subblock of the Hessian matrix has to be constructed and diagonalized, leading to a serious reduction of the computation time for the frequency analysis.

MBH can be regarded as an extension of the partial Hessian vibrational analysis approach (PHVA). Only part of the cartesian Hessian has been retained, excluding all the atoms of the passive site of the molecule that is kept fixed during the optimization. This methodology was first introduced and developed by Head and co-workers[11–14] and was further investigated by Li and Jensen[15] and Besley and Metcalf.[16] It comes to giving an infinite mass to the fixed atoms so that they are frozen at their initial position. Only the relaxed atoms can participate in the small amplitude vibrations.

The novelty of MBH with respect to PHVA lies in the fact that, in the former, the finite mass of each block is taken into consideration in the NMA, instead of giving an infinite mass to the fixed atoms. Six degrees of freedom are attributed to each block to describe its position and orientation with respect to the fully optimized part, and the global translational/rotational invariance of the potential energy surface (PES) is fully respected. Moreover, the PHVA is always limited to the case of one immobile block with infinite mass, whereas in the MBH model, parts of the molecular system can be ranged in multiple blocks which can move as rigid bodies with respect to the relaxed part of the molecule. In ref 10, both PHVA and MBH methods are submitted to a tough comparative study, while in ref 17, attention is given to the practical implementation of the MBH model and the interface with molecular modeling program packages.

One of the main applications that can be deduced from the knowledge of accurate normal-mode frequencies, is the prediction of chemical kinetics, as already mentioned. By means of the partition functions and a molecular-statistical formulation of transition state theory, the reaction rate constant $k$ of a chemical reaction can be determined.[2–5] A somewhat different approach is proposed by the group of Lin et al.[18] Basic assumption is that the Hessian elements that involve only the atoms of the active site might be more critical than the other Hessian elements. The less critical elements are approximated following some interpolation procedure, mainly for elements at the nonstationary points on the potential energy surface which are not consistently constructed by the same level (dual level scheme). Other related papers suggest proper methods to predict accurate QM/MM kinetics by incorporating quantum mechanical effects by treating vibrational motions quantum mechanically and applying multidimensional tunneling approximations into reaction rate calculations.[19,1] Recently, more sophisticated techniques concerning transition state theory have been developed including tunneling effects, quantum dynamical effects and multiple pathways (we refer to ref 20 for a review of all modern developments), but in view of the goal of this paper to validate the MBH approach in predicting kinetics, conventional TST largely suffices and tunneling and other effects will not be incorporated.

In principle, the expression of $k$ includes all normal vibrational modes in reactants and activated complex. It is inherent to both MBH and PHVA approaches that the number of frequencies is always smaller than in a standard frequency calculation. The question arises whether this reduction has a significant influence on the reaction rate constant. Here lies the scope of this work: we will demonstrate that the normal modes which disappear when defining fixed blocks have little influence on the chemical kinetics. This work aims at promoting MBH as a suitable and highly efficient tool for predicting accurate chemical kinetics parameters in large extended molecular systems where the standard full Hessian procedure fails.

Applications of MBH are numerous. They can be classified in various categories:

(i) Large biosystems consisting of thousands of atoms require a hybrid quantum mechanical/molecular mechanical (QM/MM) approach.[6–9] The whole MM region can be taken up in one or multiple blocks.

(ii) A cluster description of zeolites or other periodic systems, such as lattices, requires fixed positions of border atoms to prevent collapse of the molecule during optimization.[21–23] This represents a particular situation of partial optimization.

(iii) Reactions in solvents often require an approach with a chemical reactive site and various layers treated at different levels of theory (QM/MM or QM/QM′). The whole can even be circumvented by a bulk solvent described by a polarizable continuum model (PCM).[24,25] In MBH, the various solvent molecules are all regarded as mobile blocks which can translate and rotate freely around the active site. Only the internal structure of each solvent molecule is held fixed.

The structure of the paper is as follows. In section 2, a short outline of the theoretical methodology is given, and in section 3, the computational details are summarized. Section 4 is devoted to the validation of MBH as an adequate method to predict rate constants. Different reactions with various block choices are taken up in the test set for validation with the benchmark values (full optimization before frequency calculation). The test set includes a prototype substitution reaction, a hydrogen transfer reaction, as well as several

radical addition reactions, since these have a localized reactive site. The effectiveness of the multiple MBH has been illustrated with a more extended aminophosphonate system in section 4.3, for which solvent molecules are taken into account. The results of the MBH have been compared to those of the PHVA approach as well, and based on theoretical considerations, a modified PHVA method is presented in section 4.4, hereafter referred to as PHVA*. Finally in section 5, some conclusions are drawn.

## 2. Theoretical Background

**2.1. Partition functions.** Within the harmonic oscillator approximation, the $3N$ degrees of freedom of a $N$-atom system can be decoupled into three groups of independent motions—3 translational, 3 rotational, and $3N - 6$ vibrational motions—that all contribute to the total partition function $Q$:

$$Q = qq_{\mathrm{elec}} \tag{1}$$

where

$$q = q_{\mathrm{trans}}q_{\mathrm{rot}}q_{\mathrm{vib}} \tag{2}$$

The translational partition function reads

$$q_{\mathrm{trans}} = \left(\frac{2\pi Mk_{\mathrm{B}}T}{h^2}\right)^{3/2} V \tag{3}$$

$M$ stands for the total mass of the system, $T$ is the temperature, $k_{\mathrm{B}}$ is the Boltzmann constant, $h$ is the Planck constant, and $V$ is the volume. If $I_1$, $I_2$, and $I_3$ denote the moments of inertia of the system and $\sigma$ is the symmetry number, the rotational partition function reads

$$q_{\mathrm{rot}} = \frac{8\pi^2}{\sigma}\left(\frac{2\pi k_{\mathrm{B}}T}{h^2}\right)^{3/2}\sqrt{I_1I_2I_3} \tag{4}$$

Each vibration with frequency $\nu_i$, gives a contribution

$$q_{(\nu_i)} = \frac{\mathrm{e}^{-h\nu_i/2k_{\mathrm{B}}T}}{1 - \mathrm{e}^{-h\nu_i/k_{\mathrm{B}}T}} \tag{5}$$

to the total vibrational partition function

$$q_{\mathrm{vib}} = \prod_i q_{(\nu_i)} \tag{6}$$

The electrons also contribute to the partition function, but when the first electronic excitation energy is much greater than $k_{\mathrm{B}}T$, the first and higher excited states are assumed to be inaccessible. If $E_0$ is the energy of the ground-state level, this assumption simplifies the electronic partition function to

$$q_{\mathrm{elec}} = \mathrm{e}^{-E_0/k_{\mathrm{B}}T} \tag{7}$$

Note that the zero point energy contribution $\mathrm{e}^{-h\nu_i/2k_{\mathrm{B}}T}$ in the numerator of eq 5 is frequently left out from the vibrational partition function and incorporated in the electronic partition function.

Ab initio molecular calculations can be used to generate the molecular properties required for the evaluation of the above partition functions, such as the geometry (for the moments of inertia $I_i$), the Hessian matrix (for the vibrational frequencies $\nu_i$), and the electronic ground-state energy $E_0$.

**2.2. Reaction Rate Constant within Conventional Transition State Theory (TST).** Transition state theory has been proved to be very useful to determine the reaction rate constants.[2–5] It supposes that the transition state or activated complex is in equilibrium with the reactants, although, strictly speaking, this hypothesis is not valid since the transition state corresponds to a saddle point rather than a minimum on the PES. Within this assumption the rate constant is completely determined by the microscopic partition functions and the reaction barrier at 0 K.

For a unimolecular reaction, $A \rightarrow A^{\ddagger} \rightarrow B$ or $A \rightarrow A^{\ddagger} \rightarrow B + C$ (with the $^{\ddagger}$ superscript indicating the activated complex) the rate constant $k$ is given by:

$$k(T) = \frac{k_{\mathrm{B}}T}{h}\frac{q(\ddagger)/V}{q(A)/V}\mathrm{e}^{-\Delta E_0/k_{\mathrm{B}}T} \tag{8}$$

$\Delta E_0$ represents the molecular energy difference at 0 K between the activated complex and the reactants. The transition state frequency is assumed not to be included in the partition function $q(\ddagger)$ of the activated complex. $k$ is expressed in units $\mathrm{s}^{-1}$.

For a bimolecular reaction $A + B \rightarrow (AB)^{\ddagger} \rightarrow C$ or $A + B \rightarrow (AB)^{\ddagger} \rightarrow C + D$, the expression for the rate constant becomes

$$k(T) = \frac{k_{\mathrm{B}}T}{h}\frac{q(\ddagger)/V}{q(A)/V\, q(B)/V}\mathrm{e}^{-\Delta E_0/k_{\mathrm{B}}T} \tag{9}$$

expressed in units of cubic meters per mole second.

**2.3. MBH and PHVA.** Partially optimized geometries are nonequilibrium structures. The usual normal-mode analysis (NMA) equations $H\nu = \omega^2 M\nu$, with $H$ being the full cartesian Hessian and $M$ the Cartesian diagonal mass matrix, could be solved to obtain the frequencies, but this procedure shows some serious defects. The Hessian $H$ is the second derivative matrix of the potential energy with respect to all the Cartesian coordinates. At nonequilibrium geometries, it has only three zero-eigenvalues instead of six, implying that the rotational invariance of the potential energy surface is not manifest anymore.[26] Spurious imaginary frequencies appear. Moreover, the eigenvalues of the Hessian depend on the choice of coordinates.[27,28]

In the partial Hessian vibrational analysis (PHVA),[13,15] these defects are surmounted by giving the fixed atoms an infinite mass. The normal mode equations are then restricted to the relaxed atoms only, by taking a submatrix of the Hessian and the mass matrix:

$$H_{\mathrm{E}}\nu = \omega^2 M_{\mathrm{E}}\nu \tag{10}$$

The mobile block Hessian (MBH) model has been proposed recently by the authors[10] as an improvement of the PHVA. In the MBH model the fixed part is considered as a rigid body that is allowed to participate in the small amplitude vibrations, thus taking into account the finite mass of the fixed block. The spurious frequencies and the coordinate dependence are avoided since the system composed of optimized atoms plus block is in equilibrium. Relying on the global translational and rotational invariance, it is possible[10] to write the single block MBH normal mode equations in terms of the same submatrix $H_{\mathrm{E}}$ of the Hessian,

while the corresponding mass matrix is adapted because of the finite block mass:

$$H_E \nu' = \omega'^2 \tilde{M}' \nu' \quad (11)$$

with

$$\tilde{M}' = M_E - M_E D_E S^{-1} D_E^T M_E \quad (12)$$

The matrix $D_E$ is constructed in terms of the coordinates of the free atoms with respect to a space fixed frame. The matrix $S$ contains the information on the mass distribution, i.e. the total mass and the moments of inertia of the molecule. Details can be found in the Appendix and more extensively in ref 17.

The usefulness and applicability of the MBH approach are seriously increasing in case of extension to several mobile blocks. The multiple MBH takes into account the finite mass of each block, by including six parameters per block describing its position and orientation into the NMA equations and by mass weighting with the appropriate block mass and moments of inertia.

The multiple MBH method is for instance extremely useful when simulating chemical reactions in a solvent. Solvent molecules can easily be associated to rigid blocks with a fixed internal structure. They can move freely with respect to each other and with respect to the active site of the molecule.

At first sight the MBH is similar to the united atom concept in force fields, since groups of atoms are treated there also as a single entity.[29] However, in spite of this resemblance, the MBH is essentially different. In the MBH blocks each atom keeps its identity and continues to contribute individually to, e.g., moments of inertia, Hessian elements, steric hindrance, etc. Coarse-grained or united atom methods reduce the number of atoms and the initial all-atom potential energy surface is approximated by a parametrizized PES of lower dimension. The MBH on the other hand does not simplify the potential energy surface but freezes certain degrees of freedom when performing the vibrational analysis.

## 3. Computational Details

In order to validate both MBH and PHVA methods in their performance in reproducing accurate chemical kinetics, we compare the MBH and PHVA predictions for the reaction rate constant with benchmark values $k$.

Benchmark structures and frequencies are generated with a full geometry optimization at a high level of theory (DFT/B3LYP/6-311 g**) with tight convergence criteria such that the residual gradients on the PES are negligibly small. Consequently, a frequency calculation is carried out at the same level of theory for the whole molecular system. These equilibrium geometries permit to calculate the reaction rate with the full cartesian Hessian frequencies.

In a first analysis, frequencies and rate constants are calculated for the fully optimized geometry, while the block size is varied in the vibrational analysis. For each reaction under study, we take into consideration various choices of fixed blocks, or, various submatrices $H_E$ of the Hessian. The normal mode equations, eqs 10 and 11, are constructed and solved using the same geometry, and thereby, any perturbation resulting from geometry differences is excluded in this particular treatment. This comparative study is thus highly appropriate to investigate the influence on the rate constants of exclusion of parts of the Hessian in the frequency calculation, i.e. limiting the NMA to a partial Hessian.

In a second analysis, partial geometry optimization is performed and consequently followed by a frequency calculation. For the MBH model, the position/orientation (six degrees of freedom) of each block are optimized, in contrary to PHVA, where the atoms in the single block are kept fixed in space. Therefore, a partial optimization with multiple blocks produces a better structure than one with a single block. We remind that PHVA is always limited to a single block, whereas MBH is very suitable to treat multiple blocks.

The partial optimization is performed as follows. First, one optimizes the system at a low level of theory (HF/STO-3g) to find a plausible starting structure. Then the rigid blocks are introduced and the system is partially optimized at a high level of theory (DFT/B3LYP/6-311 g**), while keeping the rigid blocks fixed at their initial internal geometry. All calculations were carried out with the Gaussian03 software package.[30] Next, a frequency calculation is performed with the second derivatives of the potential energy using the same high level of theory. The standard full Hessian frequency analysis would give unphysical results due to the residual forces present in the partial optimized structures, as mentioned in the Introduction. Instead, the PHVA or MBH normal mode eqs 10 and 11 are constructed, as these yield physical frequencies. Obviously, the same rigid blocks are chosen as those considered in the precedent partial optimization.

A partial Hessian method such as the MBH or PHVA approach, however, reduces the number of calculated frequencies. The difference in the number of degrees of freedom between reactants and transition state determines the temperature dependence of the reaction rate, as can be easily seen by inspecting eqs 8 and 9. This difference should not change when introducing the MBH blocks. It is therefore obvious that the chosen blocks must consist of the same atoms in reactant(s) and transition state. Note that strictly speaking the internal rigid block geometry might differ between reactants and transition state, because of the first step, i.e. optimization at the low level of theory before the actual partial optimization.

Finally, we made a selection of various chemical reactions for the validation. Most of them are radical addition reactions, but also one prototype substitution reaction ($S_N2$) and the hydrogen abstraction of one of the ending carbons are included (R6 and R7, respectively). In these reactions the reactive site (the radical center) is well localized. We choose addition reactions of ethene to a large variety of radicals with different substituents. It enables us to select various types of blocks (large and heavy blocks, substituents with ring structure(s), etc) and to give some recommendations in choosing the fixed blocks and the relaxed molecular region.

An overview of the different reactions under study is depicted in Figure 1. The reactions are labeled as R1, R2, etc. and the reactants and products are numbered. The block choices in the reactants are indicated and labeled in Figures

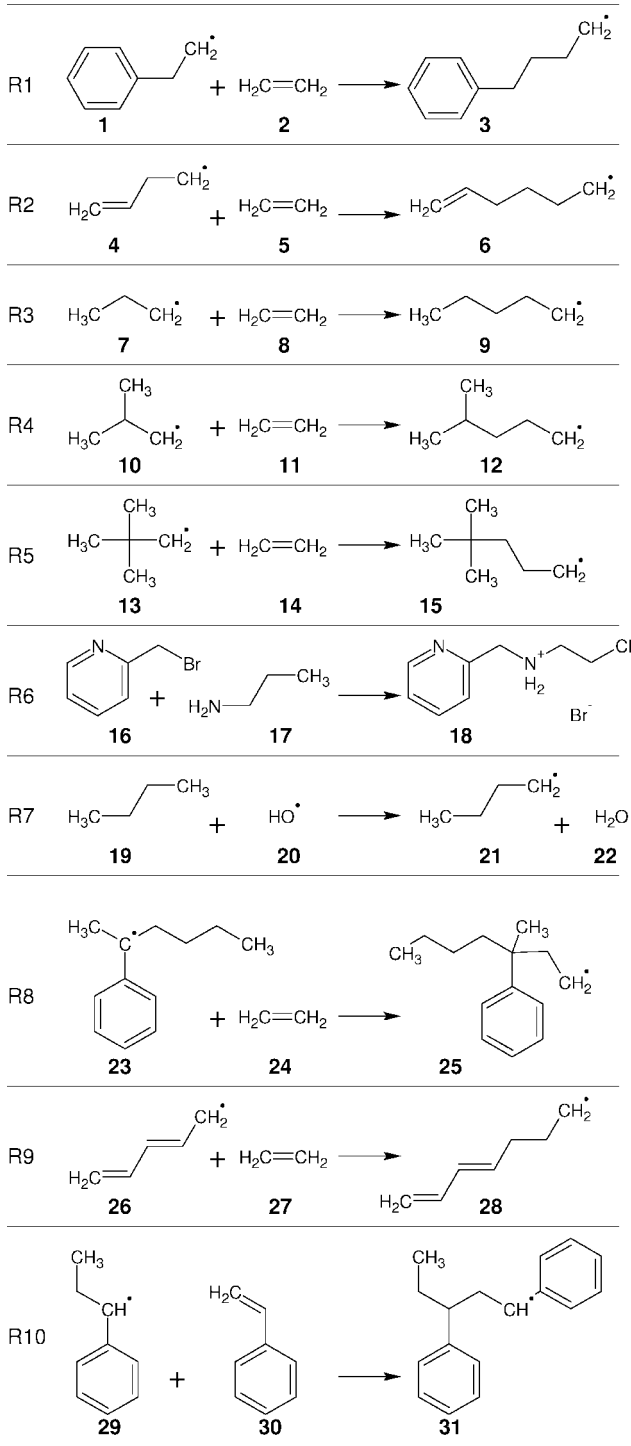


***Figure 1.*** Overview of studied reactions.

2 and 3, and the transition state and product are assumed to contain the same block(s). Blocks can be classified in various types: they can include the reactive center (which at first view appears to be a surprising choice), they can directly be connected with the reactive center by a single bond, or they are separated by more than one bond. It is also possible to combine blocks to the case of multiple blocks. In bimolecular reactions, rigid blocks can be introduced in each of the two reactants; in the activated complex and product, they form multiple blocks, hindering in principle the application of the PHVA method. To illustrate, in reaction R10, one can combine blocks a and b in the description of the two

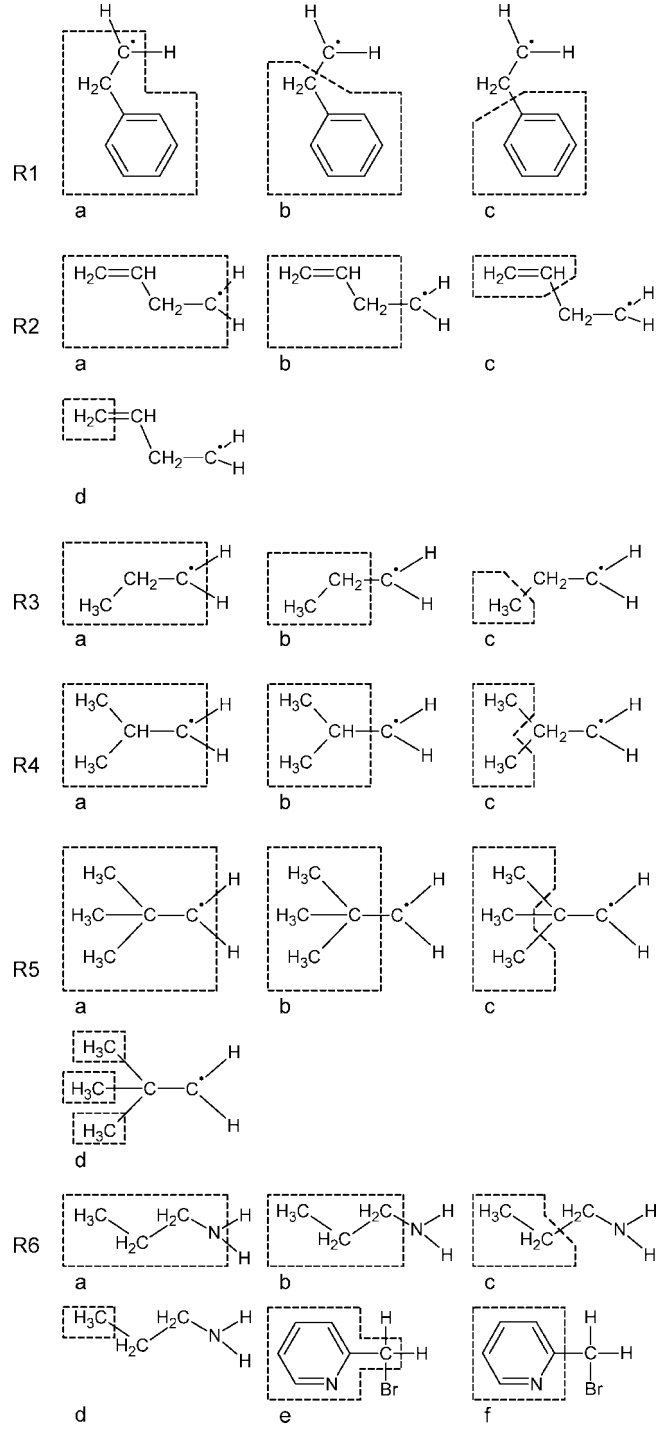


***Figure 2.*** Numbering and choice of the different blocks.

reactants. This case will be denoted as a−b and only makes sense when using MBH.

## 4. Discussion

**4.1. MBH with a Single Block.** In the MBH (PHVA) approach, the total partition function of eq 1 is used to calculate the reaction rate, but the vibrational partition function $q_{vib}$ is constructed with the MBH (PHVA) frequencies:

$$q^{\mathrm{MBH}} = q_{\mathrm{trans}} q_{\mathrm{rot}} q_{\mathrm{vib}}^{\mathrm{MBH}} \tag{13}$$

$$q^{\mathrm{PHVA}} = q_{\mathrm{trans}} q_{\mathrm{rot}} q_{\mathrm{vib}}^{\mathrm{PHVA}} \tag{14}$$

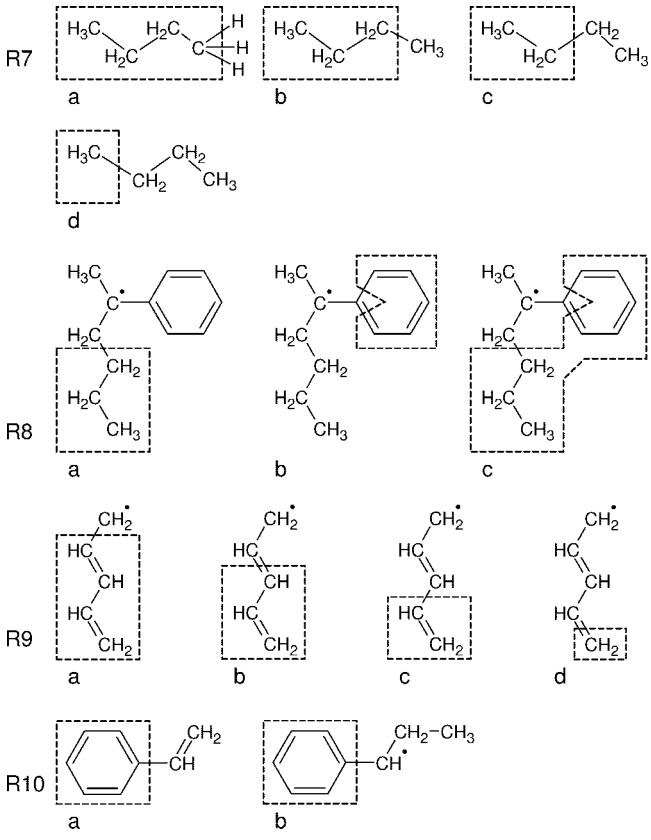


***Figure 3.*** Numbering and choice of the different blocks—continuation.

The chosen test set of chemical reactions allows an exhaustive investigation of the influence on the rate constant of the position of the rigid block, the block's mass, its distance to the reactive center, and the stochiometry of the reaction.

In Tables 1 and 2, the rate constants at $T = 300$ K are listed for the several reactions in units of cubic meters per mole second (bimolecular reactions) or inverse seconds (unimolecular reactions). In the first column, the benchmark values $k$, calculated with the full Hessian frequencies of the equilibrium structure, are tabulated for comparison. The benchmark is only available for the fully optimized structure and is calculated in absence of any block. The block size in the MBH or PHVA approach applied on a fully optimized structure is indicated by a, b, c etc. A prime is added if the geometry was obtained by partial optimization, e.g. a′, b′, c′, etc.

In a first step, we concentrate on the results obtained with the fully optimized structures. In the next step, the influence of the partial optimization will be discussed.

As can be seen in Tables 1 and 2, the overall agreement of the MBH rate constants with the benchmark values is remarkably good. The reaction rate constants are reproduced to within a factor of 2, apart from a few cases in Table 2, which are discussed further. This observation holds for a variety of reactions: for unimolecular and bimolecular reactions, for radical and nonradical reactions, and for heavy or small block masses. The deviation is within acceptable limits and is smaller than corrections induced by the level of theory,[31] internal rotations,[32,33] tunnel effects, and other factors.[34,35]

The apparent agreement of the MBH predictions with the benchmark implies that the contribution of the omitted normal modes, inherent to the MBH method, is of the same magnitude for both the transition states and reactants. Apparently the omitted modes are not essential in the determination of the rate constant. These rather unimportant modes are localized in the fixed block or spread out over the fixed block and the optimized region. The more interesting modes are located in the optimized region that contains the active site, and are well reproduced by the MBH approach. The coupling of the MBH modes with the modes localized in the fixed block is left out in the model, but a logical choice of the blocks makes this coupling irrelevant for the rate constant.

When a block is chosen too close to the active site, the coupling between MBH modes and the omitted modes is not always irrelevant anymore. In reaction R1 the rigid block a includes the reactive center, and the border of block b crosses the bond connecting the radical center. The reaction rate constant $k^{\mathrm{MBH}}$ indeed overestimates the benchmark value. Block c is a better choice because it is not directly connected to the reactive site.

In some particular cases, e.g. reaction R4 with block a or b, the MBH approach reproduces $k$ fairly well even with a direct bond between active site and block. However, one should not rely on such coincidences, and anyway, a more suitable choice of a block further away from the radical center still improves the rate estimate. As a general rule, hereafter referred to as the bond-distance rule, it is recommended not to bring the block region too close to the active site.

The mass of the rigid block does not play a crucial role in the validation of MBH in reproducing rate constants. This is best illustrated by comparing reactions R1 and R2. In R1, the fixed block c contains a phenyl group, while block c in R2 consists of an ethyl group. Results are comparable for both the forward and reverse reactions.

When we finally consider the results of the partial optimization, it is clear that the effect is rather moderate. We concentrate on the forward reaction R1 for a detailed study (Table 3). The partial optimization affects the geometry, because the rigid block conserves its initial internal geometry. This will cause differences with the benchmark geometry. In this simple example, this induces quite slight changes (some C−C distances are increased by 0.03 Å), but in more complex systems, the low level of theory geometry and partial optimized geometry may differ substantially. Or, the full optimization at the low level of theory should give a plausible internal geometry for the blocks, but the exact position/orientation of the blocks and the positions of the relaxed atoms are less important, since these are optimized during the consecutive partial optimization at the high level of theory, giving a plausible geometry of the whole system.

The ground-state configuration of a partially optimized system is obviously less bound than the fully optimized system. However, the energy increase of 2 kJ/mol, noticed in the ethylbenzene radical, is mostly compensated by a similar increase of the binding energy of the TS, hence resulting in an almost equal reaction barrier. For instance,

***Table 1.*** Calculated Rate Constants at $T = 300$ K, for Reactions R1–R4 of the Test Set[a]

| | forward | | | | | backward | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| reaction | $k$ | block | $k^{PHVA}/k$ | $k^{PHVA*}/k$ | $k^{MBH}/k$ | $k$ | block | $k^{PHVA}/k$ | $k^{PHVA*}/k$ | $k^{MBH}/k$ |
| R1 | 3.46E−02 | a | 5.36 | 1.23 | 1.36 | 1.86E−06 | a | 0.39 | 0.35 | 0.48 |
| | | b | 7.44 | 1.71 | 1.74 | | b | 0.91 | 0.83 | 0.83 |
| | | c | 4.33 | 0.99 | 1.02 | | c | 1.13 | 1.02 | 0.95 |
| | | a′ | 5.97 | 1.37 | 1.52 | | a′ | 0.37 | 0.34 | 0.46 |
| | | b′ | 7.58 | 1.73 | 1.76 | | b′ | 0.99 | 0.90 | 0.90 |
| | | c′ | 4.21 | 0.96 | 0.99 | | c′ | 1.15 | 1.04 | 0.97 |
| R2 | 2.85E−02 | a | 10.00 | 1.00 | 1.12 | 1.78E−06 | a | 0.43 | 0.36 | 0.65 |
| | | b | 14.68 | 1.47 | 1.44 | | b | 1.03 | 0.86 | 0.88 |
| | | c | 8.91 | 0.89 | 0.93 | | c | 1.09 | 0.91 | 0.96 |
| | | d | 9.62 | 0.96 | 1.00 | | d | 1.16 | 0.96 | 1.00 |
| | | a′ | 8.31 | 0.84 | 0.94 | | a′ | 0.40 | 0.34 | 0.61 |
| | | b′ | 10.83 | 1.09 | 1.07 | | b′ | 1.06 | 0.89 | 0.92 |
| | | c′ | 8.41 | 0.84 | 0.88 | | c′ | 1.07 | 0.89 | 0.96 |
| | | d′ | 9.62 | 0.96 | 1.00 | | d′ | 1.16 | 0.96 | 1.00 |
| R3 | 1.47E−02 | a | 14.68 | 0.98 | 1.23 | 1.51E−06 | a | 0.44 | 0.35 | 0.69 |
| | | b | 19.84 | 1.32 | 1.33 | | b | 1.00 | 0.81 | 0.85 |
| | | c | 14.78 | 0.98 | 0.99 | | c | 1.18 | 0.95 | 0.99 |
| | | a′ | 16.56 | 1.10 | 1.39 | | a′ | 0.41 | 0.33 | 0.65 |
| | | b′ | 21.08 | 1.40 | 1.41 | | b′ | 1.06 | 0.85 | 0.89 |
| | | c′ | 14.73 | 0.98 | 0.99 | | c′ | 1.22 | 0.98 | 1.03 |
| R4 | 1.99E−03 | a | 7.21 | 0.82 | 1.10 | 4.93E−06 | a | 0.34 | 0.28 | 0.61 |
| | | b | 9.14 | 1.04 | 1.09 | | b | 0.75 | 0.62 | 0.70 |
| | | c | 8.66 | 0.98 | 0.97 | | c | 1.27 | 1.05 | 1.08 |
| | | a′ | 7.52 | 0.86 | 1.14 | | a′ | 0.28 | 0.23 | 0.49 |
| | | b′ | 8.52 | 0.97 | 1.01 | | b′ | 0.76 | 0.63 | 0.71 |
| | | c′ | 8.57 | 0.97 | 0.96 | | c′ | 1.38 | 1.14 | 1.16 |

[a] The forward rate constants are expressed in units of cubic meters per mole second (bimolecular), and the backward rate constants are in units of inverse seconds (unimolecular). The benchmark value $k$ is given for comparison. Rate constants $k^{MBH}$ ($k^{PHVA}$, $k^{PHVA*}$) are calculated with the MBH (PHVA, PHVA*) frequencies, for several block choices. The ratios reflect the influence of the MBH (PHVA, PHVA*) treatment with respect to the benchmark value. A block without a prime indicates a fully optimized structure, and a block with a prime indicates a partially optimized structure.

in the case study, a suitable choice of the fixed block (block c′) predicts a reaction barrier that is hardly different (by 0.04 kJ/mol) from the benchmark value (see Table 3). Significant changes of reaction barriers alter the reaction rate constant to a large extent, but apparently the various reactions R1–R5 of the test set give no indication of this behavior, if one respects sufficient distance between the fixed blocks and the reactive site.

In Table 3, the kinetic parameters $A$ and $E_a$, determined within the temperature range 300−700 K, are also given. Activation energies remain almost unaffected as could be expected. Potential deviations of $k^{MBH}$ originate from the pre-exponential factor, which is mainly determined by the vibrational contribution to the partition function.

The above discussion validates the use of the MBH model to predict the rate constant on an accurate level. A plausible choice of the fixed block is the only essential ingredient a potential user of MBH should take into account to get adequate predictions of chemical kinetics. MBH is computationally attractive, makes quantum chemical calculations feasible in extended molecular systems, and preserves the true reaction mechanism.

**4.2. MBH with Multiple Blocks.** The ability of MBH to choose multiple blocks freely moving but conserving their internal structure makes it a powerful tool to a broad range of applications. This is demonstrated in reactions R5, R8, and R10 of the test set, and the results are tabulated in Tables 1 and 2.

In reaction R5, the effect of multiple blocks compared to a single block (block choice d versus c) is moderate. Reaction R8 describes a more complex system. Two individual blocks a and b can be merged to one solid block c, or they can be considered as two mobile blocks a−b. Here, the multiple MBH implies a significant improvement with respect to the single block treatment c. Block c yields ratios 3.86 and 0.44 for the forward and backward reaction, respectively, while the multiple blocks a−b give values of 1.67 and 0.76. Inspection of the MBH values obtained with the individual blocks a and b shows that the global effect of multiple blocks is mostly given by the following multiplication rule:

$$\frac{k_{a-b}^{MBH}}{k} \approx \frac{k_{a}^{MBH}}{k} \times \frac{k_{b}^{MBH}}{k} \tag{15}$$

This seems to be true for the forward and backward reaction.

A third example is reaction R10 where we choose a block in each reactant. The TS will then contain two blocks, which are treated within the multiple MBH. The overall factor is indeed fairly well reproduced by the multiplication rule (14). At least it gives an indication on the global error induced by the presence of multiple blocks. A plausible block choice is always of importance to keep the error within the limits. Unphysical block choices are for example b in R8 and c in R9. In both cases the block's border crosses a bond that is part of a delocalized system, and therefore, the $k^{MBH}$ ratios are badly reproduced, even when the bond-distance rule is respected.

***Table 2.*** Calculated Rate Constants at $T = 300$ K, for Reactions R5–R10 of the Test Set[a]

| | forward | | | | | backward | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| reaction | $k$ | block | $k^{PHVA}/k$ | $k^{PHVA*}/k$ | $k^{MBH}/k$ | $k$ | block | $k^{PHVA}/k$ | $k^{PHVA*}/k$ | $k^{MBH}/k$ |
| R5 | 1.72E−03 | a | 5.46 | 0.84 | 1.01 | 2.94E−06 | a | 0.42 | 0.36 | 0.60 |
| | | b | 6.57 | 1.01 | 1.04 | | b | 0.91 | 0.77 | 0.85 |
| | | c | 6.28 | 0.96 | 0.99 | | c | 1.03 | 0.88 | 0.95 |
| | | d | | | 1.04 | | d | | | 0.98 |
| | | a′ | 6.12 | 0.94 | 1.13 | | a′ | 0.43 | 0.36 | 0.61 |
| | | b′ | 6.47 | 0.99 | 1.03 | | b′ | 1.07 | 0.91 | 0.99 |
| | | c′ | 6.46 | 0.99 | 1.01 | | c′ | 1.19 | 1.01 | 1.08 |
| | | d′ | | | 1.16 | | d′ | | | 1.23 |
| R6 | 9.93E−12 | a | 1060.01 | 1.74 | 1.76 | 1.41E−13 | a | 0.51 | 0.49 | 0.69 |
| | | b | 1002.56 | 1.64 | 1.71 | | b | 0.81 | 0.78 | 0.78 |
| | | c | 794.07 | 1.30 | 1.33 | | c | 0.89 | 0.86 | 0.85 |
| | | d | 657.28 | 1.08 | 1.10 | | d | 1.11 | 1.08 | 1.00 |
| | | e | 2.61 | 0.29 | 0.85 | | e | 0.17 | 0.16 | 0.56 |
| | | f | 8.70 | 0.96 | 1.00 | | f | 0.83 | 0.80 | 0.80 |
| R7 | 2.85E+07 | a | 4.15 | 0.43 | 0.95 | 1.24E−04 | a | 10.67 | 2.03 | 1.73 |
| | | b | 4.66 | 0.49 | 0.99 | | b | 8.37 | 1.59 | 1.56 |
| | | c | 4.86 | 0.51 | 1.05 | | c | 5.50 | 1.04 | 1.02 |
| | | d | 4.79 | 0.50 | 1.03 | | d | 5.41 | 1.03 | 1.02 |
| | | a′ | 5.06 | 0.53 | 1.16 | | a′ | 16.69 | 3.19 | 2.71 |
| | | b′ | 5.22 | 0.55 | 1.11 | | b′ | 9.74 | 1.84 | 1.82 |
| | | c′ | 5.03 | 0.52 | 1.08 | | c′ | 5.58 | 1.06 | 1.04 |
| | | d′ | 4.85 | 0.51 | 1.04 | | d′ | 5.48 | 1.04 | 1.04 |
| R8 | 8.04E−11 | a | 2.14 | 1.14 | 1.11 | 2.97E+02 | a | 1.31 | 1.25 | 0.99 |
| | | b | 3.01 | 1.61 | 1.50 | | b | 0.89 | 0.84 | 0.77 |
| | | c | 5.91 | 3.15 | 3.86 | | c | 0.48 | 0.46 | 0.44 |
| | | a−b | | | 1.67 | | a−b | | | 0.76 |
| R9 | 1.83E−09 | a | 21.89 | 2.53 | 2.64 | 3.27E+01 | a | 1.60 | 1.31 | 2.27 |
| | | b | 14.00 | 1.62 | 1.50 | | b | 0.99 | 0.81 | 1.29 |
| | | c | 14.40 | 1.66 | 1.63 | | c | 2.29 | 1.87 | 5.14 |
| | | d | 10.24 | 1.18 | 0.98 | | d | 0.69 | 0.57 | 2.04 |
| R10 | 3.02E−06 | a | 43.51 | 0.89 | 0.96 | | | | | |
| | | b | 30.18 | 1.34 | 1.59 | | | | | |
| | | a−b | | | 1.53 | | | | | |

[a] See the footnote of Table 1 for more details.

***Table 3.*** MBH for Reaction R1 with Fully and Partially Optimized Structures[a]

| | full optim | | | partial optim | | |
|---|---|---|---|---|---|---|
| | a | b | c | a′ | b′ | c′ |
| $k^{MBH}/k$ | 1.36 | 1.74 | 1.02 | 1.52 | 1.76 | 0.99 |
| $A^{MBH}/A$ | 1.29 | 1.59 | 1.03 | 1.38 | 1.70 | 1.01 |
| $E_a^{MBH} - \Delta E_a$ | −0.14 | −0.21 | 0.01 | −0.22 | −0.02 | +0.06 |
| $\Delta E_0^{MBH} - \Delta E_0$ | 0 | 0 | 0 | −0.13 | +0.13 | +0.04 |

[a] The rate constant is given at 300 K, and kinetic parameters are fitted in the temperature range 300–700 K. $k$ and $A$ are in cubic meters per mole second, and energies are in kilojoules per mole. Benchmark values: $k = 3.46 \times 10^{-2}$ m$^3$ mol$^{-1}$ s$^{-1}$, $A = 67.22 \times 10^2$ m$^3$ mol$^{-1}$ s$^{-1}$, $E_a = 36.53$ kJ/mol, $\Delta K_0 = 24.66$ kJ/mol.

**4.3. MBH for Modeling Solvents.** Finally, we have tested the concept of multiple blocks on a more realistic and more extended example, where several explicit solvent molecules are taken into account in the computation. In this case, blocks can be chosen within the reacting molecule and/or the solvent molecules can be treated as blocks, and moreover, the influence of solvent species on both reaction barrier and frequencies, i.e. pre-exponential factor, can be tested.

We have chosen the cyclization of functionalized aminophosphonates, as a representative reaction occurring in an organic solvent (reaction R11, see Figure 4). The choice of this reaction was inspired by a recent combined experimental and theoretical study on the formation of $\beta$-lactams by some

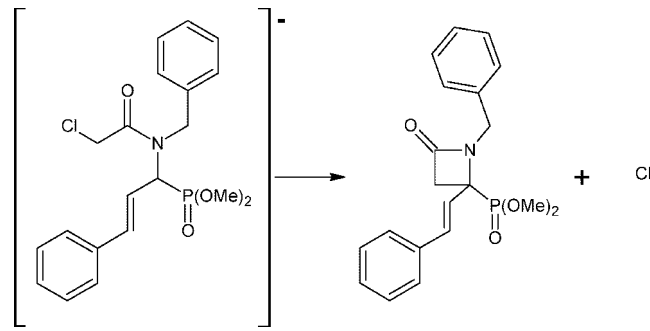


***Figure 4.*** Reaction R11.

of the authors.[36] It was found that, starting from an ambient allylic anion, ring closure occurred exclusively by 4-ring formation, without any trace of 6-ring lactams. At that time, pre-exponential factors were not calculated due to the high computational cost. This is thus an ideal example to validate the approach.

The studied system consists of the aminophosphonate anion together with a sodium counterion and solvated in dimethyl ether solvent molecules (DME). The latter were taken as model molecules for tetrahydrofuran. Three cases are considered: the reaction in the absence of explicit solvent molecules (R11), in the presence of one DME (R11 + 1DME), and in the presence of two DMEs (R11 + 2DME). The benchmark values of $k$, $A$, $E_a$, and $\Delta E_0$ are given in Table 4 for $T = 300$ K. The ratios (for $k$ and $A$) and differences

***Table 4.*** Benchmark Results for Reaction R11 without and with 1 and 2DME[a]

| | R11 | R11 + 1DME | | R11 + 2DME | |
|---|---|---|---|---|---|
| $k$ | 7.62E−15 | 6.79E−14 | (8.91) | 2.09E−12 | (274.63) |
| $A$ | 2.10E+13 | 8.43E+12 | (0.40) | 5.64E+13 | (2.69) |
| $E_a$ | 157.73 | 149.99 | (−7.73) | 146.19 | (−11.54) |
| $\Delta E_0$ | 159.85 | 151.78 | (−8.07) | 148.23 | (−11.63) |

[a] The rate constant is given at 300 K, and kinetic parameters are fitted in the temperature range 300–700 K. $k$ and $A$ are in inverse seconds, and energies are in kilojoules per mole. Ratios ($k$ and $A$) and differences ($E_a$ and $\Delta E_0$) between solvated and nonsolvated values are given between brackets.

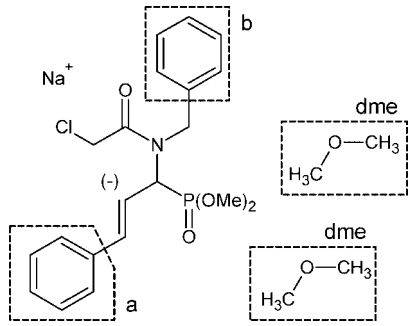


***Figure 5.*** Definition of blocks, reaction R11.

***Table 5.*** Calculated at 300 K[a]

| block | R11 | R11 + 1DME | R11 + 2DME |
|---|---|---|---|
| a | 1.71 | 1.62 | 1.43 |
| b | 0.98 | 0.97 | 0.88 |
| dme | | 1.07 | |
| b−dme | | 1.04 | |
| dme−dme | | | 0.94 |
| b−dme−dme | | | 0.83 |

[a] Several blocks choices are taken up.

(for energies) given between brackets indicate the effect of the solvation. The presence of one or two solvent molecules indeed increases the reaction rate constant by a factor of 8.91 or 274.63, respectively, with respect to the nonsolvated situation.

The relevant question is whether the MBH model is capable of reproducing the enhancement of $k$ due to the solvent. Several block choices are depicted in Figure 5, including the case of blocks within the reactant (a, b), as well as blocks consisting of solvent molecules (dme). Table 5 shows the ratios between the MBH estimates and the benchmark values of the rate constant. Block a is clearly not a good choice, which is easily understood when noting that the block's border cuts through a delocalized bond. Therefore, possible combinations of a with blocks b or dme are not considered in the table. Block b and block dme on the other hand are excellent block choices: since the ratios are close to 1.0, the $k$ enhancement 1:8.91:274.63 as reported by the benchmark is maintained and the MBH is thus clearly capable of reproducing the solvation effect. Multiple block combinations such as b−dme, dme−dme, and b−dme−dme reproduce the rate constant very well, which is in agreement with the multiplication rule as stated in eq 14. Resuming, the multiple MBH has proven to be extremely useful and effective in predicting reaction rates, both with blocks belonging to the reactant or with blocks coinciding with solvent molecules.

**4.4. PHVA and PHVA*.** Conceptually, the difference between MBH and PHVA is mainly a mass effect. In the MBH, the finite mass of the blocks is taken into consideration, while in the PHVA approach, infinite masses are associated with the atoms in the rigid body. As a result, an extension to multiple blocks has no physical meaning in PHVA. When two blocks with infinite mass are present within one molecule, the system of free atoms and blocks will behave as if the two blocks were one big block with infinite mass. The case of one block in each of the reactants of a bimolecular reaction must also be excluded. The transition state itself would have two blocks with infinite mass. Thereby, six degrees of freedom describing the relative position and orientation of the two blocks will be lost in the transition state, leading to a completely wrong temperature dependence of the reaction rate constant. From a physical point of view, it is also hard to imagine how two reactants, each containing a block with infinite mass, could ever approach each other to form the transition state. The following discussion is therefore limited to the case of a single block with fixed geometry.

