

Are Current Semiempirical Methods Better Than Force Fields? A Study from the Thermodynamics Perspective[†]

Gustavo de M. Seabra,[‡] Ross C. Walker,[§] and Adrian E. Roitberg^{*,‡}

Quantum Theory Project and Department of Chemistry, University of Florida, 2234 New Physics Building #92, P.O. Box 118435, Gainesville, Florida 32611-8435, and San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive #0505, La Jolla, California 92093-0505

Received: April 15, 2009; Revised Manuscript Received: June 16, 2009

The semiempirical Hamiltonians MNDO, AM1, PM3, RM1, PDDG/MNDO, PDDG/PM3, and SCC-DFTB, when used as part of a hybrid QM/MM scheme for the simulation of biological molecules, were compared on their abilities to reproduce experimental ensemble averages at or near room temperatures for the model system alanine dipeptide in water. Free energy surfaces in the (ϕ, ψ) dihedral angle space, $^3J(\text{H}_\text{N}, \text{H}_\alpha)$ NMR dipolar coupling constants, basin populations, and peptide–water radial distribution functions (RDF) were calculated from replica exchange simulations and compared to both experiment and fully classical force field calculations using the Amber ff99SB force field. In contrast with the computational chemist's intuitive idea that the more expensive a method the better its accuracy, the ff99SB force field results were more accurate than most of the semiempirical methods, with the exception of RM1. None of the methods, however, was able to accurately reproduce the experimental data. Analysis of the results indicate that the specific QM/MM interactions have little influence on the sampling of free energy surfaces, and the differences are well explained simply by the intrinsic properties of the various QM methods.

Introduction

Semiempirical (SE) methods to solve Schrödinger's equation have been extensively tested, compared, and adjusted for more than 20 years now.^{1–35} Those tests, however, generally focus on the method's ability to reproduce static data such as optimized geometries, heats of formation, reaction energies, and spectroscopic parameters, usually at zero temperature. Recent advances in computer processor technology, parallel programming, and availability of supercomputer clusters have allowed computational chemists to apply a broader range of methods to systems of ever increasing size and complexity, pushing the semiempirical methods beyond the limits for which they have been designed. The compromise between accuracy and speed provided by semiempirical (SE) methods now allows for significant sampling and treatment of much larger systems without complete forfeiture of quantum mechanical effects, opening the possibility of their application to computational studies of biological molecules in their native environment. For example, SE methods have already been applied in studies ranging from enzyme reactions^{36–41} to solution structures of peptides^{42–44} and even to structural studies of whole proteins.⁴⁵ Their native implementations in popular biomolecular simulation programs such as AMBER^{46–48} and CHARMM^{49,50} promise to make the use of hybrid quantum mechanics/molecular mechanics (QM/MM) methods even more widespread.

For conformational sampling, one can imagine a hierarchy of methods with different computational costs, which are generally believed to be directly proportional to the method's accuracy. On one end would be the faster force field methods, followed by the polarizable force fields, then semiempirical

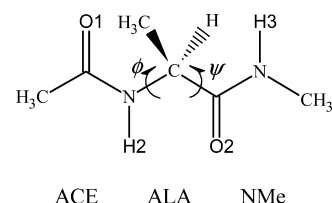


Figure 1. Scheme of the capped L-alanine dipeptide depicting the dihedral angles ϕ and ψ .

methods, and finally the much more costly ab initio and density functional methods. As computer capabilities increase and the use of semiempirical methods for larger systems becomes more accessible, it will be tempting to, at some point, just completely discard the use of empirical force fields. In such conditions, it is important to ask the question of whether those SE methods really are the most appropriate for the problems under consideration, including very large systems not included in their parametrization sets. There is no doubt QM methods are required for situations where intrinsically quantum processes such as bond breaking and forming, tunneling, or charge redistribution are important.^{51–56} However, the SE QM methods currently available have been parametrized against small molecules and reactions, usually to reproduce gas-phase data.^{2,20,27,28,32,34,57,58} The parameters thus obtained are not guaranteed to be fully transferable to biological molecules in their natural surroundings.

The present work compares the performance of a series of commonly used SE Hamiltonians when used as part of a hybrid QM/MM scheme for the simulation of biological molecules, from a thermodynamics point of view: we focus on their ability to reproduce ensemble properties at or near room temperatures, and at conditions that approach the real biological environment of such molecules. We present results for a model system composed of the alanine dipeptide (Ace-Ala-NMe, Figure 1), an alanine unit blocked by an acetyl group at the N-terminus

[†] Part of the "Walter Thiel Festschrift".

[‡] University of Florida.

[§] University of California.

* roitberg@ufl.edu.

(Ace) and an *N*-methylamide group (NMe) at the C-terminus, which is treated quantum mechanically by a semiempirical method, immersed in explicit water, treated classically. Such small peptides make very convenient models for the study of the conformational properties of biological molecules, providing important insight into the understanding of protein folding and dynamics.^{59,60} Small and flexible, those molecules are often better described as ensembles of fast interconverting conformers, providing an outstanding experimental and theoretical challenge.⁶¹ Furthermore, interactions with the environment are very important, and results from gas-phase experiments and calculations can seldom be extrapolated to the more biologically relevant condensed phase.

We compute free energy surfaces in the (ϕ, ψ) dihedral angle space, ${}^3J(\text{H}_\text{N}, \text{H}_\alpha)$ NMR dipolar coupling constants, basin populations, and peptide–water radial distribution functions (RDF), compare with available experimental data, and discuss the results on the basis of the differences between the methods. Our results show that, for the ensemble properties analyzed in this work, the latest generation force fields were able to offer results of higher accuracy than most of the QM methods examined. Furthermore, the specific effect of the QM/MM interactions is small, and the differences are well explained by intrinsic differences between the quantum mechanical methods.

Methods

Molecular Dynamics Simulations. The system composed of the alanine dipeptide in a box with 929 TIP3P water molecules was prepared with the LEaP program, distributed with the Amber package.⁴⁶ The peptide molecule was treated quantum mechanically by four methods based on the neglect of diatomic differential overlap (NDDO) approximation, namely MNDO,^{33,34} AM1,³² PM3,^{27,28} and RM1 (a reparameterization of the AM1 Hamiltonian);² by a more recent semiempirical method based on density functional theory, SCC-DFTB (second-order self-consistent-charge density functional tight binding),^{57,58} in its original formulation and with the inclusion of dispersion interactions;⁶² by two different corrections to the PM3 energy function: the peptide correction (PC),⁶³ which adds an empirical force field correction term to improve the description of planarity in peptide bonds, and the addition of the pairwise distance directed Gaussian (PDDG) functions,⁶ which differentiates between a wide range of functional groups by the addition of four new parameters per atom; and by a PDDG correction to MNDO, PDDG/MNDO. The water molecules were treated classically with the TIP3P water model,⁶⁴ and Replica Exchange Molecular Dynamics (REMD)⁶⁵ calculations were used to ensure adequate sampling of the conformational distribution of the peptides in explicit water. For comparison purposes, some calculations were also carried out with the peptide treated by two deprecated Amber force fields (ff94⁶⁶ and ff99⁶⁷), as well as the two new generation Amber force fields, ff99SB⁶⁸ and the ff03.⁶⁹

The systems were energy-minimized for 1000 steps to remove possible atomic contacts, heated to 300 K for 200 ps with Berendsen thermostat,⁷⁰ and then relaxed to 1 atm pressure for 800 ps with the Berendsen barostat,⁷⁰ followed by 1 ns equilibration at 300 K and 1 atm. The systems treated with a QM/MM approach were initially equilibrated using the same procedure and the ff99SB force field and then further relaxed at constant pressure and temperature for 500 ps (1 atm, 300 K) using the QM/MM approach under investigation. With the exception of the minimizations, all calculations used the SHAKE⁷¹ algorithm to constrain bond lengths involving hy-

drogens, a time step of 2 fs, and periodic boundary conditions with a 8.0 Å cutoff for nonbonded interactions. Long range electrostatics were treated with the particle mesh Ewald (PME) method, both by classical means and with a QM/MM implementation of PME.⁴⁷ After the initial equilibration, all calculations used the Langevin thermostat with a 2 ps⁻¹ collision frequency.⁷²

The replica exchange molecular dynamics (REMD) calculations used 32 replicas, at the following temperatures (in K) 292.9, 300.0, 307.3, 314.8, 322.4, 330.3, 338.3, 346.6, 355.0, 363.6, 372.5, 381.5, 390.8, 400.3, 410.1, 420.1, 430.3, 440.8, 451.5, 462.5, 473.7, 485.2, 497.1, 509.2, 521.5, 534.2, 547.2, 560.6, 574.2, 588.2, 602.4, and 616.8. A different random seed was used for each replica. The systems were kept at constant volume, and exchanges were attempted every 0.02 ps (10 MD steps),⁷³ for a total of 6 ns (per replica). Snapshots of the system were taken at 1000 steps (2 ps) intervals, for a total of 3000 snapshots per replica. All simulations were performed with a development version of the Amber package⁴⁶ containing the new reparameterization of the PDDG/PM3 Hamiltonian, here referred to as PDDG/PM3 (2008).⁷⁴ The analysis shown here was performed including only the snapshots corresponding to a 300.0 K temperature, ignoring the first 10% as a relaxation period.