In contrast to the MBH, the PHVA cannot be extended to treat multiple blocks. PHVA is thus only applicable within the single block approximation. An overview of the various PHVA reaction rates in Tables 1 and 2 shows that unimolecular reactions are reasonably well described using PHVA frequencies. On the other hand, bimolecular reaction rates are poorly reproduced and significant deviancies are noticed. The systematic overestimation of the reaction rate finds its origin in the appearance of spurious low frequency modes in the PHVA approach. A profound investigation of these spurious modes reveals that they represent slow translation/rotationlike movements of the whole group of free atoms. This collective motion encompasses a lot of mass, explaining why (through the mass weighting in the NMA analysis) these frequencies are low. They give a significant contribution to the vibrational partition functions, while the translational/rotational degrees of freedom, however, are already taken into account in the total partition function. The larger the total mass of the free atoms with respect to the mass of the fixed block, the more pronounced is this double counting. Hence, in unimolecular reactions, the enhancement of the vibrational partition functions due to this double counting effect is nearly similar for reactant and transition state, and the enhancement factor is canceled (see eq 8). In bimolecular reactions on the other hand, the double counting is much more prominent for the transition state than for the reactants, thus leading to an overestimated reaction rate.

In order to prevent this double counting effect, we present a corrected version of the PHVA method. In Figure 6, the ratio $q_{vib}^{MBH}/q_{vib}^{PHVA}$ between the MBH and PHVA vibrational partition functions for the reactants, TS, and products of reactions R1–R6 is plotted against a mass related factor $t$ given by

$$t = \sqrt{\frac{M_F^3 I_{F1} I_{F2} I_{F3}}{M^3 I_1 I_2 I_3}} \tag{16}$$

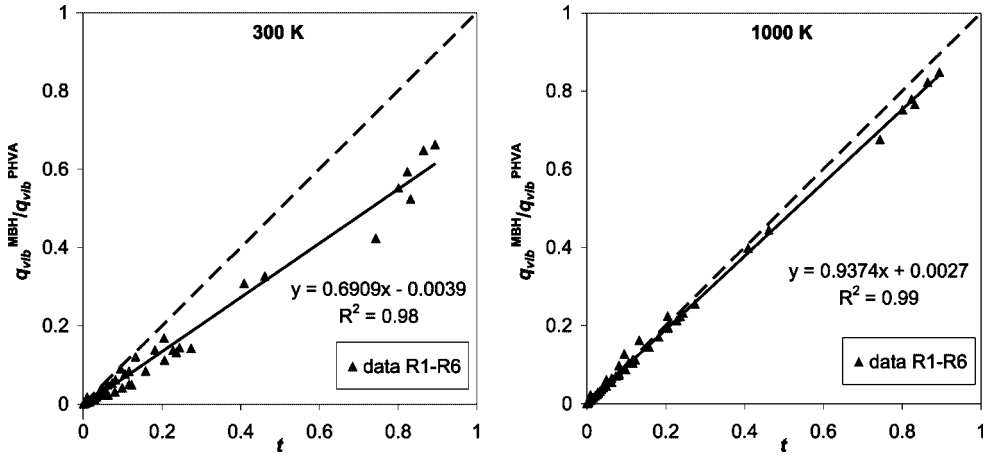


***Figure 6.*** Ratio $q_{vib}^{MBH}/q_{vib}^{PHVA}$ for reactants, TS, and products of reactions R1– R6 plotted against the mass related factor $t$ at 300 and 1000 K. The linear regression line (full) is fitted to the data with the least-squares method. The diagonal (dashed line) is added for comparison.

where $M$ is the total mass and $I_i$ ($i = 1, 2, 3$) are the moments of inertia of the molecule, while $M_F$ and $I_{Fi}$ ($i = 1, 2, 3$) are the total mass and moments of inertia of the fixed block. For higher temperatures, an almost linear behavior is observed.

$$\frac{q_{vib}^{MBH}}{q_{vib}^{PHVA}} \approx t \tag{17}$$

On the basis of eq 17, we now propose the following corrected PHVA partition function, hereafter referred to as PHVA*,

$$q^{PHVA*} = q_{trans} q_{rot} q_{vib}^{PHVA} \sqrt{\frac{M_F^{\,3} I_{F1} I_{F2} I_{F3}}{M^3 I_1 I_2 I_3}} \tag{18}$$

It is not surprising that the ratio $q_{vib}^{MBH}/q_{vib}^{PHVA}$ depends only on mass properties, because the essential difference between the MBH and PHVA approach is a reduced mass effect. The plot in Figure 6 is so overwhelming that a mathematical proof of eq 16 should be on hand and that a close similarity between PHVA* and MBH predictions for the reaction rate constants is expected. The latter is indeed confirmed by the corrected PHVA* estimates taken up in Tables 1 and 2. Concluding, PHVA* and MBH perform equally well, but the main advantage of MBH, i.e. enabling the extension the procedure to multiple blocks, still holds.

In the following, a mathematical derivation of eq 17 will be presented. In the high temperature limit ($k_B T >> h\nu$, $\forall \nu$), it is possible to relate $q_{vib}^{PHVA}$ and $q_{vib}^{MBH}$ as a simple expression containing ratios of the masses and moments of inertia. The contribution of a vibration with frequency $\nu$ to the partition function is given by eq 5 and can be approximated by $k_B T/h\nu$ if the temperature is high with respect to the vibrational temperature $h\nu/k_B$. Since the number of MBH and PHVA frequencies is equal, the ratio is independent of temperature.

$$\frac{q_{vib}^{MBH}}{q_{vib}^{PHVA}} = \frac{\Pi_\nu \nu^{PHVA}}{\Pi_\nu \nu^{MBH}} \tag{19}$$

The product of frequencies coincides with the square root of the determinant of the matrix in the NMA normal mode equations. In the PHVA model (eq 10), this is

$$\prod_\nu \nu^{PHVA} = \sqrt{\det(M_E^{-1/2} H_E M_E^{-1/2})} \tag{20}$$

while in the MBH model (eq 11), this is

$$\prod_\nu \nu^{MBH} = \sqrt{\det(\tilde{M}'^{-1/2} H_E \tilde{M}'^{-1/2})} \tag{21}$$

The ratio becomes

$$\frac{q_{vib}^{MBH}}{q_{vib}^{PHVA}} = \sqrt{\frac{\det \tilde{M}'}{\det M_E}} \tag{22}$$

We now introduce the matrix $S$, defined in the Appendix, which contains the mass information of the complete system. Eigenvalues are the total mass $M$ and the moments of inertia $I_i$. Similarly we introduce the matrix $S_F$ for the fixed atoms, with eigenvalues $M_F$ and $I_{Fi}$. Using the properties described in ref 17, the ratio can be rewritten (see the Appendix):

$$\frac{q_{vib}^{MBH}}{q_{vib}^{PHVA}} = \sqrt{\frac{\det S_F}{\det S}} \tag{23}$$

which is equivalent to expression 17 and which proves the PHVA* correction factor of eq 18.

Numerically, we find that eq 18 is not only valid for high temperatures but that its validity holds quite well for lower temperatures (300 K); see Figure 6.

An interesting property is that the mass related factor $t$ of eq 16 is also equal to the ratio of the translational/rotational partition functions of the fixed block versus global molecule

$$\sqrt{\frac{M_F^{\,3} I_{F1} I_{F2} I_{F3}}{M^3 I_1 I_2 I_3}} = \frac{q_{F,trans} q_{F,rot}}{q_{trans} q_{rot}} \tag{24}$$

The subscript $F$ refers to the fixed block atoms. Thus, an alternative formulation of the PHVA* approach is presented, where only the vibrational partition function is taken into account. The total translational and rotational partition

function of the molecule are omitted as if it were to avoid the double counting effect.

$$q^{\mathrm{PHVA}*} = q_{\mathrm{vib}}^{\mathrm{PHVA}} \tag{25}$$

Expression 25 is different from the one proposed in eq 17, but it amounts to the same result when calculating reaction rates. The factor $M_{\mathrm{F}}^{3/2}$ is the same in both TS and reactants because the same block atoms are chosen, and thus, this factor is canceled in the numerator and denominator in the expression of $k$. The factor $\sqrt{I_{\mathrm{F1}}I_{\mathrm{F2}}I_{\mathrm{F3}}}$ might slightly differ between TS and reactants, if the internal geometry of the blocks is not completely identical for the TS and reactants, but in good approximation, it is canceled as well. The factor $M^{3/2}$ cancels with the translational partition function and $\sqrt{I_1I_2I_3}$ with the vibrational partition function. Therefore, eq 25 will lead to (almost) identical results as eq 17.

## 5. Conclusion

In this work, the MBH method has been shown to act as an accurate method for the prediction of chemical kinetics in large extended molecular systems. In contrast to the PHVA approach, the MBH method also performs fairly well in bimolecular reactions. An adapted version of PHVA is presented correcting for the double counting effect of global rotation and translation inherent to the PHVA method. The surplus value of MBH with regard to PHVA* lies in the flexibility of MBH to introduce multiple rigid blocks which are freely moving with respect to each other but keeping their initial internal structure. This facility gives a lot of new perspectives in predicting chemical kinetics in very complex systems, where the introduction of one single fixed block is a too crude approximation. Partial optimization is necessary to make quantum chemical computations feasible. The possibility to introduce multiple blocks, each still having six degrees of freedom, makes an accurate reproduction of kinetics to the possibilities.

Most promising application field of MBH would be the description of chemical reactions in a solvent. Each solvent molecule may be regarded as a fixed block, keeping its internal structure, but still enabling to translate/rotate freely with respect to the chemically active part of the system. All ab initio program packages can be used on the condition that the built-in optimization routine allows constraints on internal degrees of freedom. The computational advantage of the MBH method can be exploited when the program package has the ability to calculate partial Hessians. If both features are implemented, MBH could be regarded as a groundbreaking model in the treatment of complex reactions where environment plays a crucial role.

**Acknowledgment.** This work is supported by the Fund for Scientific Research−Flanders and by the BOF funds of Ghent University.

## Appendix

Consider a molecule with $N$ masses $m_A$, $A = 1, \ldots, N$. The positions are described by Cartesian coordinates $\mathrm{r}_A \equiv \{r_{A\mu}\}_{\mu=x,y,z}$, with respect to a space-fixed frame. We will treat the case of one MBH block, consisting of $N_{\mathrm{F}}$ atoms. The remaining $N_{\mathrm{E}} = N - N_{\mathrm{F}}$ atoms are in equilibrium due to the partial optimization. An index E (F) will be used to indicate quantities where only the free (fixed) atoms are considered.

We will focus on the normal mode equations for PHVA and for MBH and, in particular, on the difference in the mass matrices, in order to study the transition from eq 22 to eq 23. The PHVA mass matrix is simply given by $M_{\mathrm{E}}$ [see ()]. The original (not yet transformed to (11)) MBH normal mode equations read

$$\tilde{H}\nu = \lambda\tilde{M}\nu \tag{26}$$

where $\tilde{M}$ and $\tilde{H}$ are the MBH mass matrix and Hessian [see ref 17].

Define now a $3N \times 6$ matrix $D$ with components

$$D_{A\mu,\,\alpha} = \begin{cases} \delta_{\mu,x} & \alpha = 1 \\ \delta_{\mu,y} & \alpha = 2 \\ \delta_{\mu,z} & \alpha = 3 \\ \Sigma_\lambda \epsilon_{\lambda\mu x} & \alpha = 4 \\ \Sigma_\lambda \epsilon_{\lambda\mu y} & \alpha = 5 \\ \Sigma_\lambda \epsilon_{\lambda\mu z} & \alpha = 6 \end{cases} \tag{27}$$

With $M$ as the diagonal $3N \times 3N$ mass matrix, the matrix $S = D^{\mathrm{T}}MD$ is introduced, and similarly, $S_{\mathrm{F}} = D_{\mathrm{F}}{}^{\mathrm{T}}M_{\mathrm{F}}D_{\mathrm{F}}$. The MBH mass matrix is then given by the block diagonal matrix

$$\tilde{M} = \begin{pmatrix} S_{\mathrm{F}} & 0_{6\times d} \\ 0_{d\times 6} & M_{\mathrm{E}} \end{pmatrix} \tag{28}$$

with $d = 3N_{\mathrm{E}}$. The normal mode equations are transformed by simultaneous block diagonalization of $\tilde{H}$ and $\tilde{M}$. The required transformation matrices are given by

$$T_1 = \begin{pmatrix} 1_{6\times 6} & 0_{6\times d} \\ x & 1_{d\times d} \end{pmatrix}; \quad T_2 = \begin{pmatrix} 1_{6\times 6} & y \\ 0_{d\times 6} & 1_{d\times d} \end{pmatrix} \tag{29}$$

with $x = D_{\mathrm{E}}$ and $y = -S^{-1}D_{\mathrm{E}}{}^{\mathrm{T}}M_{\mathrm{E}}$. The transformed MBH mass matrix and Hessian directly lead to eq 11:

$$T_2^{\mathrm{T}}T_1^{\mathrm{T}}\tilde{H}T_1T_2 = \begin{pmatrix} 0_{6\times 6} & 0_{6\times d} \\ 0_{d\times 6} & H_{\mathrm{E}} \end{pmatrix}, \quad T_2^{\mathrm{T}}T_1^{\mathrm{T}}\tilde{M}T_1T_2 = \begin{pmatrix} S & 0_{6\times d} \\ 0_{d\times 6} & \tilde{M}' \end{pmatrix} \tag{30}$$

with $\tilde{M}' = M_{\mathrm{E}} - M_{\mathrm{E}}D_{\mathrm{E}}S^{-1}D_{\mathrm{E}}{}^{\mathrm{T}}M_{\mathrm{E}}$. Or, the relevant mass matrix is $\tilde{M}'$ for MBH.

Since by construction $\det T_1 = \det T_2 = 1$, it is obvious that the following relations between determinants hold:

$$\det \tilde{M} = \det S_{\mathrm{F}} \det M_{\mathrm{E}} \tag{31}$$

$$\det (T_2^{\mathrm{T}}T_1^{\mathrm{T}}\tilde{M}T_1^{\mathrm{T}}T_2^{\mathrm{T}}) = \det \tilde{M} = \det S \det \tilde{M}' \tag{32}$$

or

$$\frac{\det \tilde{M}'}{\det M_{\mathrm{E}}} = \frac{\det S_{\mathrm{F}}}{\det S} \tag{33}$$

This proves the transition between eqs 22 and 23.

## References

(1) Gao, J. L.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467–505.

(2) Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107.

(3) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1935**, *31*, 875.

(4) Laidler, K. J. *Chemical Kinetics*; Harper Collins Pulbishers, Inc.: New York, 1987, 87–138..

(5) Mc Quarrie, D. A.; Simon, J. D. *Physical Chemistry - a molecular approach*; University Science Books: Sausalito, CA, 1997; pp 1075–1079..

(6) Warshel, A.; Levitt, M. J. *Mol. Biol.* **1976**, *103* (2), 227–249.

(7) Assfeld, X.; Rivail, J. L. *Chem. Phys. Letters* **1996**, *263* (1–2), 100–106.

(8) Gao, J. L.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102* (24), 4714–4721.

(9) Zhang, Y. K.; Lee, T. S.; Yang, W. T. *J. Chem. Phys.* **1999**, *110* (1), 46–54.

(10) Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Verstraelen, T.; Waroquier, M. *J. Chem. Phys.* **2007**, *126* (22), 224102.

(11) Jin, S. Q.; Head, J. D. *Surf. Sci.* **1994**, *318* (1–2), 204–216.

(12) Calvin, M. D.; Head, J. D.; Jin, S. Q. *Surf. Sci.* **1996**, *345* (1–2), 161–172.

(13) Head, J. D. *Int. J. Quantum Chem.* **1997**, *65* (5), 827–838.

(14) Head, J. D. *Int. J. Quantum Chem.* **2000**, *77* (1), 350–357.

(15) Li, H.; Jensen, J. H. *Theor. Chem. Acc.* **2002**, *107*, 211–219.

(16) Besley, N. A.; Metcalf, K. A. *J. Chem. Phys.* **2007**, *126* (3), 035101.

(17) Ghysels, A.; Van Neck, D.; Waroquier, M. *J. Chem. Phys.* **2007**, *127*, 164108.

(18) Lin, H.; Pu, J. Z.; Albu, T. V.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108* (18), 4112–4124.

(19) Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2001**, *114* (22), 9953–9958.

(20) Fernandez-Ramos, A.; Miller, J. A.; Klippenstein, S. J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106* (11), 4518–4584.

(21) Stevens, F.; Vrielinck, H.; Van Speybroeck, V.; Pauwels, E.; Callens, F.; Waroquier, M. *J. Phys. Chem. B* **2006**, *110* (16), 8204–8212.

(22) Lesthaeghe, D.; Delcour, G.; Van Speybroeck, V.; Marin, G.; Waroquier, M. *Microporous Mesoporous Mater.* **2006**, *96*, 350−356.

(23) Lesthaeghe, D.; De Sterck, B.; Van Speybroeck, V.; Marin, G. B.; Waroquier, M. *Angew. Chem., Int. Ed.* **2007**, *46* (8), 1311–1314.

(24) Cui, Q. *J. Chem. Phys.* **2002**, *117* (10), 4720.

(25) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105* (8), 2999–3093.

(26) Tachibana, A.; Fukui, K. *Theor. Chim. Acta* **1978**, *49* (4), 321–347.

(27) Murry, R.; Fourkas, J. T.; Wu-Xiong, L.; Keyes, T *J. Chem. Phys.* **1999**, *110*, 10410–10422.

(28) Wales, D. J. *J. Chem. Phys.* **2000**, *113*, 3926–3927.

(29) Yang, L. J.; Tan, C. H.; Hsieh, M. J.; Wang, J. M.; Duan, Y.; Cieplak, P.; Caldwell, J.; Kollman, P. A.; Luo, R. *J. Phys. Chem. B* **2006**, *110* (26), 13166–13176.

(30) 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.

(31) Van Speybroeck, V.; Van Cauter, K.; Coussens, B.; Waroquier, M. *Chemphyschem* **2005**, *6* (1), 180–189.

(32) Van Speybroeck, V.; Van Neck, D.; Waroquier, M. *J. Phys. Chem. A* **2000**, *104* (46), 10939–10950.

(33) Vansteenkiste, P.; Van Neck, D.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Phys.* **2006**, *124* (4), 044314.

(34) Sabbe, M. K.; Saeys, M.; Reyniers, M. F.; Marin, G. B.; Van Speybroeck, V.; Waroquier, M. *J. Phys. Chem. A* **2005**, *109* (33), 7466–7480.

(35) Sabbe, M. K.; Vandeputte, A. G.; Reyniers, M. F. O.; Van Speybroeck, V.; Waroquier, M.; Marin, G. B. *J. Phys. Chem. A* **2007**, *111* (34), 8416–8428.

(36) Van Speybroeck, V.; Moonen, K.; Hemelsoet, K.; Stevens, C. V.; Waroquier, M. *J. Am. Chem. Soc.* **2006**, *128* (26), 8468–8478.

CT7002836

JCTC Journal of Chemical Theory and Computation

# How Efficient Is Replica Exchange Molecular Dynamics? An Analytic Approach

Hugh Nymeyer*

*Department of Chemistry & Biochemistry, The School of Computational Science and The Institute for Molecular Biophysics, Florida State University, Tallahassee, Florida 32306-4380*

Received December 10, 2007

**Abstract:** Replica exchange molecular dynamics (REMD) has become a standard technique for accelerating relaxation in biosimulations. Despite its widespread use, questions remain about its efficiency compared with conventional, constant temperature molecular dynamics (MD). An analytic approach is taken to describe the relative efficiency of REMD with respect to MD. This is applied to several simple two-state models and to several real proteins—protein L and the B domain of protein A—to predict the relative efficiency of REMD with respect to MD in actual applications. In agreement with others, we find the following: as long as there is a positive activation energy for folding, REMD is more efficient than MD; the effectiveness of REMD is strongly dependent on the activation enthalpy; and the efficiency of REMD for actual proteins is a strong function of the maximum temperature. Choosing the maximum temperature too high can result in REMD becoming significantly less efficient than conventional MD. A good rule of thumb appears to be to choose the maximum temperature of the REMD simulation slightly above the temperature at which the enthalpy for folding vanishes. Additionally, we find that the number of replicas in REMD, while important for simulations shorter than one or two relaxation times, has a minimal effect on the asymptotic efficiency of the method.

## Introduction

Replica exchange (RE),[1–4] also known as simulated tempering, has become a standard method for enhancing relaxation in any system with rugged energy landscapes. This method has been applied to study a diverse number of systems including biomolecules, spin-glasses, lattice quantum chromodynamics, phylogenetic trees, and polymer melts. The variant known as replica exchange molecular dynamics (REMD),[5] in particular, is widely used in biosimulations.[6–45] Despite this, questions have been raised about the efficiency of REMD relative to conventional, constant temperature molecular dynamics (MD).[46,47] A number of biomolecular simulations have investigated this issue. For example, Periole and Mark[48] compared REMD with conventional simulation of a $\beta$-heptapeptide in explicit solvent and found that "for determining the relative populations at the lower temperature (275–300 K), [REMD] was at least eight times more efficient than conventional MD for this system". Zhang, Wu, and Duan[49] studied a 21-residue helix-forming peptide in implicit solvent and reported that "REMD can significantly enhance the sampling efficiency by 14.3 ± 6.4, 35.1 ± 0.2 and 71.5 ± 20.4 times at, respectively, ~360, ~300, and ~275 K in comparison to the regular MD." Sanbonmatsu and Garcia[7] studied the pentapeptide Met-enkephalin in explicit solvent and found that REMD samples "approximately five times more configurational space than constant temperature MD simulations of the same duration", which suggests an increase in efficiency of REMD by at least a factor five. Rao and Caflisch[50] studied a 20-residue antiparallel beta-sheet protein using both conventional MD and REMD. They found that the average folding time over all replicas using REMD was 0.064–0.067 $\mu$s, and the average folding time using conventional MD near the folding temperature was 0.085 $\mu$s. Seibert et al. used long REMD and MD simulations of a small beta-hairpin (Chignolin) in explicit water to estimate relative

* Phone: 850-590-8160. Fax: 850-644-0098. E-mail: hnymeyer@fsu.edu.

10.1021/ct7003337 CCC: $40.75 
Published on Web 03/15/2008

efficiency.[51] It was found that REMD simulations equilibrate in a few hundreds of nanoseconds per replica compared with a folding time using conventional MD of 1–2 $\mu$s. Extensive REMD simulations of the α-helical Trp-cage protein produced similar results:[44] using REMD, relaxation occurred after about 100 ns of simulation per replica in a system that takes a few $\mu$s to equilibrate with conventional MD.

This evidence suggests that relaxation of biological molecules with REMD is faster than in conventional MD, and for a few select systems, there are indications that REMD simulations may relax an order of magnitude or more quickly than with conventional MD. The relatively small increase in REMD efficiency observed for some biological systems has been attributed to the lack of large activation enthalpies for protein folding,[46,47] which is consistent with the analysis presented in this manuscript.

In order to better understand the relaxation behavior of REMD simulations and the effects of the temperature distribution and exchange frequency, several authors have constructed semianalytic models from master equations. For example, Zheng et al.[52] constructed a Markov state model of a protein REMD simulation to estimate its actual efficiency and the effect of different parameter choices on this efficiency. The most significant finding was that choosing the maximum temperature too large may actually decrease the efficiency of REMD; a result confirmed by the analysis presented here. Trebst et al.[53] used a similar model to improve REMD efficiency by optimizing the temperature distribution. And, simulations and Markov state modeling by Sindhikara et al.[54] were used to study the effects of exchange frequency on REMD simulations. These authors found that for most biological simulations, large gains in REMD efficiency can arise by increasing the exchange frequency. Considering these results, which have shown that the efficiency of REMD depends sensitively on the maximum temperature, temperature spacing and exchange frequency, it is not surprising that quite different results have been found regarding the efficiency of REMD compared with MD.

Semianalytic models such as these give us the ability to rapidly optimize simulation parameters; however, an analytic approach to this problem may help us to better answer a number of questions such as: how does the efficiency of REMD depend upon the particular system being studied; how does the number of replicas needed for the simulation impact its efficiency; and how does the choice of temperatures especially $T_{max}$ affect the efficiency of REMD. An analytic approach is developed in this manuscript to help address these questions. This analysis is quite general and can be adapted to analyzing the efficiency of REMD applied to nearly any system as well as analyzing the numerous variants of RE and REMD that have been developed.

## Background

Considering the extensive use of REMD and numerous descriptions of it in the existing literature, only a brief review will be provided here. REMD is an attempt to enhance sampling by allowing two simulations or *replicas* to exchange their temperatures. This additional dynamical freedom allows the system to relax more quickly than at a fixed temperature. In REMD, the probability for two replicas (1 and 2) out of $N$ replicas to exchange temperature is proportional to

$$\rho = \min\{1, e^{\Delta\beta\Delta E}\}$$

where $\Delta\beta$ is the difference in inverse temperatures between replicas 1 and 2 and $\Delta E$ is the difference in their potential energies. This exchange probability is chosen to maintain the appropriate equilibrium distribution; however, the form of this exchange probability limits the temperature spacing of neighboring replicas. It is usually necessary to increase the number of replicas $N$ as the square root of the system size.[55,10] For small peptides and proteins in explicit solvent, this typically means several tens of replicas are needed to span a few hundred degrees in temperature.

## Theory

The objective of this analysis is to determine the efficiency of REMD relative to conventional MD. Relative efficiency is defined the number of *statistically independent conformations* generated by REMD using a fixed *amount of computation* divided by the number of statistically independent conformations that can be generated by standard MD with the same amount of computation. For REMD to be useful, it should have efficiency greater than or equal to 1.

The analysis will avoid as much as possible delving into the details of molecular dynamics codes. It is assumed that the amount of computation is proportional both to the amount of simulated real time and to the number of simulations. For example, the amount of computation for $N$ simulations each of length $t$ will be the same amount of computation as a single simulation of length $Nt$. The efficiency losses due to communication among processors will be ignored. These efficiency losses may be large for some methods such as spatial decomposition. Because these losses are strongly dependent upon the particular simulated system, MD program and machine, these losses will not be included in an initial analysis. The issue of efficiency lost to communication bottlenecks will be briefly addressed later.

Two types of relative efficiency will be considered. In the first measure of relative efficiency $E_1$, it is assumed that we are interested in all temperatures ranging from $T_{min}$ to $T_{max}$, the minimum and maximum temperatures of the REMD. For this comparison, both REMD and conventional MD require $N$ separate simulations—one at each temperature. The only difference is that REMD allows for temperature exchanges and conventional MD does not.

In the second measure of relative efficiency $E_2$, it is assumed that we are only interested in the behavior at a single temperature $T$. In this case, the other temperature simulations in REMD only exist to help the sampling at $T$, and we consider the conformations generated at different temperatures to have no value. A simulation at temperatures other than $T$ can provide some additional statistics about the equilibrium properties at $T$ through reweighting techniques. The additional amount of information contributed by reweighting is strongly system dependent and normally limited to nearby temperatures, so information provided by reweighting will also not be considered. The temperature $T$ at

which we are most interested is often $T_{min}$ but may be another temperature such as the melting temperature $T_m$. The relative efficiency depends upon the temperature of interest. It will be assumed throughout the rest of the manuscript that the temperature of interest is $T_{min}$.

Relative efficiency has been defined as the ratio of the number of statistically independent samples generated by REMD and MD. But, what exactly is meant by *the number of statistically independent samples*? The difficulty is that MD and REMD both produce samples that are correlated in time and not statistically independent. For most biosimulations, the longest relaxation time characterizes the folding/unfolding conformational equilibrium. Following common practice, the number of statistically independent samples $n(t)$ of a simulation of time $t$ will be defined from the asymptotic dependence of the variance in the estimated average folded/unfolded population $\bar{f}$ with time

$$\sigma^2(\bar{f}) \propto \frac{1}{n(t)} \tag{1}$$

If $f$ has a normalized autocorrelation function $C(t)$ then it can be shown[56,57] that the number of statistically independent samples $n(t)$ generated by a simulation of length $t$ is

$$n(t) = \frac{t}{\tau_{int}} \tag{2}$$

where $\tau_{int}$ is the integrated autocorrelation time. In the limit of frequent sampling and long times[56,57]

$$\tau_{int} = 2\int_0^\infty C(t')\,dt' \tag{3}$$

For a system with a single, exponential relaxation time $\tau_{relax}$, the normalized autocorrelation time has the form $C(t) = \exp[-t/\tau_{relax}]$, and

$$\tau_{int} = 2\tau_{relax} \tag{4}$$

For systems with multiple relaxation times, it is important to fully evaluate eq 3 rather than assume that eq 4 holds. Multiple relaxation times occur frequently in nonprotein systems and for protein systems with intermediate states. Also, for simulations in which Monte Carlo rather than MD is used to generate moves, the definition of $\tau_{int}$ (eq 3) should be modified to incorporate the possibility of significant decorrelation within a single sample time.[57]

To calculate efficiency, we need to determine $n(t)$ for both conventional MD and for REMD. The efficiency of REMD relative to MD is then defined to be

$$E = \frac{n_{REMD}(t)}{n_{MD}(t)} \tag{5}$$

The values of $n_{REMD}(t)$ and $n_{MD}(t)$ depend upon the type of efficiency being calculated. For efficiency $E_1$, the independent conformations generated at all the temperatures from $T_{min}$ to $T_{max}$ are included. For efficiency $E_2$ only conformations generate at $T_{min}$ are included.

Before proceeding with a detailed analysis, it is useful to consider generally how $n_{REMD}(t)$ and $n_{MD}(t)$ arise in simple, two-state protein systems. Two-state protein systems are proteins with two macrostates: an unfolded state (U) and a folded (F) state. These two macrostates are separated by a single folding transition state barrier (TS). There exists a single unimolecular rate to transition from U to F ($k_f$) and a single unimolecular rate to transition from F to U ($k_u$). The system relaxes with a single relaxation rate given by

$$k_{relax} = k_u + k_f \tag{6}$$

Real proteins have other relaxation rates; however, in the absence of intermediates, these other relaxation rates are presumed to be several orders of magnitude larger than $k_{relax}$. In this manuscript, the analysis will be restricted to these idealized two-state proteins.

The unfolding and folding rates, $k_u$ and $k_f$, are given by a standard Arrhenius-like relationships

$$\begin{aligned} k_f &= k_0 \exp\left[-\Delta\beta\Delta G_f^{\ddagger}\right] \\ k_u &= k_0 \exp\left[-\Delta\beta\Delta G_u^{\ddagger}\right] \end{aligned} \tag{7}$$

where $\Delta G_f^{\ddagger} = G_{TS} - G_u$ is the free energy of activation for folding and $\Delta G_u^{\ddagger} = G_{TS} - G_f$ is the free energy of activation for unfolding. $\beta$ is the inverse temperature, and $k_0$ is a rate prefactor assumed to have no temperature dependence. The number of statistically independent samples generated by an MD simulation of total real time $t$ can be determined from eqs 2, 4, 6, and 7.

In a REMD simulation, each replica moves both in temperature and in its conformational space. If there are $N$ replicas in a REMD simulation, then there are $N$ trajectories in temperature and conformation space. These trajectories interact because of the requirement that no two of them can occupy the same temperature simultaneously. This interaction introduces correlations between the different replicas: the motion of replica $i$ in temperature space depends upon the potential energies that the other replicas have at any moment. If the interaction between replicas could be ignored, then the analysis of a REMD simulation would be greatly simplified because it would become $N$ statistically independent simulations of the system in temperature and conformational space.

Are there conditions under which the interactions between replicas can be minimized and perhaps ignored? Let us suppose that (i) $N$ is large and (ii) the system is close to equilibrium. These two conditions guarantee that at any time the set of replicas have nearly an equilibrium distribution in enthalpy. Then the dynamics of a single replica—say replica $i$—may be determined by assuming that the other replicas are chosen from their equilibrium distribution. Under these conditions, it is reasonable to suppose that each replica moves nearly independently on the single-replica, equilibrium free energy surface in temperature and conformational space.

For an illustration of this single-replica, equilibrium surface, let us consider the one-dimensional two-state system shown in Figure 1a. This system has a single conformational coordinate $x$. The single-replica, equilibrium free energy surface is shown in Figure 1b. This equilibrium free energy surface can be determined from the relationship between free energy and probability

$$e^{-\beta G(x,\beta)} \propto p(x,\beta) \tag{8}$$

$G(x,\beta)$ is the free energy of the microstate with conformational coordinate $x$ and inverse temperature $\beta$. $p(x,\beta)$ is the

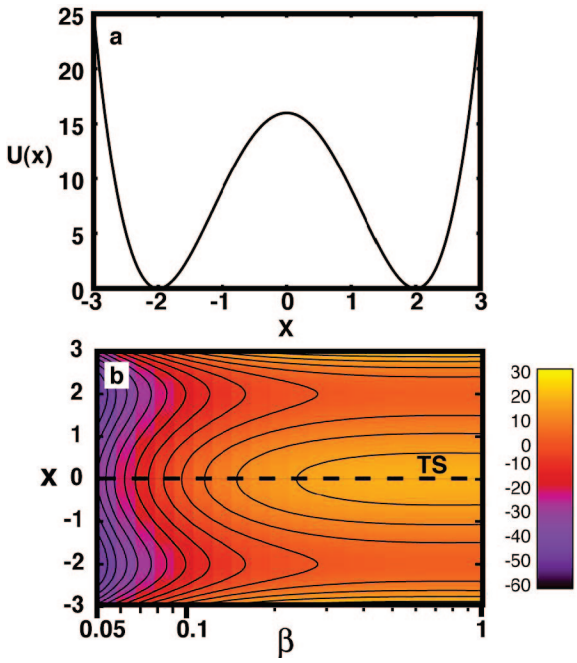


***Figure 1.*** (a) Simple one-dimensional potential of the form $U(x) = (x^2 - 4)^2$, which is a model two-state system and (b) its two-dimensional free energy contours as a function of inverse temperature $\beta$ and $x$ assuming that replicas are uniformly distributed in temperature from $T = 1$ to 200. Dimensionless units are used. Contours are drawn every 5 units of energy. This is an example of the free energy surface seen by a single replica in a REMD simulation. The relaxation rate of a replica in a REMD simulation is determined by the flux across the transition state (TS) boundary. Motion along the $\beta$ axis is normally fast compared with motion along $x$, which allows us to replace this two-dimensional free energy surface with an effective one-dimensional temperature-averaged surface along $x$ (not shown).

probability density. At a fixed inverse temperature $\beta$, the probability density $p$ at conformation $x$ is given by a Boltzmann distribution

$$p(x|\beta) = \frac{1}{\Omega_\beta} e^{-\beta H(x)} \tag{9}$$

where $H$ is the enthalpy of conformation $x$. In equilibrium, the probability distribution of an individual replica is given by

$$p(x, \beta) = p(x|\beta)p(\beta) = \frac{1}{\Omega_\beta} e^{-\beta H} p(\beta) \tag{10}$$

where $p(\beta)$ is the distribution of inverse temperatures chosen for the REMD simulation. Equations 8 and 10 lead us to a definition for the free energy of

$$e^{-\beta G(x,\beta)} = \frac{\frac{1}{\Omega_\beta} e^{-\beta H(x)} p(\beta)}{\int \frac{1}{\Omega_\beta} p(\beta)\, d\beta} \tag{11}$$

The constant in the denominator is chosen for convenience so that the integral of eq 11 over inverse temperature has a value of 1 for any state $x$ with $E = 0$. The choice of this constant does not affect our calculation of relative efficiency.

When there are many replicas near equilibrium, the number of statistically independent conformations generated by a REMD simulation should be nearly equal to $N$ statistically uncoupled replicas, where each replica has a relaxation spectrum determined by its dynamical motion on the equilibrium free energy surface $G(x, \beta)$. In other words

$$n_{\mathrm{REMD}}(t) = N\frac{t/N}{\tau_{\mathrm{int}}} = \frac{t}{\tau_{\mathrm{int}}} \tag{12}$$

because there are $N$ replicas; each replica is simulated for a time $t/N$, and each replica has the same relaxation spectrum with integrated autocorrelation time $\tau_{\mathrm{int}}$.

To determine $\tau_{\mathrm{int}}$ for two-state protein-like systems, the relaxation behavior of the system on the free energy surface $G(x, \beta)$ must be determined. Two additional simplifying assumptions will be made to facilitate this calculation. First, it will be assumed that for a two-state protein-like system, $G(x, \beta)$ will also exhibit a two-state character with a single dominant relaxation time. This is obviously true for the simple two-state system shown in Figure 1. It seems reasonable that this will be true for all two-state systems provided there remains a significant folding/unfolding barrier from $T_{\mathrm{min}}$ to $T_{\mathrm{max}}$. Second, in order to estimate that relaxation time, the free energy surface $G(x, \beta)$ will be replaced by a temperature averaged effective free energy surface $G_{\mathrm{eff}}(x)$. This replacement is justified by the fact that under most conditions at which REMD is run—especially in biological simulations—the motion in conformational space $x$ is slow compared to motion in inverse temperature $\beta$. $G_{\mathrm{eff}}(x)$ can be found by integrating eq 11 over the inverse temperature

$$e^{-G_{\mathrm{eff}}(x)} = \frac{\int \frac{1}{\Omega_\beta} e^{-\beta H(x)} p(\beta)\, d\beta}{\int \frac{1}{\Omega_\beta} p(\beta)\, d\beta} \tag{13}$$

A similar relation characterizes the effective free energy $G_{\mathrm{eff},i}$ of macrostate $i$ with free energy $G_i(\beta)$ at inverse temperature $\beta$

$$e^{-G_{\mathrm{eff},i}} = \frac{\int \frac{1}{\Omega_\beta} e^{-\beta G_i(\beta)} p(\beta)\, d\beta}{\int \frac{1}{\Omega_\beta} p(\beta)\, d\beta} \tag{14}$$

The relaxation rate (or relaxation spectrum) can be estimated from the dynamics on this temperature averaged effective free energy surface.

At this point, it is useful to consider some concrete systems in detail. The simplest two-state system has wells with identical energies and entropies separated by a high barrier with enthalpy of activation $\Delta H$. This system is essentially equivalent to that shown in Figure 1. This model is unlike most proteins, because the unfolded (U) state and folded (F) states have identical energies and enthalpies; however, this model does have characteristics similar to many smaller peptides with a discrete number of stable conformational states such as Met-enkephalin.

Let us first determine $n_{\mathrm{MD}}(t)$ for a single simulated temperature (efficiency $E_2$). From eqs 2, 4, 6, and 7, the

number of statistically independent conformations generated by a simulation of time $t$ and at inverse temperature $\beta$ is

$$n_{\mathrm{MD}}(t) = \frac{1}{2} k_{\mathrm{relax}}(\beta) t \tag{15}$$

This assumes that $\tau_{\mathrm{int}}(\beta) = 2\tau_{\mathrm{relax}}(\beta)$.

Let us now consider $n_{\mathrm{MD}}(t)$ for computing $E_1$. $n_{\mathrm{MD}}(t)$ is the number of statistically independent conformations generated at all temperatures from $T_{\mathrm{min}}$ to $T_{\mathrm{max}}$. If there are $N$ independent simulations, the total simulation time $t$ must be apportioned among the simulations at each temperature. In order to generate the same statistical error at each temperature, the time spent on each simulation should be divided unequally: more time should be spent on simulations that have a longer correlation time, and less time should be spent on simulations with a shorter correlation time. The fraction of time spent on the simulation at an inverse temperature of $\beta$ should be proportional to $\tau_{\mathrm{int}}(\beta) = 2\tau_{\mathrm{relax}}(\beta)$, so if the total simulation time is $t$, then the simulation at inverse temperature $\beta$ should ideally be run for a time

$$t(\beta) = t \frac{\tau_{\mathrm{relax}}(\beta)}{\sum_{i=1}^{N} \tau_{\mathrm{relax}}(\beta_i)} \tag{16}$$

and the number of independent conformations generated from the simulation at each $\beta$ will be

$$n(\beta) = \frac{t(\beta)}{2\tau_{\mathrm{relax}}(\beta)} = \frac{t}{2\sum_{i=1}^{N} \tau_{\mathrm{relax}}(\beta_i)} \tag{17}$$

If there are $N$ separate temperatures being simulated, the total number of conformations generated is $N$ times this amount

$$n_{\mathrm{MD}}(t) = \frac{Nt}{2\sum_{i=1}^{N} \tau_{\mathrm{relax}}(\beta_i)} = \frac{t}{2\langle \tau_{\mathrm{relax}} \rangle_{\beta}} \tag{18}$$

Let us now consider the REMD simulations. A REMD simulation consists of $N$ replicas or copies of the system. If each replica is simulated for a time $t/N$, then the total number of independent configurations generated in a REMD simulation is

$$n_{\mathrm{REMD}}(t) = \frac{t}{\tau_{\mathrm{int}}} = \frac{k_{\mathrm{REMD}} t}{2} \tag{19}$$

written in terms of the relaxation rate $k_{\mathrm{REMD}}$ of a single replica of the REMD simulation. $k_{\mathrm{REMD}}$ can be determined from the dynamics on the effective free energy surface determined using eq 14.

Each replica contributes to the statistics at each temperature, and each replica relaxes with the same rate $k_{\mathrm{REMD}}$; consequently, every temperature has the same slowest relaxation rate. This means that a REMD simulation can never be in equilibrium unless all replicas are in equilibrium together. In practice, this slow relaxation rate may not be observed in the higher-temperature replicas, because the degree to which a low-temperature replica contributes to the statistics at high temperature may be very small.

In conventional MD, simulations at one temperature do not provide any information about the behavior at another temperature (assuming reweighting methods are not considered); consequently, if one is only interested in the behavior at $T_{\mathrm{min}}$, it is inefficient to spend time simulating temperatures other than this. The data generated by these simulations is wasted. Is the same true for REMD simulations: is it less efficient to use more replicas—larger $N$? A simple argument suggests that increasing the number of replicas $N$ has a minimal impact on the efficiency of REMD. In REMD, each replica moves in conformational space and temperature space simultaneously. The distribution in this combined space is controlled by a single relaxation barrier and relaxation rate, which is the slowest relaxation rate for the whole system. The distribution at a single fixed temperature is a projection of this distribution in conformation and temperature space; consequently, the relaxation rate of this distribution must be the same (or faster) than the relaxation rate of each replica.

These above reasoning suggests that $n_{\mathrm{REMD}}(t)$ for a single temperature is the same as for all the temperatures, i.e., whether we are computing $E_1$ or $E_2$. This is equivalent to asserting that *in the limit of a long simulation* the number of replicas used for a REMD simulation is irrelevant to its efficiency. This assertion is well-recognized when all the replicas have the same temperature: dividing the simulation into $N$ separate shorter simulations generates the same amount of conformational sampling as one single simulation, provided that the system has a single dominant relaxation time.[58,59] However, this same property holds for REMD simulations because the temperature variation and conformational change are coupled: temperature relaxation is controlled by conformational change, so they have the same relaxation time.

But, what if we are not in the limit of a long simulation? Dividing the simulation in $N$ separate simulations can be a significant problem if one is unable to simulate longer than $k_{\mathrm{REMD}}$. In this case, the separate simulations are not long enough to lose memory of their initial conditions. For many biosimulations, the total simulation time is only a few multiples of the relaxation time. In these cases, REMD may be at a disadvantage; however, asymptotically the residual error should be similar whether performing $N$ simulations of length $t/N$ or one simulation of length $t$.

Putting the results together (eqs 18 and 19)

$$E_1 = \frac{t k_{\mathrm{REMD}}}{t/\langle \tau_{\mathrm{relax}} \rangle_{\beta}} = k_{\mathrm{REMD}} \langle \tau_{\mathrm{relax}} \rangle_{\beta} \tag{20}$$

The subscript relax indicates the relaxation time at a fixed inverse temperature and the subscript REMD indicates the relaxation rate of a fixed replica. The efficiency $E_2$ is given by a similar product (eqs 15 and 19)

$$E_2 = \frac{k_{\mathrm{REMD}} t}{k_{\mathrm{relax}}(\beta_{\mathrm{max}}) t} = k_{\mathrm{REMD}} \tau(\beta_{\mathrm{max}}) \tag{21}$$

To determine $k_{\mathrm{REMD}}$, the distribution of inverse temperatures in the replica simulation must be specified. For the purposes of calculation, the distribution of inverse temperatures will be assumed to be uniform from $\beta_{\mathrm{min}} = 1/T_{\mathrm{max}}$ to $\beta_{\mathrm{max}} =$

$1/T_{min}$. In other words, $p(\beta) = 1/\Delta\beta$, where $\Delta\beta = \beta_{max} - \beta_{min}$ for all values of $\beta$ from $\beta_{min}$ to $\beta_{max}$.