Dipolar Coupling Constants. Dipolar coupling constants were obtained using the Karplus relation,⁷⁵

$${}^3J(\text{H}_\text{N}, \text{H}_\alpha) = a \cos^2(\phi - 60^\circ) + b \cos(\phi - 60^\circ) + c \quad (1)$$

for the frames at 300 K, and using the parameters $a = 7.09$, $b = -1.42$, and $c = 1.55$ derived by Hu and Bax.⁷⁶ The final result was obtained as an average of all measures. The ϕ dihedral angle is shown in Figure 1, and a plot of the ${}^3J(\text{H}_\text{N}, \text{H}_\alpha)$ coupling constant as a function of the ϕ -angle is shown in the top-left plot in Figure 2.

Conformational Distributions. The conformational distributions of the peptides at 300 K were calculated according to the populations of specific regions in the Ramachandran plot. For the purposes of this study, the Ramachandran plot was divided into four regions: α , β , PP_{II}, and “other”. The boundaries between basins are depicted in the upper left image in Figure 2 and detailed in Table 1.

Free Energy Profiles. The free energies were obtained by calculating the (normalized) probability P of finding the alanine dipeptide in a conformation at a particular region in (ϕ, ψ) -space from the 300 K trajectories using 2° bins and then converting this number to free energies by $\Delta G = -RT \ln(P)Z$, where ΔG is the Gibbs free energy measured as relative to the highest probability region, R is the universal gas constant, and T is the temperature.

Results and Discussion

A number of experimental^{59,61,77–80} and theoretical,^{3,81–97} studies of alanine dipeptide indicate that the potential energy surface for alanine dipeptide in vacuum and in solution are considerably different: while in the gas phase the global minimum is believed to be a C7eq structure ($\phi \sim -83^\circ$, $\psi \sim 73^\circ$),⁸⁹ it has been suggested that interaction with water favors the polyproline-II (PP_{II}, $\phi \sim -75^\circ$, $\psi \sim 150^\circ$) conformation.^{88,98} Avbelj et al. measured the ${}^3J(\text{H}_\text{N}, \text{H}_\alpha)$ dipolar coupling constant for alanine dipeptide (along with other dipeptides) as 6.06 Hz in pH 4.9 and, by comparison with the coupling constants obtained from experiments with longer peptides and a scan of