If the population of the transition state is negligible at all temperatures, then $\Omega_\beta = 2$. The effective free energy of each well is 0, and the effective free energy of the transition state is

$$G_{eff} = -\ln \int_{\beta_{min}}^{\beta_{max}} e^{\Delta S/R} e^{-\beta\Delta H} \frac{1}{\Delta\beta} d\beta = -\ln\left[e^{\Delta S/R} \frac{e^{-\beta_{min}\Delta H} - e^{-\beta_{max}\Delta H}}{\Delta\beta\Delta H}\right] \quad (22)$$

In the limit that $\beta_{min} = \beta_{max}$, this reverts to the expected result that

$$G_{eff} = \beta_{max}[\Delta H - T_{max}\Delta S] \quad (23)$$

Equations 6, 7, and 22 then lead to

$$k_{REMD} = 2k_0 e^{-\Delta G_{eff}} = 2k_0 \frac{e^{\Delta S/R}[e^{-\beta_{min}\Delta H} - e^{-\beta_{max}\Delta H}]}{\Delta\beta\Delta H} \quad (24)$$

In the limit that $T_{max}$ is large, the relaxation rate is asymptotically equal to

$$k_{REMD} = 2k_0 \frac{e^{-\beta_{min}\Delta G(\beta_{min})}}{\Delta\beta\Delta H} \quad (25)$$

Using the calculated value of the relaxation rate from eqs 24 and 21, we can find the relative efficiency of REMD for our toy two-well system

$$E_2 = k_{REMD}\tau_{relax}(\beta_{max}) = \left[2k_0 \frac{e^{\Delta S/R}[e^{-\beta_{min}\Delta H} - e^{-\beta_{max}\Delta H}]}{\Delta\beta\Delta H}\right]\left[\frac{1}{2k_0 e^{\Delta S/R} e^{-\beta_{max}\Delta H}}\right] = \frac{e^{\Delta\beta\Delta H} - 1}{\Delta\beta\Delta H} \quad (26)$$

which is shown for a specific choice of parameters in Figure 2. A similar calculation tells us

$$E_1 = k_{relax}\langle\tau_{relax}\rangle_\beta = \left[2k_0 \frac{e^{\Delta S/R}[e^{-\beta_{min}\Delta H} - e^{-\beta_{max}\Delta H}]}{\Delta\beta\Delta H}\right] \cdot \int_{\beta_{min}}^{\beta_{max}} \frac{1}{\Delta\beta} \frac{1}{2k_0 e^{\Delta S/R} e^{-\beta'\Delta H}} d\beta' = 2\frac{[\cosh(\Delta\beta\Delta H) - 1]}{(\Delta\beta\Delta H)^2} \quad (27)$$

which is also shown in Figure 2.

At this point, a few observations are possible. First, the efficiency is strongly system dependent. Systems with relaxation controlled by large enthalpy barriers will be greatly enhanced by REMD; systems with small enthalpy barriers will be only weakly improved by REMD. This point has previously been emphasized by others.[46,47] Second, the relative efficiency is always greater than 1 for this particular system provided that there is a positive activation barrier. REMD is always more efficient than regular MD. Third, the efficiency is solely a function of $\Delta\beta\Delta H$, which is not surprising, since these are the only two characteristic energy scales in this system.

In the limit that $\Delta\beta\Delta H$ is large, $E_1$ and $E_2$ are asymptotically equal to

$$E_1 : \frac{1}{(\Delta\beta\Delta H)^2} e^{\Delta\beta\Delta H}$$
$$E_2 : \frac{1}{\Delta\beta\Delta H} e^{\Delta\beta\Delta H} \quad (28)$$

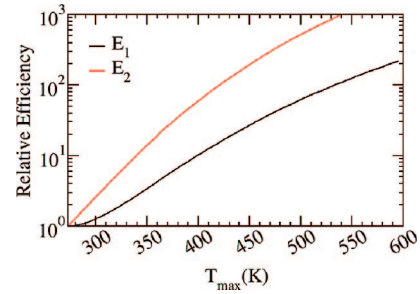


***Figure 2.*** Theoretical relative efficiencies ($E_1$ and $E_2$) of REMD for a two-state system with two wells with identical entropies and energies separated by a single barrier with an activation free energy of 10 kcal/mol. Biologically, this model is most similar to a small peptide with two stable conformational minima. Replicas are assumed to be uniformly distributed in inverse temperature with a minimum temperature of 273 K and a maximum temperature of $T_{max}$.

Naively, one might have guessed that for large activation barriers the efficiency of REMD would be equal to the ratio of relaxation rate at $T_{max}$ to the relaxation rate at $T_{min}$, that is

$$\frac{k(T_{max})}{k(T_{min})} = \frac{e^{-\beta_{min}\Delta H}}{e^{-\beta_{max}\Delta H}} = e^{\Delta\beta\Delta H} \quad (29)$$

The actual efficiency of REMD is significantly less than this for both $E_1$ and $E_2$. The factor $\Delta\beta\Delta H$ accounts for the effective fraction of replicas available to cross the transition state barrier. As the energy of the barrier increases and as the temperature spread increases, the effective fraction decreases.

A more protein-like model has two-wells (U and F) separated by a single barrier, and these wells have different entropies and energies. The same approach can be used to estimate the relative efficiency in this system. In particular, eq 14 is used to determine the effective free energy of the U, F, and TS states. These effective free energies are used with eqs 6 and 7 to determine $k_{REMD}$. This is then used in eqs 20 and 21 to determine the relative efficiencies. For more complex protein-like models, eq 14 must be evaluated numerically for each state (U, F, and TS).

It is assumed that the TS and U states have enthalpies $\Delta H_{TS}$ and $\Delta H_U$ relative to the F state and entropies $\Delta S_{TS}$ and $\Delta S_U$ relative to the F state. $\Delta S_U$ is fixed by maintaining the melting temperature at 300 K. In addition, the TS barrier is assumed to have a negligible contribution to the partition function relative to the F and U states. The principal difference between this model and real proteins is the lack of relative heat capacities among the various states. The relative efficiency $E_2$ for this protein-like model is shown in Figure 3. (Efficiency $E_1$ is not shown because it is qualitatively similar and normally of less interest.) Figure 3 shows that both the unfolding enthalpy $\Delta H_U$ and the activation enthalpy for folding $\Delta H_f^\ddagger$ can affect the relative efficiency of REMD. It can be rigorously shown that the activation entropy of folding has no effect on the relative efficiency for this model under the assumption that the contribution of the TS state to the partition function is negligible.

As in the previous model, the efficiency of REMD is always greater than conventional MD at $T_{min}$. The actual gain

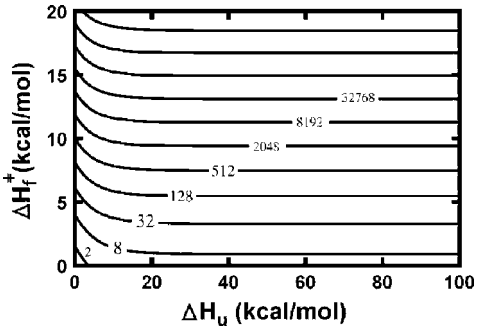


***Figure 3.*** Constant relative efficiency ($E_2$) contours for a simple, two-state, protein-like kinetic model with fixed enthalpies and entropies of activation and folding. Unlike the model of Figure 2, the two wells are not assumed to have the same energy or entropy. Contours are labeled by their relative efficiency. They are spaced logarithmically in efficiency. $\Delta H_u$ is the enthalpy of unfolding. The melting temperature is fixed at 300 K, which fixes the value of $\Delta S_u$. $\Delta H_f^{\ddagger}$ is the activation enthalpy of folding. The value of $\Delta S_f^{\ddagger}$ is chosen small enough to make the population of the transition state negligible for all values of $\Delta H_f^{\ddagger}$; the precise value of $\Delta S_f^{\ddagger}$ has no effect on the relative efficiency of REMD. Relative efficiency is computed for a REMD simulation with evenly spaced inverse temperatures spanning the temperature range from 273 to 500 K.

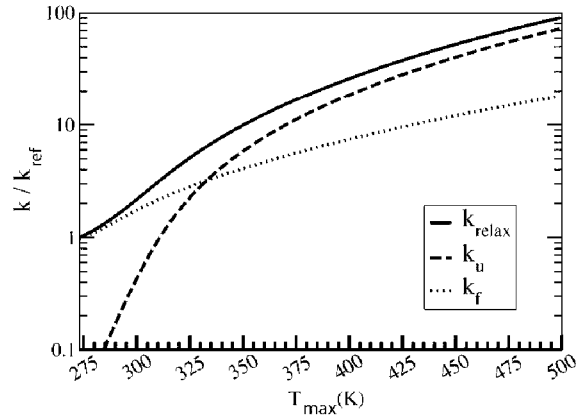


***Figure 4.*** Variation in the unfolding rate ($k_u$), the folding rate ($k_f$), and the relaxation rate ($k_{relax}$) of the replicas in a REMD simulation as a function of the maximum temperature of the simulation. Rates are shown relative to the relaxation rate at 273 K. The model used is the two-state protein-like model with no heat capacities, an unfolding enthalpy of $\Delta H_u = 20$ kcal/mol, and an activation enthalpy for folding of $\Delta H_f^{\ddagger} = 5$ kcal/mol. The entropy of the unfolded state is determined so that the melting temperature is fixed at 300 K.

is a strong function of the activation energy for folding. For even modest barriers (~5 kcal/mol) for folding, the rate enhancement of REMD can be significant. Interestingly, even in the absence of an activation enthalpy for folding, REMD can be more efficient than conventional MD because elevating the temperature can reduce the unfolding time, which contributes to the overall relaxation rate.

To gain some more insight into the relative contributions of $k_f$ and $k_u$ to the total replica relaxation rate $k_{relax}$ ($k_{REMD}$), the dependence of these rates on $T_{max}$ for one choice of model parameters ($\Delta H_u = 20$ kcal/mol; $\Delta H_f^{\ddagger} = 5$ kcal/mol) is shown in Figure 4. Because the activation enthalpy for unfolding (25 kcal/mol) is much larger than the activation enthalpy for folding (5 kcal/mol), the unfolding rate increase more rapidly with $T_{max}$ than the folding rate. At low temperatures the relaxation rate is controlled by the folding rate, so the primary effect of a REMD simulation when $T_{max}$ is low is to increase the folding rate; however, when $T_{max}$ is large, the primary effect is to increase the unfolding rate. The crossover between these two effects occurs in this model about 25 K above the melting temperature (300 K).

An important property of real proteins not included in this model is the existence of relative heat capacity differences $\Delta C_U$ and $\Delta C_{TS}$ of the U and TS states relative to the F state. These heat capacities can strongly affect the behavior because they work to reduce the activation enthalpy for folding at high temperatures. At a high enough temperature, the activation enthalpy for protein folding usually becomes negative. In this regime, the folding rate is determined solely by an entropic search for the native state. This does not correspond to a conventional Arrhenius picture of diffusion controlled by an enthalpic barrier, although it is described by a similar "Arrhenius-like" rate equation. The existence of a negative activation enthalpy strongly affects the efficiency of REMD simulation.

***Table 1.*** Measured Enthalpy, Entropy, and Heat Capacity of the Transition State and Unfolded State of Protein L Relative to the Native State at 295 K[a]

| | $\Delta H$ (kcal/mol) | $\Delta S$ (kcal/(mol K)) | $\Delta C_p$ (kcal/(mol K)) |
|---|---|---|---|
| transition state (TS) | 26.8 | 0.03 | 0.19 |
| unfolded state (U) | 20.1 | 0.05 | 0.77 |

[a] These values are used to estimate the relative efficiency $E_2$ of REMD (Figure 5).

To illustrate the effects of a nonzero heat capacity, let us consider an actual protein. Protein L is a commonly used model system for protein for folding studies. The thermodynamic properties of protein L have been measured[60] and are shown in Table 1. Following the same procedure as before, numerical integration is used to determine the relative efficiency $E_2$ as a function of the maximum temperature $T_{max}$ of the replicas. This is shown in Figure 5. It is assumed that inverse temperatures are uniformly distributed from the minimum to maximum inverse temperature.

Unlike simpler models for which relative efficiency is always greater for a larger $T_{max}$, protein L shows a maximum efficiency for an intermediate $T_{max}$. This maximum efficiency occurs slightly above the temperature at which the activation enthalpy for folding vanishes, which appears to be a good rule of thumb for choosing a value of $T_{max}$. At temperatures for which the activation enthalpy is negative, increasing temperature would be expected to be counterproductive, slowing the search rate; in contrast, for any positive activation enthalpy, increasing the temperature should make barrier crossing more rapid. The same behavior is exhibited by the F13W/G29A mutant of the B domain of staphylococcal protein A (protein A) when modeled from its thermodynamic data[61] (data not shown). The relative efficiency $E_2$ for a REMD simulation of protein A starts at a value of 1, peaks at 6.4 when $T_{max}$ is 326 K, and declines as $T_{max}$ is increased further. The peak efficiency occurs close to

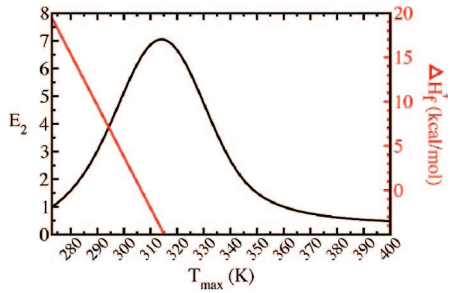


***Figure 5.*** Relative efficiency $E_2$ of a REMD simulation of protein L with inverse temperatures uniformly distributed from 1/273 K to $1/T_{max}$. Although REMD can produce significant increases in efficiency, the amount of rate enhancement varies strongly with the maximum temperature. In particular, a maximum temperature that is too high can actually result in a decrease in the relaxation rate because of the existence of heat capacities that reduce the activation energy of folding. The maximum relative efficiency of REMD occurs when $T_{max}$ is chosen slightly above the temperature at which the activation free energy of folding disappears. This appears to be a good rule of thumb for choosing $T_{max}$ of a REMD simulation.

and slightly above the temperature, 318 K, at which the activation enthalpy vanishes.

The strong dependence of efficiency on the choice of $T_{max}$ and on the enthalpy of folding is the most plausible explanation for the large differences between the estimated efficiency of REMD simulations. Close attention should be paid to the type of problem to which one is applying REMD as well as the choice of $T_{max}$. Estimation of the optimal $T_{max}$ can be made from experimental measurements of the thermodynamic properties of the folded, unfolded, and transition states; however, one should be cautious that these values are approximately reproduced using the molecular dynamics force field. These results also suggest that the performance advantages of REMD on systems with explicit solvent may be very different from the performance advantages with implicit solvent, because, in many implicit solvent models, entropic nonpolar solvation effects are represented as enthalpic contributions to the free energy leading to different activation enthalpies.

The folding and unfolding contributions to the relaxation rate can be determined for the REMD simulation of protein L (Figure 6). Like the simpler protein model without relative heat capacities, the primary rate enhancement of REMD arises at lower values of $T_{max}$ from increases in the folding rate. However, unlike the simpler model, enhancement of the unfolding rate makes a relatively minor contribution to the overall rate enhancement of REMD.

The analysis that has been presented is predicated upon the fact that motions in temperature are fast relative to conformational changes. Many biomolecular simulations have been run in a suboptimal manner in which temperature exchanges are relatively infrequent.[54] An analysis similar to the one presented here may be applicable; however, in this suboptimal regime one must examine the relaxation behavior on the full $G(x, \beta)$ surface. Additionally, the existence of additional correlations between replicas under very high

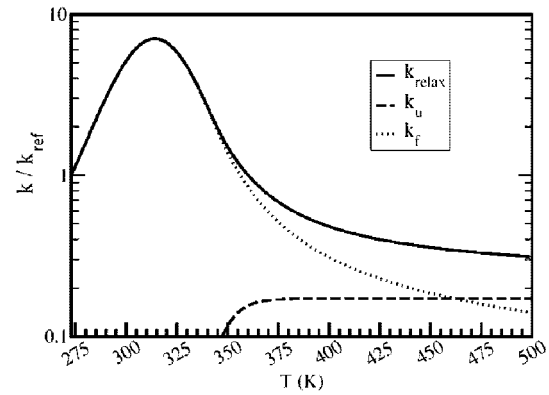


***Figure 6.*** Variation in the unfolding rate ($k_u$), the folding rate ($k_f$), and the relaxation rate ($k_{relax}$) of the replicas in a REMD simulation of protein L as a function of the maximum temperature of the simulation. Rates are shown relative to the relaxation rate at 273 K.

exchange rates that might lead to slowing cannot be definitively ruled out.

Up to this point, the issue of additional computational losses from spatial decomposition in conventional MD has not been addressed. In essence, the results so far have been a comparison of REMD with MD on a single processor. When calculations are run on $N$ processors simultaneously, conventional MD incurs additional losses due to communication bottlenecks that REMD will not. For conventional MD, the actual speedup with a conventional decomposition approach will not normally be $N$ but $\gamma N$, where $\gamma$ is equal to the ratio of actual speedup to theoretically possible speedup. For this reason, even if the theoretical efficiency of REMD as we have calculated it is no better than conventional MD, the actual efficiency of REMD on $N$ processors is often greater than conventional MD by a factor of $1/\gamma$.

## Summary and Conclusion

In summary, an analytic approach has been developed for analyzing REMD simulations. This analysis is applicable for conditions under which individual replicas are only weakly interacting. Under these conditions, a REMD simulation can be analyzed as a collection of $N$ noninteracting simulations moving in both conformational space and in temperature. The dynamics in conformation and temperature for a single replica is determined by the structure of the equilibrium free energy surface $G(x, \beta)$. Beginning from this surface (defined via eq 11), one can in principle determine the relaxation spectrum and the integrated autocorrelation time, which can be used to determine the relative efficiency of REMD.

To simplify calculations, the timescale separation between temperature motion and conformational change can be used to reduce the equilibrium free energy surface $G(x,\beta)$ to an effective free energy surface $G_{eff}(x)$ by integrating out the temperature degrees of freedom. Applying this approach to a number of two-state protein models demonstrates that the relative efficiency of REMD is a strong function of the activation enthalpy; however, REMD is always more efficient for these models than conventional, constant temperature MD provided that the activation enthalpies for folding and unfolding are positive

throughout the temperature range. Analysis of models with thermodynamic parameters of real proteins—protein L and the B domain of protein A—indicates that the peak REMD efficiency occurs for a value of $T_{max}$ slightly above the temperature at which the activation enthalpy for folding vanishes. If $T_{max}$ is increased beyond this, the relative efficiency of REMD begins to drop. Eventually, the relative efficiency of REMD can become negative, making this a critically important parameter choice for a REMD simulation.

Assuming that protein L and the B domain of protein A are representative of the small proteins for which REMD is used, REMD appears to be significantly more efficient than conventional MD. For the optimal choice of $T_{max}$ REMD is respectively 7 and 6.4 times more efficient than conventional MD for these proteins. This does not include efficiency losses due to parallelization, which in practical circumstances strongly favor REMD simulations. Additionally, REMD has the benefit of providing accurate temperature dependent statistics to determine the relative enthalpies of different conformational states.

Because the dynamics of replicas are controlled by a single transition state barrier on the $G(x, \beta)$ surface, the asymptotic efficiency of REMD is nearly independent of the number of replicas. However, because biological simulations are usually run for only a few multiple of the relaxation rate, dividing the simulation time among the multiple replicas can be detrimental.

The analysis presented in this manuscript is solely for two-state proteins. A similar analysis can be carried out for proteins with intermediates or nonprotein systems. For this analysis, it is important to carefully determine the integrated autocorrelation time, because a single relaxation time is no longer dominant. Also, when motion in temperature is not fast compared to conformational change, the full $G(x, \beta)$ surface should be used for analysis rather than the effective surface $G_{eff}(x)$. Replica methods involving parameters other than temperature can also be analyzed in a similar manner.

**Acknowledgment.** The author gives many thanks to Bernd Berg for careful, critical readings of the initial manuscript. Additional thanks go to Huan-Xiang Zhou for comments and encouragement.

## References

(1) Swendsen, R. H.; Wang, J. S. Replica Monte-Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.

(2) Marinari, E.; Parisi, G. Simulated Tempering - a New Monte-Carlo Scheme. *Europhys. Lett.* **1992**, *19*, 451–458.

(3) Geyer, C. J.; Thompson, E. A. Annealing Markov-Chain Monte-Carlo with Applications to Ancestral Inference. *J. Am. Stat. Assoc.* **1995**, *90*, 909–920.

(4) Hukushima, K.; Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.

(5) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(6) Garcia, A. E.; Sanbonmatsu, K. Y. Exploring the energy landscape of a beta hairpin in explicit solvent. *Proteins Struct. Funct. Genet.* **2001**, *42*, 345–354.

(7) Sanbonmatsu, K. Y.; Garcia, A. E. Structure of Met-Enkephalin in explicit solvent using replica molecular dynamics. *Biophys. J.* **2001**, *80*, 399A–399A.

(8) Zhou, R.; Berne, B. J. Can a continuum solvent model reproduce the free energy landscape of a beta -hairpin folding in water. *Proc. Natl Acad. Sci. U.S.A.* **2002**, *99*, 12777–82.

(9) Garcia, A. E.; Onuchic, J. N. Folding a protein in a computer: An atomic description of the folding/unfolding of protein A. *Proc. Natl Acad. Sci. U.S.A.* **2003**, *100*, 13898–13903.

(10) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. Peptide folding simulations. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168–174.

(11) Nymeyer, H.; Garcia, A. E. Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized born approximation with explicit solvent. *Proc. Natl Acad. Sci. U.S.A.* **2003**, *100*, 13934–13939.

(12) Pitera, J. W.; Swope, W. Understanding folding and design: Replica-exchange simulations of "Trp-cage" fly miniproteins. *Proc. Natl Acad. Sci. U.S.A.* **2003**, *100*, 7587–7592.

(13) Rhee, Y. M.; Pande, V. S. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.* **2003**, *84*, 775–86.

(14) Garcia, A. E.; Onuchic, J. N. All atom simulation of the folding/unfolding thermodynamics of protein A. *Biophys. J.* **2004**, *86*, 345A–345A.

(15) Jas, G. S.; Kuczera, K. Equilibrium structure and folding of a helix-forming peptide: Circular dichroism measurements and replica-exchange molecular dynamics simulations. *Biophys. J.* **2004**, *87*, 3786–3798.

(16) Zhou, R. H. Exploring the protein folding free energy landscape: coupling replica exchange method with P3ME/RESPA algorithm. *J. Mol. Graphics Model.* **2004**, *22*, 451–463.

(17) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. Protein folding pathways from replica exchange simulations and a kinetic network model. *Proc. Natl Acad. Sci. U.S.A.* **2005**, *102*, 6801–6806.

(18) Best, R. B.; Clarke, J.; Karplus, M. What contributions to protein side-chain dynamics are probed by NMR experiments? A molecular dynamics simulation analysis. *J. Mol. Biol.* **2005**, *349*, 185–203.

(19) Chen, J.; Won, H. S.; Im, W.; Dyson, H. J.; Brooks, C. L., 3rd. Generation of native-like protein structures from limited NMR data, modern force fields and advanced conformational sampling. *J. Biomol. NMR* **2005**, *31*, 59–64.

(20) Griffin, M. A.; Friedel, M.; Shea, J. E. Effects of frustration, confinement, and surface interactions on the dimerization of an off-lattice beta-barrel protein. *J. Chem. Phys.* **2005**, *123*, 174707.

(21) Im, W.; Brooks, C. L., 3rd. Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proc. Natl Acad. Sci. U.S.A.* **2005**, *102*, 6771–6.

(22) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–54.

(23) Malolepsza, E.; Boniecki, M.; Kolinski, A.; Piela, L. Theoretical model of prion propagation: a misfolded protein induces misfolding. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7835–40.

(24) Nanias, M.; Chinchio, M.; Oldziej, S.; Czaplewski, C.; Scheraga, H. A. Protein structure prediction with the UNRES force-field using Replica-Exchange Monte Carlo-with-Minimization; Comparison with MCM, CSA, and CFMC. *J. Comput. Chem.* **2005**, *26*, 1472–86.

(25) Nguyen, P. H.; Mu, Y.; Stock, G. Structure and energy landscape of a photoswitchable peptide: a replica exchange molecular dynamics study. *Proteins Struct. Funct. Genet.* **2005**, *60*, 485–94.

(26) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C. K.; Li, M. S. Free energy landscape and folding mechanism of a beta-hairpin in explicit water: a replica exchange molecular dynamics study. *Proteins Struct. Funct. Genet.* **2005**, *61*, 795–808.

(27) Nishino, M.; Sugita, Y.; Yoda, T.; Okamoto, Y. Structures of a peptide fragment of beta2-microglobulin studied by replica-exchange molecular dynamics simulations - towards the understanding of the mechanism of amyloid formation. *FEBS Lett.* **2005**, *579*, 5425–9.

(28) Nymeyer, H.; Woolf, T. B.; Garcia, A. E. Folding is not required for bilayer insertion: replica exchange simulations of an α-helical peptide with an explicit lipid bilayer. *Proteins Struct. Funct. Genet.* **2005**, *59*, 783–90.

(29) Paschek, D.; Gnanakaran, S.; Garcia, A. E. Simulations of the pressure and temperature unfolding of an α-helical peptide. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6765–70.

(30) Pokarowski, P.; Droste, K.; Kolinski, A. A minimal proteinlike lattice model: an α helix motif. *J. Chem. Phys.* **2005**, *122*, 214915.

(31) Roe, D. R.; Hornak, V.; Simmerling, C. Folding cooperativity in a three-stranded beta-sheet model. *J. Mol. Biol.* **2005**, *352*, 370–81.

(32) Sugita, Y.; Okamoto, Y. Molecular mechanism for stabilizing a short helical peptide studied by generalized-ensemble simulations with explicit solvent. *Biophys. J.* **2005**, *88*, 3180–90.

(33) Vreede, J.; Crielaard, W.; Hellingwerf, K. J.; Bolhuis, P. G. Predicting the signaling state of photoactive yellow protein. *Biophys. J.* **2005**, *88*, 3525–35.

(34) Chen, J.; Im, W.; Brooks, C. L., 3rd. Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field. *J. Am. Chem. Soc.* **2006**, *128*, 3728–36.

(35) Ho, B. K.; Dill, K. A. Folding very short peptides using molecular dynamics. *PLoS Comput. Biol.* **2006**, *2*, e27.

(36) Jang, S.; Kim, E.; Pak, Y. Free energy surfaces of miniproteins with a betabetaalpha motif: replica exchange molecular dynamics simulation with an implicit solvation model. *Proteins Struct. Funct. Genet.* **2006**, *62*, 663–71.

(37) Jang, S.; Shin, S. Amyloid beta-peptide oligomerization in silico: dimer and trimer. *J. Phys. Chem. B* **2006**, *110*, 1955–8.

(38) Mu, Y.; Nordenskiold, L.; Tam, J. P. Folding, misfolding, and amyloid protofibril formation of WW domain FBP28. *Biophys. J.* **2006**, *90*, 3983–92.

(39) Ohkubo, Y. Z.; Thorpe, I. F. Evaluating the conformational entropy of macromolecules using an energy decomposition approach. *J. Chem. Phys.* **2006**, *124*, 024910.

(40) Ravindranathan, K. P.; Gallicchio, E.; Friesner, R. A.; McDermott, A. E.; Levy, R. M. Conformational equilibrium of cytochrome P450 $\mu_B$-3 complexed with N-palmitoylglycine: a replica exchange molecular dynamics study. *J. Am. Chem. Soc.* **2006**, *128*, 5786–91.

(41) Soto, P.; Baumketner, A.; Shea, J. E. Aggregation of polyalanine in a hydrophobic environment. *J. Chem. Phys.* **2006**, *124*, 134904.

(42) Trebst, S.; Troyer, M.; Hansmann, U. H. Optimized parallel tempering simulations of proteins. *J. Chem. Phys.* **2006**, *124*, 174903.

(43) Zhang, J.; Qin, M.; Wang, W. Folding mechanism of beta-hairpins studied by replica exchange molecular simulations. *Proteins Struct. Funct. Genet.* **2006**, *62*, 672–85.

(44) Paschek, D.; Nymeyer, H.; Garcia, A. E. Replica exchange simulation of reversible folding/unfolding of the Trp-cage miniprotein in explicit solvent: On the structure and possible role of internal water. *J. Struct. Biol.* **2007**, *157*, 524–533.

(45) Garcia, A. E.; Paschek, D. Simulation of the pressure and temperature folding/unfolding equilibrium of a small RNA hairpin. *J. Am. Chem. Soc.* **2008**, *130*, 815–7.

(46) Zuckerman, D. M.; Lyman, E. A second look at canonical sampling of biomolecules using replica exchange simulation. *J. Chem. Theory Comput.* **2006**, *2*, 1200–1202.

(47) Zuckerman, D. M.; Lyman, E. A second look at canonical sampling of biomolecules using replica exchange simulation. *J. Chem. Theory Comput.* **2006**, *2*, 1200.

(48) Periole, X.; Mark, A. E. Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent. *J. Chem. Phys.* **2007**, *126*.

(49) Zhang, W.; Wu, C.; Duan, Y. Convergence of replica exchange molecular dynamics. *J. Chem. Phys.* **2005**, *123*, 1–9.

(50) Rao, F.; Caflisch, A. Replica exchange molecular dynamics simulations of reversible folding. *J. Chem. Phys.* **2003**, *119*, 4035–4042.

(51) Seibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *J. Mol. Biol.* **2005**, *354*, 173–183.

(52) Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. Simulating replica exchange simulation of protein folding with a kinetic network model. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15340–15345.

(53) Trebst, S.; Troyer, M.; Hansmann, U. H. E. Optimized parallel tempering simulations of proteins. *J. Chem. Phys.* **2006**, *124*, –.

(54) Sindhikara, D.; Meng, Y.; Roitberg, A. E. Exchange Frequency in Replica Exchange Molecular Dynamics. *J. Chem. Phys.* **2008**, *128*, 024103.

(55) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116*, 9058–67.

(56) Flyvbjerg, H.; Petersen, H. G. Error-Estimates on Averages of Correlated Data. *J. Chem. Phys.* **1989**, *91*, 461–466.

(57) Berg, B. A. *Markov chain Monte Carlo simulations and their statistical analysis: with web-based Fortran code*; World Scientific: Hackensack, NJ, 2004; pp 196–235..

(58) Voter, A. F. Parallel replica method for dynamics of infrequent events. *Phys. Rev. B* **1998**, *57*, R13985–R13988.

(59) Shirts, M. R.; Vijay, S. P. Mathematical Analysis of Coupled Parallel Simulations. *Phys. Rev. Lett.* **2001**, *86*, 4983–4987.

(60) Scalley, M. L.; Baker, D. Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10636–40.

(61) Dimitriadis, G.; Drysdale, A.; Myers, J. K.; Arora, P.; Radford, S. E.; Oas, T. G.; Smith, D. A. Microsecond folding dynamics of the F13W G29A mutant of the B domain of staphylococcal protein A by laser-induced temperature jump. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3809–3814.

CT7003337

JCTC Journal of Chemical Theory and Computation

# Theoretical Investigation of the Geometries and UV−vis Spectra of Poly(L-glutamic acid) Featuring a Photochromic Azobenzene Side Chain

Pierre-François Loos,*[,†] Julien Preat,[‡,§] Adèle D. Laurent,[‡] Catherine Michaux,[‡,⊥] Denis Jacquemin,[‡,||] Eric A. Perpète,[‡,||,¶] and Xavier Assfeld[†]

*Equipe de Chimie et Biochimie Théoriques, UMR 7565 CNRS-UHP, Institut Jean Barriol (FR CNRS 2843), Faculté des Sciences et Techniques, Nancy-Université, B.P. 239,54506 Vandoeuvre-les-Nancy Cedex, France, and Groupe de Chimie Physique Théorique et Structurale, Facultés Universitaires Notre-Dame de la Paix, rue de Bruxelles, 61, B-5000 Namur, Belgium*

Received August 20, 2007

**Abstract:** The geometries and UV−vis spectra of azobenzene dyes grafted as a side chain on poly(L-glutamic acid) have been investigated using a combination of quantum mechanics/molecular mechanics (QM/MM) and time-dependent density functional theory (TD-DFT) methods at the TD-PBE0/6-311+G(d,p)//B3LYP/6-311G(d,p):Amber ff99 level of theory. The influence of the secondary structure of the polypeptide on the electronic properties of both the trans and cis conformations of azobenzene dyes has been studied. It turns out that the grafted dyes exhibit a red-shift of the $\pi \rightarrow \pi^*$ absorption energies mainly due to the auxochromic shift induced by the peptidic group used to link the chromophoric unit to the polypeptide and that specific interactions between the glutamic side chain and the azobenzene moiety lead to a large blue-shift of the $n \rightarrow \pi^*$ transition.

## 1. Introduction

During the two last decades, the interest in azobenzene (AB) derivatives has become increasingly important due to their wide range of industrial applications. In particular, AB based molecules represent 60–70% of the world production of industrial "absorption" dyes[1] and have been recently found to be promising materials for media storage devices[2–5] and molecular motors.[6,7] These processes exploit the photochromic abilities of the AB dyes that are involved in a reversible photoisomerization from the trans (TAB) to the cis (CAB) isomer (Figure 1). In biological systems, the photoreversible isomerization of a molecule attached to a macromolecular system, such as the visual pigment rhodopsin,[8–14,39] induces conformational changes that in turn lead to a physiological response of the protein. Following this idea, Pieroni et al. have prepared poly(L-glutamic acid) with photochromic AB side chains[15–17] (Figure 2), that can exist in disordered forms (random coil) or in regularly folded structures (like α-helix or $\beta$-sheets), similar to biological systems. The photochromic polymers were prepared from high molecular weight poly(L-glutamic acid) ($\bar{M}_v = 200\,000$) and samples containing

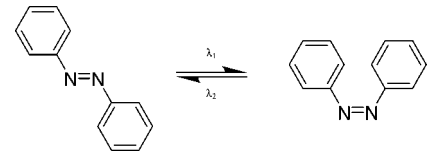


***Figure 1.*** Reversible photochromic isomerization of the azobenzene derivative.

* Corresponding author. E-mail: Pierre-Francois.Loos@cbt.uhp-nancy.fr.

† Nancy-Université.

‡ Facultés Universitaires Notre-Dame de la Paix.

§ Fellow of the Belgian Fund for the Formation to Research in Industry and Agriculture (FRIA).

⊥ Postdoctoral researcher of the Belgian National Fund for Scientific Research.

|| Research associate of the Belgian National Fund for Scientific Research.

¶ Invited professor at the Université Henri Poincaré (Nancy, France) from January to May 2007.

10.1021/ct700188w CCC: $40.75 
Published on Web 03/07/2008

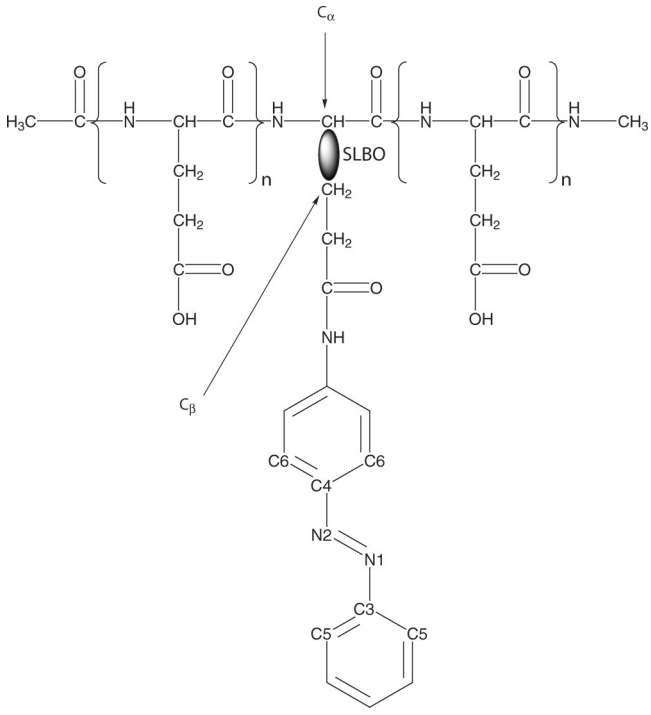


***Figure 2.*** QM/MM partitioning of the poly(L-glutamic acid) with the photochromic AB side chain. The QM/MM boundaries are located between the $C_\alpha$ and the $C_\beta$ atoms.

13–56% mol of azo groups were studied at different pH values. It has been shown that irradiation at 350 nm produces an isomerization from the trans to the cis isomer ($\pi \rightarrow \pi^*$ transition), whereas the reverse reaction is obtained using a 450 nm irradiation ($n \rightarrow \pi^*$ transition), or via dark adaptation.[15]

In water, this reversible process is accompagnied by large photoinduced structural changes of the polypeptide secondary structure, which is detected by circular dichroism (CD) spectroscopy, as well as a drastic modification of the absorption spectrum of the AB dye. Depending on the experimental conditions (pH and mole percent of azo group), the polypeptide presents a random coil, an α-helix, or a $\beta$-structure CD pattern with respect to the relative amount of each structure.[18] In acid pHs (4.7−5.0), the dark-adapted samples containing 16 and 21 mol % of azo groups indicate the presence of an appreciable amount of α-helix, while the 36 mol % samples exhibit the CD curve of $\beta$-structure. When the pH is increased to alkaline values (pH 8), all polypeptides undergo a conformational transition to random coil structures. In water, light produces the isomerization of the azo side chains and a remarkable effect on the CD bands, which is also influenced by the pH value and/or the azo content. For example, the 16 mol % sample does not exhibit any variation of the CD spectrum, while the 36 mol % azo-polypeptide undergoes a $\beta$-coil transition at pH 6.5. For the 21% azopolypeptide, the helix content is increased by irradiation below pH 6.3.

Prior to study the effect of experimental conditions (pH and azo content) and the dynamical behavior leading to the modification of the polypeptide conformation, this work aims at investigating the effects of the polypeptide structure on the geometries and the absorption spectra of the two AB isomers (i.e., TAB and CAB).

From our point of view, to obtain accurate UV−vis spectra of macromolecular systems, one needs to take up two challenges:

• Due to the size of the system, the modeling of macromolecules with pure quantum mechanics is still out of reach for modern computational resources.

• To obtain accurate UV−vis spectra, one has to describe precisely the excited state(s) of the chromophoric unit.

Some propositions have been made to solve these challenges separately. For example, the size problem can be overcome with the hybrid quantum mechanics/molecular mechanics methods (QM/MM) which are available to treat such large systems.[19–29] For the second challenge, time-dependent density functional theory (TD-DFT) calculations[30] indeed yield accurate determination of the absorption energies associated with these excited states for a wide range of chromophoric units[31–36] and, especially, for the AB derivatives.[37,38,43]

One should note that only a limited number of studies combining QM/MM and TD-DFT methods have been previously published, such as the meticulous studies of Elstner et al.[13,14] and Vreven et al.,[39] who compare several computational approaches for the determination of the absorption shifts in retinal proteins. Rothlisberger and co-workers also report TD-DFT/MM calculations[40] on solvated acetone[41] and aminocoumarins[42] based on Car–Parrinello simulations. Altought Elstner et al. have shown that TD-DFT calculations fail dramatically in the case of protonated Schiff base chromophores and neutral polyenes[13,14] due to the local approximation of the exchange-correlation functional,[44] we would like to mention the successful work of Crecca and Roitberg on the isomerization mechanism of azobenzene and disubstitued azobenzene derivatives.[43] Using DFT and TD-DFT calculations, they have studied the isomerization pathway of several azobenzene derivatives. The different barriers of the potential energy surface have been found in good agreement with experimental results.

In this article, we report the calculation of UV−vis spectra of TAB and CAB units for various stable secondary and supersecondary structures (motifs) of the poly(L-glutamic acid). First, we investigate the geometries and the UV−vis spectra of both isomers in gas phase and in ethanol. The results are compared to experimental data to validate the theoretical scheme and to obtain some reference data in order to point out the specific effects of the polypeptide surroundings on the geometry and absorption wavelengths of the AB moiety. In particular, we investigate different polypeptide structures containing α-helix and $\beta$-sheet conformations, as well as other supersecondary structures like $\beta$-hairpin, $\beta$-strand, and $\beta-\alpha-\beta$ and $\alpha-\beta-\alpha$ motifs.

## 2. Methodology

The QM/MM calculations are performed by means of the local self-consistent field (LSCF) method developed by Rivail, Assfeld, and co-workers,[47–56] that is implemented in a modified version of the Gaussian 03 package[57] linked to

***Table 1.*** Structures and UV−vis Spectra of TAB and CAB in Various Media, Obtained at the TD-PBE0/6-311+G(d,p)// B3LYP/6–311G(d,p) Level of Theory[a]

| | medium | method | *d*(N=N) | *d*(C−N) | ∠(C−N=N) | ∠(C−C−N) | τ(C−C−N=N) | $\lambda^{n\rightarrow\pi^*}$ | $\lambda^{\pi\rightarrow\pi^*}$ | ref |
|---|---|---|---|---|---|---|---|---|---|---|
| TAB | gas phase | TD-DFT//DFT | 1.253 | 1.418 | 115.1 | 124.6, 115.5 | 0.0, 180.0 | 480(0.00) | 327(0.78) | 37 |
| | | X-ray 1997 (82 K) | 1.259 | 1.431 | 114.1 | | 21.0 | | | 71 |
| | | GED 2001 (407 K) | 1.260 | 1.428 | 113.7 | 124.8 | 0.0, 180.0 | | | 72 |
| | | exp 1981 | | | | | | 444 | 303 | 73 |
| | | exp 1982 | | | | | | 440(380) | 301(21300) | 69 |
| | ethanol | TD-DFT//DFT | 1.255 | 1.419 | 115.5 | 124.7, 115.5 | 0.0, 180.0 | 477(0.00) | 344(0.90) | this work |
| | | exp 1953 | | | | | | 443(510) | 320(21300) | 74 |
| | | exp 1961 | | | | | | 437(510) | 320(17300) | 75 |
| CAB | gas phase | TD-DFT//DFT | 1.243 | 1.436 | 124.1 | 122.9, 116.5 | 51.0, −138.1 | 478(0.03) | 292(0.08) | this work |
| | | X-ray 1971 | 1.253 | 1.449 | 121.9 | 122.5, 117.3 | 53.3 | | | 67 |
| | | exp 1982 | | | | | | 425 | 265 | 69 |
| | ethanol | TD-DFT//DFT | 1.246 | 1.436 | 124.2 | 122.7, 116.5 | 50.8, −138.0 | 467(0.05) | 305(0.09) | this work |
| | | exp 1953 | | | | | | 433(1518) | 281(5260) | 74 |
| | | exp 1973 | | | | | | 443(1514) | 281(5248) | 76 |
| TAB ⇌ CAB | gas phase | TD-DFT//DFT | −0.010 | 0.018 | 9.0 | −1.7, 1.0 | 51.0, 41.9 | −2 | −35 | this work |
| | | X-ray | −0.006 | 0.018 | 7.8 | | 32.3 | | | 67, 71 |
| | | exp | | | | | | −19 → −15 | −38 → −35 | 69, 73 |
| | ethanol | TD-DFT//DFT | −0.009 | 0.017 | 8.7 | −2.0, 1.0 | 50,8, 42.0 | −10 | −39 | this work |
| | | exp | | | | | | −10 → +6 | −39 | 74−76 |

[a] The changes in structural parameters and UV−vis spectra following to the reversible photoisomerization of both isomers are also reported. Distances are in angstroms, while the valence and the dihedral angles are in degrees. $\lambda^{n\rightarrow\pi^*}$ and $\lambda^{\pi\rightarrow\pi^*}$ are the wavelengths (in nanometers) of the first $n\rightarrow\pi^*$ and $\pi\rightarrow\pi^*$ transitions, respectively. The oscillator strengths are given in parenthesis.

the Tinker software[58] for the MM calculations. In the LSCF framework, a doubly occupied strictly localized bond orbital (SLBO) is employed to link the QM and the MM parts. The SLBO is obtained from a preliminary computation on a model molecule featuring the chemical bond of interest.[59] In the present case, the QM/MM frontier is located between the $C_\alpha$ and the $C_\beta$ atoms of the residue where the azobenzene dye is grafted (Figure 2).

The MM surrounding is described with the Amber ff99 force field,[60,61] and we have considered the protonated form of the glutamic acid in order to avoid the spurious polarization due to negative charges on the glutamate anions. Then, we choose to mimic the effect of the counterions by the protonated state of the glutamic side chain. This strategy has been employed in several studies for biological systems[45,46] to mimic the situation in which the counterion is tightly bounded to the negatively charged moiety. The classical charge of the $C_\alpha$ frontier atom has been set to 0.0365 e instead of 0.0145 e to ensure the overall electroneutrality of the MM part. The van der Waals parameters for the QM atoms are set to the values defined for the corresponding atom type of the force field. Moreover, the N-terminus and C-terminus are capped with an acetyl and a *N*-methylamide group, respectively.

Concerning the QM calculations, the B3LYP functional[62] combined to the 6-311G(d,p) basis set has been used for geometry optimization of the chromophoric unit. After a full geometry optimization of the entire system, TD-DFT calculations have been performed to evaluate the UV−vis spectrum with the fitted-parameter-free PBE0 functional[63,64] and the 6-311+G(d,p) basis set. Previous studies have shown that, for the largest majority of organic dyes,[31–38] this theoretical scheme provides reliable results for geometrical parameters, as well as for the UV−vis spectra. To take into account solvent effects on the model systems, the polarizable continuum model (PCM) with UAKS atomic radii has been used.[65] For TD-DFT calculations using SCRF solvation model, the nonequilibrium PCM method was selected.[66]

## 3. Results

**3.1. TAB and CAB in the Gas Phase and in Ethanol.** Table 1 reports the structural parameters and the UV−vis spectra of the TAB and CAB forms in the gas phase and in ethanol. As reported in a previous study,[37] direct comparisons of our results to experimental data might be impeded because X-ray structures are determined in the solid phase, whereas gas electron diffraction (GED) measurements are performed at relatively high temperature. Nevertheless, concerning the TAB isomer in the gas phase, B3LYP yields results with a maximal deviation of 0.013 Å for the bond lengths and 1.4° for the valence angles, if the most recent X-ray diffraction and GED experiments are used as references. Moreover, the GED measurements perfectly predict the planar gas-phase geometry of the TAB. For the CAB form, low-temperature X-ray experiments or GED measurements are not available in the literature, but the X-ray structure from ref 67 is in good agreement with our DFT results with a maximum deviation of 0.013 Å, 2.2°, and 2.3° for the bond lengths, valence, and dihedral angles, respectively.