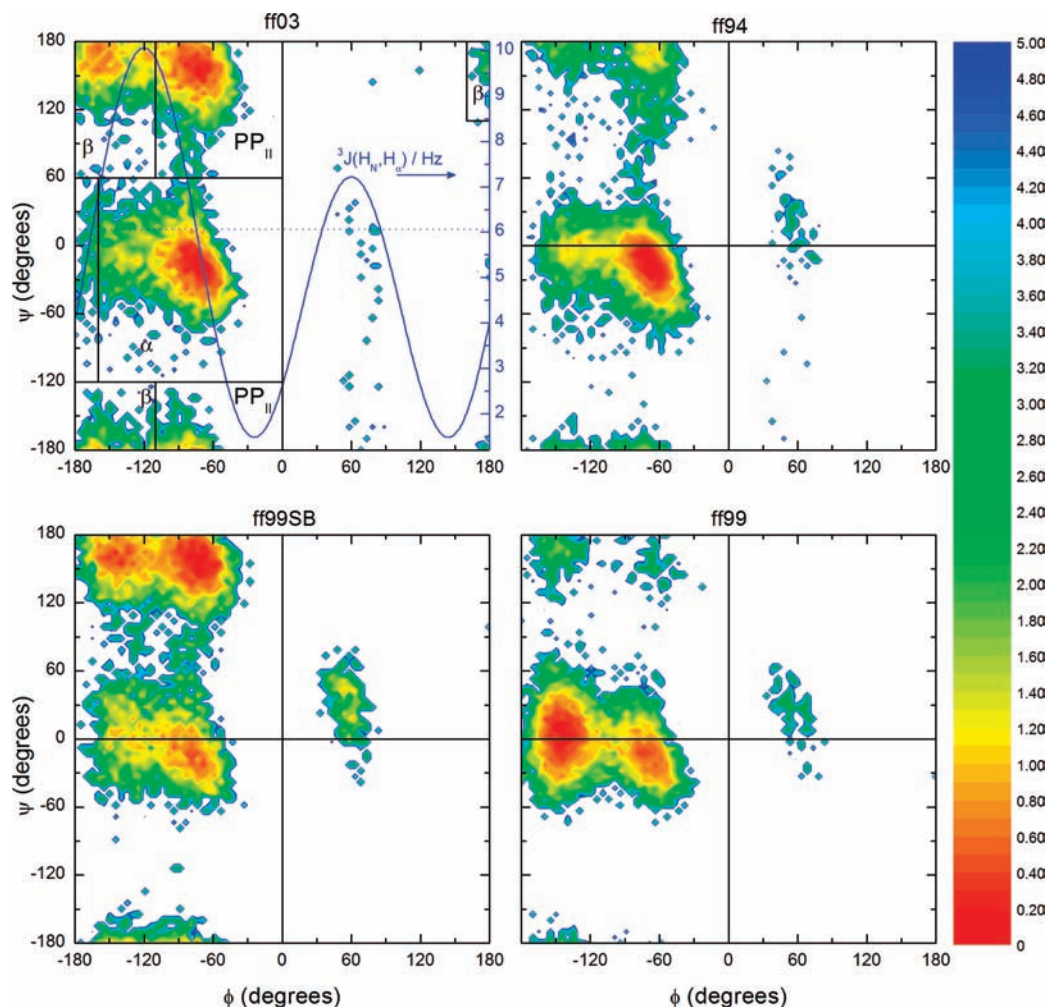


Figure 2. Free energy surfaces obtained from the 300 K replicas of the REMD simulations, for the different force field methods. The graph at the upper left also depicts the basin divisions used in this work for population analysis, and a plot of eq 1 in blue, linked to the right y-axis. The dashed red line indicates the experimental dipolar coupling constant.⁵⁹

TABLE 1: Basin Divisions Used for Calculations of the Population Distribution

basin	ϕ (deg)	ψ (deg)
α	$-160 < \phi < 0$	$-120 < \psi < 60$
β	$-180 < \phi < -160$ $160 < \phi < 180$	$-180 < \psi < -120$ or $60 < \psi < 180$ $110 < \psi < 180$
PP _{II}	$-110 < \phi < 0$	$-180 < \psi < -120$ or $60 < \psi < 180$

the “coil library”,^{99,100} concluded that the intrinsic backbone preferences determining protein folding preferences are already fully present in the dipeptide.⁵⁹ Graf et al. used a combination of molecular dynamics calculations and NMR experiments to obtain backbone (ϕ , ψ) distributions for a series of alanine polypeptides (Ala_n, $n = 3-7$), concluding the PP_{II} basin to be the most populated one in all cases, and sampling of the α -region becomes significant only for larger peptides. The preference of the alanine residue for the PP_{II} conformation is also supported by Schweitzer-Stenner et al., who estimated a PP_{II} population for alanine at least 60% in H-(AAKA)-OH.^{101,102} Examining alanine dipeptide by vibrational spectroscopy, Grdadolnik et al. used a decomposition of the amide-III region and of Raman skeletal vibrations to infer the population of the PP_{II} basin as between 60–76%, depending on the method of determination.¹⁰³

The calculated $^3J(\text{H}_N, \text{H}_\alpha)$ dipolar couplings are listed in Table 2. For comparison, the experimental results of Avbelj et al.⁵⁹ are also included. The basin populations obtained for the

TABLE 2: $^3J(\text{H}_N, \text{H}_\alpha)$ NMR Dipolar Couplings for Alanine Dipeptide, in Hz, Calculated as an Average from the 6 ns of MD Simulation^a

method	$^3J(\text{H}_N, \text{H}_\alpha)$
ff94	6.20 ± 0.08
ff99	7.80 ± 0.07
ff03	6.69 ± 0.08
ff99sb	7.35 ± 0.08
MNDO	7.67 ± 0.07
AM1	8.25 ± 0.07
PM3	8.14 ± 0.07
RM1	6.77 ± 0.09
PDDG/MNDO	7.76 ± 0.07
PDDG/PM3 (2002)	7.85 ± 0.07
PDDG/PM3 (2008)	8.06 ± 0.07
PM3 + MM correction	8.24 ± 0.07
SCC-DFTB	8.07 ± 0.08
SCC-DFTB + dispersion	8.16 ± 0.08
exp ⁵⁹	6.06 ± 0.05

^a The error margin is shown as the 95% confidence intervals calculated using the Student’s *t*-value for an infinite number of measurements. The experimental error margin is an estimate based on different parameterizations of the Karplus equation.⁵⁹

different basins from the 300 K trajectory, as well as the experimental results of Grdadolnik et al.¹⁰³ are shown in Table 3. Figure 2 shows the free energy profiles at 300 K obtained for the classical mechanical methods. The free energy surfaces

TABLE 3: Conformational Distribution of Alanine Dipeptide, Shown as Fractional Populations of the Different Conformational Basins

method	α	β	PP _{II}	other
ff94	0.84	0.04	0.11	0.01
ff99	0.91	0.03	0.01	0.05
ff03	0.45	0.19	0.35	0.01
ff99SB	0.32	0.24	0.40	0.04
MNDO	0.07	0.27	0.63	0.03
AM1	0.57	0.19	0.21	0.04
PM3	0.14	0.51	0.33	0.02
RM1	0.26	0.17	0.53	0.04
PDDG/MNDO	0.33	0.15	0.52	0.01
PDDG/PM3 (2002)	0.16	0.40	0.41	0.02
PDDG/PM3 (2008)	0.08	0.43	0.48	0.01
PM3 + MM correction	0.19	0.38	0.42	0.02
SCC-DFTB	0.40	0.34	0.19	0.06
SCC-DFTB/dispersion	0.48	0.30	0.18	0.05
IR ¹⁰³	0.11	0.29	0.60	0.00
Raman ¹⁰³	0.18	0.06	0.76	0.00

obtained with the 4 NDDO parametrizations (MNDO, AM1, PM3, and RM1) are shown in Figure 3. Figure 4 compares the free-energy surfaces obtained with the three different modifications of the PM3 method, PM3-MM, PDDG/PM3 (2002), and PDDG/PM3 (2008), and Figure 5 compares the results obtained by SCC-DFTB, standard and with the addition of dispersion interactions.

It is interesting to note that the only simulation capable of reproducing the experimental dipolar couplings with an error

of less than 5% used the oldest parameter set considered here, ff94. This is an artificial effect due to the well-known oversampling on the α -region by this force field,^{104,105} which can be clearly seen in Table 3 and Figure 2. The coupling constant as calculated from eq 1 assigns values around 6.0 Hz for a ϕ -angle of $\sim 75^\circ$, as can be seen in Figure 2. This result also helps emphasize that the $^3J(H_N, H_\alpha)$ coupling constant alone is incapable of fully distinguishing between basins. The ff99 force field also overpopulates the α -basin with an α -population of about 91%, but the sampling is concentrated at lower values of the ϕ -angle, as shown in Figure 2, leading to a larger value for the coupling constant. A similar effect can be seen in the results from the more recent ff03 force field: although it is capable of reproducing the experimental coupling constant within 7%, it still shows the largest population in the α -basin ($\sim 45\%$). This α -overpopulation issue has been partially resolved on the latest force field, ff99SB, which is the only force field to show the PP_{II} basin as the most populated (Table 3). However, the population of the α -basin is still too high in comparison to any of the vibrational spectroscopy results. The ff99SB force field also shows a relatively large population of the β -basin, implying a slight bias of the force field toward a more extended structure, which explains the higher calculated dipolar coupling.

The results from the QM methods vary just as much as for the different MM force fields. As shown in Table 2, with the exception of RM1, most QM methods lead to grossly overestimated dipolar coupling constants. With respect to the populations (Table 3), the only QM method to find a PP_{II} population