Experimentally, the absorption UV−vis spectrum consists of low lying ($n\rightarrow\pi^*$) bands between 380 and 520 nm which is a symmetry forbidden transition in the case of the trans isomer. The spectrum also exhibits a $\pi\rightarrow\pi^*$ transition around 330 nm for TAB and 275 nm for CAB.[68]

For the UV−vis spectra of TAB, the TD-PBE0 absorption energies are in rather good agreement with the experimental values. Although the excitation energies are underestimated by 20–40 nm in gas phase or in ethanol, the hypsochromic/bathochromic shift for the $n\rightarrow\pi^*/\pi\rightarrow\pi^*$ are well-reproduced by the theoretical calculations. Experimentally, one notes a small shift between +3 and −7 nm for the $n\rightarrow$

$\pi^*$ transition, to be compared to the −3 nm predicted by TD-PBE0 calculations. For the $\pi \rightarrow \pi^*$ excitation, we predict a bathochromic shift of +17 nm that is very close to the +17 and +19 nm experimentally obtained. We refer the reader to ref 37 for more details about the geometries and UV−vis transition of TAB.

Concerning the CAB absorption spectra, the deviation between theoretical and experimental results is of the same order of magnitude than those obtained for the trans isomer. The $n \rightarrow \pi^*$ excitation becomes a dipole-allowed transition due to the nonplanar geometry of the CAB derivative. Theoretically, taking into account the solvent effects, we observe an hypsochromic shift of −11 nm and a bathochromic shift of +13 nm for the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions, respectively. Compared to the TAB derivative, the solvent induces an increase of the hypsochromic shift for the $n \rightarrow \pi^*$ transition, while the $\pi \rightarrow \pi^*$ excitation energy remains unchanged with a clear decrease of the oscillator strength. The bathochromic effect of the solvent on the $\pi \rightarrow \pi^*$ excitation is well-reproduced by the present theoretical approach. Indeed, experimental results lead to a bathochromic shift of +16 nm. On the other hand, the red-shift of 8–18 nm, experimentally observed on the $n \rightarrow \pi^*$ transition, is not reproduced by TD-PBE0, which reports a blue-shift of −11 nm. However, the gas-phase UV−vis spectra of CAB is subject to caution: it is calculated at relatively high and variable temperature (181–322 °C) by means of a difference spectrum. This spectrum, obtained by a flash photolysis technique, corresponds to the difference between the absorption spectra of a stable compound (TAB), beforehand determined, and a metastable specie (CAB).[69]

Table 1 also reports the modification of the geometrical parameters and absorption wavelengths going with the photoisomerization of the AB dye (see Figure 1). The change of the C−N bond length is nicely reproduced, while the variations of the $d$(N=N) and ∠(C−N=N) values are slightly overestimated by the theoretical scheme. The largest discrepancy between theory and experiment comes from the overestimation of the variation of the $\tau$(C−C−N=N) value by 10–20°, but these dihedral angle values might be very sensitive to the packing effect of the crystal. Nevertheless, the most recent GED experiment on the TAB gives a $\tau$(C−C−N=N) value of 0.0°, that is a deviation between theory and experiment of 0.2° only. For the shift of the absorption wavelengths, the experimental and theoretical approaches predict blue-shift of the $n \rightarrow \pi^*$ and the $\pi \rightarrow \pi^*$ excitation energies in both the gas phase and ethanol. Moreover, TD-PBE0 provides energetic shifts in the range of the experimental figures, but for the shift of the $n \rightarrow \pi^*$ excitation in the gas phase. This observation blames once more the experimental $\lambda^{n \rightarrow \pi^*}$ value obtained in the case of CAB in the gas phase.

According to previous computational studies, the TD-DFT excitation energies are more red-shifted compared to other theoretical approaches like CC2 or SOPPA methods,[70] leading to a slightly larger discrepancy between the theoretical and experimental absolute wavelengths. However, this type of post Hartee−Fock methods are very time-consuming and, consequently, difficult to use for larger systems, especially when solvent effects must be taken into account. In this case, DFT calculations using hybrid functionals such as B3LYP or PBE0 present a valuable accuracy/CPU balance. On top of that, these theoretical schemes have been proved reliable to reproduce the experimental shifts corresponding to solvatochromic effects and photoisomerization processes of the AB dye. We therefore apply this methodology in order to study the effect of the polypeptide structure on the absorption spectra of grafted AB moiety.

**3.2. Poly(L-glumatic acid) with an AB Side Chain.** Since the actual QM/MM method requires one SLBO to connect the QM to the MM part, it is important to check that it does not disturb the QM properties. The influence induced by the SLBO is analyzed in order to verify that no artificial shifts of the absorption energies in TD-DFT calculations are introduced. To do so, we computed the absorption wavelength of the QM part capped with a methyl group, at the QM level, with and without the SLBO (see Figure 2). On the basis of these preliminary calculations, one can conclude that the SLBO creates an error of 0.09 and 0.22 nm on the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions, respectively. Moreover, a previous study has demonstrated that the SLBO induces only slight modifications of the QM geometry.[59]

The geometrical parameters of the poly(L-glumatic acid) with AB side chain are listed in Table 2. We also report the optimized geometries and the number of glutamic residues corresponding to several secondary structures of the polypeptide (see Figure 3 for TAB and Figure 4 for CAB). Due to the importance of the N=N double bond length, the chromophoric unit of the AB dye, we first investigated its modifications. The various MM surroundings imply an increase of 0.002–0.004 Å, compared to the isolated TAB structure in the gas phase. Concerning the cis isomer of the poly(L-glumatic acid), the increase of the N=N distance is not systematic. Indeed, for the $\beta$-sheet, $\beta$-hairpin, and $\beta$-strand conformations of the MM part, the $d$(N1=N2) values are equal to or smaller (−0.001 Å) than the values of the isolated CAB structure. For the other secondary structures, the MM part induces an increase of the N=N bond length by 0.002–0.005 Å. The same effects are induced by the solvent (Table 1) for both isomers. The C−N distances are also affected by the polypeptide structure, as well as the angle values around the N=N chromophoric unit for both conformers. Contrary to the other geometrical parameters for which the effect of the MM surrounding is negligible, the dihedral angles are more affected by the polypeptide structure, leading to a nonplanar form of the TAB, regardless of the conformation of the MM part. For example, the α-helix conformation and the α−$\beta$−α motif induce a $\tau$(N1=N2−C4−C5) value of ≃20°. For the CAB derivatives, the α-containing structures (α-helix, α−$\beta$−α, and $\beta$−α−$\beta$) lead to a large modification of the $\tau$(N1=N2−C4−C6) values, yielding an increase of the N=N double bond as pointed above.

The UV−vis spectra of the poly(L-glumatic acid) containing AB side chain are reported in Table 3. For the TAB derivatives, an overall blue-shift (3–29 nm) and red-shift (23–36 nm) with respect to the isolated TAB molecule in the gas phase ($\Delta\lambda^{AB}$ values) are observed for

***Table 2.*** Structural Parameters of Isolated TAB and CAB and Grafted As Side Chain of Poly(L-glutamic acid) Obtained at the B3LYP/6-311G(d,p) Level of Theory[a]

| | MM structure | *d* (N1=N2) | *d* (C3−N1) | *d* (N2−C4) | ∠(C3−N1 =N2) | ∠(N1=N2 −C4) | ∠(C5−C3−N1) | ∠(N2−C4−C6) | τ(C5−C3− N1=N2) | τ(N1=N2− C4−C6) |
|---|---|---|---|---|---|---|---|---|---|---|
| TAB | | 1.253 | 1.418 | 1.418 | 115.1 | 115.1 | 124.6, 115.5 | 124.6, 115.5 | 0.0, 180.0 | 0.0, 180.0 |
| | α-helix | 1.255 | 1.416 | 1.412 | 115.2 | 114.7 | 124.0, 116.1 | 124.4, 116.8 | −23.5, 159.3 | −12.2, 168.3 |
| | β−α−β | 1.257 | 1.417 | 1.410 | 115.2 | 115.2 | 124.6, 115.6 | 124.7, 116.3 | −7.8, 174.3 | −3.2, −179.6 |
| | α−β−α | 1.255 | 1.419 | 1.409 | 114.3 | 116.1 | 124.3, 116.0 | 125.9, 115.2 | −8.0, 173.2 | 2.4, −175.6 |
| | β-sheet | 1.255 | 1.417 | 1.412 | 115.1 | 115.3 | 124.7, 115.5 | 125.2, 115.9 | −1.3, 178.5 | −0.9, 178.1 |
| | β-hairpin | 1.255 | 1.417 | 1.411 | 115.1 | 115.3 | 124.7, 115.6 | 125.1, 115.9 | 0.3, −179.7 | 0.2, −179.8 |
| | β-strand | 1.255 | 1.415 | 1.415 | 115.1 | 115.7 | 124.3, 115.7 | 124.4, 116.4 | 17.2, −164.1 | −18.3, 165.5 |
| CAB | | 1.243 | 1.436 | 1.436 | 124.1 | 124.1 | 122.9, 116.5 | 122.9, 116.5 | 51.0, −138.1 | 51.0, −138.1 |
| | α-helix | 1.245 | 1.433 | 1.430 | 124.4 | 125.1 | 122.0, 117.4 | 126.3, 114.6 | −57.6, 131.4 | −35.2, 154.0 |
| | β−α−β | 1.248 | 1.428 | 1.427 | 125.7 | 127.9 | 122.5, 117.1 | 129.2, 112.5 | −56.8, 133.0 | −20.6, 166.6 |
| | α−β−α | 1.246 | 1.431 | 1.429 | 125.5 | 126.2 | 121.2, 118.2 | 127.9, 113.5 | 65.2, −124.4 | 21.0, −166.9 |
| | β-sheet | 1.242 | 1.441 | 1.434 | 123.2 | 123.5 | 119.8, 119.4 | 122.1, 118.1 | 73.5, −116.7 | 56.4, −133.9 |
| | β-hairpin | 1.242 | 1.437 | 1.437 | 122.8 | 123.5 | 119.7, 119.4 | 121.3, 118.6 | 64.8, −125.6 | 58.8, −133.1 |
| | β-strand | 1.243 | 1.431 | 1.436 | 123.5 | 125.0 | 122.9, 116.4 | 122.5, 117.4 | 51.9, −136.0 | 51.3, −139.4 |

[a] The MM part is described with the Amber ff99 force field. Distances are in angstroms, while the valence and the dihedral angles are in degrees. See Figure 2 for the atom labeling.

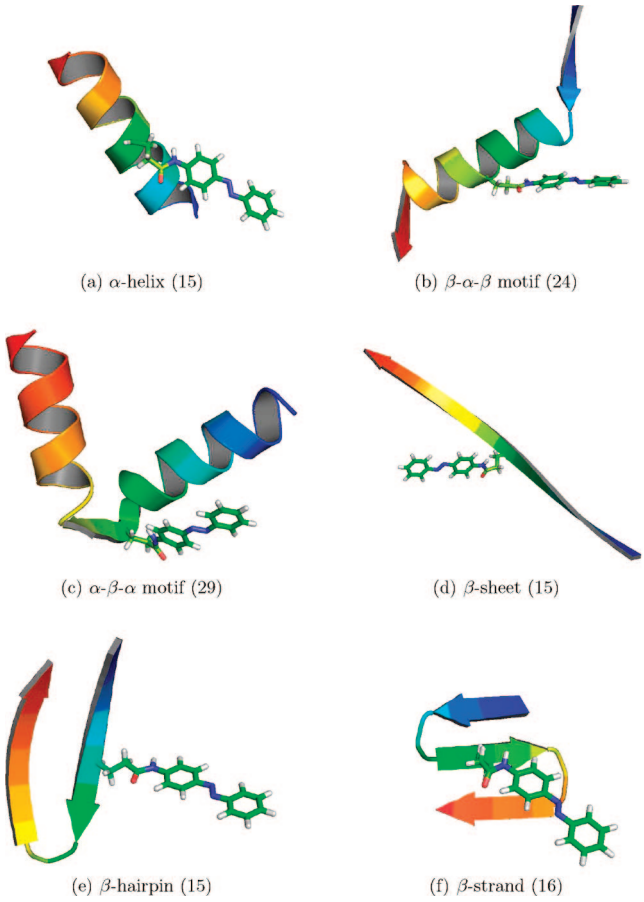


***Figure 3.*** QM/MM optimized geometries of the poly(L-glutamic acid) with the photochromic TAB side chain. The QM part is depicted as a stick model, while the MM structure has a cartoon representation. We also report the number of glutamic residues in parenthesis.

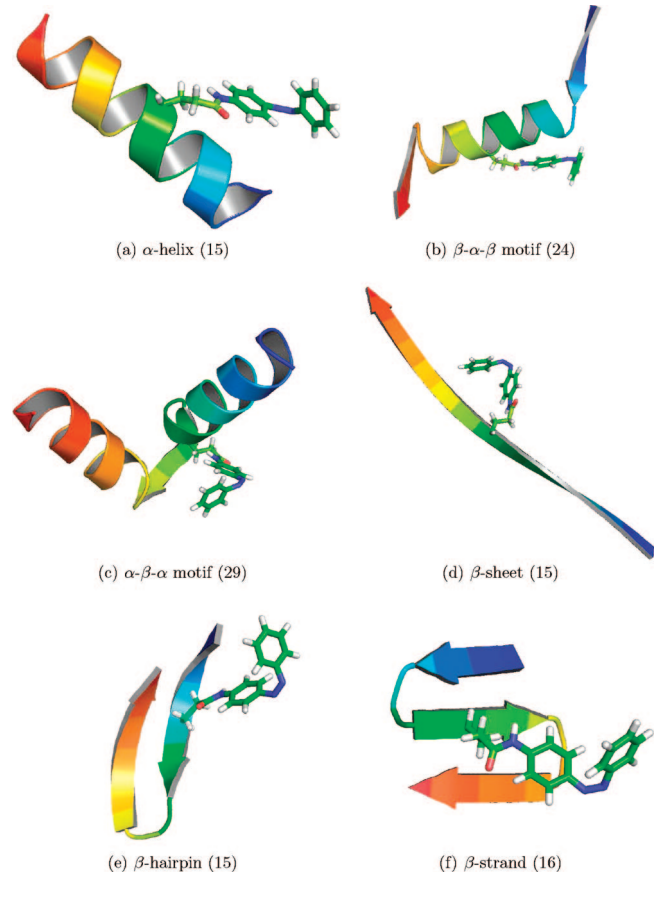


***Figure 4.*** QM/MM optimized geometries of the Poly(L-glutamic acid) with the photochromic CAB side chain. The QM part is depicted as a stick model, while the MM structure has a cartoon representation. We also report the number of glutamic residues in parenthesis.

the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions, respectively. This is in full agreement with previous results on substituted TAB derivatives.[38] It is noteworthy that the $n \rightarrow \pi^*$ transition in TAB is no more symmetry-forbidden and significant, though small oscillator strengths are observed, especially for the α-helix conformation and the β−α−β motif. The overall red-shift of the $\pi \rightarrow \pi^*$ band is also observed in the case of the CAB derivatives (21–32 nm), notwithstanding that the blue-shift of the $n \rightarrow \pi^*$ is not systematic: the conformations mainly featuring β structures (β-sheet, β-hairpin, and β-strand) show a blue-shift of −20, −15, and −21 nm, respectively, whereas the α-containing structures lead to a red-shift of 10–14 nm.

***Table 3.*** UV−vis Spectra of the TAB and CAB Grafted on the Side Chain of Poly(L-glutamic acid) Obtained at the TDPBE0/6-311+G(d,p) Level of Theory[a]

| | MM structure | $\lambda^{n\rightarrow\pi^*}$ | $\Delta\lambda^{AB}$ tot | $\Delta\lambda^{elec}$ | $\Delta\lambda^{nuc}$ | $\Delta\lambda^{aux}$ | $\lambda^{\pi\rightarrow\pi^*}$ | $\Delta\lambda^{AB}$ tot | $\Delta\lambda^{elec}$ | $\Delta\lambda^{nuc}$ | $\Delta\lambda^{aux}$ | $\Delta\lambda^{max}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TAB | α-helix | 469(0.04) | −11 | 0 | −7 | −4 | 350(0.94) | 23 | 0 | −7 | 30 | 119 |
| | β−α−β | 477(0.01) | −3 | 0 | 1 | −4 | 363(1.00) | 36 | 3 | 3 | 30 | 114 |
| | α−β−α | 475(0.00) | −5 | 1 | −2 | −4 | 361(1.02) | 34 | 2 | 2 | 30 | 114 |
| | β-sheet | 477(0.00) | −3 | 0 | 1 | −4 | 357(0.98) | 30 | 2 | −2 | 30 | 120 |
| | β-hairpin | 476(0.00) | −4 | 1 | −1 | −4 | 359(0.87) | 32 | 0 | 2 | 30 | 117 |
| | β-strand | 451(0.00) | −29 | −21 | −4 | −4 | 358(0.75) | 31 | 5 | −4 | 30 | 93 |
| CAB | α-helix | 488(0.05) | 10 | −1 | 2 | 9 | 314(0.35) | 22 | −1 | 0 | 23 | 174 |
| | β−α−β | 492(0.06) | 14 | −16 | 21 | 9 | 324(0.47) | 32 | 5 | 4 | 23 | 168 |
| | α−β−α | 492(0.04) | 14 | −2 | 7 | 9 | 316(0.50) | 24 | 3 | 4 | 23 | 176 |
| | β-sheet | 458(0.04) | −20 | 0 | −29 | 9 | 315(0.12) | 23 | 1 | −1 | 23 | 143 |
| | β-hairpin | 463(0.05) | −15 | 0 | −24 | 9 | 313(0.15) | 21 | −6 | 4 | 23 | 150 |
| | β-strand | 457(0.08) | −21 | −25 | −5 | 9 | 323(0.18) | 31 | 4 | 4 | 23 | 134 |

[a] The MM part is described with the help of the Amber ff99 force field. $\lambda^{n\rightarrow\pi^*}$ and $\lambda^{\pi\rightarrow\pi^*}$ are the wavelengths (in nanometers) of the first $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions, respectively. The oscillator strengths are given in parenthesis. We also report the energetic shift (in nanometers) between the TD-DFT/MM values and the AB in gas phase ($\Delta\lambda^{AB}$). The $\Delta\lambda^{AB}$ value is decomposed in three components coming from the electronic polarization of the MM charges ($\Delta\lambda^{elec}$), the nuclear polarization of the MM surrounding ($\Delta\lambda^{nuc}$), and the auxochromic shift due to the substitution of the AB moiety ($\Delta\lambda^{aux}$), such as $\Delta\lambda^{AB} = \Delta\lambda^{elec} + \Delta\lambda^{nuc} + \Delta\lambda^{aux}$.

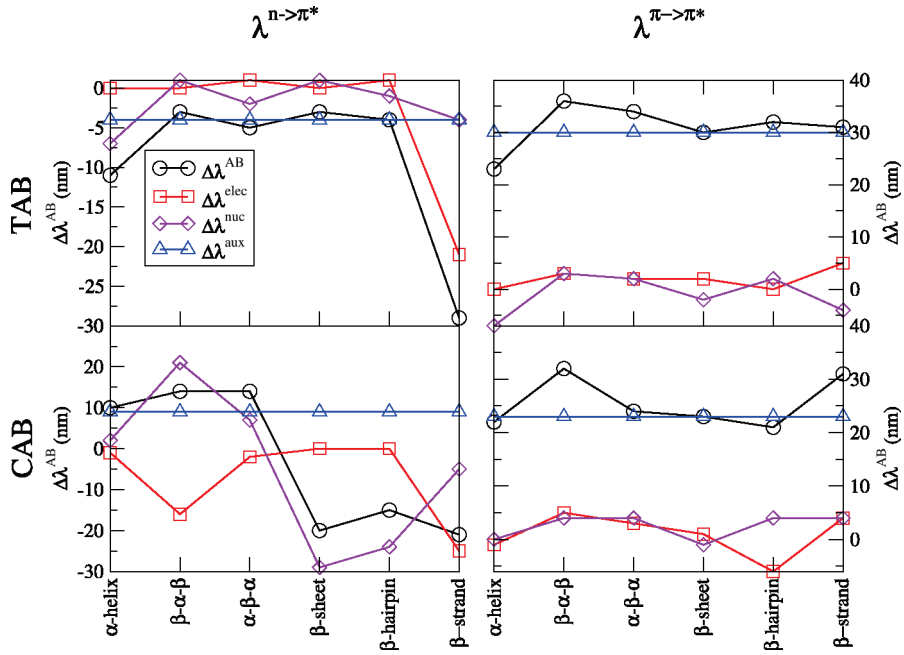


***Figure 5.*** Energetic shift (in nanometers) between the TD-DFT/MM values and the AB in the gas phase ($\Delta\lambda^{AB}$). The $\Delta\lambda^{AB}$ value is decomposed in three components coming from the electronic polarization of the MM charges ($\Delta\lambda^{elec}$), the nuclear polarization of the MM surrounding ($\Delta\lambda^{nuc}$), and the auxochromic shift due to the substitution of the AB moiety ($\Delta\lambda^{aux}$), such as $\Delta\lambda^{AB} = \Delta\lambda^{elec} + \Delta\lambda^{nuc} + \Delta\lambda^{aux}$.

In order to better understand the physical meaning of these results, the $\Delta\lambda^{AB}$ values are split up in three components gathered in Table 3 and depicted in Figure 5:

• The polarization of the electronic wave function due to the MM classical point charges of the force field ($\Delta\lambda^{elec}$). These quantities are determined by the difference between the TD-DFT/MM calculations obtained with electronic embedding (EE) and the results of the TD-DFT/MM calculations in which the point charges are set to zero with the same geometry.

• The geometrical modification of the QM part geometries with respect to the capped primary system (CPS) induced by the MM surroundings ($\Delta\lambda^{nuc}$). The CPS corresponds to the QM part capped with an hydrogen atom (Figure 2). These energetic shifts are obtained by the difference between the TD-DFT/MM results with the MM point charges set to zero and the TD-DFT values on the optimized structures of the CPS. This effect corresponds to the polarization of the MM surrounding on the nuclei positions.

• $\Delta\lambda^{aux}$ which corresponds to the shift implied by the auxochromic groups (including the amide group and an alkyl chain) used to graft the photochromic AB to the polypeptide backbone. This quantity, independent of the polypeptide conformation, is defined as the difference of the excitation energies between the CPS and the AB moiety.

According to these definitions, the overall shift of the absorption energy when going from the isolated AB mol-

ecules to the full system is defined by: $\Delta\lambda^{AB} = \Delta\lambda^{elec} + \Delta\lambda^{nuc} + \Delta\lambda^{aux}$.

For the $\pi \rightarrow \pi^*$ transitions in the TAB derivatives, the main component of the red-shift comes from the auxochromic contribution $\Delta\lambda^{aux}$. Indeed, the polarization due to the MM environment on the electronic wave function and the nuclei positions leads to only small modifications of the $\lambda^{\pi\rightarrow\pi^*}$ values. However, further investigations of the $\Delta\lambda^{elec}$ values show that the polarization of the MM point charges also implies a red-shift of 0–5 nm. For the $n \rightarrow \pi^*$ excitation, the situation is slightly different with a much smaller auxochromic effect (blue-shift of −4 nm). We point out the effect of the EE which implies a large blue-shift (−21 nm) in the case of the $\beta$-hairpin structure. This effect corresponds to a stabilization of the ground-state compared to the excited state and originates from a specific interaction of the carboxylic acid hydrogen atom of the glutamic acid side chain with the lone pair of the N═N chromophoric unit of the TAB moiety (see Figure 6). This highligths the key role that can be played by the glumatic acid side chain of the polypeptide. Similar to the specific solvent–solute interactions, the intramolecular interaction between the chromophoric unit and the polypeptide may induce a strong shift of the excitation energies and especially for the $n \rightarrow \pi^*$ transition. Indeed, these two orbitals correspond to a localized phenomenon: the $n$ nonbonding (Figure 6a) and the $\pi^*$ antibonding (Figure 6c) orbitals are mainly localized on the chromophoric units of the AB dye. The $n$ orbital corresponds to the nitrogen lone pair, while the $\pi^*$ orbital is mainly located on the diazo bond. The $\lambda^{\pi\rightarrow\pi^*}$ value is less affected by these specific interactions ($\Delta\lambda^{elec} = +5$ nm) due to the nature of the $\pi$ bonding orbital, which is delocalized over the whole molecule (Figure 6b).

The investigation of the $\pi \rightarrow \pi^*$ excitations in the CAB derivatives leads to conclusions similar to the case of the trans isomers. The red-shift of the absorption energies is mainly due to the substitution effect on the AB moiety. For the $n \rightarrow \pi^*$ transition, the $\beta$-containing structure exhibit a blue-shift of the excitation wavelengths. The $\beta$-strand case is explained by the same phenomenon than for the TAB derivatives, whereas the results of the $\beta$-sheet and $\beta$-strand conformations shows a clear dependence upon the modification of the AB geometries, as illustrated by the $\Delta\lambda^{nuc}$ values. As previously mentionned, both $\beta$-sheet and $\beta$-strand conformations exhibit the smallest N═N bond length and the largest $d$(C—N) values (Table 2). The effect, albeit less pronounced, is still present in the $\beta$-strand structure, whereas the opposite phenomenon is observed for the $\beta-\alpha-\beta$ motif. The increase of the N═N chromophoric unit leads to a $\Delta\lambda^{nuc}$ value of +21 nm. However, this MM conformation does not induce a large red-shift of the absorption energies due to a balanced effect with the $\Delta\lambda^{elec}$ value, as a consequence of the proximity of a glutamic side chain to the AB moiety.

## 4. Conclusions

On the basis of a theoretical approach combining QM/MM and TD-DFT calculations, we have reported the obtention of the UV−vis spectra of poly(L-glutamic) acid modified with chromophoric AB side chain at the TD-PBE0/6-311+G(d,p)//B3LYP/6-311G(d,p) level of theory using the Amber ff99 force field to treat the MM part. The results on the AB units, in both the gas phase and in ethanol, indicate that this theoretical approach is able to provide geometries and absorption wavelengths that are in good agreement with the available experimental data. In particular, we have shown that the TD-PBE0 results reproduce the energetic shifts due to the solvatochromic effects and the photoisomerization process of the AB dye.

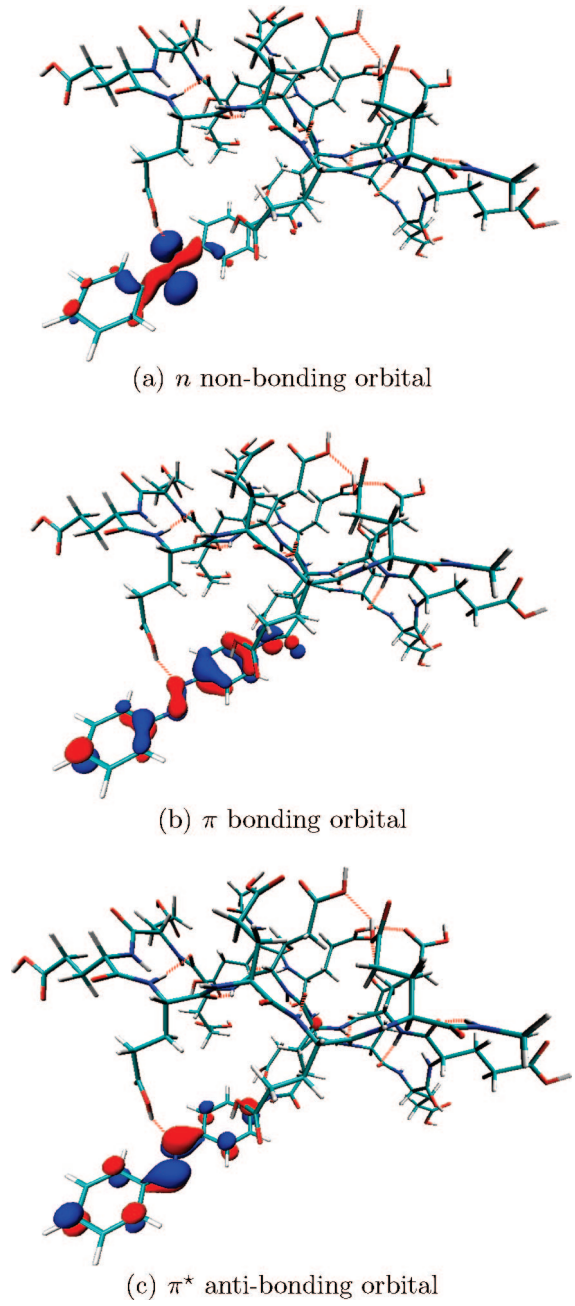


***Figure 6.*** Isosurface (0.05 au$^{-3/2}$) of the frontier orbitals involved in the $n \rightarrow \pi^\star$ and $\pi \rightarrow \pi^\star$ transitions of the TAB grafted on a polypeptide with a $\beta$-strand conformation.

In order to foresee the effect of the polypeptide structure, we have studied the polarization effect on the UV−vis spectra of the AB derivatives of several stable conformations featuring regularly folded structures, such as α-helix and $\beta$-sheet, on the UV−vis spectra of the AB derivatives. It

has been shown that the $\pi \rightarrow \pi^*$ absorption energies are globally red-shifted. This effect has been ascribed to a large auxochromic shift mainly due to the amide group used to graft the AB moiety to the carboxylic group of the glutamic side chain. Moreover, specific interactions between the glutamic side chain and the azobenzene moiety, such as hydrogen bond or large geometrical changes of the QM part, may lead to a large blue-shift on the $n \rightarrow \pi^*$ transition.

We are currently investigating the dynamic behavior of the AB photoisomerization in the framework of multiscale QM/MM dynamics in order to point out the major phenomena that could lead to strong modifications of the polypeptide secondary structure.

**Acknowledgment.** Two of the authors (X.A. and P.-F.L.) are grateful to Dr. Nicolas Ferré for helpful and fruitful discussions and for the some modifications of the code. They also acknowledge financial support from the Jean Barriol Institute (FR CNRS 2843). J.P. acknowledges the FRIA (Belgian "Fonds pour la formation à la Recherche dans L'Industrie et dans l'Agriculture") for his Ph.D. grant. C.M. thanks the Belgian National Fund for her postdoctoral researcher position. D.J. and E.A.P. thank the Belgian National Fund for their respective research associate positions. Part of the calculations have been performed on the Interuniversity Scientific Computing Facility (ISCF) which are installed at the Facultés Universitaires Notre-Dame de la Paix (Namur, Belgium). The authors gratefully acknowledge the financial support of the FNRS-FRFC and the "Loterie Nationale" for the convention number 2.4578.02 and of the FUNDP. The authors thank the CNRS-CGRI collaboration (Centre National de la Recherche Scientifique−Commissariat Général aux Relations Internationales de la Communauté Française de Belgique) for its financial support (agreement no. 18195). E.A.P. acknowledges a 4 month invited professor position at the Université Henri Poincaré (Nancy, France).

## References

(1) Zollinger, H. In *Color Chemisrty, Syntheses, Properties and Applications of Organic Dyes and Pigments*, 3rd ed.; Wiley-VCH: Weinheim, 2003.

(2) Sudesh Kumar, G.; Neckers, D. C. *Chem. Rev.* **1989**, *89*, 1915–1925.

(3) Tamai, N.; Miyasaka, H. *Chem. Rev.* **2000**, *100*, 1875–1890.

(4) Natansohn, A.; Rochon, P. *Chem. Rev.* **2002**, *102*, 4139–4176.

(5) Yu, Y.; Nakano, M.; Ikeda, T. *Nature* **2003**, *425*, 145.

(6) Balzani, V.; Credi, A.; Raymo, F. M.; Stoddart, J. F. *Angew. Chem., Int. Ed.* **2000**, *39*, 3348–3391.

(7) Qu, C. J.; Wang, Q. C.; Ren, J.; Tian, H. *Org. Lett.* **2004**, *6*, 2085–2088.

(8) Mathies, R. A.; Lugtenburg, J. In *Handbook of Biological Physics*; Stavenga, D. G., de Grip, W. J., Pugh, E. N., Eds.; Elsevier Science B. V.: New York, 2000; Vol. 3.

(9) Ferré, N.; Olivucci, M. *J. Am. Chem. Soc.* **2003**, *125*, 6868–6869.

(10) Andruniow, T.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17908–17913.

(11) Coto, P. B.; Sinicropi, A.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17154–17159.

(12) Frutos, L. M.; Andruniow, T.; Santoro, F.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7764–7769.

(13) Wanko, M.; Garavelli, M.; Bernardi, F.; Niehaus, T. A.; Frauenheim, T.; Elstner, M.*J. Chem. Phys.* **2004**, *120*, 1674–1692.

(14) Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

(15) Pieroni, O.; Houben, J. L.; Fissi, A.; Constantino, P.; Ciardelli, F. *J. Am. Chem. Soc.* **1980**, *102*, 5913–5915.

(16) Houben, J. L.; Fissi, A.; Bacciola, D.; Rosato, N.; Pieroni, O.; Ciardelli, F. *Int. J. Biol. Macromol.* **1983**, *5*, 94–100.

(17) Ciardelli, F.; Pieroni, O.; Fissi, A.; Houben, J. L. *Biopolymers* **1984**, *23*, 1423–1437.

(18) Pieroni, O.; Fissi, A.; Angelini, N.; Lenci, F. *Acc. Chem. Res.* **2001**, *34*, 9–17.

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

(20) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718–730.

(21) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.

(22) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodoseek, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534–10547.

(23) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46–54.

(24) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290–9295.

(25) DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *166*, 9578–9584.

(26) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468–1494.

(27) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632–650.

(28) Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656–6665.

(29) Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991–4004.

(30) Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.

(31) Preat, J.; Jacquemin, D.; Perpete, E. A. *Chem. Phys. Lett.* **2005**, *415*, 20–24.

(32) Preat, J.; Jacquemin, D.; Wathelet, V.; André, J.-M.; Perpete, E. A. *J. Phys. Chem. A* **2006**, *110*, 8144–8150.

(33) Jacquemin, D.; Preat, J.; Wathelet, V.; Fontaine, M.; Perpete, E. A. *J. Am. Chem. Soc.* **2006**, *128*, 2072–2083.

(34) Preat, J.; Loos, P.-F.; Assfeld, X.; Jacquemin, D.; Perpete, E. A. *Int. J. Quantum Chem.* **2007**, *107*, 574–585.

(35) Preat, J.; Loos, P.-F.; Assfeld, X.; Jacquemin, D.; Perpete, E. A. *J. Mol. Struct. (THEOCHEM)* **2007**, *808*, 85–91.

(36) Preat, J.; Jacquemin, D.; Wathelet, V.; André, J.-M.; Perpete, E. A. *Chem. Phys.* **2007**, *335*, 177–186.

(37) Briquet, L.; Vercauteren, D. P.; Perpete, E. A.; Jacquemin, D. *Chem. Phys. Lett.* **2006**, *417*, 190–195.

(38) Briquet, L.; Vercauteren, D. P.; André, J.-M.; Perpete, E. A.; Jacquemin, D. *Chem. Phys. Lett.* **2007**, *435*, 257–262.

(39) Vreven, T.; K.; Morokuma, K. *Theor. Chem. Acc.* **2003**, *109*, 125–132.

(40) Moret, M.-E.; Tapavicza, E.; Guidoni, L.; Rohrig, U. F.; Sulpizi, M.; Tavernelli, I.; Rothlisberger, U. *Chimia* **2005**, *59*, 493–498.

(41) Roehrig, U. F.; Frank, I.; Hutter, J.; Laio, A.; VandeVondele, J.; Rothlisberger, U. *Chem. Phys. Chem* **2003**, *4*, 1177–1182.

(42) Sulpizi, M.; Carloni, P.; Hutter, J.; Rothlisberger, U. *Phys. Chem. Chem. Phys.* **2003**, *5*, 4798–4805.

(43) Crecca, C. R.; Roitberg, A. E. *J. Phys. Chem. A* **2006**, *110*, 8188–8203.

(44) Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009–4037.

(45) Gu, J.; Xie, Y.; Schaefer, H. F., III *J. Am. Chem. Soc.* **2005**, *127*, 1053–1057.

(46) Gu, J.; Xie, Y.; Schaefer, H. F., III *J. Am. Chem. Soc.* **2005**, *127*, 1250–1252.

(47) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100–106.

(48) Ferré, N.; Assfeld, X.; Rivail, J.-L. *J. Comput. Chem.* **2002**, *23*, 610–624.

(49) Ferré, N.; Assfeld, X. *J. Chem. Phys.* **2002**, *117*, 4119–4125.

(50) Ferré, N.; Assfeld, X. *J. Mol. Struct. (Theochem)* **2003**, *632*, 83–90.

(51) Moreau, Y.; Loos, P.-F.; Assfeld, X. *Theor. Chem. Acc.* **2004**, *112*, 228–239.

(52) Moreau, Y.; Assfeld, X. *Lecture Ser. Comput. Comput. Sci.* **2005**, *3*, 1–9.

(53) Fornili, A.; Loos, P.-F.; Sironi, M.; Assfeld, X. *Chem. Phys. Lett.* **2006**, *427*, 236–240.

(54) Loos, P.-F.; Assfeld, X. *J. Chem. Theory Comput.* **2007**, *3*, 1047–1053.

(55) Loos, P.-F.; Assfeld, X. *Int. J. Quantum Chem.* **2007**, *107*, 2243–2252.

(56) Loos, P.-F.; Assfeld, X. *AIP Conf. Proc.* **2007**, *963*, 308–315.

(57) 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 0.05; Gaussian, Inc.: Wallingford, CT,2004.

(58) Ponder, J. W. *Tinker*, version 4.2; Washington University: St. Louis, MO, 2004.

(59) Fornili, A.; Moreau, Y.; Sironi, M.; Assfeld, X. *J. Comput. Chem.* **2006**, *27*, 515–523.

(60) 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.

(61) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

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

(63) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029–5036.

(64) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(65) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3094.

(66) Cossi, M.; Barone, V. *J. Chem. Phys.* **2001**, *115*, 4708–4717.

(67) Mostad, A.; Romming, C. *Acta Chem. Scand.* **1971**, *25*, 3561–3568.

(68) Nagele, T.; Hoche, R.; Zinth, W.; Wachtveitl, J. *Chem. Phys. Lett.* **1997**, *272*, 489–495.

(69) Andersson, J.-A.; Petterson, R.; Tegner, L. *J. Photochem.* **1982**, *20*, 17–32.

(70) Fliegl, H.; Kohn, A.; Hattig, C.; Ahlrichs, R. *J. Am. Chem. Soc.* **2003**, *125*, 9812–9827.

(71) Harada, J.; Ogawa, K.; Tomoda, S. *Acta Cryst. B* **1997**, *53*, 662–672.

(72) Tsuji, T.; Takashima, H.; Takeuchi, H.; Egawa, T.; Konaka, S. *J. Phys. Chem. A* **2001**, *105*, 9347–9353.

(73) Millefiori, S.; Millefiori, A. *J. Chem. Soc., Faraday Trans. II* **1981**, *77*, 245–258.

(74) Birnhaum, P. P.; Linford, J. H.; Style, D. W. G. *Trans. Faraday Soc.* **1953**, *49*, 735–744.

(75) Gore, P. H.; Wheeler, O. H. *J. Org. Chem.* **1961**, *26*, 3295–3298.

(76) Grasselli, J. G. In *CRC Atlas of Spectral Data and Physical Constants for Organic Compounds*; CRC Press: Cleveland, OH, 1973.

CT700188W

JCTC Journal of Chemical Theory and Computation

# Mechanism of the Ethylene Polymerization at Very High Pressure

Martina Mugnai,[†] Marco Pagliai,[†] Gianni Cardini,*[,†,‡] and Vincenzo Schettino[†,‡]

*Laboratorio di Spettroscopia Molecolare, Dipartimento di Chimica, Universitá di Firenze, Via della Lastruccia 3, Sesto Fiorentino, and European Laboratory for Nonlinear Spectroscopy (LENS), Via Nello Carrara, Sesto Fiorentino, Italy*

Received October 19, 2007

**Abstract:** The reaction of ethylene in condensed phases under high pressure has been investigated by ab initio molecular dynamics. Both disordered and crystalline samples have been simulated, and some insights on the reaction mechanism have been obtained. System size effects have been investigated for the disordered samples. A polymerization reaction occurs by an ionic mechanism. In both the disordered and the crystal phases, the reaction products obtained (linear chains in the disordered systems and branched chains in the crystal) are in qualitative agreement with the experiments.

## 1. Introduction

Unsaturated hydrocarbons have been known for a long time to spontaneously react when subjected to very high pressure and to transform into polymeric materials.[1–3] Recent experiments[4,5] have shown that, under appropriate control of the experimental conditions, the reaction can proceed along well-defined pathways, and high-quality (conformationally pure and crystalline) polymers can be obtained with potentially interesting physical and chemical properties. In the case of butadiene, the high-pressure polymerization of the liquid, under laser irradiation, has been reported to produce pure trans polybutadiene.[4,6] It has also been reported that the high-pressure polymerization of ethylene can result in a fully crystalline high-density polyethylene,[5] as revealed by spectroscopic and X-ray diffraction characterization. It has been shown that, when the compressed liquid monomer reactants are laser-irradiated at appropriate wavelengths, the pressure reaction threshold is quite significantly lowered, thus opening perspectives of practical technological applications of these novel synthetic procedures.

The mechanisms of high-pressure polymerization, and more in general high-pressure reactions, are not completely understood, even though some general concepts are well established.[1–3] In the first instance, high-pressure reactions will be governed by geometrical constraints. In fact, at very high pressures, significant potential barriers build up and hinder the molecular mobility and reorientations, and reactions involving minimum molecular displacements are favored (topochemical principle).[7] However, it is believed that the leading factor determining the reaction pathway is the substantial change of the electronic structure of the reactants occurring at high pressure.[8] This will include shifts and broadening of the electronic energy levels and, consequently, mixing of the ground and excited states and the population of higher electronic states that can be efficiently accomplished thermally. These effects can be conveniently investigated by first principles molecular dynamics that simultaneously considers the nuclear motion and the electron density fluctuations. Polymerization reactions have been deeply studied, mainly as catalytic processes,[9–11] and recently, also the problem of stereoselectivity has been modeled using ab initio molecular dynamics.[12]

Previous Car–Parrinello molecular dynamics simulations of reactions under high pressure[13–16] have enlighted a few problems. First of all, the computed pressure is higher than the experimental counterpart due to the lack of long-range dispersive contributions (negative), since this kind of interaction is strongly underestimated by currently available exchange-correlation functionals in the generalized gradient approximation framework.[17] Furthermore, to observe a reaction induced by pressure in the typical time scale of an

* Corresponding author e-mail: gianni.cardini@unifi.it.
[†] Universitá di Firenze.
[‡] European Laboratory for Nonlinear Spectroscopy.

10.1021/ct700275v CCC: $40.75 
Published on Web 02/27/2008

ab initio simulation, the system is overpressurized to speed up the process.

In the present work, we report on Car–Parrinello[18] molecular dynamics simulation of the polymerization at a high pressure of ethylene both in the liquid and in the solid phase. This is an interesting study case for high-pressure polymerization. In fact, while from the liquid,[5] as mentioned above, a linear polymer with no branching is obtained, the reaction of the solid[19] produces a substantially branched polymer at higher pressures, thus exposing the effects of geometrical constraints on the reaction mechanism. It will be shown that, even if the size of the simulated sample is small, the essentials of the high-pressure polymerization mechanisms can be extracted from the simulation in good agreement with experiments. In particular, it will be shown that, in the crystal, the anisotropy of the compressibility is responsible for the change from a linear to a nonlinear growth of the polymer.

## 2. Computational Details

Ab initio molecular dynamics calculations have been performed on ethylene samples in the liquid and crystal phases. Calculations in the liquid were carried out in cubic boxes of different sizes with 27 and 52 molecules to check for the effects of the sample size. The two initial samples were obtained by fully randomized samples obtained with classical molecular dynamics for rigid molecules[20] interacting with a Buckingham potential.[21] The ab initio simulations were started with a box of 13 Å for the sample with 27 molecules. Due to the higher memory requirements, the ab initio simulation for the larger sample was started with an initial box of 12 Å. In the case of the crystal, the initial simulation box is made up of a $3 \times 2 \times 3$ unitary monoclinic ($P = 20$ bar; $T = 90$ K; $P2_1/n$; lattice constants: $a = 4.613$, $b = 6.610$, $c = 4.037$ Å, $\beta = 94.54$ °, $Z = 2$)[22] cell with 36 molecules.

All of the calculations have been carried out by using BLYP[23,24] functional in the framework of the Car–Parrinello molecular dynamic method (CPMD).[18,25] Periodic boundary conditions (PBC) have been used. In order to check whether the reaction mechanism is ionic or radicalic, the simulations have been carried out with both the restricted and spin density formalisms. In the latter case, the simulations have also been repeated starting from a configuration close to the onset of the reaction, performing a random initialization of the wave function to prevent the electron coupling which could affect the observed reaction mechanism. As in previous studies,[15,26] a 40 Ry cutoff has been used for the plane-wave expansion limited to the Γ point. In the case of the spin density calculations in the crystal, the cutoff has been extended up to 60 Ry. Norm conserving Martins−Troullier[27] (radius = 1.0 au) and Car−von Barth[28] pseudo-potentials have been adopted for carbon and hydrogen atoms, respectively.

The evaluation of the pressure (calculated by using the stress tensor) can be affected by a large uncertainty due to Pulay forces effects, which are present in a variable-cell simulation. Therefore, we found it more convenient to refer to the volume or density of the simulated sample.

A thermal bath has been applied to the system by using the Nosé−Hoover chain method.[29–31] As usual, the bath parameters have been chosen by a trial-and-error procedure, after inspection of the power spectra of the atomic velocities autocorrelation function, in order to ensure a good coupling to the vibrational motions and a decoupling between electronic and nuclear degrees of freedom.