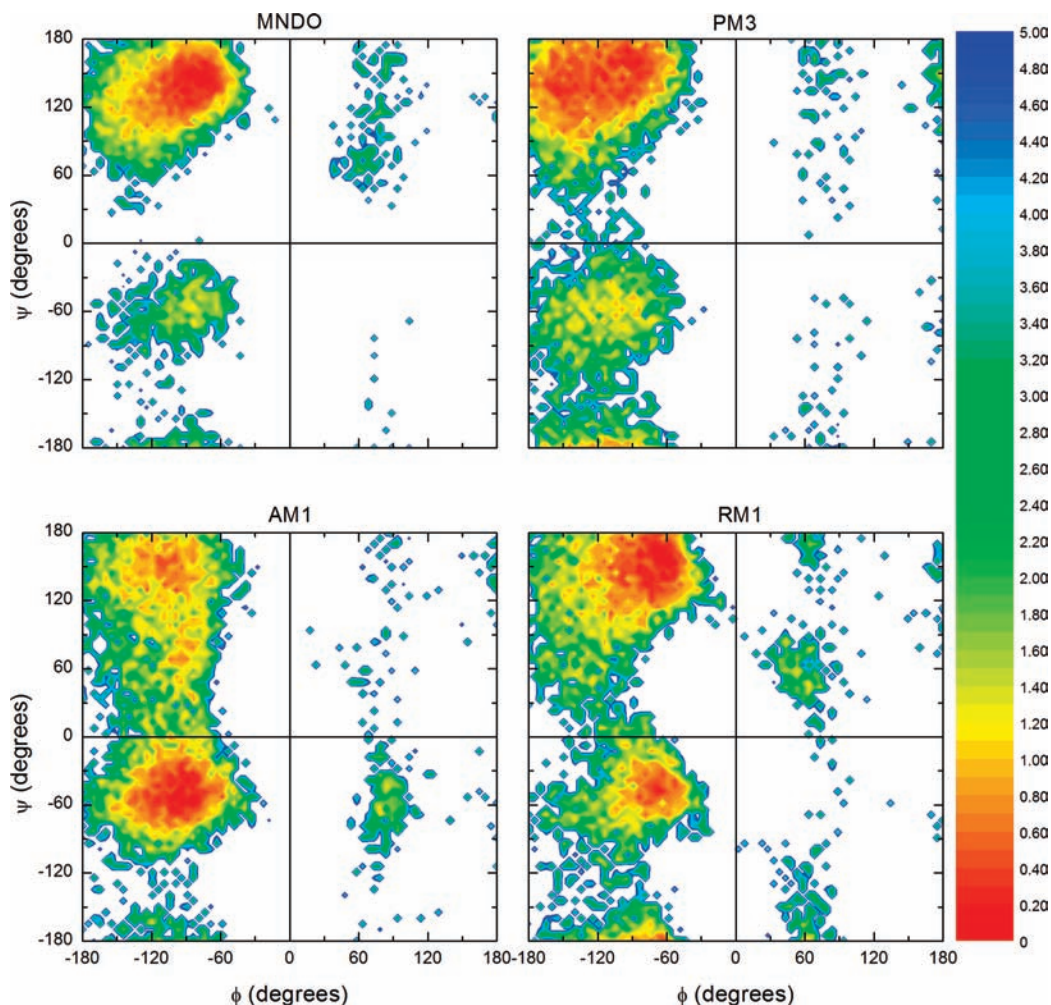


Figure 3. Free energy surfaces obtained from the 300 K replicas of the REMD simulations, for the different MNDO parametrizations.

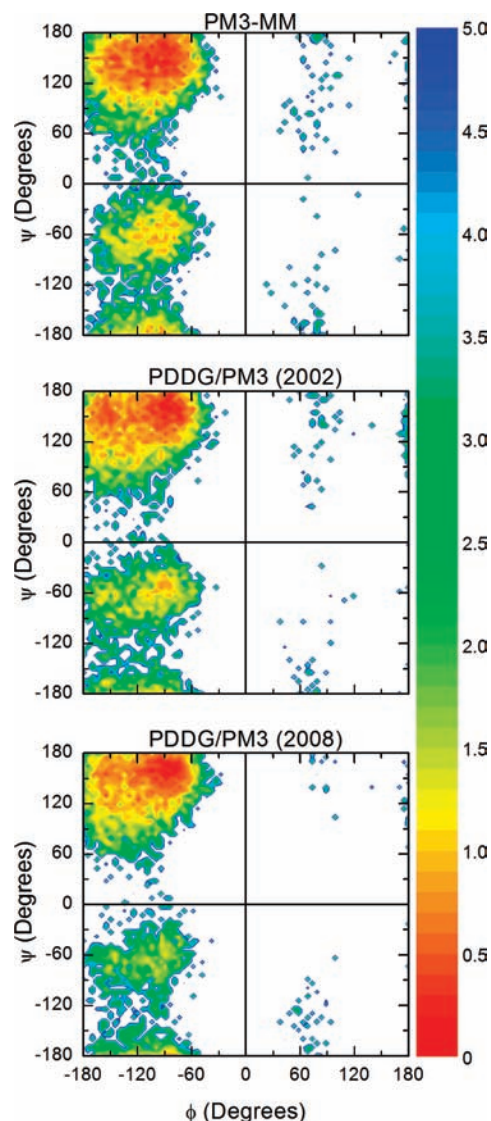


Figure 4. Free energy surfaces obtained from the 300 K replicas of the REMD simulations, for the different variations of the PM3 method.

within the experimental range is MNDO (Figure 3), the oldest and most primitive of all the semiempirical methods used. This population, however, is concentrated at a ϕ -range too low: the average ϕ -angle is around -83.0° (Table 4), yielding a dipolar coupling constant that is too high (7.67 Hz, Table 2). Although supposed to yield better results, the second parametrization of the MNDO Hamiltonian, AM1, completely loses the agreement with experiment, and about 59% of the structures sampled are in the α -region. The sampling of the β / PP_{II} -region is again improved in the next parametrization (PM3), with about 48% population in the β -basin, and 34% on the PP_{II} -basin. The addition of the peptide correction (PM3/MM) inverts those populations, and PP_{II} becomes the most populated basin with about 43% of the structures in this region, and only 36% in the β -region. However, this correction also increases sampling in the α -region, which goes from 14 to 19%. The addition of the PDDG functions, in general, also improves the populations of the PP_{II} basin for PM3 but reduces those populations for the MNDO case. The latest parametrization of the MNDO Hamiltonian, RM1, yields the lowest dipolar coupling constant among the QM methods (6.77 Hz) and also samples the PP_{II} region for most of the structures (51%) but still shows a high population in the α -basin (28%).

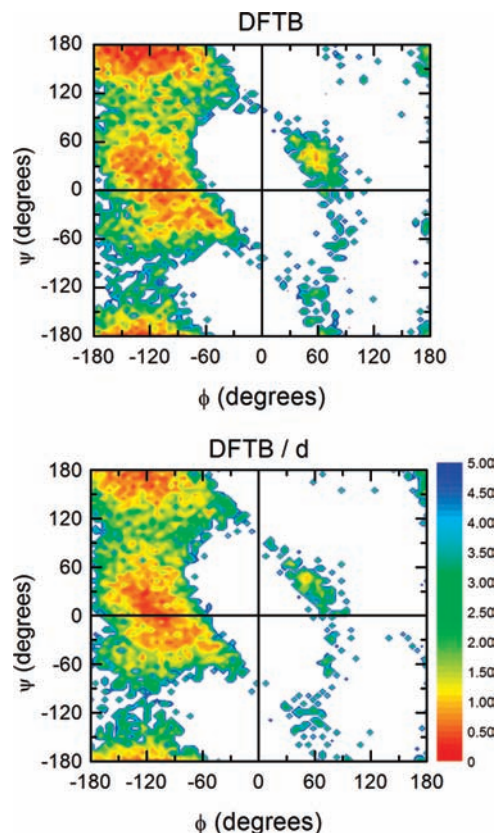


Figure 5. Free energy surfaces obtained from the 300 K replicas of the REMD simulations, for the SCC-DFTB and SCC-DFTB/d methods.

In principle, the differences between the calculated values could arise from at least two different sources: the treatment of the QM/MM interactions and intrinsic differences among the QM methods. Analysis of the results indicates the latter to be the case (the former is discussed below). As shown above, MNDO was the only QM method to (fortuitously) find a PP_{II} population within the same region as the experiment. The main issue with MNDO is the underestimation of weak interactions such as H-bond and electrostatic interactions, caused by a spurious repulsion around the van der Waals distance.²⁴ This in turn leads to the overestimation of steric repulsions and repulsions between nonbonded atoms when their distance is around the van der Waals distance,²⁰ which would also translate to unusually high rotational barriers. A problem for conventional calculations, this effect here seems to work to increase the population of the β / PP_{II} area since, for a fixed ϕ -angle, a larger ψ -angle leads to less steric interactions.

Evidence in favor of the above explanation is offered by the AM1 results. The spurious repulsions shown by MNDO were remediated during the generation of the AM1 parameters, by the introduction of spherical harmonic Gaussians in the core repulsion function (CRF). This change greatly improved the general results, and AM1 proved a much superior method in relation to MNDO. However, lacking the steric repulsions, AM1 introduces a large bias toward the α -basin (or rather removes an artificial bias toward the PP_{II} basin). This effect can be seen clearly in Figure 3 and Table 3. As shown in Table 4, this also leads to a reduction in the average ψ -angle of about 10° for both α and PP_{II} basins.

The bias toward the α -basin was partially corrected during the PM3 reparameterization, as can be seen by the population shift to the upper region of the Ramachandran plot (Figure 3).

TABLE 4: Average Angles Sampled by the Different Methods in Each Basin^a

method	ϕ (deg)	ψ (deg)	method	ϕ (deg)	ψ (deg)
α			β		
ff99SB	-97.6 ± 1.8	-6.0 ± 1.5	ff99SB	-137.8 ± 1.1	136.9 ± 5.1
MNDO	-93.5 ± 3.1	-50.0 ± 3.5	MNDO	-131.3 ± 1.5	115.1 ± 3.3
AM1	-101.3 ± 1.1	-40.4 ± 1.6	AM1	-129.3 ± 3.3	114.1 ± 5.9
PM3	-112.2 ± 2.5	-53.3 ± 3.4	PM3	-133.2 ± 2.4	112.5 ± 3.9
RM1	-84.9 ± 2.0	-38.7 ± 2.5	RM1	-128.6 ± 3.9	106.2 ± 6.9
PDDG/MNDO	-89.4 ± 1.3	-47.7 ± 1.4	PDDG/MNDO	-125.8 ± 1.2	106.0 ± 9.9
PDDG/PM3 (2002)	-107.4 ± 2.4	-56.1 ± 2.7	PDDG/PM3 (2002)	-136.1 ± 2.6	111.5 ± 4.4
PDDG/PM3 (2008)	-105.0 ± 2.8	-61.1 ± 3.5	PDDG/PM3 (2008)	-134.7 ± 2.0	118.5 ± 3.9
PM3/MM	-105.9 ± 2.0	-52.4 ± 2.7	PM3/MM	-133.6 ± 1.6	102.9 ± 5.4
DFTB	-106.2 ± 1.6	-4.8 ± 2.1	DFTB	-131.0 ± 3.3	78.1 ± 7.9
PP_{II}			Other		
ff99SB	-76.5 ± 0.9	142.7 ± 2.7	ff99SB	35.2 ± 12.4	25.0 ± 4.5
MNDO	-83.0 ± 0.7	132.5 ± 1.8	MNDO	64.4 ± 11.3	79.2 ± 18.3
AM1	-88.5 ± 1.1	115.3 ± 4.7	AM1	39.6 ± 18.3	-37.8 ± 12.0
PM3	-87.1 ± 0.9	115.4 ± 5.6	PM3	17.0 ± 28.8	14.5 ± 23.8
RM1	-75.7 ± 0.9	116.1 ± 4.2	RM1	41.4 ± 13.2	3.0 ± 19.1
PDDG/MNDO	-85.9 ± 0.7	130.0 ± 3.9	PDDG/MNDO	-5.7 ± 57.9	54.0 ± 41.2
PDDG/PM3 (2002)	-85.0 ± 0.8	125.0 ± 4.2	PDDG/PM3 (2002)	-57.6 ± 32.8	-9.2 ± 23.8
PDDG/PM3 (2008)	-86.1 ± 0.8	125.5 ± 4.1	PDDG/PM3 (2008)	-28.9 ± 42.3	-70.1 ± 35.8
PM3/MM	-87.3 ± 0.8	114.8 ± 4.8	PM3/MM	-4.2 ± 34.0	-8.5 ± 31.8
DFTB	-86.1 ± 1.5	100.9 ± 8.5	DFTB	11.8 ± 14.5	-0.5 ± 9.9

^a The error margin is shown as the 95% confidence intervals calculated using the Student's t -value for an infinite number of measurements.