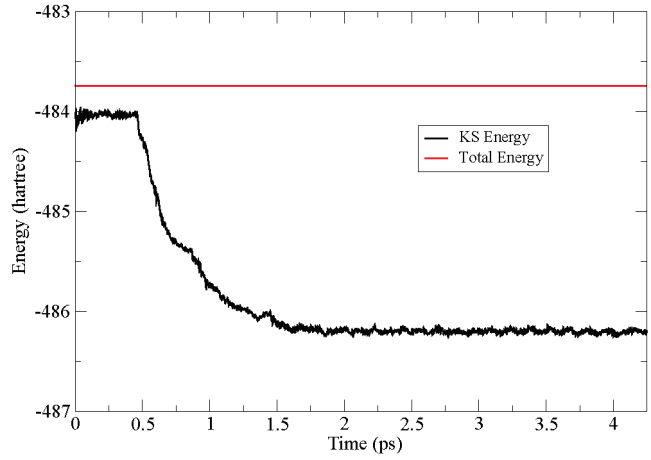


***Figure 1.*** Evolution of the Kohn–Sham energy in the crystal sample during the reaction.

In the case of the reaction in the crystal, the bath has also been applied to the electrons (a bath frequency of 1500 and 4500 $cm^{-1}$ has been chosen for the ions and the electrons, respectively; a kinetic energy of 0.04 au, corresponding to the average value of a simulation for unreacting crystals, has been chosen for the electrons). The goodness of this choice can be appreciated from Figure 1, showing that the Hamiltonian does not change appreciably along the simulation run even if the Kohn–Sham (KS) energy show a large variation.

To facilitate the analysis of the reaction mechanism, the simulations were performed in the NVT ensemble ($T = 300$ K, bath frequency 400 $cm^{-1}$, three thermostats). To compress the system, simulations in the NPT ensemble were performed (isotropic cell simulations with a cell mass of 1200 au) with an increasing external pressure between subsequent NVT simulations.

The various steps of the reaction pathway have been characterized calculating the maximally localized Wannier functions,[32–34] which are obtained by minimizing the spread of the Wannier functions ($w_n$) in direct space:

$$S = \sum_{n=1}^{N} \left( \langle w_n | r^2 | w_n \rangle - \langle w_n | \mathbf{r} | w_n \rangle^2 \right) \quad (1)$$

The implementation of this formula in the CPMD code is in the reciprocal space:[35]

$$S = \frac{2}{(2\pi)^2} \sum_{n}^{N_{\text{states}}} \sum_{I} \omega_I (1 - |z_{I,n}|) + O\left[ \left( 2\pi \hat{g}_I^{\text{t}} \times \overleftrightarrow{\mathbf{H}}^{-1} \right)^2 \right] \quad (2)$$

$$z_{I,n} = \int_V \mathrm{d}r \exp[i\mathbf{G}_I \times \mathbf{r}] |w_n(\mathbf{r})|^2 \quad (3)$$

where $\mathbf{H}$ is the transformation matrix from the crystallographic to the orthogonalized reference system and $t$, as

usual, indicates the transpose. The reciprocal vectors are $\mathbf{G}_I = 2\pi\ (\overleftrightarrow{\mathbf{H}}^{-1})^{\ t} \times \hat{g}_I$, where $\hat{g}_I$ represents the smallest independent Miller indeces, and they are determined along with their weights, $\omega_I$,[35] by

$$\sum_{\mu=1}^{3} \overleftrightarrow{\mathbf{H}}^{t}_{\alpha\mu} \overleftrightarrow{\mathbf{H}}_{\mu\beta} = g_{\alpha\beta} = \sum_{I} \omega_I \hat{g}_{\alpha I} \hat{g}_{\beta I} \tag{4}$$

In eq 4, $\omega_I$ represents weights defined on the basis of the crystal symmetry, and the index $I$ runs to a maximum value of 6 for triclinic crystals.

An analysis of the maximally localized Wannier functions has been performed adopting both a spin density and a restricted approach to verify that $\alpha$ and $\beta$ spins were characterized by the same center and to rule out the radicalic mechanism.

## 3. Results and Discussion

**3.1. Reactivity in the Liquid Phase.** A major problem in ab initio studies of polymerization reactions under high pressure can be the dependence of the computed properties on the small system size necessarily adopted in the simulations. In particular, it is important to determine if and how geometrical properties, dipole moments, reaction mechanisms, and final products are affected by this parameter. To this aim, two ethylene samples made up of 27 (sample A) and 52 (sample B) molecules have been studied at the same density. Initially, the conditions for reactivity of the smaller sample A were found. The simulation box was compressed to 8.99 Å a side, which corresponds to a density of about 1.97 g cm$^{-1}$.

The reaction can be followed through the changes of the KS behavior reported in Figure 2, sample A.

Initially, the KS energy is almost constant, indicating that the system does not react. After 0.75 ps, an energy drop is observed, but the reaction immediately stops and only after a further 0.80 ps starts again, giving a polymeric chain. When the KS energy of the B sample (Figure 2) is observed, the absence of the initial plateau and an immediate energy drop can be noted. This can be explained considering that the polymerization reaction under high pressure is induced by thermal fluctuations which become more probable upon increasing the sample size. This leads to the hypothesis that, when the sample size is increased, the reaction can occur at lower densities. A simulation with the 52-molecule sample has also been carried out at a density of 1.91 g cm$^{-3}$ (box side 11.27 Å; sample C). Also, in this sample, a reaction is observed but the associated energy decrease is smaller than for the other samples. In fact, in this case, the reaction immediately stops. The explanation of this behavior will be discussed in the following.

The compression procedure has been performed in steps. After each compression step, the system has been studied at constant volume. The data reported in the present work represent a selection made of significant phase space points to show the trend with increasing density. For sample A, three points have been selected for densities where reactions have not been observed (0.68, 1.47, and 1.72 g cm$^{-3}$). The fourth point (density of 1.97 g cm$^{-3}$) corresponds to the

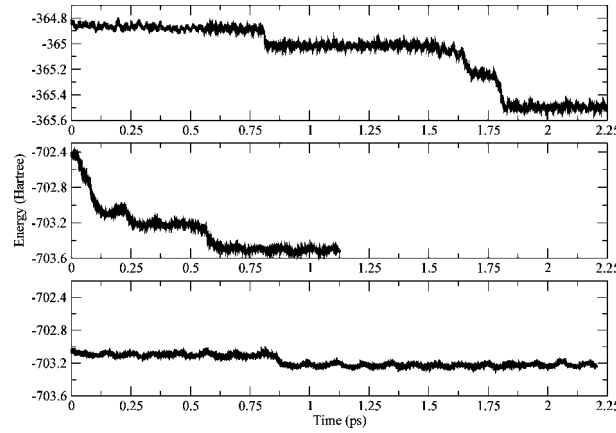


***Figure 2.*** Kohn–Sham energy for three different samples. Parts A, B, and C (see text) go from top to bottom. Samples A and B show more evident drops in energy, corresponding to different reaction events occurring in the sample.

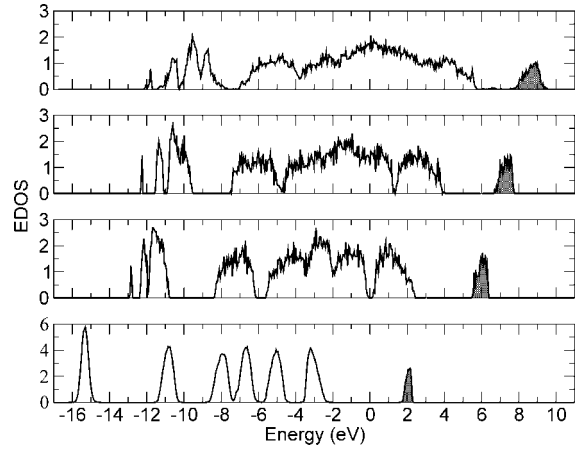


***Figure 3.*** Kohn–Sham eigenvalues for decreasing cell size. The size of the cubic box for samples with 27 molecules from the bottom to the top are 12.8, 9.93, 9.42, and 8.99 Å.

density where the reaction has been observed. For these densities, the electronic density of states (EDOS) has been calculated, and the results for sample A are shown in Figure 3. For the higher-density sample, the calculation of the EDOS was reported before the occurrence of the first reactive event.

The EDOS has not been found to depend significantly on the sample size, except for a normalization factor. It can be seen from the figure that the EDOS continuously broadens with increasing density, thus making the mixing of states possible even by thermal fluctuations. At the highest density, the appearance of states in the HOMO–LUMO gap can be noted. It is to be stressed that these occupied and unoccupied states appear before the occurrence of the first reactive event and are essentially related to the strong molecular distortions that occur in a prereactive system.

The changes in the electron distribution induced by the pressure can be monitored by the electric dipole moment. Since for the reaction mechanism the dipole moment before the onset of the reaction is of interest, calculations were carried only for samples A and C, where the reaction occurs after a sufficiently long simulation time. The distribution of the dipole moment is shown in Figure 4.

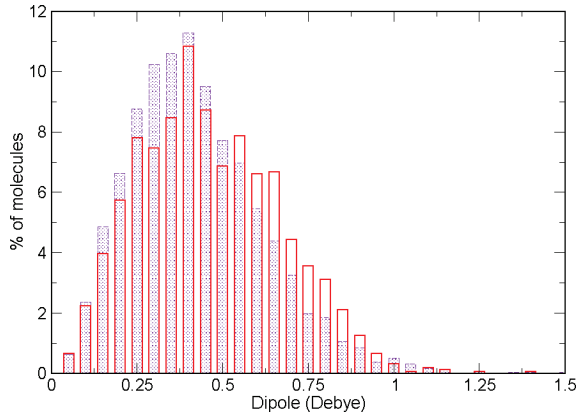


***Figure 4.*** Molecular dipole distribution calculated before reaction in the A (red) and C (blue) samples.

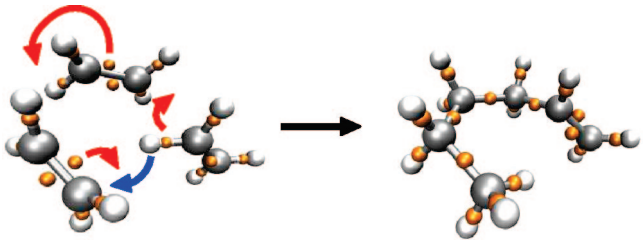

***Figure 5.*** Reaction mechanism for the formation of a trimer. The small spheres depict the position of the WFCs.

The small difference between the two samples should not be ascribed to their size but rather to the different densities. The charge separation effect and transformation of an initially apolar into a polar molecule is a typical response of a system to pressure since this ensures a closer packing of the units (electrostriction). The present results confirms previous observations in high-pressure simulations of propylene[26] and butadiene.[15] The sample polarization suggests that the high-pressure polymerization mechanism is ionic. This conclusion is reinforced by the calculation of the centers of the maximally localized Wannier functions as shown in Figure 5.

The generation of a molecular dipole moment implies a transient geometrical deformation anticipating the change from $sp^2$ to $sp^3$ hybridization of the C atoms. This intermediate transformation preceding the reaction is independent from sample size. As shown by Boero et al.[12] in their study of propene polymerization, the use of Wannier function centers (WFC) is very useful to depict the reaction mechanism with the breaking and forming of chemical bonds.

As to the reaction products obtained in the simulation, we focus the discussion on the samples at higher density (A and B). Only a linear head-to-tail condensation of the monomeric units is observed. This is a relevant result since it is in agreement with experimental findings,[5] showing that the high-pressure polymerization of liquid ethylene produces a linear polymer and, under laser irradiation, even a completely crystalline material. The termination of the chain growth invariably occurs with the formation of $-CH{=}CH_2$ and $CH_3{-}CH_2-$ end groups, which implies an intramolecular hydrogen transfer in the forming oligomer. At the end of the simulation, 44% and 38% of the molecules have reacted in samples A and B, respectively. As a matter of fact, in several experiments, it has been found that at any given pressure the reaction may proceed only incompletely.[2] The incomplete sample transformation should most likely be ascribed to the drop, as the reaction proceeds, in pressure below a reaction threshold. This has been confirmed by a constant pressure simulation on one of the samples with 27 molecules where the reaction further proceeded with the formation of longer chains, and when the simulation was completed, the final products contained a chain with 18 and a chain with 16 carbon atoms, with two smaller oligomers (with four and six C atoms) and only five unreacted molecules. No attempts have been made to check the stability of these chains upon relaxation of the pressure since the molecules obtained appear to be stable species and, furthermore, the molecules extend through the PBC, and this can give rise to spurious effects when a decompression is performed.

Considering the reaction products in more detail, in sample A, two pentamers and one dimer are obtained, while in sample B a decamer, a pentamer, a trimer, and a dimer are produced. In the case of sample C, only the formation of a linear trimer (1-hexene) is observed. This simply implies that in this case the sample density is very close to the reaction threshold.

**3.2. Reactivity in the Crystal.** The molecular dynamics simulation shows that upon applying pressure the compressibility of the ethylene crystal is highly anisotropic. At the end of the NPT simulation, a final volume of 800 $Å^3$ (the simulation cell parameters are 9.825, 8.7, and 9.375 Å and $\beta = 93.4°$) was reached with a contraction of the $b$ axis by 34.2% in comparison with a contraction by 29.3% and 22.5% for the $a$ and $c$ axes, respectively. The high-pressure structure is however still monoclinic with only a small variation of the $\beta$ angle. The shortest intermolecular C−C distance between equivalent molecules is initially along the $a$ axis and, from the initial value of 3.66 Å, reduces to 2.36 Å, whereas the shortest C−C distance between nonequivalent molecules reduces to 2.567 Å from 3.88 Å. These short intermolecular distances will allow a very fast reaction induced by thermal motion.

The high-pressure structure was the starting configuration for a subsequent NVT simulation. After 0.5 ps, a large single-step variation of the Kohn–Sham energy is observed and a reaction occurs involving two steps. First, a dimer is formed, and later, there is a growth of the chain. Initially, the reaction involves molecules on equivalent sites along the $a$ crystallographic axis, leading to the formation of linear chains. In a later stage, nonequivalent molecules in the unit cell also participate in the reaction, with the formation of a branched polymeric chain. This result is in very nice agreement with experiments at high pressures (3.5 GPa)[19] showing, by a kinetic analysis according to Avrami's law,[36,37] that the polymer growth is linear in the initial stages and becomes nonlinear at later stages. In Figure 6, a snapshot of the simulation is reported, showing the formation of a linear and a branched polymeric chain. The two chains are longer than the simulation box parameters and therefore propagate through the periodic boundary conditions; this can be a

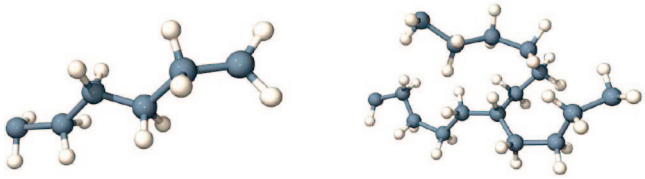

***Figure 6.*** Linear and branched chains obtained in the high-pressure reaction simulations in crystalline ethylene.

source of spurious effects if the stability of the chains is studied upon relaxation of the pressure.

An indication of the electronic reorganization and of the reaction mechanism is obtained from the Wannier functions. In Figure 7a, a molecular orbital localized on the ethylene molecules on equivalent sites is represented. The orbital shows the presence of electron density between neighboring carbons laying along the same axis. This behavior represents an explanation for the formation of the linear chains. In fact, from a configuration of the crystal before the reaction occurs (see Figure 7b), it is possible to note once again the presence of this orbital (even if at higher energy). However, the electron density is localized also between two carbon atoms of ethylene molecules in nonequivalent sites. This allows for the propagation of the reaction along two different selected directions and determines the formation of branched chains. The above outlined mechanism can be considered typical of a crystal and can be attributed to collective motions of the molecules; this make this reaction mechanism quite different from the one described in the disordered system.

## 4. Conclusion

The reaction of ethylene induced by pressure has been found to proceed by an ionic mechanism and to be induced by thermal fluctuations. This suggests that the reaction can be obtained after the same short time at lower densities using larger samples with respect to those employed in this work. As expected, the KS density of states has been found to be strongly modified at high densities. The reactive system is characterized by a strong reduction of the HOMO–LUMO gap and a collapse of the higher occupied states in only one band. The molecules under high pressure are strongly deformed and show appreciable dipole moments supporting the electrostriction interpretation of the sample response to pressurization and an ionic reaction mechanism.

In the liquid, the formation of only linear polymeric chains is observed, in nice agreement with experiments. The length of the chain is limited by the reduction of pressure due to the condensation of the monomeric units and by the sample size. In the crystal, the anisotropy of the compressibility is evidenced, and the shortest C−C intermolecular contacts along the *a* axis and along the unit cell diagonal become comparable. As a consequence, the reaction in the solid proceeds in two steps. Initially, the polymerization along the *a* axis produces linear chains, but in a second stage, the onset of condensation along the diagonal leads to branched chains. The close agreement with experiments shows that the use of ab initio simulations in the Car–Parrinello approach is a quite reliable method to study the mechanism of this kind of reaction at high pressures.

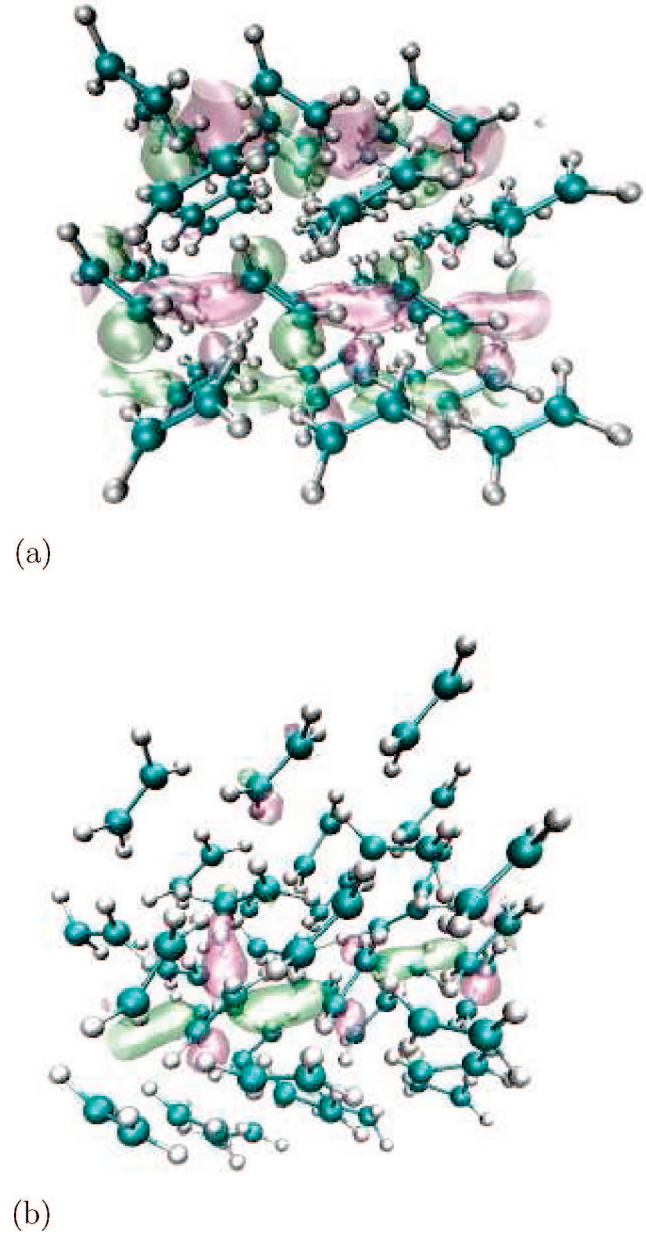


***Figure 7.*** Bonded localized Wannier orbital (a) in the crystal at high pressure and (b) before the reaction starts.

In both systems studied by the analysis of the Wannier centers and the KS eigenvalues, it has been possible to rule out the radicalic mechanism under the thermodynamic conditions examined. The ionic mechanism has been found as in previous simulations on similar systems.[14,15] In the crystal, the polymerization occurs with a collective motion of the molecules.

**Acknowledgment.** The authors would like to thank Prof. Michele Parrinello, for helpful discussions. This work was supported by the Italian Ministero dell'Istruzione, dell'Universita e della Ricerca Scientifica e Tecnologica (MIUR).

## References


(1) Schettino, V.; Bini, R. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1951–1965.

(2) Schettino, V.; Bini, R.; Ceppatelli, M.; Ciabini, L.; Citroni, M. *Adv. Chem. Phys.* **2005**, *131*, 105–242.

(3) Schettino, V.; Bini, R. *Chem. Soc. Rev.* **2007**, *36*, 869–880.

(4) Citroni, M.; Ceppatelli, M.; Bini, R.; Schettino, V. *Science* **2002**, *295*, 2058–2060.

(5) Chelazzi, D.; Ceppatelli, M.; Santoro, M.; Bini, R.; Schettino, V. *Nat. Mater.* **2004**, *3*, 470.

(6) Bini, R. *Acc. Chem. Res.* **2004**, *37*, 95–101.

(7) Cohen, M. *Angew. Chem., Int. Ed.* **1975**, *14*, 386.

(8) Drickramer, H.; Frank, C. *Electrostatic transitions and the high pressure chemistry and physics of solids*; Chapman and Hall: London, 1973.

(9) Boero, M.; Parrinello, M.; Terakura, K. *J. Am. Chem. Soc.* **1998**, *120*, 2746–2752, and references therein.

(10) Vanka, K.; Xu, Z.; Seth, M.; Ziegler, T. *Top. Catal.* **2005**, *34*, 143–164.

(11) Kim, S.; Somorjai, G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15289–15294.

(12) Boero, M.; Parrinello, M.; Hüffer, S.; Weiss, H. *J. Am. Chem. Soc.* **2000**, *122*, 501–509, and references therein.

(13) Bernasconi, M.; Chiarotti, G. L.; Focher, P.; Parrinello, M.; Tosatti, E. *Phys. Rev. Lett.* **1997**, *78*, 2008–2011.

(14) Mugnai, M.; Cardini, G.; Schettino, V. *J. Chem. Phys.* **2004**, *120*, 5327–5333.

(15) Mugnai, M.; Cardini, G.; Schettino, V. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2004**, *70*, 020101(1–4).

(16) Ciabini, L.; Gorelli, F. A.; Santoro, M.; Bini, R.; Schettino, V.; Raugei, S. *Nat. Mater.* **2007**, *6*, 39–43.

(17) Raugei, S.; Cardini, G.; Schettino, V. *Mol. Phys.* **1998**, *95*, 477–482.

(18) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.

(19) Chelazzi, D.; Ceppatelli, M.; Santoro, M.; Bini, R.; Schettino, V. *J. Phys. Chem. B* **2005**, *109*, 21658.

(20) Ciccotti, G.; Ryckaert, J. *Comput. Phys. Rep.* **1986**, *4*, 345–392.

(21) Williams, D. E. *J. Chem. Phys.* **1967**, *47*, 4680.

(22) Press, W.; Eckert, J. *J. Chem. Phys.* **1976**, *65*, 4362.

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

(24) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.

(25) Hutter, J.; Alavi, A.; Deutch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckerman, M.; Parrinello, M. CPMD; MPI für Festkörperforschung and IBM Zurich Research Laboratory: Stuttgart, 1995.

(26) Mugnai, M.; Cardini, G.; Schettino, V. *J. Chem. Phys.* **2004**, *120*, 5327.

(27) Troullier, N.; Martins, J. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 1993–2006.

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

(29) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

(30) Hoover, W. G. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695.

(31) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.

(32) Mazari, N.; Vanderbilt, D. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *56*, 12848–12865.

(33) Silvestrelli, P.; Mazari, N.; Vanderbilt, D.; Parrinello, M. *Solid State Commun.* **1998**, *107*, 7–11.

(34) Silvestrelli, P. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 9703–9706.

(35) Berghold, G.; Mundy, C.; Romero, A.; Hutter, J.; Parrinello, M. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2000**, *61*, 10040–10048.

(36) Avrami, M. *J. Chem. Phys.* **1939**, *7*, 1103.

(37) Avrami, M. *J. Chem. Phys.* **1940**, *8*, 212.


CT700275V

JCTC Journal of Chemical Theory and Computation

# Molecular Dynamics of Surface-Moving Thermally Driven Nanocars

Alexei V. Akimov,[†,§] Alexander V. Nemukhin,[†,‡] Alexander A. Moskovsky,[†,‡] Anatoly B. Kolomeisky,[*,§] and James M. Tour[§]

*Department of Chemistry, M.V. Lomonosov Moscow State University, Leninskie Gory, 1/3, Moscow, 119991, Russian Federation, N.M. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, ul. Kosygina, 4, Moscow, 119994, Russian Federation, and Department of Chemistry, Rice University, Houston, Texas 77005*

Received October 3, 2007

**Abstract:** We developed molecular models describing the thermally initiated motion of nanocars, nanosized vehicles composed of two to four spherical fullerene wheels chemically coupled to a planar chassis, on a metal surface. The simulations were aimed at reproducing qualitative features of the experimentally observed migration of nanocars over gold crystals as determined by scanning tunneling microscopy. Coarse-grained-type molecular dynamics simulations were carried out for the species "Trimer" and "Nanotruck", the simplified versions of the experimentally studied nanomachines. Toward this goal, we developed a version of the rigid body molecular dynamics based on the symplectic quaternion scheme in conjunction with the Nose−Poincare thermostat approach. Interactions between rigid fragments were described by using the corrected CHARMM force field parameters, while several empirical models were introduced for interactions of nanocars with gold crystals. With the single adjusted potential parameter, the computed trajectories are consistent with the qualitative features of the thermally activated migration of the nanocars: the primary pivoting motion of Trimer and the two-dimensional combination of translations and pivoting of Nanotruck. This work presents a first attempt at a theoretical analysis of nanocars' dynamics on a surface by providing a computationally minimalist approach.

## Introduction

Fullerene-based nanometer-sized molecular structures, called nanocars, which have demonstrated a wheel-like rolling motion along metal surfaces,[1–3] are among the artificial molecular systems that mimic some of the properties of biological molecular motors. The latter include active enzyme molecules that move progressively along such tracks as protein filaments or nucleic acids by transforming chemical energy into mechanical work.[4] Nanocars are single-molecule vehicles composed of two to four spherical fullerene wheels that are chemically coupled to a planar chassis and sometimes bearing a loading bay. Their thermally initiated migration on the surface of gold crystals was determined by scanning tunneling microscopy techniques.[3]

To support the eventual goals of these studies, specifically the understanding and control of surface transport by such entities, it is important to develop theoretical models of nanocars and explore modern computational approaches for detailed analyses of their motion. Molecular dynamics (MD) simulations provide a convenient tool for the investigation of such complex dynamic systems. Theoretical computations are supposed to reproduce the observed features of the thermally activating surface movements of nanocars.[2,3] According to these experimental findings, a four-wheeled nanocar remained stationary on the gold surface up to 170 °C, and as the temperature increased to 200–225 °C, the molecule began to move in two dimensions (2D) through a combination of both translation and pivoting. When a three-wheeled nanocar was heated to 225 °C, only occasional

* Corresponding author e-mail: tolya@rice.edu.
† M.V. Lomonosov Moscow State University.
§ Rice University.
‡ Russian Academy of Sciences.

10.1021/ct7002594 CCC: $40.75 
Published on Web 03/14/2008

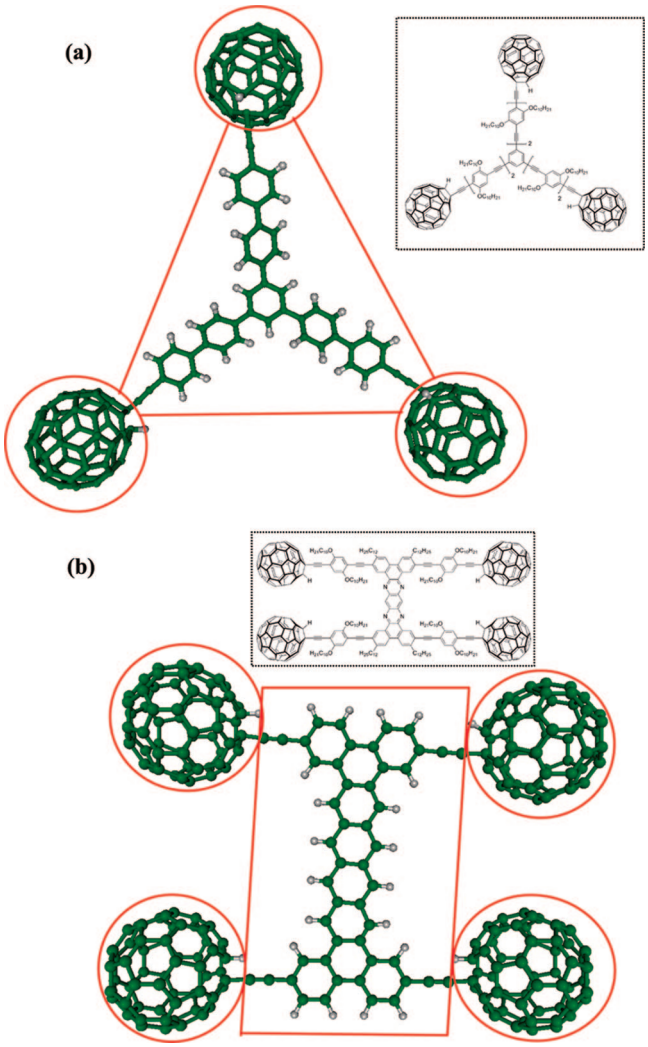


***Figure 1.*** Molecular models of nanocars used in molecular dynamics simulations: (a) Trimer and (b) Nanotruck. The insets show chemical structures of the parent nanocars.[3] Partitioning the molecules into rigid fragments (four for Trimer and five for Nanotruck) is clarified by the red lines.

surface diffusion was observed, and the majority of motion was pivoting in space around a central pivot point. The MD models that are capable of replicating these physical movements may be used in the future to understand and optimize the fabrication of the next generation of nanocars.

In this work, we consider two hydrocarbon analogs of fullerene-based nanocars originally designated[3] as "Trimer" (a three-wheeled vehicle) and "Nanotruck" (a four-wheeled vehicle). The panels of Figure 1 illustrate the models, showing chemical structures of the parent nanocars in the insets and the corresponding analogs considered here and shown in ball-and-stick representations. Starting with the structures of the experimentally studied species[3] (insets in Figure 1), we introduced simplifying modifications to the molecular structure in order to reduce computational expenses in this first attempt to model nanocar dynamics. Specifically, we kept the alkynyl connectors (−C≡C−) to the fullerene wheels, but we replaced the other alkynyl moieties with direct aryl−aryl linkages. As discussed below, we maintained the rigidity of the parent chassis (insets in Figure 1) by using the rigid body MD simulations. The most significant features of the parent nanocars, the rolling motion of the wheels, should be the same in our model vehicles.

For simulations of the thermally activated surface transport of the nanocar models, we applied an original version of the rigid-body molecular dynamics, an extension of the previous approach.[5] This method allowed us to concentrate on the most essential dynamical features of nanovehicles on the surface: the wheels were allowed to rotate, but the wheels and chassis were represented by rigid fragments. As shown in Figure 1, four rigid fragments (three wheels and a chassis) were assumed for Trimer, and five rigid fragments (four wheels and a chassis) were considered for Nanotruck. Application of rigid-body MD helped us to drastically reduce an array of internal coordinates, to simplify forms of interaction potentials, to decrease the amount of potential parameters, and to afford fairly long MD trajectories.

In preliminary steps, we implemented and tested several algorithms of rigid-body MD in the canonical ensemble (NVT) and found that the symplectic quaternion scheme described, in particular, by Miller et al.,[6] in conjunction with the Nose−Poincare thermostat approach,[7] was an optimal choice for constant temperature molecular dynamics calculations of nanocars. It should be noted that, in this work, similarly to other molecular dynamics investigations, we did not study how the presence of a thermostat influences the dynamics of nanocars.

All MD runs were performed for an integration time step of 1 fs. The trajectories were typically about 5 ns lengths. Interactions between rigid fragments were modeled by the site–site potential composed of the Lennard-Jones and Coulomb functions in eq 1.

$$E_{\alpha\beta} = \sum_{i\in\alpha}\sum_{j\in\beta}\left\{4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{r_{ij}}\right\} \qquad (1)$$

Here, $i$ and $j$ are the indices of atoms belonging to fragments $\alpha$ and $\beta$, respectively, $r_{ij}$ is the distance between atoms $i$ and $j$, and $q_i$ and $q_j$ are partial charges on the respective atoms. Potential parametrization was basically consistent with the CHARMM27 force field[8,9] but corrected in this work as explained below. The hydrogen atoms were assigned to type HA(1) with the parameters $\epsilon = 0.022$ kcal/mol and $\sigma = 1.32$ Å; the carbon atoms were assigned to type CA(11), except those connecting fullerene wheels with the chassis (see Figure 2). For the latter, type CC(12) was assumed. The Lennard-Jones parameters of all carbon atoms were taken as $\epsilon = 0.070$ kcal/mol and $\sigma = 1.9924$ Å. The partial charges on atoms were determined in *ab initio* quantum chemical calculations in the RHF/6-31G** approximation by using the PC GAMESS program[10] as the Mulliken charges for the compounds benzene, biphenyl, biphenylacetylene, and phenylfullerenylacetylene, occurring as fragments of model nanovehicles (Figure 1). The atom types and computed partial charges are shown in Figure 2. The remaining fullerene carbon atoms were assumed to be CA(11)-type atoms with zero partial charge.

The parameters required for the bonded interactions when using the CHARMM force fields were selected as follows. For the CA(11)−CC(12) stretching term: $k_b = 320.0$ kcal/mol and $r_0 = 1.400$ Å. For the CA(11)−CC(12)−CA(11)

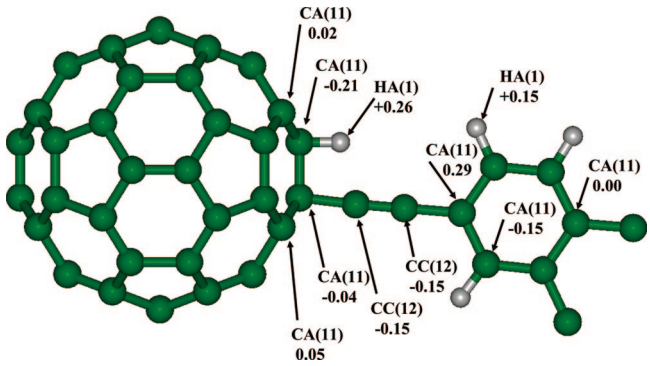


***Figure 2.*** An assignment of the atoms in the model nanocars (Figure 1) to the CHARMM force field types and their partial charges as computed by *ab initio* calculations in this work.

bending term: $k_a = 40.0$ kcal/mol and $\theta_0 = 170.0°$. For the CA(11)−CC(12)−CC(12) bending term: $k_a = 40.0$ kcal/mol and $\theta_0 = 120.0°$. The torsional terms were excluded because of their negligible contribution to the total energy. Some precautions were introduced to the computer code for the bending potential in order to exclude singularities of the potential when the CA(11)−CC(12)−CC(12) angle approached 180°. We verified that such a constructed total interaction potential was consistent with the small rotational barrier (1.0 kcal/mol) of the fullerene-wheel structures without a surface, as estimated in ref 3.

Several models have been considered for the nanocar−surface interaction potentials assuming that each atom of a nanocar experiences an external force from the resting metal. All models are based on the pairwise 12–6 Lennard-Jones potential function in eq 2:

$$E_W = 4\epsilon\left[\left(\frac{\rho}{d}\right)^{12} - \left(\frac{\rho}{d}\right)^{6}\right] \tag{2}$$

where $d$ is the distance from a particular atom of a nanocar to a specific point in the metal. We considered parameters of these interaction potentials close to the values reasonable for a description of the gold crystals. In particular, the lattice parameter $a = 4.07$ Å was taken from the *fcc* gold lattice.[11] We found in preliminary calculations that the geometry parameter $\rho$ in eq 2 did not strongly affect the essential features of nanocar dynamics. Therefore, we selected the value corresponding to the gold van der Waals atomic radius 1.66 Å.[11]

The most uncertain parameter in eq 2 is the potential well depth $\epsilon$. An initial selection was the value 0.47 kcal/mol corresponding to the molecular mechanical parameters of metals ($Ca^{2+}$, $Ba^{2+}$) from the AMBER force field set;[12] however, we allowed its large variations within 2 orders of magnitude, from 4.7 to 0.094 kcal/mol. In recent MD simulations of gold nanoparticles inside carbon nanotubes, Schoen et al. used the 12–6 Lennard-Jones potential function with parameters $\epsilon = 0.3$ kcal/mol and $\rho = 3$ Å for interactions between gold and carbon atoms, conditions reasonably consistent with our choice. Narrowing the range of the suitable values of $\epsilon$ and better theoretical justification of $\epsilon$ are subjects of our current work.

Within the simplest *Model-W*, each atom of a nanocar interacts with a structureless metal surface (or with a rigid *wall*) calculated by the potential (eq 2), where $d$ is the distance from an atom to the surface plane. Correspondingly, the only nonzero components of forces acting on the machine's atoms are directed along the normal to the surface. Summation over all atoms of a vehicle defines a total interaction potential $E_W$. As discussed below, this model is capable of explaining the thermal activation of Trimer's pivoting but fails to describe translations of Nanotruck.

The more advanced *Model-L* assumes certain features of a *lattice* atomic structure of metals. When this model is applied, the current coordinates of a nanocar's atom are projected on the surface plane, and the nearest to this point from the surface atom of the *fcc* lattice serves as a reference node to construct the interacting shell. To this end, several (9 or 25) nearest lattice points from the surface layer and from the one underneath the surface contribute to the interaction potential calculated as a sum of the terms defined by eq 2. The procedure is repeated for all atoms of a nanocar. This is the most computationally expensive approach for molecular dynamics calculations.

*Model-P* takes into account the 2D *periodicity* of the surface potential. The corresponding total potential $E_P$ acting on a nanocar is related to the wall-like potential $E_W$ by eq 3:

$$\begin{aligned} &E_P(x, y, z) = f(x, y)\, E_W(z) \\ &f(x, y) = f(x + na, y + ma) \\ &f(x, y) = 1 + \frac{\left[\sin^2\left(\frac{\pi x}{a}\right) + \sin^2\left(\frac{\pi y}{a}\right)\right]}{2} \end{aligned} \tag{3}$$

where coordinates $z$ are counted along the normal to the metal surface, coordinates $x$ and $y$ are in the surface plane, $n$ and $m$ are integers, and $a$ is the lattice parameter. Application of this model allows us to qualitatively describe the translational motion of nanocars.

We also considered a model accounting for charge polarization of the metal. In this approach, we introduced, for each atom with a nonzero partial charge in the nanocar's structure, its image inside the metal with a charge of the opposite sign. Then, the Coulombic contributions from the nanocar−metal interactions were added to the total interaction potential $E_W$. Since no noticeable improvements over other models were found, we do not further discuss the results of this approximation.

The most essential results of MD simulations are as follows. In complete accord with experimental observations,[2,3] pivoting around the central point was the primary movement for Trimer. With *Model-L*, surface rotations mixed with weak displacements of Trimer were observed. At temperatures lower than 300 K (27 °C), the nanocar remained stationary; at $T = 400$ K (127 °C), the mobility was initiated, and at $T = 500$ K (227 °C), we could observe an apparent rotational rolling motion. The sensitivity of the results to the parameter $\epsilon$ of eq 2 is demonstrated in Table 1. These data show that surface dynamics of model nanocars depend on the parameters of this simple potential (eq 2) in a complicated fashion; we do not observe monotonous behavior of the computed trajectories by varying $\epsilon$.

We illustrate the pivoting motion of Trimer on the gold surface in Figure 3 as well as in the animated movies

***Table 1.*** Dependence of the Mode of Surface Movement of Trimer (Figure1a) on the Parameter $\epsilon$ of Potential eq 2[a]

| temperature, K | $\epsilon$ = 4.7 kcal/mol | $\epsilon$ = 0.47 kcal/mol | $\epsilon$ = 0.094 kcal/mol |
|---|---|---|---|
| 300 | weak rotation, no displacement | no rotation, no displacement | weak rotation, weak displacement |
| 500 | rotation, no displacement | weak rotation, weak displacement | rotation, displacement |
| 600 | rotation, no displacement | | |

[a] Other parameters of potential eq 2 are as follows: $\rho$ = 1.66 Å and $a$ = 4.07 Å.

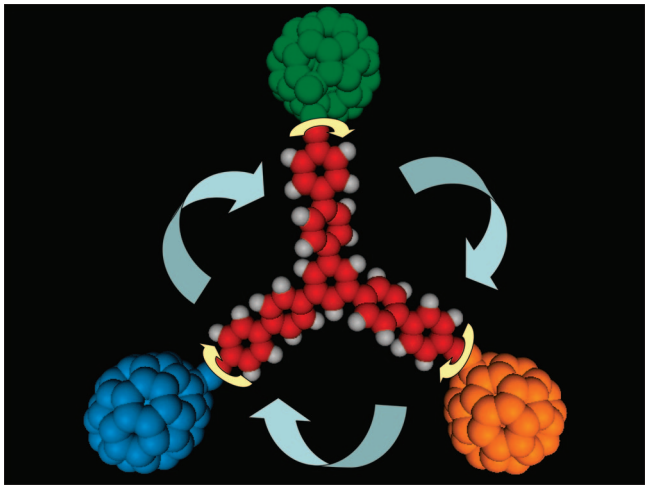

***Figure 3.*** The pivoting mode of surface movement of Trimer. The animated movies are presented in the Supporting Information (trimer_300K.avi and trimer_500K.avi).

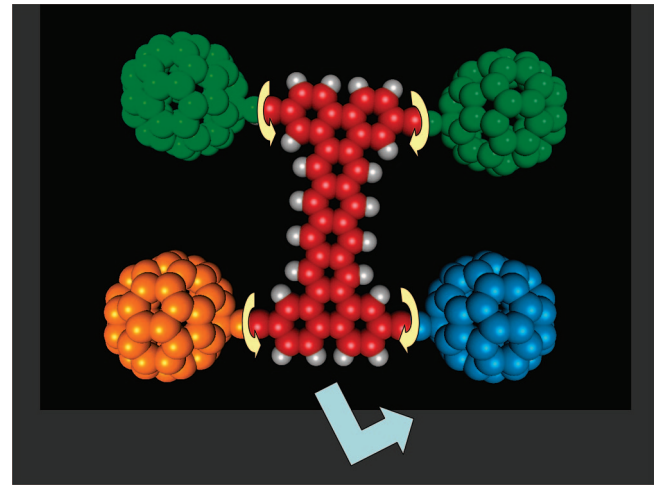

***Figure 4.*** The mode of surface movement of Nanotruck. The animated movies are in the Supporting Information (ntruck_300K.avi and ntruck_500K.avi).

presented as Supporting Information: trimer_300K.avi ($T$ = 300 K) and trimer_500K.avi ($T$ = 500 K). These data refer to simulations with *Model-L* with the potential parameters $\rho$ = 1.66 Å, $a$ = 4.07 Å, and $\epsilon$ = 4.7 kcal/mol. Application of *Model-P* permits the observation of a combination of strong rotations and non-unidirectional displacements of Trimer along the surface.

If we define the rotational velocity, $k_{rot}$, of Trimer as the ratio of the (noninteger) number of rotation circles (for example, 5.5 at $T$ = 600 K) to the trajectory length (5 ns) and assume the temperature dependence of the rotational velocity in the Arrhenius form (eq 4):

$$k_{rot} = A\, e^{-\frac{E_a}{RT}} \qquad (4)$$

then plotting the values $k_{rot}$ versus $T^{-1}$ (300 K $\leq T \leq$ 800 K) allows us to estimate the slope of the graph and, therefore, an activation energy barrier $E_a$ for thermally driven Trimer rotation. With the trajectories of *Model-W,* such an estimate results in $E_a$ values between 3 and 4 kcal/mol.

For the four-wheeled Nanotruck, an application of *Model-W* was an unsuccessful attempt to detect the 2D movement of the vehicle. However, applications of *Model-P* or the more expensive *Model-L* allowed us to observe a combination of both translation and pivoting. Within the latter models, the use of potential parameter $\epsilon$ = 4.7 kcal/mol in eq 2, corresponding to the strongest interaction of Nanotruck with the metal surface, precluded migration of the nanovehicle even at higher temperatures (400 and 500 K). For the smaller values of $\epsilon$, a desired type of motion (two-dimensional combination of both translation and pivoting) could be observed. We illustrate the movement of Nanotruck on the gold surface in Figure 4 as well as in the animated movies

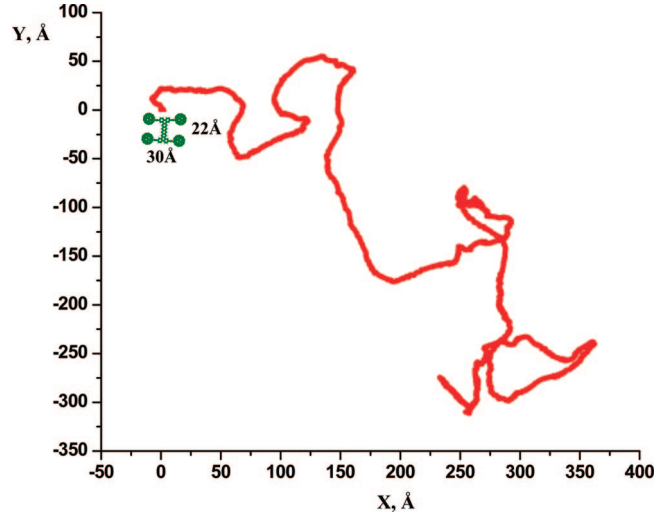


***Figure 5.*** Two-dimensional trajectory of Nanotruck in the surface plane ($X$,$Y$) as computed with *Model-L* at $T$ = 500 K. The shape of the moving molecule is shown for comparison in green.

in the Supporting Information: ntruck_300K.avi ($T$ = 300 K) and ntruck_500K.avi ($T$ = 500 K). These data refer to simulations with *Model-L* with the potential parameters $\rho$ = 1.66 Å, $a$ = 4.07 Å, and $\epsilon$ = 0.47 kcal/mol.

In Figure 5, we show a typical 2D trajectory of Nanotruck at $T$ = 500 K as computed with *Model-L* with the potential parameter $\epsilon$ = 0.47 kcal/mol. According to these MD simulations, the nanovehicle can experience large-scale migrations, especially compared to its size, shown in green in Figure 5.

Our results indicate that lattice structure and periodicity of the surface play a critical role in the dynamics of thermally

driven nanocars. This suggests the following mechanism of transport: nanocars are bound to the surface by strong bonds (covalent or electrostatic), and they move along the surface by overcoming the barriers with the help of thermal fluctuations; that is, the dynamics of nanocars are an activated process. The experimentally determined existence of a temperature threshold below which no motion is observed, and the strong temperature dependence of the dynamics, supports the computational findings here.

To summarize the results of this work, we emphasize that molecular models have been developed for surface-moving analogs of fullerene-based nanocars, which successfully reproduce experimentally observed[2,3] thermal activation and dynamical features of the nanovehicles. This paper presents a first attempt in the theoretical analysis of these complex systems by providing a computationally minimalist approach that allows one to understand the essential features of nanocar dynamics upon a surface. Our theoretical analysis suggests that the dynamics of nanocars is an activated process that strongly depends on the structure and periodic properties of the surface. In addition, we provide an estimate for the activation energy of thermally driven rotation of Trimer (3–4 kcal/mol) that can be measured in experiments to test the validity of our approach.

**Acknowledgment.** The authors thank S.S. Konyukhov for help with the rigid body molecular dynamics algorithms. This work was partially supported by grants from the Russian Foundation for Basic Research (projects #06-03-33009 and #05-07-90146), from the SKIF-GRID program, the NSF NIRT (ECCS-0708765), and Penn State MRSEC.

**Supporting Information Available:** Animated movies of typical trajectories of Trimer and Nanotruck. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Shirai, Y.; Morin, J.-F.; Sasaki, T.; Guerrero, G. M.; Tour, J. M. Recent progress on nanovehicles. *Chem. Soc. Rev.* **2006**, *35*, 1043–1055.

(2) Shirai, Y.; Osgood, A. J.; Zhao, Y.; Kelly, K. F.; Tour, J. M. Directional Control in Thermally Driven Single-Molecule Nanocars. *Nano Lett.* **2005**, *5*, 2330–2334.

(3) Shirai, Y.; Osgood, A. J.; Zhao, Y.; Yao, Y.; Saudan, L.; Yang, H.; Yu-Hung, C.; Alemany, L. B.; Sasaki, T.; Morin, J.-F.; Guerrero, J. M.; Kelly, K. F.; Tour, J. M. Surface-Rolling Molecules. *J. Am. Chem. Soc.* **2006**, *128*, 4854–4864.

(4) Kolomeisky, A. B.; Fisher, M. E. Molecular Motors: A Theorist's Perspective. *Annu. Rev. Phys. Chem.* **2007**, *58*, 675–695.

(5) Moskovsky, A. A.; Vanovschi, V. V.; Konyukhov, S. S.; Nemukhin, A. V. Implementation of the replica-exchange molecular dynamics method for rigid bodies. *Int. J. Quantum Chem.* **2006**, *106*, 2208–2213.

(6) Miller, T. F.; Eleftheriou, M.; Pattnaik, P.; Ndirango, A.; Newns, D.; Martyna, G. J. Symplectic quaternion sheme for biophysical molecular dynamics. *J. Chem. Phys.* **2002**, *116*, 8649–8659.

(7) Bond, S. D.; Leimkuhler, B. J.; Laird, B. B. The Nose-Poincare Method for Constant Temperature Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 114–134.

(8) MacKerrell, A. D., Jr.; Bashford, D.; Bellott, R. L.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(9) Foloppe, N.; MacKerrell, A. D., Jr. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21*, 86–104.

(10) Nemukhin, A. V.; Grigorenko, B. L.; Granovsky, A. A. Molecular Modeling with the PC GAMESS. *Moscow Univ. Chem. Bull.* **2004**, *45*, 75–102.

(11) Winter, M. Chemistry: WebElements Periodic Table, Professional Edition; The University of Sheffield and WebElements Ltd: Sheffield, U.K. http://www.webelements.com/webelements/elements/text/Au/xtal.html (accessed Sept 10, 2007).

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

CT7002594

JCTC Journal of Chemical Theory and Computation

# Implicit Solvent Models and the Energy Landscape for Aggregation of the Amyloidogenic KFFE Peptide

Birgit Strodel and David J. Wales*

*University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, U.K.*

Received November 9, 2007

**Abstract:** This study compares the performance of four implicit solvent models in describing peptide aggregation. The solvent models are the effective energy function-1 (EEF1) and three generalized Born (GB) models: one following the original implementation of Still (GB1), the analytical continuum electrostatics (ACE) potential, and GB with "simple switching" (GBSW). For each solvent model the first step of aggregation, namely dimerization, is investigated for the KFFE peptide, which is one of the shortest peptides known to form amyloid fibrils in vitro. Using basin-hopping for global optimization and replica exchange molecular dynamics simulations, we conclude that of the four solvent models considered, the EEF1 potential provides the most reliable description for the formation of KFFE amyloid precursors. It produces results that are closest to the experimental findings of a partial $\beta$-strand conformation for the KFFE peptide in solution along with the formation of fibrils exhibiting antiparallel $\beta$-strand structure. The ACE and GB1 potentials also show a significant $\beta$-propensity for the KFFE peptide but fail to produce stable KFFE dimers. The GBSW potential, on the other hand, supports a very stable antiparallel dimer structure, but in a turn rather than a $\beta$ conformation.

## 1. Introduction

Implicit solvent models are based on the assumption that ensemble averages of interactions between the solute and explicit solvent molecules can be approximated by a mean field formalism, which is expressed as a function of the solute configuration alone. Explicit solvent molecules can then be omitted, thereby reducing the computational cost due to the reduced system size and the absence of solvent relaxation. Several approaches exist to describe the solute−solvent interactions by their mean field behavior. In many of these models the solvation free energy, $\Delta G_{solv}$, is decomposed into electrostatic (el) and nonpolar (np) contributions, $\Delta G_{solv} = \Delta G_{el} + \Delta G_{np}$. The nonpolar contribution to $\Delta G_{np}$ includes the energetic penalty for forming a cavity in the solvent and the solute−solvent van der Waals (vdW) dispersion interactions.[1] In most implicit solvent models $\Delta G_{np}$ is treated entirely empirically by approximating it as product of the solvent-exposed surface area of the solute and a phenomenological surface tension coefficient. For the calculation of the electrostatic contribution the solvent is treated as a homogeneous, high-dielectric medium, which surrounds the cavity containing the low-dielectric solute with explicit charges at the atomic centers. The Poisson–Boltzmann (PB) equation provides a rigorous formulation for the calculation of $\Delta G_{el}$ and can be solved numerically for the electrostatic potential throughout space. However, the cost involved in obtaining accurate PB results and the difficulties in obtaining continuous first derivatives have limited the application of PB theory in biomolecular simulations; in most applications, the method is used as a benchmark for other implicit solvent models.

An alternative approach is provided by the generalized Born (GB) formalism, which is based on the same continuum electrostatic model for the solvent as the PB theory, but is computationally more efficient and can reproduce the PB results accurately. The starting point is the Born formula,

$$\Delta G_{el} = -\frac{\tau q^2}{2R} \quad \text{with} \quad \tau = \frac{1}{\epsilon_p} - \frac{1}{\epsilon_s} \tag{1}$$

which is the analytical solution of the PB equation for a single charge $q$ in a sphere of radius $R$ with (low) dielectric constant $\epsilon_p$, embedded in a solvent with (high) dielectric

* Corresponding author. E-mail: dw34@cam.ac.uk.

10.1021/ct700305w CCC: $40.75 
Published on Web 03/14/2008

constant $\epsilon_s$. In the GB model, this formula is generalized for a polyatomic system occupying a more complex shape. One of the most reliable and widely used GB formulations is that of Still et al.,[2]

$$\Delta G_{el} = -\frac{\tau}{2}\sum_{i,j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i^{GB}R_j^{GB}\exp(-r_{ij}^2/4R_i^{GB}R_j^{GB})}} \quad (2)$$

where $r_{ij}$ is the distance between atoms $i$ and $j$, $q_i$ and $q_j$ are their respective charges, and $R_i^{GB}$ denotes the effective Born radius. The Born radius can be understood as the distance from an atom to the dielectric boundary, and successful application of the GB model depends upon an appropriate evaluation of this quantity. In most GB methods the so-called Coulomb field approximation (CFA) is invoked, which is exact for a charge in the center of a spherical cavity and assumes that the dielectric displacement is Coulombic in form and remains so even when the external dielectric is altered in the solvation process. The CFA allows us to calculate the self-electrostatic energy, $\Delta G_{el,i}$, of atom $i$, and comparison with the Born equation (1) results in

$$\frac{1}{R_i^{GB}} = -\frac{2\Delta G_{el,i}}{\tau q^2} = \frac{1}{R_i} - \frac{1}{4\pi}\int_{solute,r>R_i}\frac{1}{r^4}\,dV \quad (3)$$

where $R_i$ is the atomic radius, e.g., vdW radius, used to define the solute cavity filling the volume $V$, over which the integral is calculated.

Most of the GB methods differ in how the volume integral in eq 3 is calculated. It can be approximated as a discrete sum of overlapping spheres[3–6] or Gaussians[7] or by carrying out the integration numerically, either by reformulating the volume integral into a surface integral,[8] using a cubic integration lattice,[9] or employing integration techniques adapted from density functional theory.[10] The success of a GB model in reproducing PB results depends on how the solvent boundary at the molecular surface is defined for the evaluation of the volume integral, and various improvements addressing this issue have been proposed in recent years.[5,11] Another limitation of GB models, which might lead to deviations from the PB result, is the CFA, which is not valid for charges in nonspherical cavities. In the latest GB models, empirical higher-order correction terms that extend eq 3 beyond the CFA have been suggested, and it was shown that this approach further improves the solvation energies.[6,10,12,13]

In this study, we investigate the performance of three different GB models, which are implemented in CHARMM.[14] Two of them, called GB1[15] and ACE,[7] use the pairwise summation of atomic volumes to approximate the volume integral (3), while the third, GBSW,[11] employs a direct evaluation of the molecular volume. Alternative implicit solvent models concentrate on the screening of electrostatic interactions in the solute due to solvation. One of these methods is the effective energy function EEF1, which is also implemented in CHARMM and modulates the screening as a function of the solvent-excluded volume. The EEF1 solvation model is also considered in the present contribution.

Implicit solvent models have been used successfully for a variety of applications, such as distinguishing native and near-native structures from nonnative decoys[16–21] and studying polypeptide folding and unfolding in atomic detail.[22–24] In recent years, there have also been a number of comparative studies of the various implicit solvent models aiming to identify their advantages and limitations for particular applications. The outcome of these studies is as different as the systems investigated and it seems that, at least for the moment, no single model performs best in all situations. Steinbach, for instance, found that EEF1 performed better than ACE and the surface-area model SASA[25] in identifying near-native states for the three peptides trp-cage, BS1 and U(1-17)T9D.[21] Stultz concluded that only the explicit solvent model TIP3P[26] was able to correctly predict the experimental end-to-end distance distribution for a twelve-residue peptide from rat tyrosine hydroxylase, while none of the implicit solvent models EEF1, ACE, GB1, and SASA were successful.[27] Huang and Stultz investigated whether the three implicit solvent models EEF1, GB1, and GBSW could reproduce the set of potential energy minima obtained from explicit solvent simulations with the TIP3P water model for the six-residue peptide PHF6.[28] They concluded that all three implicit solvent models were in good agreement with the results from explicit solvent simulations and that EEF1 provided the best representation of the most favored conformations. However, in an assessment of various implicit solvent models for the description of peptide−surface interactions, EEF1 performed badly and ACE was best when compared to results obtained from density functional theory combined with the self-consistent reaction field implicit solvent model.[29]

The objective of the present work is to compare the results for EEF1, ACE, GB1, and GBSW when applied to peptide aggregation. To this end, we consider the tetrapeptide KFFE, which is one of the shortest peptides known to form amyloid fibrils in vitro.[30] In this study, we focus on the first step in the aggregation process, namely dimerization. In a study by Baumketner and Shea,[31] this process was investigated using the GB1 solvent model and compared to experiment.[30] In other simulations the formation mechanism of larger KFFE oligomers, ranging from tetramers to heptamers, was analyzed in detail using a combination of all-atom molecular dynamics (MD) simulations in explicit solvent and the activation-relaxation technique coupled with a coarse-grained potential.[32–35] The focus of the current work, however, is not to investigate the mechanism of dimerization but rather to study the influence of the potentials on the conformations of the dimer. Implicit solvent models have already been used in simulations examining the process of oligomer formation of amyloidogenic peptides,[31,36–39] but we are not aware of any study that has checked whether implicit solvent models are capable of capturing the correct mechanistic details and oligomeric structures. Compared to investigations of the conformational equilibria of a single peptide, simulations of peptide oligomerization entail an additional layer of complexity. Not only do the solvation forces and intramolecular forces have to be represented correctly but intermolecular forces come into play. In this sense, the KFFE peptide is an

interesting test case, since it is short with opposite charges at the termini, leading to strong competition between Coulomb interactions within the peptide, between the peptides, and between the peptide and the solvent. In addition, the two middle residues of KFFE are aromatic and hydrophobic and, hence, probably promote the dimerization as a result of hydrophobic and $\pi-\pi$ interactions. Such competing forces all have to be accounted for in the potential and any imbalance is likely to be particularly noticeable for relatively short and highly charged peptides.

To assess the performance of the four implicit solvent models in question in describing the aggregation of KFFE, we used basin-hopping[40–43] (BH) global optimization and replica exchange molecular dynamics[44] (REMD) to explore the energy landscape of the KFFE monomer and dimer. In the next section, more details of the BH approach and REMD are provided, followed by the results, which vary significantly between the implicit solvent models. Our discussion focuses on the origin of these differences and points out possible opportunities for improvement. In Appendix A, we describe the solvent models, and in Appendix B, we outline the details of a REMD simulation using explicit TIP3P water,[26] which we have performed for comparison. In Appendix C, we discuss problems regarding the numerical convergence that we encountered during calculations with the GBSW implicit solvent model.

## 2. Methods

Experimentally,[30] the KFFE peptide has charged termini, $NH_3^+$-Lys-Phe-Phe-Glu-$COO^-$, and we performed simulations of the corresponding system using the CHARMM software (version c31a2).[14] Four implicit water models, EEF1,[45] ACE,[7] GB1,[15] and GBSW[11] were used in combination with three protein empirical potentials. The united-atom CHARMM19 parameters[46] were employed for the ACE and GB1 solvent models, a modified version of CHARMM19 was used with the EEF1 solvent model,[45] and the all-atom CHARMM22 potential[47] along with the CMAP modifications[48–50] was employed for the GBSW solvent model. These three choices will be referred to as C19, C19E, and C22, respectively. The four implicit solvent models are described in detail in Appendix A.

**2.1. Basin-Hopping Global Optimization.** The basin-hopping (BH) approach to global optimization[40–43] can be viewed as a generalization of the Monte Carlo plus energy minimization procedure of Li and Scheraga.[51] Moves are proposed by perturbing the current geometry and are accepted or rejected based upon the energy difference between the local minima obtained following minimization from the two instantaneous configurations. In effect, the potential energy surface is transformed into the basins of attraction[43,52,53] of all the local minima, so that the energy for configuration $\mathbf{r}$ is

$$\tilde{E}(\mathbf{r}) = \min\{E(\mathbf{r})\} \quad (4)$$

where min denotes minimization. Large steps can be taken to sample this transformed landscape, since the objective is to step between local minima. Furthermore, there is no need to maintain detailed balance when taking steps, because the BH approach attempts to locate the global potential energy minimum and is not intended to sample thermodynamic properties. Basin-hopping has already been employed to find the global minimum of peptides and proteins.[54–60] The moves for perturbing the current geometry of the KFFE peptides were taken in the dihedral angle space $\phi$ and $\psi$.[61] A number of these backbone dihedrals were selected and then twisted by a maximum angle, which was initially set to 30°. To select the dihedrals, we followed earlier work[61] and chose different twisting probabilities depending on the position along the peptide chain. The relative probabilities were highest for the two ends of the chain, lowest for the middle of the chain, and varied linearly between the ends and the middle. The probabilities for the ends and the middle of the chain were set to 0.4 and 0.2 for all BH runs.

Since one objective of this study was to find the global potential minimum for dimers, rigid body moves for whole peptide chains were implemented within the BH algorithm. The simplest of these moves were rigid body translation and rotation of one peptide chain, rotating around the local center of mass. Rigid body moves were performed every 10 or 20 BH steps. In most basin-hopping runs, the maximum values for translation and rotation were initially set to 1.0 Å and 60°, respectively. However, all move sizes for rigid-body and dihedral-angle moves were automatically adjusted every 50 steps during each BH run to produce an average acceptance ratio of 0.3. To further increase the efficiency of the BH algorithm for peptide oligomers, we also implemented a so-called internal rigid body move. The idea is to randomly select one of the four possible end-to-end vectors between the two peptide chains, rotate one of the peptides around it through a random angle, and rescale the end-to-end distance by a random factor within user defined limits. Because this type of move generally leads to a substantial change in geometry, it was only applied every 100 to 500 BH steps. Furthermore, it is very likely that after such an internal rigid body move the energy of the new local minimum will be much higher than that of the original minimum. If the energy exceeded a user defined value, the step was rejected and a fresh internal rigid body move was attempted until an acceptable move was found. For the resulting geometry small rigid body translation and rotation, as well as dihedral moves, were applied for a fixed number of steps between 100 and 200, without employing the acceptance criterion, to allow the dimer to relax. If a suitable minimum was found, the previous minimum was replaced with it; if it was not, the geometry was reset to the structure from which the internal rigid body move was initiated.

**2.2. Replica Exchange Molecular Dynamics.** We used replica exchange molecular dynamics (REMD),[44] as implemented in the MMTSB tool set,[62] to explore the conformational space of the KFFE dimer modeled with each of the CHARMM force fields and solvent models. In the REMD protocol multiple simulations of the same system (replicas) are run simultaneously at different temperatures. Every $\tau$ time steps an attempt is made to swap temperatures between two different replicas $i$ and $j$. The exchanges are accepted with probability $\min\{1, \exp(-\Delta)\}$, where $\Delta = (1/k_B T_i - 1/k_B T_j)(E_j - E_i)$, and $k_B$ is the Boltzmann constant, $T_i$ and $E_i$ denote

the temperature and potential energy of replica $i$, respectively, and $j$ is typically replica $i + 1$ at the same time step. The temperatures for the replicas were exponentially spaced between a minimum value $T_{min}$ and a maximum value $T_{max}$. This procedure allows for an improved sampling of the conformational space at low temperatures, since crossing potential energy barriers is facilitated at higher temperatures, and the resulting conformational changes migrate into the lower $T$ replicas.[63] The dynamics were propagated using the Langevin method, with a time step of $\delta t = 1$ fs and a friction coefficient of $\zeta = 20$ ps$^{-1}$. The SHAKE algorithm[64,65] was used to freeze out bond vibrations involving hydrogen atoms. A time step of 1 fs combined with SHAKE guarantees stable trajectories for simulations with implicit solvent models.[66] The data collected during each REMD simulation were analyzed using the weighted histogram analysis method.[67]

For MD simulations of the KFFE dimer with implicit solvent, it was necessary to confine the peptides by repulsive walls to model constant concentration conditions and to prevent the peptides from drifting away. We used a spherical container, and the confining potential experienced by an atom located at $\mathbf{r}$ is then[68]

$$V_c(r) = \frac{4\pi eR}{5r}\left[\left(\frac{\sigma}{r-R}\right)^{10} - \left(\frac{\sigma}{r+R}\right)^{10}\right] \quad (5)$$

where $r = |\mathbf{r}|$ and $\mathbf{r} = \mathbf{0}$ is the center of the sphere with radius $R$. The parameters $\sigma$ and $e$ were chosen to be 1 Å and 1 kcal mol$^{-1}$, respectively.[31] The potential (5) guarantees short-range repulsive interactions between peptide atoms and the inner walls of the container. The container radius was chosen to be $R = 25$ Å, which corresponds to a molecular concentration of 50 mM. This concentration is about two orders of magnitude higher than the concentration that was used in the experiment by Tjernberg et al. to investigate the aggregation of KFFE.[30] In the current theoretical study, lower concentrations were not possible, because two of the implicit solvent models prevented the KFFE peptides from dimerizing when a container radius of $R = 43$ Å or more was used, which corresponds to a concentration of 10 mM or lower. This problem is discussed in more detail in section 3.2.

## 3. Results

**3.1. KFFE Monomer.** *Conformational Space of the Monomer.* Before analyzing the performance of the various potentials for the dimerization of the KFFE peptide, it is instructive to investigate their impact on the conformations of the KFFE monomer. We therefore started with a BH run for each of the four solvent models to locate the corresponding global potential minimum. In each case, it was sufficient to perform a single BH run consisting of 5000 steps at a temperature of 300 K. It is likely that the $\beta$-propensity of a peptide plays a crucial role in determining whether or not that peptide forms amyloid aggregates.[30,69] To investigate the $\beta$-propensity of the KFFE peptide, the $C_\alpha$-rmsd from the $\beta$-strand illustrated in Figure 1A was calculated for each accepted local minimum during the BH run, and the structure lowest in potential energy with an rmsd < 0.5 Å was saved along with the global potential minimum. The EEF1, GB1, and GBSW potentials produce a type VIII $\beta$-turn[70] for the global potential minimum, which is shown in Figure 1B. This structure is stabilized by salt bridges between the oppositely charged N and C termini, as well as hydrogen bonds between these two residues. The comparison between strand and turn reveals that for the GBSW potential the $\beta$-strand is disfavored most, lying 12.2 kcal mol$^{-1}$ above the turn structure, whereas for EEF1 and GB1 this difference is only 7.1 and 6.3 kcal mol$^{-1}$, respectively. The ACE potential has the $\beta$-strand as the global potential minimum, which is, however, only 0.7 kcal mol$^{-1}$ lower in energy than a turn-like structure shown in Figure 1C.

To further explore the conformational space of the KFFE monomer, we performed REMD simulations with six replicas exponentially distributed between 250 and 450 K, giving a uniform acceptance ratio for exchanges with values between 0.35 and 0.45 for all REMD simulations. The simulations consisted of 100 000 replica exchange cycles, each involving 1000 MD time steps, resulting in a total simulation time of 100 ns for each potential. All replica runs were started from random structures, selected from a previous high temperature MD run. Before data were collected, 2000 MD steps were applied for equilibration. In the production phase, we saved the $C_\alpha$-rmsd from the $\beta$-strand, rmsd$_\beta$, and the turn, rmsd$_t$. The resulting free energy surface (FES) is shown in Figure 1 for $T = 300$ K as a function of rmsd$_\beta$ and rmsd$_t$. The surfaces for the EEF1 and GB1 potentials are similar. Both are dominated by a broad minimum for strand-like structures, whereas the turn configuration plays a minor role. Instead, a structure intermediate between the strand and turn (Figure 1C) is the second-most populated structure for the monomer modeled by the EEF1 and GB1 potentials. For the ACE potential, the latter structure is the most pronounced free energy minimum, followed by the $\beta$-strand, whereas the turn is hardly populated. The FES for the KFFE peptide modeled by the GBSW potential is very distinct from the EEF1, GB1, and ACE surfaces. The only pronounced free energy minimum here is the turn motif, whereas the $\beta$-strand and the intermediate structure are $\gtrsim$4 kcal mol$^{-1}$ higher in free energy.

The conclusion from the REMD results is that the EEF1, ACE, and GB1 potentials produce a considerable $\beta$-propensity for the KFFE peptide, in accord with experiment.[30] Comparison with the BH results reveals that the $\beta$-strand is favored by entropy and the turn is stabilized by potential energy. In the case of the GBSW potential, this stabilization is larger than the energy gain from the increased entropy for the $\beta$-strand, resulting in a very low $\beta$-propensity of the KFFE peptide.

*Energy Analysis for the $\beta$-Strand and Turn.* To understand the origin of the low $\beta$-propensity of the KFFE peptide modeled by the GBSW potential, we considered the total energy, $E_{tot}$, decomposed into the solvation free energy, $\Delta G_{solv}$, and the vacuum potential energy, $V_{vac} = V_{bond} + V_{el} + V_{vdW}$. Here, $V_{bond}$ represents all the bonded interactions, $V_{vdW}$ is the van der Waals interaction treated by a 6−12 potential, and the electrostatic interactions $V_{el}$ are modeled by a Coulomb potential of atom-centered point charges. These contributions are listed in Table 1 for the most stable $\beta$-strand and turn obtained for each potential. We note,

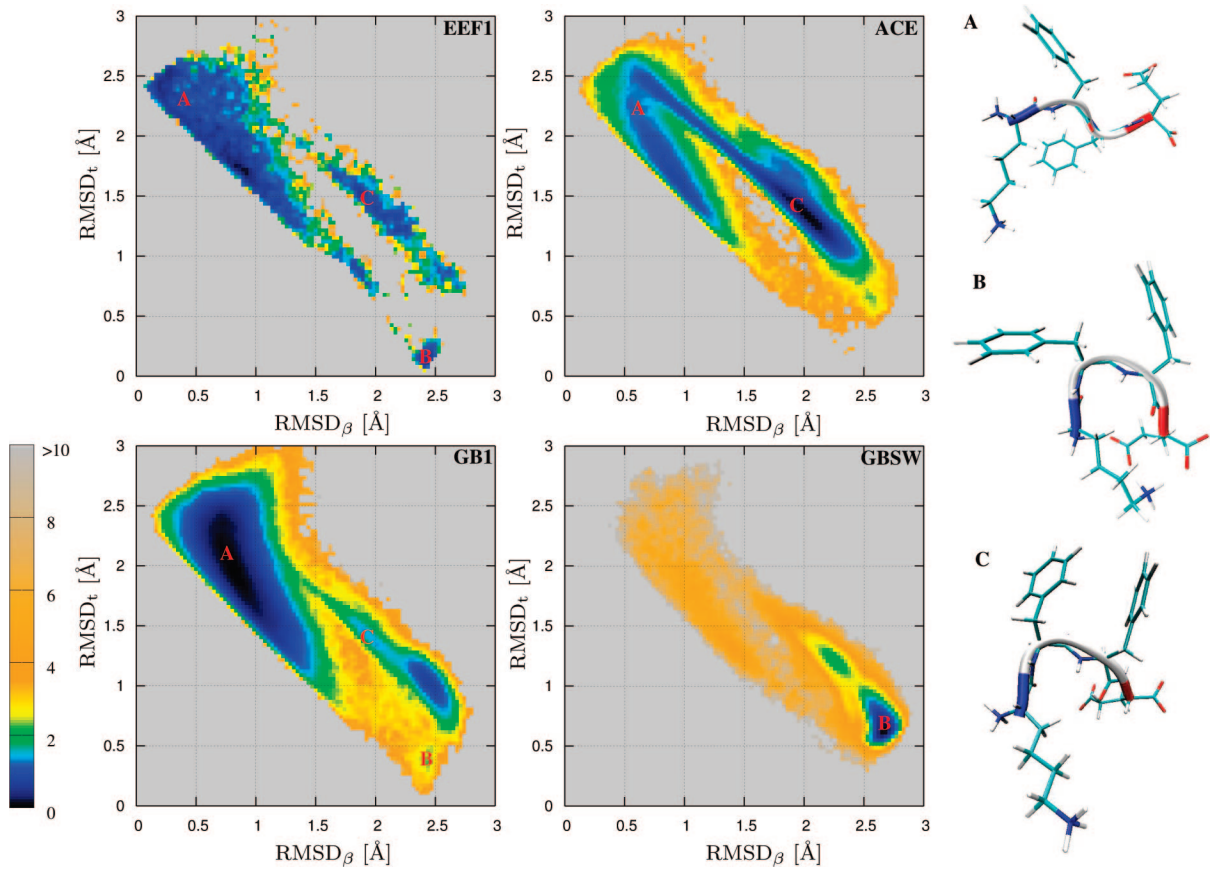


***Figure 1.*** Free energy surfaces at $T = 300$ K in terms of the $C_\alpha$-rmsd from the $\beta$-strand, rmsd$_\beta$, and from the turn, rmsd$_t$, for the KFFE monomer described by the implicit solvent models EEF1, ACE, GB1, and GBSW. The energy scale (in kilocalories per mole) is given on the left. The structures for the free energy minima are shown on the right, which are (A) a $\beta$-strand, (B) a turn, and (C) an intermediate structure between strand and turn.

***Table 1.*** Total Energy, $E_{tot}$, Decomposed into the Solvation Free Energy, $\Delta G_{solv}$, and Vacuum Energy, $V_{vac} = V_{bond} + V_{el} + V_{vdW}$ for the $\beta$-Strand and Turn of the KFFE Monomer[a]

| solvent model | $\beta$-strand | | | | | turn | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $E_{tot}$ | $\Delta G_{solv}$ | $V_{bond}$ | $V_{el}$ | $V_{vdW}$ | $E_{tot}$ | $\Delta G_{solv}$ | $V_{bond}$ | $V_{el}$ | $V_{vdW}$ |
| EEF1[b] | −156.1 | −87.7 (56%) | 7.1 | −58.8 | −16.6 | −163.2 | −80.3 (49%) | 12.2 | −75.2 | −19.8 |
| ACE | −475.1 | −304.8 (64%) | 5.0 | −154.2 | −21.1 | −474.4 | −208.9 (44%) | 6.9 | −250.9 | −21.5 |
| GB1 | −483.1 | −325.8 (67%) | 6.2 | −145.4 | −18.2 | −489.4 | −193.4 (39%) | 7.5 | −277.4 | −26.1 |
| GBSW | −211.4 | −302.7 (143%) | 20.4 | 61.2 | 9.7 | −223.6 | −153.6 (69%) | 18.6 | −104.4 | 15.8 |

[a] All values are in kilocalories per mole. The values in parentheses are the percentage contribution of $\Delta G_{solv}$ to $E_{tot}$. [b] The electrostatic energies, $V_{el}$, are not really vacuum values for the EEF1 potential, because the C19E force field has charge screening incorporated into its parameters.

however, that for the EEF1 potential $V_{el}$ is not the vacuum energy, since by design it contains the charge screening by the solvent. For each potential, the $\beta$-strand is favored by solvation and the turn by internal stabilization, but the relative values are different between the potentials. Most distinct is the GBSW potential, which produces a positive value of $V_{vac}$ for the $\beta$-strand, and thus the stabilization of this structure entirely derives from $\Delta G_{solv}$. We note that the vacuum potential energy, $V_{vac}$, is really a property of the underlying force field, i.e. C19E, C19, and C22, rather than the solvent model. Hence, the ACE and GB1 potentials, both used together with C19, give similar results for the energy contributions. For all three force fields, $V_{bond}$ is always positive for both structures. For C19E and C19 the vdW interactions stabilize the $\beta$-strand and turn by about 20 kcal mol$^{-1}$, whereas for C22 both structures are disfavored in terms of vdW interactions. In each case, the electrostatic interactions provide the largest contribution to $V_{vac}$ and stabilize the structures, except for the $\beta$-strand when modeled with C22. The Coulomb interactions are, however, diminished by the charge screening due the surrounding solvent, as the comparison between the values of $V_{el}$ for the C19E and C19 force fields reveals. From these values, we estimate that the aqueous solvent reduces the Coulomb interactions by 90−100 kcal mol$^{-1}$ in the $\beta$-strand, where the terminal charges are quite far apart. The salt bridges in the turn are screened substantially, reducing their stability by 180−200 kcal mol$^{-1}$.

To study the effect of solvent charge screening for each residue, we have decomposed $V_{el}$ into its residue−residue contributions, as shown in Figure 2. As expected, the stabilization of the turn modeled by the C19 (ACE and GB1)

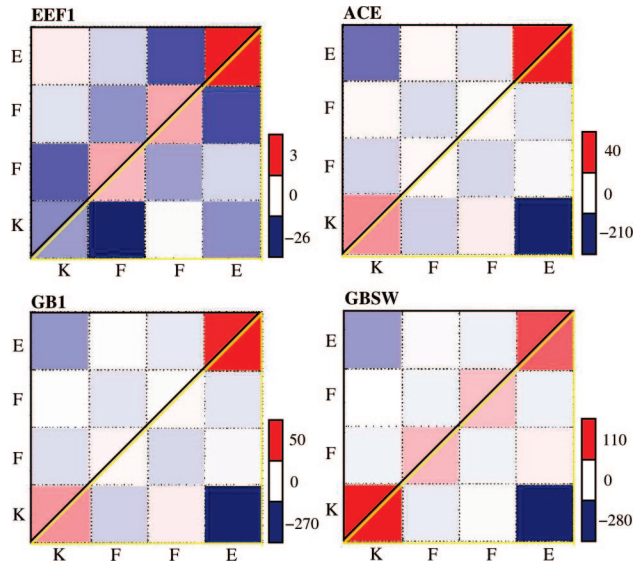


***Figure 2.*** Residue–residue map of the electrostatic interaction energy, $V_{el}$, for the $\beta$-strand (upper triangle, black) and turn (lower triangle, yellow) conformations of the KFFE monomer described by the implicit solvent models EEF1, ACE, GB1, and GBSW. The energy scale (in kilocalories per mole) is given for each map on the right.

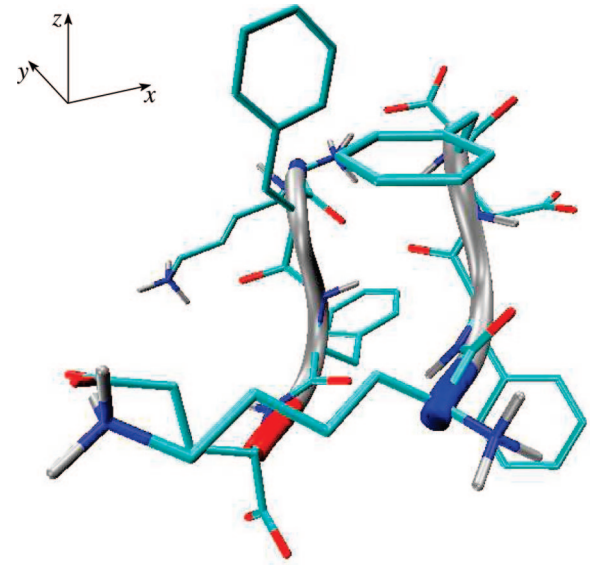


***Figure 3.*** Reference antiparallel $\beta$-sheet structure for the KFFE dimer.

and C22 (GBSW) force fields is almost exclusively due to salt bridges between the oppositely charged Lys and Glu residues. This attraction is reduced by repulsive self-interactions arising from the double charges at the unblocked termini. Here, the united-atom force field C19 predicts a stronger repulsion for the negative C terminus than for the positive N terminus, whereas for the all-atom C22 force field the opposite is true. Given that at the N terminus the two positive charges are further apart than the two negative charges at the C terminus, one would expect a larger repulsion at the C terminus, as predicted by the C19 force field. The electrostatic attraction between the termini is also present in the in vacuo $\beta$-strand. It is, however, reduced due to the increased distance between them. The loss of intramolecular electrostatic interactions in the $\beta$-strand is compensated by electrostatic peptide−solvent interactions. The main difference in solvation energy for the $\beta$-strand and turn stems from the electrostatic, and not the nonpolar, contribution to $\Delta G_{solv}$. The solvent accessible surface areas of these two structures are indeed very similar, namely 620 Å$^2$ for the $\beta$-strand and 580 Å$^2$ for the turn, and give rise to similar nonpolar solvation free energies in the range of 10−20 kcal mol$^{-1}$. The inclusion of charge screening in the C19E force field substantially changes the Coulomb interaction between the residues. The most pronounced difference from the in vacuo result is the greatly reduced stability of the salt bridges between the termini in the turn structure and their complete screening in the $\beta$-strand. Instead, the electrostatic interactions between neighboring residues become important, which are stronger than those between the spatially more distant terminal residues. Like the other two force fields C19 and C22, C19E predicts a repulsive self-interaction for the C terminus, but an attractive self-interaction for the N terminus. The attraction between like charges in aqueous solution could be explained by the stabilizing effect of the solvent-mediated hydrogen-bonded bridges between them, which could overcome the Coulombic repulsion.[71,72] However, a stabilization of 10 kcal mol$^{-1}$, as found for the $NH_3^+$–Lys interaction, seems excessive and it is known that for some relative orientations of like charged side chains the EEF1 model predicts an attraction that is too strong.[73]

**3.2. KFFE Dimer.** *Results from Global Optimization.* For each potential, six independent BH simulations, each starting from a random structure from an initial high-temperature run, were performed. For each of the six runs, a cascade of several BH searches, each consisting of 10 000 steps, was carried out. The first search was started at a temperature of 5000 K, which was gradually reduced after each step down to 250 K at the end of the run. The lowest energy structure from each run was then taken as the starting geometry for the next run and the temperature for the following BH runs was fixed with values ranging from 270 up to 400 K. Various combinations of dihedral angle moves and rigid body moves were applied in the following BH runs. If after three successive BH runs, no structure of lower energy was found and the BH searches were terminated. For each solvent model, at least three of the six BH runs located the same structure of lowest potential energy, which is therefore assumed to be the global potential minimum.

Several different structures with energies only 0.5−2 kcal mol$^{-1}$ above the global potential minimum have been located for all four of the solvent models considered. For each potential the structure of the global potential minimum is shown (together with other structures from REMD) in Figure 5. These structures are quite different from each other. The EEF1 and GBSW models favor an antiparallel orientation for the two KFFE peptides, but only the EEF1 potential supports a low-lying sheetlike structure with a $C_\alpha$-rmsd of 1.2 Å from the ideal antiparallel $\beta$-sheet shown in Figure 3. As for the monomer, the GBSW potential prefers the turn structure for the KFFE dimer, with a global potential minimum consisting of two turns aligned antiparallel. In the global potential minimum located for the GB1 solvent model the two peptides are also in turn conformations, but rotated

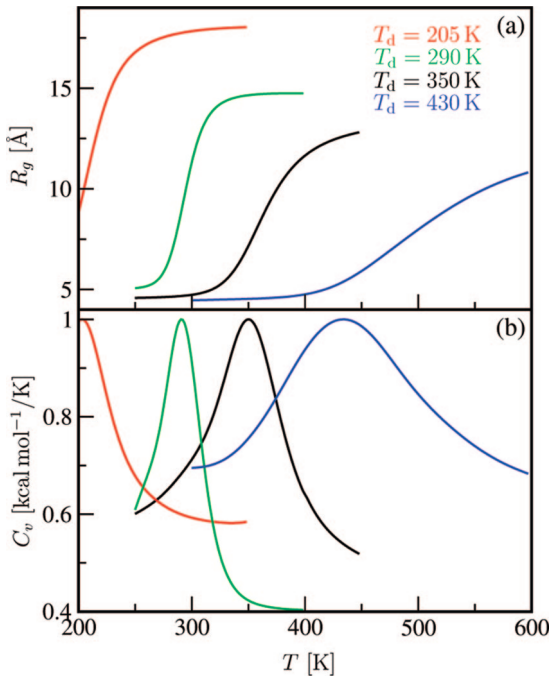


***Figure 4.*** Temperature dependence of (a) the radius of gyration, $R_g$, and (b) the heat capacity, $C_V$, computed from replica exchange molecular dynamics simulations for the KFFE dimer described by the implicit solvent models EEF1 (black), ACE (red), GB1 (green), and GBSW (blue). The dissociation temperature, $T_d$, can be inferred from the peak in $C_V$ and is given in panel a for each of the solvent models according to the color code. In the simulations, the peptides were confined to a spherical container with radius $R = 25$ Å.

relative to one another by about 90°. The global potential minimum for the ACE model consists of a $\beta$-strand and a turn with the C terminus of the strand pointing towards the middle of the turn. The dissimilarity of these four structures is further highlighted if one relaxes the global potential minimum from each solvent model using the other three potentials and then ranks their energies against that of the true global potential minimum. With one exception, the potential energies obtained always lie between 10 and 30 kcal mol$^{-1}$ above the global potential minimum in question. The exception is the global minimum for GB1, which lies only 0.5 kcal mol$^{-1}$ above the true global minimum for the GBSW solvation model, when relaxed with the latter force field.

*REMD Simulations and Dissociation Temperature.* All REMD simulations in implicit solvent were started by placing two KFFE peptides in random orientations for each replica. Before data collection, 20 000 MD steps were performed for equilibration, followed by a production phase of 100 ns at each temperature, involving 50 000 replica exchange cycles. The acceptance ratio for exchanges was uniform in the temperature range considered, with values between 0.35 and 0.5 for all REMD simulations. The temperature range for the REMD simulations was selected via short trial runs to ensure that the estimated dissociation temperature of the dimer, $T_d$, lay well within the interval in question. Specifically, $T_{min}$ and $T_{max}$ were set to 250 and 450 K for EEF1, 200 and 350 K for ACE, 250 and 400 K for GB1, and 300 and 600 K for GBSW. Eight independent replicas were considered for the GBSW potential and six replicas for the other three potentials. The principal quantities monitored during the REMD simulations were the total energy, $E_{tot}$, the peptide–peptide interaction energy, $V_{int}$, the radius of gyration, $R_g$, and the C$_\alpha$-rmsd from the antiparallel $\beta$-sheet in Figure 3.

The heat capacity of the system, $C_V$, is plotted together with the radius of gyration, $R_g$, as a function of temperature in Figure 4. The dissociation temperature, $T_d$, can be inferred from the peak in $C_V$. Above this temperature $R_g$ is controlled by the confining sphere as well as the equilibrium between the monomers and dimer. Figure 4a indicates that for higher dissociation temperatures the equilibrium is shifted toward the dimer, leading to lower values for $R_g$ above $T_d$. For the four solvent models, we find dissociation temperatures of 350 K for EEF1, 205 K for ACE, 290 K for GB1, and 430 K for GBSW. Hence, the ACE potential shows the lowest propensity for aggregation and GBSW the highest.

In a similar REMD simulation of the KFFE dimer modeled by the GB1 potential, Baumketner and Shea found a dissociation temperature of 325 K.[31] The deviation of 35 K from our result can be explained by the smaller confining sphere, with $R = 17$ Å, used in their simulation compared to $R = 25$ Å in our study, as well as by the different temperature range chosen for the replicas. When we repeated the REMD simulation for the GB1 potential with exactly the same settings as in ref 31 we also found $T_d = 325$ K and obtained the same FES for $T = T_d$ as shown in Figure 2d of ref 31. With the larger container radius of $R = 25$ Å, the same representation for the FES, i.e., plotted as function of interaction energy between the KFFE peptides and $R_g$ at the dissociation temperature $T_d = 290$ K, shows a greater free energy barrier for dimerization. Using an even larger container radius of $R = 43$ Å in the REMD simulations, which corresponds to a peptide concentration of 10 mM (still two orders of magnitude above the experimental concentration of 200−300 $\mu$M),[30] the ACE and GB1 potentials did not support a free energy minimum for the dimer. Hence, they clearly underestimate the aggregation propensity of KFFE. However, for the EEF1 and GBSW models the KFFE peptides still dimerize for container radii of 43 and 55 Å, the latter corresponding to a peptide concentration of 1 mM. The effect of the container radius on the dissociation temperature for the EEF1 potential is to produce a gradual decrease from $T_d = 365$ K for $R = 17$ Å to $T_d = 325$ K for $R = 55$ Å. Further studies will be needed to determine the effect of confinement on peptide aggregation. In the current work $R = 25$ Å was chosen as the default confining radius, since it allows aggregation to occur for each potential considered.

*Structural Analysis of the Free Energy Minima.* In Figure 5, the FES is plotted for the KFFE dimer as a function of $R_g$ and the C$_\alpha$-rmsd from the antiparallel $\beta$-sheet in Figure 3. Temperatures below $T_d$ (300 K for EEF1, 200 K for ACE, 260 K for GB1, and 350 K for GBSW) were chosen for detailed analysis of the conformational space. Typical structures for the most populated free energy minima, along with the structure of the global potential minimum previously

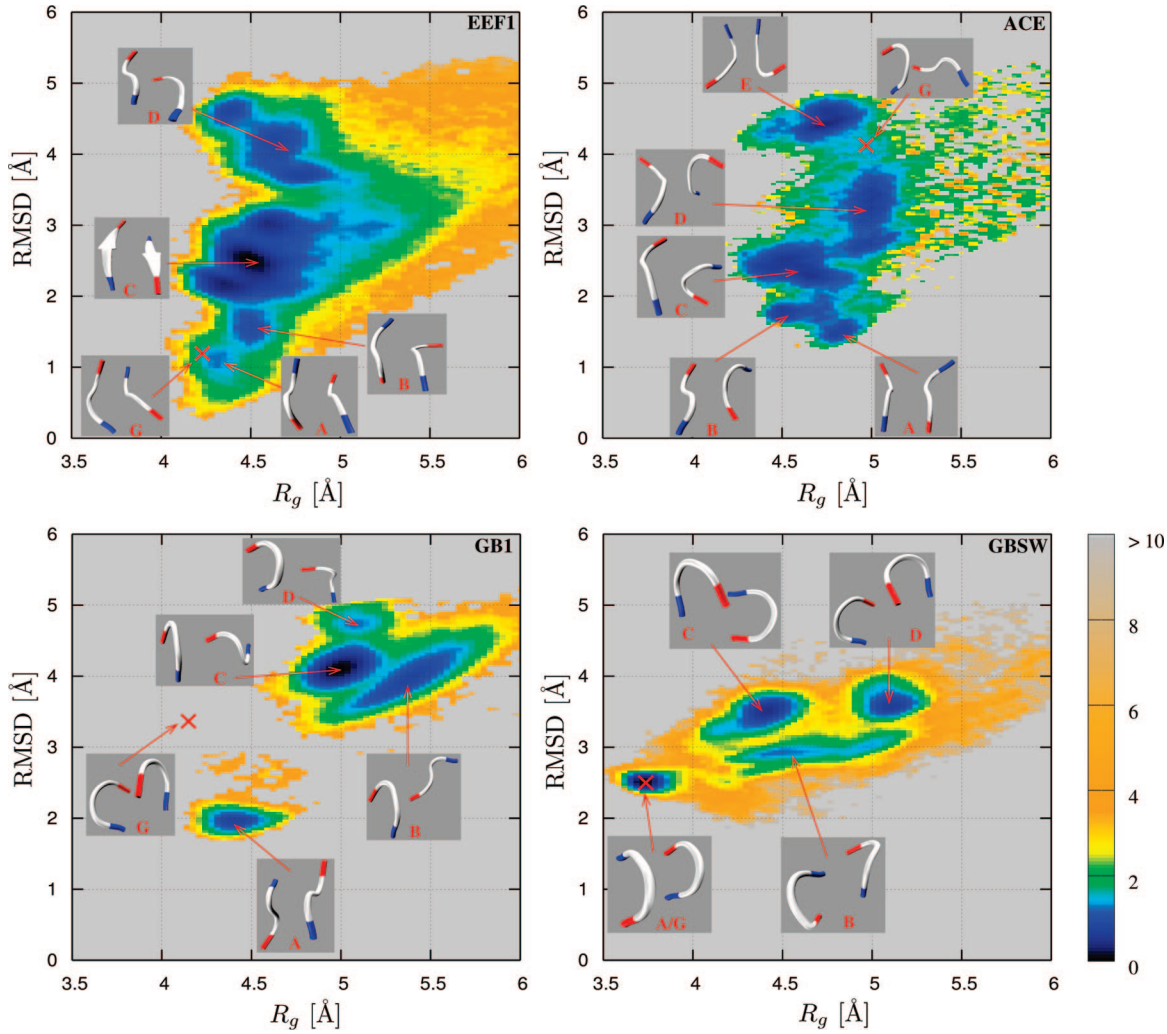


***Figure 5.*** Free energy surfaces in terms of the radius of gyration, $R_g$, and $C_\alpha$-rmsd from the ideal antiparallel $\beta$-sheet (Figure 3) for the KFFE dimer described by the implicit solvent models EEF1, ACE, GB1, and GBSW. The free energy surfaces are shown for temperatures below the dissociation temperature; specifically, $T = 300$ K for EEF1, 200 K for ACE, 260 K for GB1, and 350 K for GBSW. The energy scale (in kilocalories per mole) is given on the right. Typical structures for the free energy minima are shown and labeled A−E. The structure of the global potential minimum, labeled G, found from basin-hopping is also shown together with its location on the free energy surface.

found from BH, are also presented in Figure 5. All of the free energy minima for EEF1 consist of sheetlike structures, which are aligned in an antiparallel fashion, aside from conformation EEF1-D. The most favorable geometry is structure EEF1-C, which is a twisted $\beta$-sheet. The twist causes a rmsd deviation of 2.5 Å from the untwisted reference $\beta$-sheet. The global potential minimum, EEF1-G, almost coincides with the free energy minimum EEF1-A, suggesting that the conformational space of the KFFE dimer described by the EEF1 potential is dominated by potential energy rather than entropy.

The FES for the ACE potential is also dominated by structures involving $\beta$-strands and other extended conformations of the KFFE peptide. The structure closest to the antiparallel $\beta$-sheet is the free energy minimum ACE-A. In one of the free energy minima, ACE-E, the two peptides are aligned parallel, despite the repulsion resulting from the like charges being next to each other, which must be compensated by other effects. The FES for GB1 exhibits four well-defined minima. The structure of GB1-A is closest to the antiparallel $\beta$-sheet, but one of the strands is flipped over. The same structure was identified in the study of Baumketner and Shea.[31] The other three minima are quite loose, with $R_g \gtrsim 5$ Å, and consist of one KFFE peptide adopting a turn conformation and the other a strand geometry, which points with its C terminus toward the middle of the turn. The structure of the global potential minimum, GB1-G, which was also identified by Baumketner and Shea,[31] is far away from any free energy minimum. From the findings for the ACE and GB1 solvation models we conclude that for both potentials solvation and entropy have a large impact on the conformations populated by the KFFE dimer. All four minima of the FES for GBSW are composed of two KFFE peptides in turn conformations. This finding again reflects the very low $\beta$-propensity of the GBSW model for KFFE. The free energy minima differ in the relative orientation of the two turns. Minimum GBSW-A is also the global potential minimum where the two peptides are aligned antiparallel. The coincidence of the global potential minimum with a free energy minimum provides evidence that stable

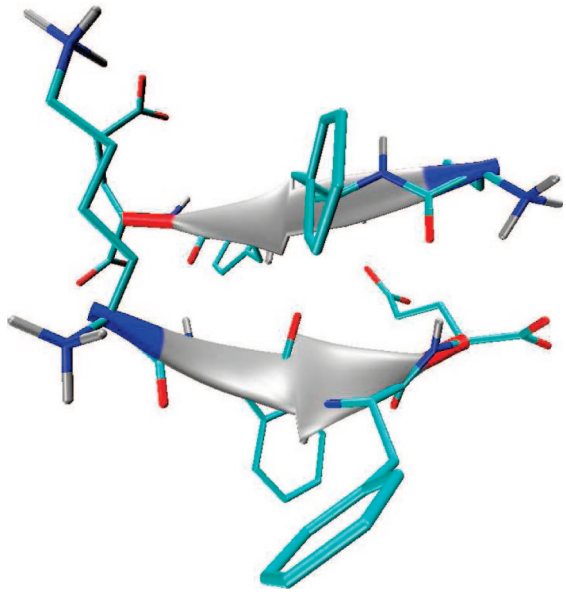

***Figure 6.*** Twisted antiparallel $\beta$-sheet conformation found in replica exchange simulations of the KFFE dimer using the EEF1 implicit solvent model. It is referred to as structure EEF1-C.

KFFE dimers are populated, and it explains the high dissociation temperature found from the REMD simulation for GBSW.

*Energy Analysis for the Free Energy Minima.* To shed more light on how the intrapeptide, interpeptide, and peptide−solvent interactions trigger aggregation, all the structures presented in Figure 5 have been relaxed to local potential energy minima. The resulting total potential energies, $E_{tot}$, were decomposed into the following contributions: $V_{vac,1}$ and $V_{vac,2}$ for the intramolecular vacuum potential energies of the two KFFE peptides, $V_{int}$ for the peptide−peptide interaction energy, and the solvation free energy, $\Delta G_{solv}$. The results of this decomposition scheme are presented in Figure 7 for the relevant structures, as well as for the twisted antiparallel $\beta$-sheet, structure EEF1-C, which is shown in more detail in Figure 6. Based on experimental findings, this conformation should be the structural element of the KFFE amyloid fibril,[30] and we wish to understand why it is disfavored by the GB models under consideration. The five structures derived from the EEF1 potential are all very similar and are stabilized by the same forces. Solvation (excluding the screening of the peptide charges) always contributes about 45% to the total energy, and the other 55% is equally distributed between intra- and intermolecular terms. The results are not as straightforward for ACE and GB1. Most of the corresponding free energy minima are not very well stabilized by peptide−peptide interactions and are instead optimized for solvation. Exceptions are the structures ACE-B and ACE-C, for which $V_{int}$ contributes about 25% to $E_{tot}$. However, these structures are in equilibrium with the three other free energy minima, which are only marginally or not at all stabilized by $V_{int}$, and thus can easily dissociate. That the peptide−peptide interaction is not a strong driving force for aggregation for ACE can also be seen for the global potential minimum, where solvation contributes 53% and $V_{int}$ only 10% to $E_{tot}$. The antiparallel $\beta$-sheet, on the other hand, is significantly stabilized by the interaction between the two strands, yet is clearly disfavored by the ACE solvation energy. The same findings as for the structures ACE-A through ACE-E apply for the structures GB1-A through GB1-D, i.e., they are stabilized by solvation rather than the peptide−peptide interaction and are thus susceptible to dissociation. For the global potential minimum GB1-G, on the other hand, $V_{int}$ contributes 34% to $E_{tot}$, but this structure is clearly disfavored by solvation and entropy. A similar picture holds for the EEF1-C structure when described by the GB1 potential. The GBSW structures are all stabilized by peptide−peptide interactions and by the solvation energy, leading to stable KFFE dimers. As already found in the study of the KFFE monomer, the peptides alone are quite unstable, with either small negative or even positive values for the energies $V_{vac,i}$. These destabilizing forces are due to the large electrostatic repulsion at the charged termini within the C22 force field. They are most destabilizing for the $\beta$-conformation of the KFFE peptide, thus preventing the antiparallel $\beta$-sheet from forming within the GBSW description.

*Residue–Residue Interactions in the $\beta$-Sheet.* The peptide−peptide interactions present in the twisted $\beta$-sheet in Figure 6 can be decomposed into electrostatic, $V_{int}^{el}$, and vdW, $V_{int}^{vdW}$, contributions, which are listed in Table 2. For each potential the vdW interactions are less stabilizing than the electrostatic forces. The unscreened $\beta$-sheet is dominated by attractive Coulomb interactions arising from the salt bridges between the unblocked Lys and Glu residues from adjacent strands. However, after the inclusion of screening due to the solvent, one finds that $V_{int}^{vdW}$ is of a similar magnitude to $V_{int}^{el}$, as the result for EEF1 shows. The comparison of $V_{int}^{el}$ between the C19E and C19 force fields reveals that the solvent reduces the electrostatic interpeptide interactions by 220−230 kcal mol$^{-1}$. If one assumes the same amount of screening for the GBSW solvent model, the Coulomb interactions would remain three or four times more important for the $\beta$-sheet described by the C22 force field.

Analysis of the vdW interactions is of interest to help answer questions about the importance of the $\pi-\pi$ interactions between the Phe residues in adjacent strands.[30] To this end, we have decomposed $V_{int}^{vdW}$ into its residue−residue contributions, and the results are presented in Figure 8. Since there is no difference in the description of the vdW interactions between the C19 and C19E force fields, and since the equilibrated structures of the EEF1-C $\beta$-sheet differ only marginally for EEF1, ACE, and GB1, we show only one residue−residue map representing the C19(E) force fields and another one for the C22 force field from the GBSW calculation. The comparison between these two plots reveals that the vdW interactions are quite different. For instance, with the C19(E) force fields the most stabilizing interaction is between Lys and Glu, with a value of –5.0 kcal mol$^{-1}$. The same residue pairing gives a repulsive vdW interaction of 2.5 kcal mol$^{-1}$ for the C22 force field and instead finds the strongest such interactions for the Phe−Lys and Phe−Phe pairs with values of –2.5 kcal mol$^{-1}$. From the results in Figure 8, we conclude that $\pi-\pi$ interactions

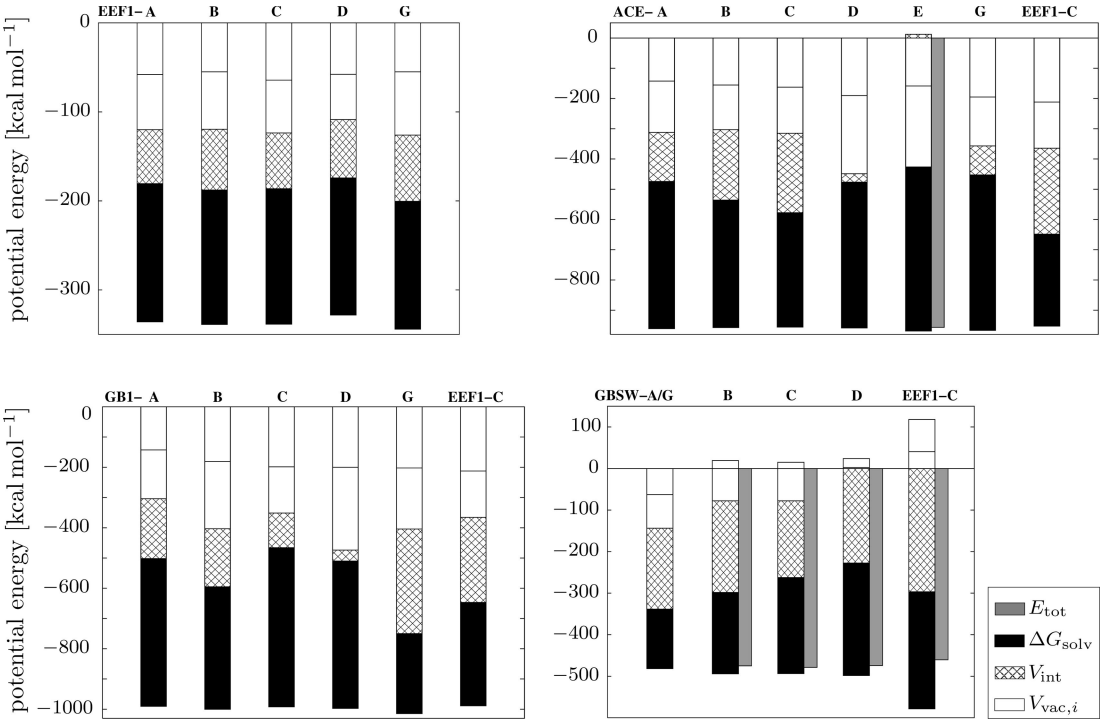


***Figure 7.*** Energy contributions to the EEF1-C $\beta$-sheet (Figure 6) for the implicit solvent models EEF1, ACE, GB1, and GBSW. The total potential energy, $E_{tot}$, is decomposed into the vacuum energies, $V_{vac,i}$, of the two KFFE peptides, their interaction energy, $V_{int}$, and the solvation free energy, $\Delta G_{solv}$. If one of the contributions is positive, $E_{tot}$ is plotted explicitly, otherwise it is $E_{tot} = V_{vac,1} + V_{vac,2} + V_{int} + \Delta G_{solv}$.

***Table 2.*** Contributions to the Interaction Energy, $V_{int}$, between the KFFE Peptides in the twisted $\beta$-Sheet (Figure 6)[a]

| | $V_{int}^{el}$ | $V_{int}^{vdW}$ |
|---|---|---|
| EEF1[a] | −34.5 | −28.1 |
| ACE | −252.3 | −31.7 |
| GB1 | −250.3 | −31.5 |
| GBSW | −282.3 | −14.4 |

[a] The electrostatic energies, $V_{int}^{el}$, are not really vacuum values for the EEF1 potential, because the C19E force field has charge screening incorporated into its parameters.

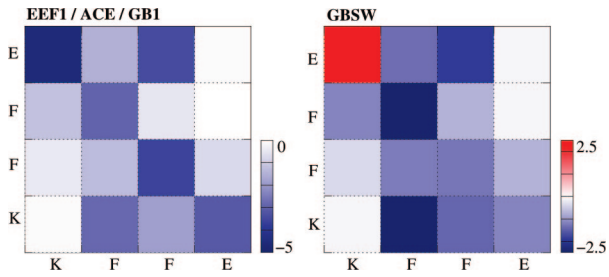


***Figure 8.*** Residue–residue map of the van der Waals interaction energy between the two KFFE peptides in the EEF1-C $\beta$-sheet (Figure 6). The map on the left shows the result for the CHARMM19 force field used together with EEF1, ACE, and GB1 solvent models. The map on the right shows the result for the CHARMM22 force field used together with the GBSW solvent model. The energy scale (in kilocalories per mole) is given for both maps on the right.

play a role in the stabilization of the antiparallel $\beta$-sheet for KFFE, but are not more important than other vdW interactions.

## 4. Discussion and Conclusions

The process of aggregation is a delicate balance between the competing solvation forces and intra- and intermolecular forces. In the case of the KFFE peptide this competition is seen experimentally from the following results.[30] (i) If one removes the terminal charges of the unblocked KFFE peptide by acetylating the N terminus and amidating the C terminus the fibril formation is significantly reduced. (ii) KFFK and EFFE do not form fibrils when incubated individually, but aggregate when incubated together in equimolar concentrations. (iii) KFFE and KVVE form amyloid fibrils whereas KLLE and KAAE do not. These findings reveal that charge attraction plays a crucial role for fibril formation. However, Coulombic forces are not the only factor governing amyloid aggregation, otherwise KLLE and KAAE should also form fibrils. Other factors include the $\beta$-propensity and hydrophobic interactions. The KFFE and KVVE peptides show partial $\beta$-conformations in solution, whereas KLLE and KAAE do not.[30] KFFE is also the most hydrophobic peptide and KAAE the least from the four peptides studied experimentally. The objective of any theoretical model would be to reproduce these experimental findings and explain them in atomistic detail. However, most of the empirical potentials that allow us to study such peptides including solvent effects in a reasonable length of time, have certain shortcomings leading to an imbalanced description of the aggregation process.

The EEF1 potential produces results that are closest to the experimental findings of partial $\beta$-strand conformation for the KFFE peptide in solution and the formation of fibrils exhibiting $\beta$-strand structure.[30] Thus, of the four energy

functions considered, the EEF1 potential provides the best balanced description for the secondary structure of the KFFE peptide and the detailed interplay between solvation forces and intermolecular forces. Other advantages of the EEF1 energy function are that it is very fast to evaluate and numerically stable. Our conclusion is in line with other comparative studies assessing the performance of various implicit solvent models implemented in CHARMM, which also also found that the EEF1 model is one of the most reliable potentials in identifying experimental structures or minima from explicit solvent simulations.[21,28] The success of the EEF1 potential in describing protein aggregation probably lies in incorporating the solvent screening by reducing the side chain charges. The resulting C19E parameters are also used together with the implicit solvent model SASA, which is based on the solvent-accessible surface area to describe the main effects of the aqueous solvent on the solute.[25] This potential has recently been used in MD simulations of various amyloidogenic peptides to predict aggregation pathways and the resulting $\beta$-sheet structures,[37,74,75] and to identify $\beta$-aggregation "hotspots" in amyloidogenic proteins.[38] It correctly predicted the in-register parallel packing of three GNNQQNY peptides,[37] whereas when applied to polypeptide sequences experimentally known not to form amyloid structures no ordered $\beta$-aggregates were observed.[37,74]

The main limitation of the EEF1 solvent model that we are aware of, is the empirical implementation of the charge screening by neutralizing the charged termini and sidechains. While in many cases this approximation gives reasonable results, in some instances the Coulomb interaction between charges is underestimated. This failure is most pronounced for repulsive interactions between like charges, which the EEF1 solvent model often predicts to be attractive as, for example, in the interaction between the two positive charges in the unblocked N terminus of KFFE (Figure 2). This issue has already been addressed by Masunov and Lazaridis in a study of the interactions between ionizable amino acid side chains.[73] In a recent study of protein−surface interactions, EEF1 also overestimated the electrostatic screening for a negatively charged peptide approaching a negatively charged surface.[29] A possible way to address this problem could be to derive the partial charges for the amino acids from electronic structure calculations combined with an implicit solvent model.

The ACE and GB1 models predict similar results for both the KFFE monomer and dimer due to the similarity of the two potentials. Both solvent models were used together with the C19 force field and both GB models use a pairwise approach summing over atomic volumes to approximate the volume integral (3) for the calculation of the effective Born radii within the CFA. The ACE and GB1 potentials predict a similar $\beta$-propensity for the KFFE peptide as EEF1, but fail to produce stable KFFE dimers. It seems that the balance between the competing forces triggering aggregation lies in favor of solvation, preventing stable dimers from forming and thus underestimating the aggregation propensity. The two KFFE peptides can be forced to form a dimer by confining them into a rather small sphere, which still results in a dissociation temperature well below the experimental value. The comparison between the EEF1 and ACE/GB1 results reveal that a similar description of the conformational space of a single peptide does not guarantee the same behavior of the potentials with respect to peptide aggregation.

The overestimation of solvation energies for ACE and GB1 contrasts with the common assumption that GB models based on the CFA tend to underestimate the solvation energy, due to an overestimation of the effective Born radii compared to the exact ones calculated by solving the PB equation.[10] However, in a previous comparison it was found that the performance of a GB method also varies with the conformation of the protein.[76] For a test set of 120 near-native, misfolded and unfolded structures of chicken villin headpiece, GB1 (used together with C22) overestimates the solvation energy for extended structures, while the solvation energies for nativelike conformations are not favorable enough. This finding would explain our observation that the extended KFFE peptide does not form stable dimers and instead prefers solvent exposed configurations for GB1. The results for the ACE solvent model published in ref 76 do not support our finding of an overestimation of the solvation energy for the extended KFFE peptide. However, in ref 76 the ACE model was used together with C22 and a new parameter set with zero hydrogen volumes. Hence the ACE models in our study and in ref 76 may be too different to produce consistent results. Our conclusions are supported by two other studies (using the C19 force field), which both find that salt bridges are screened too much within the ACE potential,[27,29] and that charged residues prefer solvent exposed orientations for ACE and GB1.[27] It is not clear whether this failure is due to the ACE and GB1 solvent models or to the in vacuo performance of the C19 force field.

The GBSW model supports a very stable antiparallel dimer structure, but with the KFFE peptides in a turn rather than a $\beta$ conformation. Compared to experiment, GBSW clearly disfavors the $\beta$-state, which can be attributed to an overestimation of the electrostatic forces present in the KFFE peptide. The effective electrostatic forces are stronger than the CMAP corrections to the C22 potential. Thus it is unlikely that further corrections to the backbone dihedral cross terms, as introduced recently for the CHARMM22/CMAP/GBSW potential, would change the current result.[77] For the uncapped KFFE peptide modeled by the C22 potential, significant repulsive forces are found at the doubly charged termini. In the turn structure, these repulsions are more than compensated by attraction between the adjacent N and C termini. In the $\beta$-strand, however, the termini are too far apart from each other to counteract these repulsions. The apparently excessive Coulomb forces could be due to either an overestimation of the partial charges in the ionizable side chains and termini in the C22 force fields, or to an underestimation of the charge screening by GBSW. We have found several indications to support the former possibility. (i) We have checked that the GBSW result for the KFFE monomer does not change if the terminal charges are further screened by invoking the GBSW solvent model with a salt concentration of 0.05 and 5.0 M, or if they are removed by acetylating the N terminus and amidating the C terminus. In

each case the $\beta$-strand remains substantially disfavored. (ii) From the study in ref 76 it is known that GBSW tends to overestimate the solvation energy for extended structures,[76] which should favor the $\beta$-strand for the KFFE peptide, but not the turn. (iii) A 50 ns REMD simulation of the KFFE dimer using the TIP3P explicit water model,[26] combined with the C22 force field, produces an FES similar to that from the GBSW simulation. Details of this simulation are given in Appendix B. The only difference is that minima GBSW-A and GBSW-C are less populated in the explicit water simulation. From these results we conclude that the partial charges for the charged termini and side chains in the C22 force field are probably overestimated, rather than the screening by GBSW being underestimated. An adjustment of these charges could be beneficial for both the performance of the C22 force field and for GBSW, as well as for other GB solvent models. However, we acknowledge that the calibration of a force field is a complex process and the change of partial charges would entail extensive reparametrization of the force field. We finally note that the GBSW solvent model is computationally less efficient compared to EEF1, but also compared to ACE and GB1, due to the numerical evaluation of the integral in eq 3. In addition, we observed that the default values for the number of grid points in the numerical integration do not produce well converged results, as discussed in Appendix C.

The nonelectrostatic interactions, such as dispersion, between the two KFFE peptides in the antiparallel $\beta$-sheet are described quite differently by the C19(E) and C22 force fields, as shown in Figure 8. For instance, the most attractive vdW force between Lys and Glu for the C19(E) potential is repulsive for C22. In both energy functions the $\pi-\pi$ interaction between phenylalanines is no more stabilizing than other vdW interactions. Hence, the present force fields do not suggest a dominant role for $\pi$ interactions in amyloid formation.[78] However, it is known from experiment that charge attraction is not the only factor in the aggregation process, as the comparative study of KFFE, KVVE, KLLE, and KAAE shows.[30] Interestingly, the more hydrophobic KLLE does not form fibrils while KVVE does. To resolve such subtleties, and to quantify the impact of hydrophobicity and $\beta$-propensity on the aggregation process, more accurate potentials are probably needed. A comparison between the results from ab initio electronic structure calculations and empirical potentials would be helpful to reliably identify the most stabilizing forces in the different dimers. However, such calculations involving intermolecular interactions have to be performed with care, especially if accurate treatment of dispersion is required.

Our results allow us to infer the likely behavior of the various force fields and solvent models in describing the aggregation processes of the KFFE peptide beyond dimerization, although we have only studied the latter process in the present contribution. The structure of the KFFE dimer is not yet known experimentally and it is not clear whether it serves as a nucleation site for fibrillization, or whether it rather represents an off-pathway kinetic intermediate. Based on our results it seems unlikely that stable amyloid fibrils would be obtained for KFFE described by the C22 force field/GBSW solvent model, due to the very low $\beta$-propensity in this case. The C19 force field combined with the ACE and GB1 solvent models, on the other hand, clearly underestimate the aggregation propensity of the KFFE peptide. Thus out of the four potentials in question, EEF1 is probably the most promising representation of the KFFE peptide and its aqueous surroundings for future studies of the oligomerization processes beyond dimerization, which eventually lead to fibrillization.

**Acknowledgment.** The authors are grateful to Dr J.-E. Shea and Dr A. Baumketner for providing us with the structures of the KFFE dimer from their theoretical study for comparison. B.S. gratefully acknowledges the BBSRC for financial support.

## Appendix A. Solvent Models

**EEF1.** The EEF1 model is a solvent-exclusion model based on the assumption that the solvation free energy of a protein is a sum of group contributions, which are determined from values for small, fully solvent-exposed model compounds minus the reduction in solvation because of the presence of surrounding groups:

$$\Delta G_{\mathrm{solv}} = \sum_i \Delta G^{\mathrm{ref}}_{\mathrm{solv},i} - \sum_i \sum_{j \neq i} f_i(r_{ij}) V_j \tag{6}$$

Here, $\Delta G^{\mathrm{ref}}_{\mathrm{solv},i}$ is the solvation free energy of the reference compound $i$, $r_{ij}$ is the distance between $i$ and $j$, $V_j$ is the volume of group $j$, and $f_i(r_{ij})$ is the solvation free energy density, which is assumed to be a Gaussian function. To account for the screening of the interactions between charges due to the solvent, ordinarily charged protein groups (ionic side chains and termini) are neutralized and a distance-dependent dielectric constant is employed. The EEF1 model was used together with a modified version of the united-atom CHARMM19 parameters, which contain the neutralized side chains and termini. For the calculation of the nonbonded interactions, the same cutoffs were used for which the EEF1 model was parametrized, i.e., both vdW and electrostatic interactions were cut off at 9 Å with a switching function between 7 and 9 Å.

**GB1.** The GB method, which we refer to here as GB1, follows the pairwise approach, summing over atomic volumes to approximate the volume integral (3) for the calculation of the Born radii. It uses a linearized version of Still's original formula[4] for the self-electrostatic free energy of atom $i$,[15]

$$\Delta G_{\mathrm{el},i} = \tau \left[ \frac{1}{\lambda}\left(\frac{-166}{R_i}\right) + P_1\left(\frac{166}{R_i^2}\right) + \sum_j^{\mathrm{bond}} \frac{P_2 V_j}{r_{ij}^4} + \sum_j^{\mathrm{angle}} \frac{P_3 V_j}{r_{ij}^4} + \sum_j^{\mathrm{nonbond}} \frac{P_4 V_j C}{r_{ij}^4} \right] \tag{7}$$

where $V_j = 4/3\pi R_j^3$ is the atomic volume of atom $j$ and $C$ is a close-contact function that adjusts radii for nonbonded atoms close to atom $i$. The reduction of atomic volumes due to neighboring atoms is accounted for by the adjustable parameters $\gamma$ and $P_k$, which were refitted by Dominy and Brooks to solvation energies from PB calculations for a

database of peptides.[15] As in the study by Baumketner and Shea,[31] we used GB1 together with the CHARMM19 parameters and no truncation of the nonbonded interactions was performed.

**ACE.** In their GB implementation, Schaefer and Karplus developed an analytical and continuous pairwise atomic expression for the electrostatic solvation energy, which is known as the analytical treatment of continuum electrostatics (ACE).[7] In the ACE model, the volume integral over the solute in eq 3 is rewritten as an integral over all space and the integrand is multiplied by a step function, $P(\mathbf{r})$, whose value is one in the solute's interior and zero elsewhere. This function can be expressed as a sum of atomic terms, $P(\mathbf{r}) = \Sigma_i P_i(\mathbf{r})$, and the density functions $P_i(\mathbf{r})$ describing the atomic volume are given by Gaussians, whose width and height control the smoothness of the atomistic solute volume description. The resulting expression for the self-electrostatic energy is

$$\Delta G_{el,i} = -\frac{\tau q_i^2}{2R_i} + \sum_{j\neq i} \frac{\tau q_i^2}{\omega_{ij}} \exp(-r_{ij}^2/\sigma_{ij}^2) + \frac{\tau q_i^2 \tilde{V}_j}{8\pi} \left( \frac{r_{ij}^3}{r_{ij}^4 + \mu_{ij}^4} \right)^4 \tag{8}$$

where the parameters $\omega_{ij}$ and $\sigma_{ij}$ determine the width and height of the Gaussian, which dominates $\Delta G_{el,i}$ in the short-range domain. At long-range $\Delta G_{el,i}$ is dominated by the last term in eq 8, which vanishes for $r_{ij} = 0$ due to the parameter $\mu_{ij}$. For the volumes $\tilde{V}_j$ we used the parameter set based on the Voronoi volumes for the CHARMM19 force field,[79] which we employed together with the ACE solvent model. The calculation of the nonbonding interactions was performed without a cutoff.

**GBSW.** The GBSW model[11] is one of the most recent GB models and involves a continuous and smooth switching function to define the molecular surface. To improve the calculated Born radii, an empirical correction term, $\Delta G^1_{el,i}$, is added to the Coulomb field term, $\Delta G^0_{el,i}$.[10] The self-electrostatic free energy can then be expressed as

$$\begin{aligned} \Delta G^1_{el,i} &= a_0 \Delta G^0_{el,i} + a_1 \Delta G^1_{el,i} \\ &= -\frac{\tau q_i^2}{2} \left[ a_0 \left( \frac{1}{R_i} - \frac{1}{4\pi} \int_{\text{solute},r>R_i} \frac{1}{r^4}\, dV \right) + \right. \\ &\qquad \left. a_1 \left( \frac{1}{R_i^4} - \frac{1}{4\pi} \int_{\text{solute},r>R_i} \frac{1}{r^7}\, dV \right)^4 \right] \end{aligned} \tag{9}$$

from which the Born radii can be calculated according to eq 3. The volume integrals in eq 9 are evaluated numerically using Gaussian-Legendre quadrature[80] for the radial grid points and Lebedev quadrature[81] for the angular grid points. For each atom, we used 50 angular integration points and 24 radial integration points up to 20 Å. For the switching function at the dielectric boundary, we chose a smoothing length of 0.6 Å and used the values $a_0 = -0.180$ and $a_1 = -1.817$ for the coefficients in (9), as recommended for this choice of smoothing length.[11] We employed GBSW together with the CHARMM22 force field including the CMAP modifications,[48–50] and the nonbonded interactions were calculated without cutoffs.

## Appendix B. Details of the REMD Simulation with TIP3P Water

The TIP3P water model[26] was used together with the CHARMM22/CMAP force field. The peptides were solvated in a preequilibrated cubic box of TIP3P water molecules extending at least 10 Å from any solute atom. Any water molecule that was too close to the peptides, i.e. within 2 Å, was removed. The solvated system was initially minimized for 4000 steps using the adapted basis Newton-Raphson algorithm[80] to remove any bad contacts. The minimized system was gradually heated from 50 to 700 K in 20 K increments using short 10 ps MD runs. During heating, the backbone atoms of the peptides were restrained to their initial positions with a force constant of 5 kcal mol$^{-1}$ to allow relaxation of the water molecules. All MD simulations with TIP3P employed the SHAKE algorithm,[64,65] with a geometric tolerance of 0.0001 Å to constrain covalent bonds involving hydrogen atoms, thus allowing a time step of 2 fs in the verlocity Verlet algorithm[82,83] to integrate Newton's equations. For MD runs at constant temperature, the *NVT* ensemble was used together with the Nose-Hoover thermostat.[84] For the nonbonding interactions a cutoff of 10 Å was used, while vdW interactions were switched to zero between 7 and 9 Å, and the particle-mesh Ewald technique was employed to calculate the long-ranged electrostatic energies along with a force-shifting function at a cutoff distance of 9 Å.[85,86]

The edge length of the cubic water box was chosen to be 40.6 Å, yielding a total of 2130 water molecules, which corresponds to the same peptide concentration of 50 mM as in the implicit solvent runs. For the initialization of the REMD simulation, a random conformation for the KFFE dimer was chosen followed by solvation, minimization and gradual heating. Every 20 K the structure was saved and taken as the starting point for the replica that was closest in temperature. The REMD simulation consisted of 32 replicas, with temperatures exponentially spaced between 300 and 600 K. Prior to the production phase, each replica was allowed to further equilibrate in a 5 ns MD run. During the subsequent production phase, 25 000 replica exchange cycles of 2 ps length were applied, resulting in a total production length of 50 ns at each temperature. The acceptance ratio was found to be quite uniform throughout the considered temperature range with values between 0.1 and 0.25.

## Appendix C. Numerical Convergence of the GBSW Solvent Model

When performing global optimization and minimization of structures with the GBSW potential, we became aware of convergence problems arising from the numerical volume integration for the calculation of the Born radii. The recommended values are 24 radial grid points for the Gaussian-Legendre quadrature and 38 angular grid points for the Lebedev quadrature.[11] However, these values do not seem to give well converged energies for our purposes in the present system, as demonstrated in Figure 9 for the antiparallel $\beta$-sheet of the KFFE dimer shown in Figure 3. This structure was rotated around the $x$-axis in increments of 10° up to 90° and the energy before and after minimization

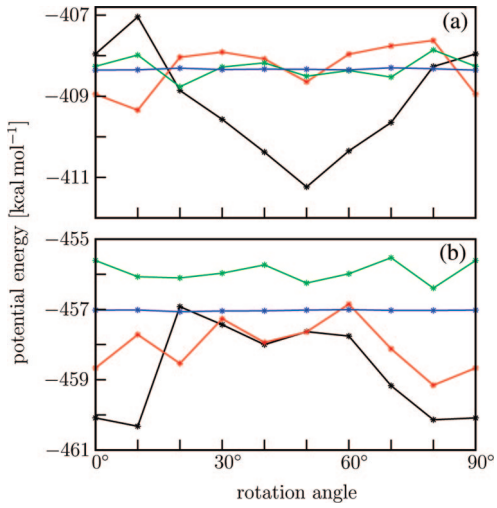


***Figure 9.*** Convergence behaviour of the GBSW solvent model is analyzed for the ideal $\beta$-sheet of the KFFE dimer (Figure 3). This structure was rotated around the *x*-axis from 0° up to 90° in increments of 10° and the energy was calculated (a) before and (b) after minimization using the GBSW implicit solvent model with 24 radial grid points and 38 (black), 50 (red), 110 (green), and 590 (blue) angular Lebedev grid points.

was calculated using 24 radial grid points and 38, 50, 110, and 590 angular grid points.[81] It can be seen in Figure 9b that after minimization the energy varies by more than 4 kcal mol$^{-1}$ for the default value of 38 angular grid points and by more than 2 kcal mol$^{-1}$ for 50 Lebedev grid points. The resulting minimized structures are also different from each other. Another problem is that the energy variations before and after minimization are not correlated, i.e. orientations with high energies before minimization can become lowest in energy after minimization, and vice versa, as the comparison between Figure 9a and b reveals. The results improve on increasing the number of Lebedev grid points, with energies varying less than 1 kcal mol$^{-1}$ for 110 and less than 0.1 kcal mol$^{-1}$ for 590 grid points. However, for methods that rely on the unambiguous identification of stationary points of the potential energy surface, such as discrete path sampling,[87,88] even a set of 590 angular grid points does not give sufficiently converged results. Furthermore, the KFFE dimer with 160 atoms is a rather small system, and the magnitude of the energy variation with orientation scales linearly with increasing system size. For this reason, and because of the linear scaling of the computer time required for volume integration with the number of integration points and atoms, the solution to this problem probably does not lie in increasing the number of integration points. We note that these issues are probably unimportant in conventional Monte Carlo or molecular dynamics simulations.

For calculations that do not require energy gradients, we could simply orient the angular integration grid so that the Cartesian axes coincide with the principal axes of the molecule. This approach provides a rotationally invariant energy, but not the correct gradients. The calculation of solvation forces becomes more complicated because the integration points $\mathbf{r}_{mn}$ in eqs 23 and 24 in ref 11 are now dependent on all the atomic positions, and the calculation of the derivatives of the integration points $\mathbf{r}_{mn}$ with respect to each atom $\alpha$ would be necessary. The dependence of the integration points $\mathbf{r}_{mn}$ on the atomic positions $\{\mathbf{r}_\alpha\}$ results from the diagonalization of the inertia tensor of the molecule to obtain the rotation matrix that orients the angular grid points with respect to the principal axes. Hence, the derivatives $\partial\mathbf{r}_{mn}/\mathbf{r}_\alpha$are quite involved, and we have not pursued the option of orienting the grid. If precisely converged stationary points are required, it may be better to replace the numerical volume integration with an analytical implementation, as in the parameter-free pairwise descreening GB models,[5,6] while keeping the correction to the Coulomb field approximation.[6,10,13] Such an approach also has the advantage that the evaluation of the Born radii would probably be faster than the current implementation based on numerical integration.

## References

(1) Chen, J.; Brooks, C. L., III *Phys. Chem. Chem. Phys.* **2008**, *10*, 471–481.

(2) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(3) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122–129.

(4) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.

(5) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479–499.

(6) Tjong, H.; Zhou, H.-X. *J. Phys. Chem. B* **2007**, *111*, 3055–3061.

(7) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578–1599.

(8) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983–10990.

(9) Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. A* **1997**, *101*, 8098–8106.

(10) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *J. Chem. Phys.* **2002**, *116*, 10606–10614.

(11) Im, W.; Lee, M. S.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1691–1702.

(12) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1348–1356.

(13) Grycuk, T. *J. Chem. Phys.* **2003**, *119*, 4817–4826.

(14) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(15) Dominy, B. N.; Brooks, C. L. *J. Phys. Chem. B* **1999**, *103*, 3765–3773.

(16) Dominy, B. N.; Brooks, C. L. *J. Comput. Chem.* **2001**, *23*, 147–160.

(17) Feig, M.; Brooks, C. L. *Proteins* **2002**, *49*, 232–245.

(18) Felts, A. K.; Gallicchio, E.; Wallqvist, A.; Levy, R. M. *Proteins* **2002**, *48*, 404–422.

(19) Zhu, J.; Zhu, Q.; Shi, Y.; Liu, H. *Proteins* **2003**, *52*, 598–608.

(20) Fiser, A.; Feig, M.; Brooks, C. L.; Sali, A. *Acc. Chem. Res.* **2002**, *35*, 413–421.

(21) Steinbach, P. J. *Prot. Struct. Func. Bioinf.* **2004**, *57*, 665–677.

(22) Bursulaya, B. D.; Brooks, C. L. *J. Phys. Chem. B* **2000**, *104*, 12378–12383.

(23) Karanicolas, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3432–3437.

(24) Ohkubo, Y. Z.; Brooks, C. L. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13916–13921.

(25) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Func., Gen.* **2002**, *46*, 24–33.

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

(27) Stultz, C. M. *J. Chem. Phys B* **2004**, *108*, 16525–16532.

(28) Huang, A.; Stultz, C. M. *Biophys. J.* **2007**, *92*, 34–45.

(29) Sun, Y.; Latour, R. A. *J. Comput. Chem.* **2006**, *27*, 1908–1922.

(30) Tjernberg, L.; Hosia, W.; Bark, N.; Thyberg, J.; Johansson, J. *J. Biol. Chem.* **2002**, *277*, 43243–43246.

(31) Baumketner, A.; Shea, J.-E. *Biophys. J.* **2005**, *89*, 1493–1503.

(32) Wei, G.; Mousseau, N.; Derreumaux, P. *J. Phys. Cond. Mat.* **2004**, *16*, 5047–5054.

(33) Wei, G.; Mousseau, N.; Derreumaux, P. *Biophys. J.* **2004**, *87*, 3648–3656.

(34) Melquiond, A.; Boucher, G.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2005**, *122*, 174904.

(35) Melquiond, A.; Mousseau, N.; Derreumaux, P. *Prot. Struct. Func. Bioinf.* **2006**, *65*, 180–191.

(36) Meinke, J. H.; Hansmann, U. H. E. *J. Chem. Phys.* **2007**, *126*, 014706.

(37) Gsponer, J.; Haberthür, U.; Caflisch, A. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5154–5159.

(38) Cecchini, M.; Curcio, R.; Pappalardo, M. M.; Melki, R.; Caflisch, A. *J. Mol. Biol.* **2006**, *357*, 1306–1321.

(39) Strodel, B.; Whittleston, C. S.; Wales, D. J. *J. Am. Chem. Soc.* **2007**, *129*, 16005–16014.

(40) Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.

(41) Doye, J. P. K.; Wales, D. J. *Phys. Rev. Lett.* **1998**, *80*, 1357–1360.

(42) Wales, D. J.; Scheraga, H. A. *Science* **1999**, *285*, 1368–1372.

(43) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, 2003.

(44) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(45) Lazaridis, T.; Karplus, M. *Proteins: Struct., Func., Gen.* **1999**, *35*, 133–152.

(46) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902–1921.

(47) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(48) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(49) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.

(50) Feig, M.; MacKerell, A. D.; Brooks, C. L. *J. Phys. Chem. B* **2003**, *107*, 2831–2836.

(51) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6611–6615.

(52) Mezey, P. G. *Potential Energy Hypersurfaces*; Elsevier: Amsterdam, 1987.

(53) Wales, D. J. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 653–657.

(54) Derreumaux, P. *J. Chem. Phys.* **1997**, *106*, 5260–5270.

(55) Derreumaux, P. *J. Chem. Phys.* **1997**, *107*, 1941–1947.

(56) Miller, M. A.; Wales, D. J. *J. Chem. Phys.* **1999**, *111*, 6610–6616.

(57) Mortenson, P. N.; Wales, D. J. *J. Chem. Phys.* **2001**, *114*, 6443–6454.

(58) Mortenson, P. N.; Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2002**, *117*, 1363–1376.

(59) Carr, J. M.; Wales, D. J. *J. Chem. Phys.* **2005**, *123*, 234901.

(60) Verma, A.; Schug, A.; Lee, K. H.; Wenzel, W. *J. Chem. Phys.* **2006**, *124*, 044515.

(61) Mortenson, P. N.; Wales, D. J. *J. Chem. Phys.* **2001**, *114*, 6443–6454.

(62) Feig, M.; Karanicolas, J.; Brooks, C. L. *J. Mol. Graph. Mod.* **2004**, *22*, 377–395.

(63) Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.

(64) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327–341.

(65) Andersen, H. C. *J. Comp. Phys.* **1983**, *52*, 24–34.

(66) Chocholoušová, J.; Feig, M. *J. Comput. Chem.* **2006**, *27*, 719–729.

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

(68) Klimov, D. K.; Newfield, D.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8019–8024.

(69) Pellarin, R.; Caflisch, A. *J. Mol. Biol.* **2006**, *360*, 882–892.

(70) Hutchinson, E. G.; Thornton, J. M. *Protein Sci.* **1994**, *3*, 2207–2216.

(71) Dang, L. X.; Pettitt, B. M. *J. Am. Chem. Soc.* **1987**, *109*, 5531–5532.

(72) Friedman, G. L. *Faraday Discuss. Chem. Soc.* **1988**, *85*, 1–11.

(73) Masunov, A.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 1722–1730.

(74) Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. *J. Chem. Phys.* **2004**, *121*, 10748.

(75) Paci, E.; Gsponer, J.; Salvatella, X.; Vendruscolo, M. *J. Mol. Biol.* **2004**, *340*, 555–569.

(76) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 265–284.

(77) Chen, J.; Im, W.; Brooks, C. L. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.

(78) Gazit, E. *FASEB J.* **2002**, *16*, 77–83.

(79) Schaefer, M.; Bartels, C.; Leclerc, F.; Karplus, M. *J. Comput. Chem.* **2001**, *22*, 1857–1879.

(80) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in FORTRAN*, 2 ed.; Cambridge University Press: Cambridge, 1992.

(81) Lebedev, V. I.; Laikov, D. N. *Doklady Math.* **1999**, *59*, 477–481.

(82) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.

(83) Verlet, L. *Phys. Rev.* **1967**, *165*, 201–214.

(84) Evans, D. J.; Holian, B. L. *J. Chem. Phys.* **1985**, *83*, 4069–4074.

(85) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667–683.

(86) Bogusz, S.; Cheatham, T.; Brooks, B. *J. Chem. Phys.* **1998**, *108*, 7070–7084.

(87) Wales, D. J. *Mol. Phys.* **2002**, *100*, 3285–3306.

(88) Wales, D. J. *Int. Rev. Phys. Chem.* **2006**, *25*, 237–282.

CT700305W

JCTC Journal of Chemical Theory and Computation

# Transition-State Docking of Flunitrazepam and Progesterone in Cytochrome P450

Patrik Rydberg,[†,‡] Sine Myrup Hansen,[†] Jacob Kongsted,[‡] Per-Ola Norrby,[§] Lars Olsen,[†] and Ulf Ryde*[,‡]

*Biostructural Research Group, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen Ø, Denmark, Department of Theoretical Chemistry, Lund University, Chemical Centre, P.O. Box 124, SE-221 00 Lund, Sweden, and Department of Chemistry, University of Gothenburg, Kemigården 4, SE-412 96 Göteborg, Sweden*

Received November 16, 2007

**Abstract:** We have developed a method to dock a transition-state structure into the active site of an enzyme. Such an approach is more discriminative than standard docking when looking for substrates of an enzyme, because a transition state has more sterical restrictions than a nonreactive state. We use an accurate and tailored force field for the transition-state for the hydroxylation reaction in cytochrome P450, obtained with the Q2MM method. We apply this method to the docking of two drugs, progesterone and flunitrazepam, to the active sites of two human cytochromes P450, 2C9 and 3A4. We obtain a qualitative agreement compared to experiments, both for hydrogen atoms bound to the same carbon atom (for which the force-field energies are directly comparable) and for general sites on the drug molecules, if the method is combined with an estimate of the intrinsic reactivity of the various sites. However, the method does not rank all the sites correctly. It is not significantly improved if the proteins are allowed to relax locally or if it is combined with the MM/PBSA approach, which fully accounts for the protein flexibility and explicitly treats solvation and entropy effects. On the other hand our method performs better than standard docking with the GOLD software or predictions of metabolic sites with the MetaSite software.

## Introduction

The effect of drugs is determined not only by their action on their target proteins but also by the concentration of the drug at the target, among other things. This concentration is a function of many different factors, e.g. the amount administered, the uptake in the body, the excretion from the body, and the modification and degradation of the drug in the body. Thus, the drug metabolism and pharmacokinetics are very important for the success of a drug candidate and much effort has been devoted to their measurement and prediction.

Of particular interest in this aspect is the group of enzymes called cytochromes P450 (CYPs). They are oxidative liver enzymes that degrade foreign substances in the body. In fact, they metabolize ~90% of the drugs on the market.[1] The active site of these enzymes consists of a haem group, i.e. an iron ion in the center of a porphyrin ring. Below the ring, the metal is coordinated to a cysteine ligand, whereas the upper site is open to the coordination of a small ligand, like water or $O_2$. During the catalytic cycle, a highly reactive Fe(V)═O complex (formally) is formed, called compound I. It has the potential to oxidize many chemical groups, e.g. by aliphatic hydroxylation, aromatic hydroxylation or epoxidation, dealkylation, N, S, or SO oxidation.

* To whom correspondence should be addressed. Phone: (+46) 46 2224502. Fax: (+46) 46 2224543. E-mail: Ulf.Ryde@teokem.lu.se.

[†] University of Copenhagen.

[‡] Lund University.

[§] University of Gothenburg.

10.1021/ct700313j CCC: $40.75 
Published on Web 03/04/2008

Much effort has been directed to the prediction of CYP reactivity of drug candidates. It has been realized that this consists of two topics, viz. estimation of the intrinsic reactivity of the various groups in the drug and calculation of the steric accessibility of the groups to the oxoferryl group of compound I,[2] whereas dynamic effects during binding seem to be less important.[3] The intrinsic reactivity has been thoroughly studied for the most interesting reaction types, both with density functional theory (DFT) methods[4,5] and by more approximate, but faster methods.[6,7] The steric effects have also been studied, e.g. by solvent accessibility, docking, and QSAR methods.[8,9] In some cases, both types of effects were considered.[2,8,10,11]

However, nobody seems to have employed the fact that, for substrates, a transition state must form. Such a state has quite severe sterical restrictions (for aliphatic hydroxylation, the reacting hydrogen atom on the substrate should be ~1.25 Å from the oxoferryl group with Fe−O−H and O−H−C angles of ~121° and 171°).[12] Thus, docking of the transition state should be more discriminative than a normal docking of the substrate into the active site. In fact, it has been shown that docking of high-energy intermediates to enzymes can improve the prediction of their function.[13,14] Therefore, we in this paper develop a transition-state docking procedure and test it on two human CYPs.

Docking programs employ an empirical energy function to determine and optimize the interaction energy between the drug candidate and the active site, typically in the form of a molecular mechanics (MM) force field. This is a problem for docking of transition states, because force fields normally are only developed for equilibrium states—for transition states, which are first-order saddle points on the potential surface, special optimization algorithms are needed, which are less robust than those for minima (i.e., they cannot guarantee that a transition state is found from every starting position, and they normally require information about the curvature of the potential around the saddle point). Four types of methods have been employed in the optimization of transition states with MM methods.[15,16] In this paper, we use the simplest approach, Q2MM,[17] in which the transition state is converted to an equilibrium state during the parametrization (the negative frequency is replaced with a high positive frequency). Thereby, any minimization code can be used and a transition state is found from any reasonable structure, which is important for a stable docking code.

We use a general Q2MM force field for the transition state of aliphatic hydroxylation, which we recently developed and tested.[12] It was obtained for a training set of 14 transition-state structures of small model molecules with most chemical groups commonly encountered in drugs, optimized at the DFT level, and it was tested for a set of 10 additional molecules. We use this force field to dock two drugs, progesterone and flunitrazepam, into the active sites of two human CYPs, 3A4 and 2C9. We test whether we can predict which sites of the drugs will be metabolized by the respective enzyme by combining the docked results with a measure of the intrinsic reactivity of each site.[5] We also test if the predictions of the method can be improved by optimizing the geometry of the active site side chains or by including dynamic, solvation, hydrophobic, and entropy effects by the MM/PBSA (molecular mechanics Poisson–Boltzmann surface area) approach.[18] Finally, we compare the results with two other methods to predict reactivity in the CYPs, standard docking with the GOLD software[19] and CYP reactivity predictions by MetaSite.[8]

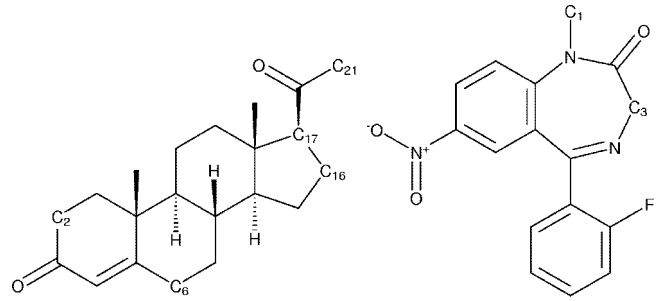


***Figure 1.*** Substrates studied, progesterone (a) and flunitrazepam (b), with the considered carbon atoms indicated.

## Methods

**Systems Studied.** Our calculations are based on the crystal structures of CYP 2C9 and 3A4 (PDB entries 1R9O and 1TQN).[20,21] The protein part of the system was set up in the same way as in our previous study of the water dynamics in the active-site cavity of these proteins.[22] To describe the transition state, we added an oxygen atom with a bond length of 1.76 Å to the iron ion, opposite to the sulfur atom of the cysteine ligand.

We studied two substrates, progesterone and flunitrazepam (shown in Figure 1). They were described with the general AMBER force field,[23] and the transition state was modeled by our recently developed transition-state force-field parameters.[12]

We studied the transition states for hydroxylation of the 2$\beta$, 6$\beta$, 16$\alpha$, 16$\beta$, 17$\alpha$, and 21 positions on progesterone, and the 1 and 3 positions on flunitrazepam. The $\beta$ hydrogen atoms in progesterone are directed toward the viewer in Figure 1.

**Transition-State Docking.** All calculations with the transition-state force field were done with the AMBER software suite, version 9.[24] Molecular mechanics minimizations were run with a distance-dependent dielectric constant of 4$r$ and an infinite cutoff for nonbonded interactions. The temperature was kept constant at 300 K using the Berendsen weak-coupling algorithm[25] with a time constant of 1 ps. We also tested a generalized Born (GB) solvation model,[26,27] but this increased the calculation time too much.

First, we generated a starting position with the reactive hydrogen atom close to the oxoferryl group by manually docking the substrates into the active site in a reasonable conformation. This structure was refined by a molecular mechanics optimization, in which the protein and the haem group were kept fixed and only the substrate allowed to move. To fully sample all possible conformations, we then did a full conformational analysis of the substrates by systematically rotating (with local software) the four rotable bonds (Fe−O, O−H, H−C, and C17−C20 in progesterone or the bond to the fluorophenyl ring in flunitrazepam) 6-fold

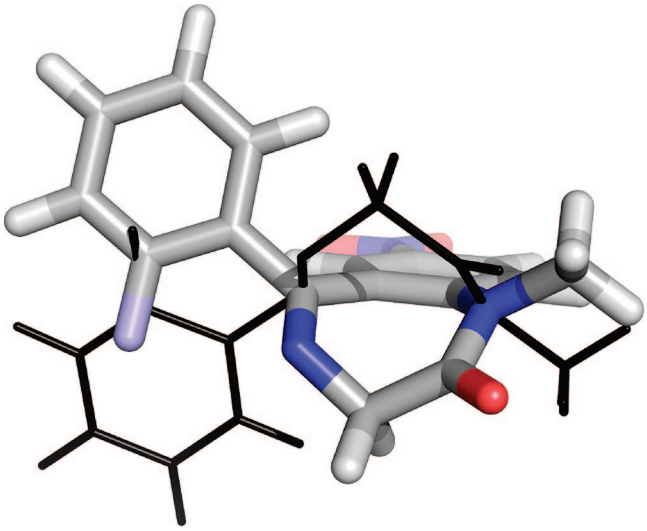

***Figure 2.*** Two flunitrazepam conformations: conformation 1 in color and conformation 2 in black lines.

(8-fold for the Fe−O bond) with a fixed protein structure, generating 1728 conformers of the substrates. All these structures were optimized with molecular mechanics, keeping the protein and haem group fixed. From this conformational analysis, we took the structure with the lowest energy for further analysis of the binding and estimation of the binding affinities.

For flunitrazepam, there are two possible conformations of the seven-membered ring, as is shown in Figure 2. The two conformations are degenerate in quantum chemical calculations. Therefore, we tested both conformations, but report only the results of the conformation that gives the lowest docked energy (conformation 1 for all calculations, except for the reaction of C1 in CYP 3A4).

In some calculations, we studied the effects of the protein flexibility by performing an additional geometry optimization of the structure with the lowest energy, with all side chains around the active site free to move. These side chains were 100–102, 104, 106, 108, 113, 114, 197, 200, 201, 204, 205, 208, 209, 233, 234, 236, 237, 240, 292, 293, 295–297, 299–301, 304, 361, 362, 366, 474, 476, 477, and 479 in 2C9, and 105, 106, 108, 119, 120, 212, 213, 215, 241, 301, 304, 305, 309, 369, and 370 in 3A4.

**Binding Energies.** The binding energy of a ligand (L) to a protein (P) is the free energy of the reaction:

$$P + L \rightarrow PL \tag{1}$$

where PL is the complex between the protein and the ligand. Therefore, we could estimate the binding by

$$E_{bind}' = E_{PL} - E_P - E_L \tag{2}$$

where $E_{PL}$, $E_L$, and $E_P$ are the MM energies of PL, P, and L. However, for a transition-state, in which there is a bond between the reactive hydrogen atom and the oxoferryl group of compound I, $E_{PL}$ contains bonded energy terms (one bond, two angles, and six or seven dihedrals) that are not present in either $E_L$ or $E_P$. A simple and intuitive way to compensate for this is to subtract the energy of these nine or ten interaction terms in a fully optimized small model of the substrate and the haem group, $E_{tsbond}$.

$$E_{bind}'' = E_{PL} - E_P - E_L - E_{tsbond} \tag{3}$$

Thereby, all bonded terms appear on both sides in eq 3, making the energies comparable. Moreover, $E_{tsbond}$ provides an estimate of how much the transition state is strained relative to the best possible unstrained structure.

Ideally, the small model for which we calculate $E_{tsbond}$ should be the same model for which the intrinsic reactivity is estimated (see below), because, for larger substrates, this energy will already include some steric effects (the large haem group will restrict the approach of some sites to the reactive oxoferryl group). Therefore, we have used the Fe(porphine)($SCH_3$)(O)(substrate) model, which was used in our QM calculations.[5]

Strictly speaking, all three species in eq 1 should be studied in their optimum states in aqueous solution. However, as a first approximation, we have kept the geometry of the protein fixed to save time and avoid the risk of ending up in different local minima for the various complexes. On the other hand, we optimize the geometry of the isolated ligand, using an implicit solvent model with a distance-dependent dielectric constant ($\mu = 4r$), giving the energy $E_{L,opt}$. Thus, we end up with the estimate:

$$E_{bind} = E_{PL} - E_P - E_{L,opt} - E_{tsbond} \tag{4}$$

To this binding affinity of the transition state, we should add an estimate of the intrinsic reactivity of each site (when comparing different reactive sites),[2] $E_{QM}$. In our previous paper on aliphatic hydroxylation, we found that activation barriers calculated with density functional theory employing the B3LYP functional using a methoxy radical gives a reliable estimate of the reaction barrier ($E_{QM}$).[5] The $E_{QM}$ values for progesterone were published in that paper and the flunitrazepam data can be found in the Supporting Information. Thus, our estimated activation energy for the various reactive sites is

$$E_{est} = E_{bind} + E_{QM} \tag{5}$$

Absolute values of $E_{est}$ do not have any specific meaning, but relative values of different reactive sites of the same drug in one protein should indicate their relative reactivities, a positive value indicating a higher activation barrier and therefore a poorer substrate.

**MM/PBSA Calculations.** The energies in eqs 2–4 are pure MM energies, and they are obtained with a primitive solvation model. In an attempt to improve these, we have employed the MM/PBSA approach[18] for the docked structure with the lowest energy. In this approach, each of the three (free) energies on the right-hand side of eq 2 is estimated as a sum of four terms:

$$E = \langle E_{MM} \rangle + \langle G_{Solv} \rangle + \langle G_{np} \rangle - T\langle S_{MM} \rangle \tag{6}$$

where $G_{Solv}$ is the polar solvation energy of the molecule, estimated by the solution of the Poisson–Boltzmann (PB) equation,[18] $G_{np}$ is the nonpolar solvation energy (including the cost of making a cavity in the solvent, solvent entropy, the hydrophobic effect, and solvent–solute dispersion and repulsion), estimated form the solvent-accessible surface area of the molecule,[28] $T$ is the temperature, $S_{MM}$ is the entropy of the molecule, estimated from a normal-mode analysis of

harmonic frequencies calculated at the molecular mechanics (MM) level, and $E_{MM}$ is the MM energy of the molecule, i.e. the sum of the internal energy of the molecule (i.e., the bonded terms, $E_{int}$) and the electrostatics ($E_{es}$) and van der Waals interactions ($E_{vdW}$):

$$E_{MM} = E_{int} + E_{es} + E_{vdW} \quad (7)$$

All the terms in eq 6 are averages of energies obtained from 20 snapshots taken from molecular dynamics (MD) simulations. In order to reduce the time-consumption and to obtain stable energies, the same geometry is normally used for all three reactants (complex, ligand, and receptor), i.e. only the PL complex is simulated by MD.[29]

The MM/PBSA calculations followed the same protocol and settings as in ref 30, and here we only outline the differences. In the MD simulations, the SHAKE algorithm[31] was not used, and consequently, a smaller time step (0.5 fs) had to be used. However, the duration of the MD simulations for both the equilibration (275 ps) and production runs (200 ps) were the same as in ref 30. All MD simulations used periodic boundary conditions with an octahedral box, extending at least 9 Å outside the protein (in total, ~40 400 and 46 500 atoms in CYP 2C9 and 3A4, respectively). Each MM/PBSA calculation was based on 20 snapshots extracted from the MD production run. The components of the binding energy in eq 6 were obtained using the mm_pbsa module of Amber 8.0.[24] The polar solvation energy was calculated with the Poisson–Boltzmann model, calculated with the DelPhi II software,[32] or with the default Generalized Born model in Amber 8.0 ($GB^{OBC}$)[27] (we use the acronym MM/GBSA for the latter results).

It has previously been shown that the entropic contribution to the free energy obtained within the MM/PBSA method may lead to large fluctuations and thereby a significant standard deviation in the predicted binding affinities.[30] The reason for this is that the protein is truncated and then minimized before the calculation of the frequencies from which the vibrational contribution to the entropy is derived. We have recently formulated an alternative method in which a fixed buffer region is introduced into the minimization and frequency calculations,[33] thereby stabilizing the vibrational entropic term. This new method is employed in this paper.

Finally, we have added the same $E_{tsbond}$ correction for the extra bond in the transition-state complex and the $E_{QM}$ correction for the intrinsic reactivity of each site as in the $E_{est}$ estimate to all the MM/PBSA and MM/GBSA energies.

**GOLD and MetaSite Calculations.** For comparison, we also performed a set of standard docking calculations, using the GOLD software,[19] version 3.1. All water molecules except the one bound to the haem iron were deleted from the crystal structure (and also ligands, if any), and hydrogen atoms were added to the protein, but not to the water molecule (in order to mimic compound I). The Fe–O distance was modified to 1.62 Å. We used a flat-bottomed harmonic distance restraint of 2.4–2.7 Å between this oxygen atom and the hydrogen atom to be abstracted with a spring constant of 50.0 kcal/(mol Å$^2$) to define the reactive site. This distance was chosen from an optimized complex of methane and compound I, in which it is 2.55 Å. This is similar to the approach used by Vermeulen and co-workers, in which only docked poses for which the reactive carbon atom was within 6 Å of the iron ion were considered.[34]

Finally, we also predicted the reactivity of the various sites with the MetaSite software, version 2.7.5.[8] These calculations employed the default settings and the parameters for the CYP 3A4 and CYP 2C9 models implemented in the software.

***Table 1.*** Calculated $E_{est}$ Energies (kJ/mol) for the Reaction of Flunitrazepam in Human CYP 2C9 and 3A4[a]

| | 2C9 | | 3A4 | |
|---|---|---|---|---|
| | C1 | C3 | C1 | C3 |
| $E_{est}$ | −27 | 86 | −87 | −83 |
| $V_{max}$ | 0.38 | | 1.54 | 15.7 |

[a] For comparison, experimental $V_{max}$ data are also included (min$^{-1}$).[37]

## Results and Discussion

**Flunitrazepam.** Experimentally, it is known that the CYPs convert flunitrazepam to desmethylflunitrazepam and 3-hydroxyflunitrazepam.[35] These metabolites are formed after initial hydrogen abstraction from the aliphatic carbons C1 and C3, respectively, and several different CYP isoforms can catalyze these reactions. For example, CYP 2A6 and 3A4 catalyze both reactions, whereas CYP 2B6, 2C9, and 2C19 only cause demethylation.[36]

We have docked the transition states corresponding to the hydroxylation reactions on C1 and C3 into CYP 2C9 and 3A4, performing a systematic conformational search, as described in the Methods section. For the most stable of the 1728 considered conformations, we calculated the estimated binding affinity by eq 5. The results are listed In Table 1. It can be seen that there is a significant correlation between $E_{est}$ and the experimental data. For CYP 3A4, $E_{est}$ is around −85 kJ/mol for both reactions, whereas for CYP 2C9 $E_{est}$ is negative for the 1-demethylation, but positive for the 3-hydroxylation. This is in good agreement with the experimental observation that CYP 2C9 does not catalyze the reaction at C3, whereas CYP 3A4 catalyzes both reactions. However, our method predicts that the C1 site should be more reactive than the C3 site in CYP 3A4 (by 4 kJ/mol), although experiments indicate the opposite (by 6 kJ/mol), so the predictions are not fully quantitative.

**Progesterone.** It is well-known that steroids in general and progesterone in particular are metabolized by several human CYPs at different positions.[37–39] All the various metabolites of progesterone are alcohols formed by hydroxylation of aliphatic carbons, which makes it a good test case for our force-field and transition-state docking procedure. Four different metabolites are produced by CYP 3A4, viz. the 2$\beta$-, 6$\beta$-, 16$\alpha$-, and 21-hydroxylated progesterone, whereas CYP 2C9 primarily catalyzes the reaction at the 21 position (with minor metabolites from the 6$\beta$ and 16$\alpha$ positions).

To test our force field, we first studied the $E_{PL}$ energies (i.e., the docked energies) of the 16$\alpha$ and the 16$\beta$ hydroxylations, because for hydrogen atoms bound to the same carbon atom, the Q2MM energies are directly comparable.

***Table 2.*** Calculated $E_{est}$ Energies (kJ/mol) for the Reaction of Progesterone in Human CYP 2C9 and 3A4[a]

| | 2C9 | | | | | | 3A4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 21 | 16α | 6β | 2β | 16β | 17 | 6β | 16α | 2β | 21 | 16β | 17 |
| $E_{est}$ | 82 | 289 | 335 | 178 | 533 | 695 | 136 | 60 | 17 | −68 | 139 | 346 |
| $V_{max}$ | 0.51 | 0.08[b] | 0.04 | | | | 33 | 8.7 | 8[b] | 1.1 | | |

[a] In addition, experimental $V_{max}$ data is included ($min^{-1}$).[40] [b] Estimated from single-point measurements, assuming that the $K_M$ values are equal to those of the 6β hydroxylation.

Quite satisfactorily, our results show that the $E_{PL}$ energies are 80 and 246 kJ/mol lower for the H16α atom than for the H16β atom in CYP 3A4 and 2C9, respectively, clearly explaining why only the 16α product is observed.[39] These energy differences come from the bonded energy terms (bonds, angles, and dihedrals) and the van der Waals energies, about 50% from each, whereas electrostatics has only a minor influence. This shows that the preference to react at H16α comes from a better fit in the active site.

However, to compare all the other reactive sites, we instead need to study the $E_{est}$ energies. These are listed and compared to experimental $V_{max}$ values in Table 2. It can be seen that most of the results are qualitative correct: $E_{est}$ is large and positive (139–695 kJ/mol) for the two sites (16β and 17) that do not react in any of the two proteins. A similar high value (178 kJ/mol) is observed also for the 2β position in CYP 2C9, the product of which is also not observed. The other sites have lower $E_{est}$ energies in CYP 3A4 (−68 to +136 kJ/mol), and these reactions are also observed experimentally. The experimentally most reactive site in CYP 2C9, C21, has the lowest $E_{est}$, 82 kJ/mol. However, $E_{est}$ for the nonreactive 2β position (178 kJ/mol) is smaller that of the 6β and 16α positions (335 and 289 kJ/mol), which are observed experimentally, although only as minor metabolites. Moreover, the quantitative correlation between $E_{est}$ and $V_{max}$ is poor for CYP 3A4.

**Improvements to the Transition-State Docking.** The approach used in our transition-state docking method, although similar to other docking approaches, is admittedly quite primitive, ignoring many important contributions to the true free energy of binding.[40] In particular, we keep the protein fixed in both the docking and the energy calculation.

This is the most common approach in docking, and it normally gives good results. However, it does not include the flexibility of the protein, which could be a reason why we cannot rank all sites correctly. A first attempt to include the flexibility of the protein would be to allow some amino acids to relax during the docking.

We tried such an approach by performing a geometry optimization of the amino-acid side chains in the active for the best docked pose of each reactive site, as is described in the Methods section. The results are described in Tables 3 and 4. It can be seen that the estimated activation energies ($E_{flex}$) all become more favorable as expected and also much more similar and therefore more realistic. Unfortunately, this means that it becomes harder to discern which sites are metabolized and which are not. For flunitrazepam, flexible transition-state docking performs excellently, pointing out the reactive sites in both proteins. However, for progesterone, the results are worse. $E_{flex}$ clearly and correctly shows that the 16β site is not reactive in any of the proteins (by 77–155 kJ/mol). However, for other nonreactive sites (C17 in both proteins and 2β in CYP 2C9), $E_{flex}$ is equal or more negative than for some of the reactive sites, although the energy difference is not large 0–9 kJ/mol. Moreover, the reactive sites are ranked in a completely erroneous order (but again with an energy difference of only 9–16 kJ/mol). Thus, such a partly flexible method does not lead to any consistent improvement of the transition-state docking.

***Table 3.*** Comparison of the Results of Transition-State Docking with a Rigid ($E_{est}$) or Partly Flexible ($E_{flex}$) Protein, As Well As the MM/GBSA, MM/PBSA, MetaSite, and GOLD Results for Flunitrazepam[a]

| | 2C9 | | 3A4 | | | |
|---|---|---|---|---|---|---|
| | C1 | C3 | C1 | C3 | $r^2$ | #miss |
| $E_{est}$ | −27 | 86 | −87 | −83 | 0.56 | 0 |
| $E_{flex}$ | −177 | −162 | −179 | −179 | 0.62 | 1 |
| MM/PBSA | 130 | 320 | 174 | 391 | 0.95 | 1 |
| MM/GBSA | 82 | 237 | 78 | 219 | 0.85 | 1 |
| $E_{QM}$ | 48 | 43 | 48 | 43 | 0.86 | 1 |
| MetaSite | 1.69 | 1.07 | 2.13 | 1.50 | 0.18 | 1 |
| GOLD | 38.4 | 37.2 | 49.9 | 47.5 | 0.42 | 1 |
| $V_{max}$ | 0.38 | | 1.54 | 15.7 | | |

[a] Energies are in kilojoules per mole (a low value indicates a reactive site), MetaSite and Gold scores are in arbitrary units (a high value indicates a reactive site), and $V_{max}$ is in inverse minutes. $r^2$ is the correlation coefficient between the respective data and the activation energy calculated from $V_{max}$. #miss is the number of qualitatively incorrect predictions.

However, the energy function used in our transition-state docking, eqs 2–4, involves many approximations. In particular, many important energy terms are missing, e.g. solvation, dynamics, entropy, hydrophobic effects, etc.[40] We have therefore tried to improve the results of our transition-state docking approach by including all these terms in a well tested and computationally effective way, viz. by the use of the MM/PBSA method.[18] In this approach, the standard MM energy is supplemented by energies for the entropy, polar and nonpolar solvation, and all energies are calculated for a number of snapshots taken from a molecular dynamics simulation of the complex, thereby taking into account also dynamics effects and the protein flexibility. The calculations were performed only on the most favorable docked conformation for each site. Naturally, the MM/PBSA method was extended to the transition-state docking by including the $E_{QM}$ correction for the intrinsic reactivity of each site and the $E_{tsbond}$ correction of the additional bond in the transition state. The results of such calculations for the docking of flunitrazepam and progesterone into CYP 3A4 and 2C9 are also included in Tables 3 and 4.

For progesterone in CYP 2C9, MM/PBSA performs rather well, predicting that the three nonreactive sites 2β, 16β, and 17 have higher energies (161–176 kJ/mol) than the two most

***Table 4.*** Comparison of the Results of Transition-State Docking with a Rigid ($E_{est}$) or Partly Flexible ($E_{flex}$) Protein, as well as the MM/GBSA, MM/PBSA, MetaSite, and GOLD Results for Progesterone[a]

| | 2C9 | | | | | | | | 3A4 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 21 | 16α | 6β | 2β | 16β | 17 | $r^2$ | #miss | 6β | 16α | 2β | 21 | 16β | 17 | $r^2$ | #miss |
| $E_{est}$ | 82 | 289 | 335 | 178 | 533 | 695 | 0.99 | 3 | 136 | 60 | 17 | −68 | 139 | 346 | 0.96 | 0 |
| $E_{flex}$ | −209 | −216 | −218 | −209 | −132 | −218 | 1.00 | 4 | −214 | −224 | −224 | −230 | −59 | −219 | 0.91 | 2 |
| MM/PBSA | 151 | 149 | 173 | 161 | 167 | 176 | 0.43 | 3 | 196 | 150 | 197 | 221 | 198 | 234 | 0.20 | 2 |
| MM/GBSA | 88 | 68 | 110 | 97 | 110 | 110 | 0.08 | 4 | 101 | 47 | 107 | 83 | 79 | 113 | 0.04 | 2 |
| $E_{QM}$ | 55 | 41 | 56 | 53 | 44 | 61 | 0.04 | 3 | 55 | 41 | 56 | 53 | 44 | 61 | 0.00 | 4 |
| MetaSite | 1.45 | 1.38 | 1.25 | 1.29 | 1.43 | 1.33 | 0.83 | 5 | 1.62 | 1.71 | 1.49 | 1.85 | 1.79 | 1.70 | 0.45 | 5 |
| GOLD | 39.1 | 35.3 | 3.11 | 15.4 | 7.11 | 19.7 | 0.60 | 4 | 16.2 | 32.4 | 35.3 | 43.7 | 32.2 | −18.8 | 0.89 | 2 |
| $V_{max}$ | 0.51 | 0.08[b] | 0.04 | | | | | | 33 | 8.7 | 8[b] | 1.1 | | | | |

[a] Energies are in kilojoules per mole (a low value indicates a reactive site), MetaSite and Gold scores are in arbitrary units (a high value indicates a reactive site), and $V_{max}$ is in inverse minutes. $r^2$ is the correlation coefficient between the respective data and the activation energy calculated from $V_{max}$. #miss is the number of qualitatively incorrect predictions. [b] Estimated from single-point measurements, assuming that the $K_M$ values are equal to those of the 6β hydroxylation.

reactive sites (149–151 kJ/mol). Unfortunately, the least reactive site, 6β, has a higher energy 173 kJ/mol, and the ranking between the 21 and 16α sites is incorrect (by 7 kJ/mol). Likewise, for progesterone in CYP 3A4, MM/PBSA predicts a higher energy for the reactive 21 site (221 kJ/mol) than for one of the nonreactive site (16β with 198 kJ/mol). The MM/GBSA predictions are similar or slightly worse. For flunitrazepam, both MM/PBSA and MM/GBSA are poor, predicting that the C3 site is appreciably less reactive than the C1 site in both enzymes. Thus, the MM/PBSA method does not give any improvement compared to the simpler $E_{est}$ estimate.

The MM/PBSA energies are dominated by a favorable van der Waals energy (∼170 kJ/mol) and an unfavorable solvation energy (∼180 kJ/mol for PB and ∼90 kJ/mol for GB). Interestingly, the electrostatic energy is also unfavorable, so it is not canceled by the solvation term, as is normally observed. This may explain the poor performance of MM/PBSA for the CYPs, and it is probably connected to the fact that the active site is completely hidden in the center of the enzyme. The difference between the C1 and C3 sites in both enzymes is completely caused by the electrostatic term. The entropy term is always unfavorable (∼80 kJ/mol), and the nonpolar solvation is favorable and nearly constant (22 kJ/mol). The internal energy, which comes entirely from the 6–7 extra bonded interactions in the transition state, is unfavorable and small (∼10 kJ/mol). It gives an indication of how strained each transition state is. In fact, it is perfectly correlated to $V_{max}$ for progesterone in CYP 3A4 (all six sites are ranked correctly), but the ranking is much worse for CYP 2C9 or for flunitrazepam. The average statistical standard deviation of the MM/PBSA estimates is 6 kJ/mol and slightly lower for MM/GBSA, which also may explain the poor and varying results.

Finally, we have included in Tables 3 and 4 also the intrinsic reactivity of each site, estimated by a methoxy-radical model ($E_{QM}$). It can be seen that this factor alone has a poor predictive power: For example, it can never explain why the C3 site in flunitrazepam is reactive in CYP 3A4, but not in 2C9. The same applies to C21 in progesterone, and it can also be noted that $E_{QM}$ attains its second lowest value for the nonreactive 16β site. Thus, the steric (docking) energies are very important for a predictive method.

**Comparison to Other Methods.** We have seen that transition-state docking is rather successful in discriminating reactive and nonreactive sites of both flunitrazepam and progesterone in CYP 3A4 and 2C9, but it often fails to rank the sites correctly. In order to decide how useful such an approach is, we need to compare its performance with other available methods. Therefore, we have studied the same drugs and enzymes with two other methods, GOLD, a standard docking (and scoring) program,[19] and MetaSite,[8] an integrated method to predict the reactivity of each site of a drug molecule against several human CYPs. The results of these methods are shown in Tables 3 and 4 for flunitrazepam and progesterone, respectively (full data on the GOLD and MetaSite calculations for all the sites on the substrates are available in the Supporting Information).

To facilitate the comparison, we have included two quality criteria in the table. First, we give the correlation coefficient between the theoretical and experimental data ($r^2$). Our methods provide estimates to the activation energy. Therefore, we cannot directly compare to $V_{max}$ data; instead, $V_{max}$ has to be recalculated to activation energies ($-RT\ln(hV_{max}/kT)$). Unfortunately, no experimental data is available for the nonreactive sites ($V_{max} = 0$ gives an infinite activation energy). Therefore, we also included a more qualitative criterion, viz. the number of qualitative misses in the predictions (#miss). For flunitrazepam we consider the method qualitatively wrong if the predictions for the two sites are too dissimilar for CYP 3A4 (for which both sites are reactive) or too similar for 2C9 (for which only C1 is reactive), where two predictions are considered similar if they differ by less than 10% (any value up to ∼30% gives the same results). For progesterone, we define the number of misses as the number of nonreactive sites that have a higher $E_{est}$ value than any of the reactive sites plus the number of reactive sites that have a lower $E_{est}$ value than any of the nonreactive sites (and similar for the other methods).

Using these criteria, we can see from Table 3 that for flunitrazepam, $E_{est}$ is the only method that gives qualitatively correct results for both enzymes: MetaSite gives a too large difference between 1-demethylation and 3-hydroxylation in 3A4, whereas GOLD does not give any significant difference between the two sites in 2C9. This is also reflected in the

correlation coefficient, which is 0.56, 0.18, and 0.42 for the three methods, respectively.

For progesterone, $E_{est}$ gave qualitatively correct predictions for CYP 3A4 (#miss = 0), but it had a qualitative problem with the 2$\beta$ site in CYP 2C9 (giving #miss = 3). However, MetaSite had even worse problems: It gives the second highest score to the nonreactive 16$\beta$ position and the lowest score to the reactive 6$\beta$ position in CYP 2C9 (#miss = 5), and it gives even poorer predictions for CYP 3A4 (the two nonreactive sites are ranked as number two and four; #miss = 5). GOLD docking performs somewhat better, but it gives the lowest score to the reactive 6$\beta$ site in CYP 2C9 (#miss = 4) and the second lowest score for most reactive site in CYP 3A4 (#miss = 2). Likewise, the correlation coefficients for $E_{est}$, 0.99 and 0.96, are consistently higher than those for MetaSite (0.83 and 0.45) and GOLD (0.60 and 0.89). Thus, according to these criteria, our transition-state docking approach gives significantly better results than other available methods.

## Conclusions

In this paper, we have developed a method for transition-state docking of putative substrate molecules into the active site of an enzyme, and we have applied it for the prediction of metabolism pattern of two typical drugs (progesterone and flunitrazepam) in two human CYPs (3A4 and 2C9). Transition-state docking is a novel method to increase the discriminative power of a docking procedure by requiring that a productive transition state must form between the drug and the protein. It differs from previous approaches using constrained docking[34,41] or hypothetic "high-energy intermediates"[13,14] in that we employ a full and accurate force field, tailored for the transition state the hydroxylation reaction in the CYP, rather than a single distance or a modified substrate structure. Thereby, we directly estimate if a transition state actually can be obtained for a certain substrate.

To this end, an MM force field for the substrate is needed and this was developed for the hydroxylation of aliphatic carbon atoms by CYPs in a previous article.[12] This force field gives excellent docked structures of the two drugs in the two enzymes, when it is combined with a systematic search of the conformational space.

However, we also need an estimate of the activation energy of each of the docked structures. To estimate the intrinsic reactivity of each reactive site on the drug we have used DFT calculations with a methoxy radical model of the haem group. Such calculations can be performed within a few hours for most druglike molecules.[5] A qualitative model, [5] which predicts the activation energies from the chemical environment of the reacting hydrogen atom, gives only slightly worse results (the two estimates differ by 3–18 kJ/mol).

Second, we need to estimate how well the transition state fits into the protein. This means that we should estimate the binding affinity of the drugs to the proteins, with the modification that the transition states involve a partial bond between the protein and the drug (an O−H bond to the oxoferryl group). This is a serious complication, because it means that standard methods of binding affinity cannot be used[40] and that the MM energies of different reactive sites are not comparable.

Depending on the energy of interest, this problem can be solved in different ways. For reactions involving different hydrogen atoms bound to the same carbon atom (e.g., H16$\alpha$ and H16$\beta$ in progesterone), the force fields contain exactly the same terms and, therefore, the MM energies are directly comparable. Our results show that we can predict the reactivity of such sites successfully and the results can be directly interpreted in contributions from the various MM terms (bonded terms, van der Waals interactions, and electrostatics).

However, for more general applications, we have developed a method to compare activation barriers for different reactive sites in different proteins, combining both the intrinsic reactivity and steric effects and correcting for the extra bonded terms in the transition state by comparing them to the ideal terms in a small optimized model complex, eq 5. From the results in Tables 1 and 2, it can be seen that the method in all except one case gives qualitatively, but not always quantitatively, correct results (i.e., we can predict which sites are reactive, but we cannot rank them correctly).

Thus, the method works reasonably, but not perfectly. On the other hand, it is significantly better than alternative methods, viz. standard docking and scoring with GOLD[19] and metabolic predictions by MetaSite.[8] We have tried to improve the method by local optimization of the amino-acid side chains in the active site or by including full protein flexibility and improving the energy function with solvation, hydrophobic, and entropy effects, using the MM/PBSA approach.[18] Unfortunately, neither of these tests led to any significant improvement.

There are several possible ways to further improve the approach. First, we could try to use more conformations than the best one with local protein optimization or by MM/PBSA refinement. However, this would be much more expensive in terms of computer resources. Second, the Q2MM method for optimizing transition states essentially fixes the position of the transition state, allowing only for minimal variations in the geometry. This will exaggerate steric effects. Other methods, e.g. SEAM or the empirical valence bond method,[15] can be expected to model variations around the transition state more realistically, but they require special software. We currently work on various approaches to improve the method and we also try to develop similar methods to study other important CYP reactions with this approach.

**Acknowledgment.** This investigation has been supported by funding from the research school in pharmaceutical science (FLÄK), the Benzon foundation, the Carlsberg foundation, the Novo Nordisk Plus Prize, the Swedish Research Council, the Villum Kann Rasmussen Foundation, and by computer resources of LUNARC at Lund University.

**Supporting Information Available:** Hydrogen abstraction barriers for C1 and C3 in flunitrazepam calculated with B3LYP. Docking scores for all positions of progesterone. MetaSite scores for all positions in flunitrazepam and progesterone. This material is available free of charge via the Internet at http://pubs.acs.org.

## References


(1) Hodgson, J. ADMET - turning chemicals into drugs. *Nat. Biotechnol.* **2001**, *19*, 722–726.

(2) Harris, D.; Loew, G. Prediction of regiospecific hydroxylation of camphor analogs by cytochrome P450. *J. Am. Chem. Soc.* **1995**, *117*, 2738–2746.

(3) Rydberg, P. *Theoretical Studies of Cytochrome P450*; Lund University: Lund, 2007.

(4) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.* **2005**, *105*, 2279–2328.

(5) Olsen, L.; Rydberg, P.; Rod, T. H.; Ryde, U. Prediction of activation energies for hydrogen abstraction by cytochrome P450. *J. Med. Chem.* **2006**, *49*, 6489–6499.

(6) Jones, J. P.; Korzekwa, K. R. Predicting the rates and regioselectivity of reactions mediated by the P450 superfamily. *Cytochrome P450, B* **1996**, *272*, 326–335.

(7) de Graaf, C.; Vermeulen, N. P. E.; Feenstra, K. A. Cytochrome P450 in silico: An integrative modeling approach. *J. Med. Chem.* **2005**, *48*, 2725–2755.

(8) Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. MetaSite: Understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem.* **2005**, *48*, 6970–6979.

(9) Zamora, I.; Afzelius, L.; Cruciani, G. Predicting drug metabolism: A site of metabolism prediction tool applied to the cytochrome P4502C9. *J. Med. Chem.* **2003**, *46*, 2313–2324.

(10) Park, J. Y.; Harris, D. Construction and assessment of models of CYP2E1: Predictions of metabolism from docking, molecular dynamics, and density functional theoretical calculations. *J. Med. Chem.* **2003**, *46*, 1645–1660.

(11) Afzelius, L.; Arnby, C. H.; Broo, A.; Carlsson, L.; Isaksson, C.; Jurva, U.; Kjellander, B.; Kolmodin, K.; Nilsson, K.; Raubacher, F.; Weidolf, L. State-of-the-art tools for computational site of metabolism predictions: Comparative analysis, mechanistical insights, and future applications. *Drug Metab. Rev.* **2007**, *39*, 61–86.

(12) Rydberg, P.; Olsen, L.; Norrby, P. O.; Ryde, U. General Transition-State Force Field for Cytochrome P450 Hydroxylation. *J. Chem. Theory Comput.* **2007**, *3*, 1765–1773.

(13) Hermann, J. C.; Ghanem, E.; Li, Y.; Raushel, F. M.; Irwin, J. J.; Shoichet, B. K. Predicting substrates by docking high-energy intermediates to enzyme structures. *J. Am. Chem. Soc.* **2006**, *128*, 15882–15891.

(14) Hermann, J. C.; Marti-Arbona, R.; Dedorov, A. A.; Fedorov, E.; Almo, C. A.; Shoichet, B. K.; Raushel, F. M. Structure-based activity prediction for an enzyme of unknown function. *Nature* **2007**, *448*, 775–779.

(15) Jensen, F.; Norrby, P. O. Transition states from empirical force fields. *Theor. Chem. Acc.* **2003**, *109*, 1–7.

(16) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.

(17) Norrby, P. O. Molecular mechanics as a predictive tool in asymmetric catalysis. In *Transition State Modeling for Catalysis*; Truhlar, D. G., Morokuma, K., Eds.; American Chemical Society: Washington, DC., 1999; pp 163–172.

(18) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(19) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

(20) Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The structure of human cytochrome P4502C9 complexed with flurbiprofen at 2.0-angstrom resolution. *J. Biol. Chem.* **2004**, *279*, 35630–35637.

(21) Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-angstrom resolution. *J. Biol. Chem.* **2004**, *279*, 38091–38094.

(22) Rydberg, P.; Rod, T. H.; Olsen, L.; Ryde, U. Dynamics of water molecules in the active-site cavity of human cytochromes P450. *J. Phys. Chem. B* **2007**, *111*, 5445–5457.

(23) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general AMBER force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

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

(25) Berendsen, H. J. C.; Postma, J. P. M.; van gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(26) Bashford, D.; Case, D. A. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.

(27) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins* **2004**, *55*, 383–394.

(28) Hermann, R. B. Theory of hydrophobic bonding. 2. Correlation of hydrocarbon solubility in water with solvent cavity surface-area. *J. Phys. Chem.* **1972**, *76*, 2754–2759.

(29) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* **2004**, *86*, 67–74.

(30) Weis, A.; Katebzadeh, K.; Soderhjelm, P.; Nilsson, I.; Ryde, U. Ligand affinities predicted with the MM/PBSA method: Dependence on the simulation method and the force field. *J. Med. Chem.* **2006**, *49*, 6596–6606.

(31) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-integration of Cartesian equations of motion of a system with constraints - molecular-dynamics of N-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(32) Rocchia, W.; Alexov, E.; Honig, B. Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. *J. Phys. Chem. B* **2001**, *105*, 6507–6514.

(33) Kongsted, J.; Ryde, U. Improved method for entropy term in the MM/PBSA approach. *Chem. Phys. Lett.*, submitted.

(34) de Graaf, C.; Oostenbrink, C.; Keizers, P. H. J.; van der Wijst, T.; Jongejan, A.; Vemleulen, N. P. E. Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2006**, *49*, 2417–2430.

(35) Coller, J. K.; Somogyi, A. A.; Bochner, F. Quantification of flunitrazepam's oxidative metabolites, 3-hydroxyflunitrazepam and desmethylflunitrazepam, in hepatic microsomal incubations by high-performance liquid chromatography. *J. Chromatogr. B Biomed. Sci. Appl.* **1998**, *719*, 87–92.

(36) Hesse, L. M.; Venkatakrishnan, K.; von Moltke, L. L.; Shader, R. I.; Greenblatt, D. J. CYP3A4 is the major CYP isoform mediating the in vitro hydroxylation and demethylation of flunitrazepam. *Drug Metab. Dispos.* **2001**, *29*, 133–140.

(37) Waxman, D. J.; Lapenson, D. P.; Aoyama, T.; Gelboin, H. V.; Gonzalez, F. J.; Korzekwa, K. Steroid-hormone hydroxylase specificities of 11 cDNA-expressed human cytochrome-P450S. *Arch. Biochem. Biophys.* **1991**, *290*, 160–166.

(38) Richardson, T. H.; Jung, F.; Griffin, K. J.; Wester, M.; Raucy, J. L.; Kemper, B.; Bornheim, L. M.; Hassett, C.; Omiecinski, C. J.; Johnson, E. F. Universal approach to the expression of human and rabbit cytochrome P450S of the 2C subfamily in Escherichia-coli. *Arch. Biochem. Biophys.* **1995**, *323*, 87–96.

(39) Yamazaki, H.; Shimada, T. Progesterone and testosterone hydroxylation by cytochromes P450 2C19, 2C9, and 3A4 in human liver microsomes. *Arch. Biochem. Biophys.* **1997**, *346*, 161–169.

(40) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed. Engl.* **2002**, *41*, 2645–2676.

(41) Kirton, S. B.; Kemp, C. A.; Tomkinson, N. P.; St.-Gallay, S.; Sutcliffe, M. J. Impact of incorporating the 2C5 crystal structure into comparative models of cytochrome P450 2D6. *Proteins* **2002**, *49*, 216–231.

CT700313J