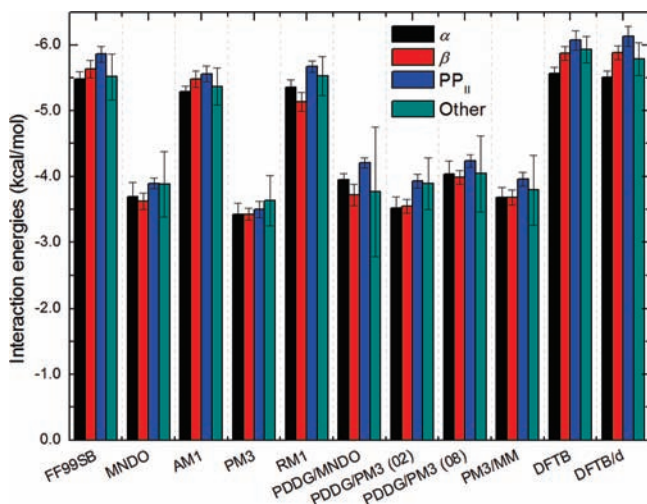


Figure 6. Average QM/MM interaction energies calculated with each method for the different basins. The (fully classical) ff99SB results are included for comparison. The 95% confidence intervals, indicated by the error bars above the columns, were calculated as $t(\sigma)/(\sqrt{N})^{1/2}$, where σ is the interaction energy's standard deviation, N is the number of structures inside that basin, and t is 1.960, the Student's t -value for an infinite number of measurements.

However, PM3 still overpopulates the ϕ -region around -110° , and it incorrectly predicts the β -basin to be the most populated. Another problem that affects mainly the PM3-based models is a tendency for too low rotational barriers around bonds with partial double bond character, such as the peptide bond. As a consequence, PM3 predicts about 45% of the structures to have at least one of the peptide bonds in the cis configuration. An ad-hoc correction, referred to here as the MM correction in the PM3/MM method, was introduced by Stewart in 1990 and included in the MOPAC program^{25,63} but is rarely used in QM/MM studies of biological molecules. Addition of this correction prevents the cis–trans isomerization around the peptide bond, introduces a very small but noticeable barrier between the β and PP_{II} basins, and slightly shifts the β/PP_{II} equilibrium back toward the latter. However, the population maximum is still at

TABLE 5: Minimized Distances and Interaction Energies for the QM/MM Water Dimer, Calculated with Different QM Methods^a

method	r_{OO} (Å)	IE (kcal/mol)
TIP3P	2.73	6.75
MNDO	3.08	3.12
AM1	2.98	3.48
PM3	2.94	3.62
PDDG/PM3 (02)	2.93	3.73
PDDG/PM3 (08)	2.92	3.79
PM3/MM	2.94	3.62
RM1	2.95	3.72
DFTB	2.77	4.97
DFTB/d	2.77	4.72
Exp	2.97–2.98	5.4 \pm 0.7

^a The QM water has the oxygen atom as the H-acceptor in the hydrogen bond. Fully classical (TIP3P) and experimental results are included for comparison.

too low values of ϕ ($\sim 90^\circ$), yielding a high dipolar coupling constant (Figure 4).

Figure 4 compares the effect of two slightly different versions of the PDDG corrections with the original PM3 formulation. The PDDG (pairwise distance directed Gaussian) correction, originally published in 2002,⁶ involves the addition of an extra function to the core repulsion terms of PM3 and MNDO, followed by a reoptimization of the parameter set. This function is composed of three or four weighted pairwise Gaussians (for the case of homodimer or heterodimer atom pairs, respectively), effectively adding four adjustable parameters per element, increasing the flexibility of the parameter set and allowing the function to differentiate between various functional groups. After the original publication in 2002, the PDDG parameters were further reoptimized, resulting in the parametrization here denoted by PDDG/PM3(2008).⁷⁴ The PDDG/PM3(2002) set is available in the Amber 10 distribution,⁴⁶ and the PDDG/PM3(2008) will be included in the next version of the Amber package. It is clear that the addition of the PDDG functions shifts the populations slightly to the PP_{II} region, and the latest set has an even stronger effect, which is corroborated by the increase in PP_{II} population shown in Table 3. As shown in Table 4, this

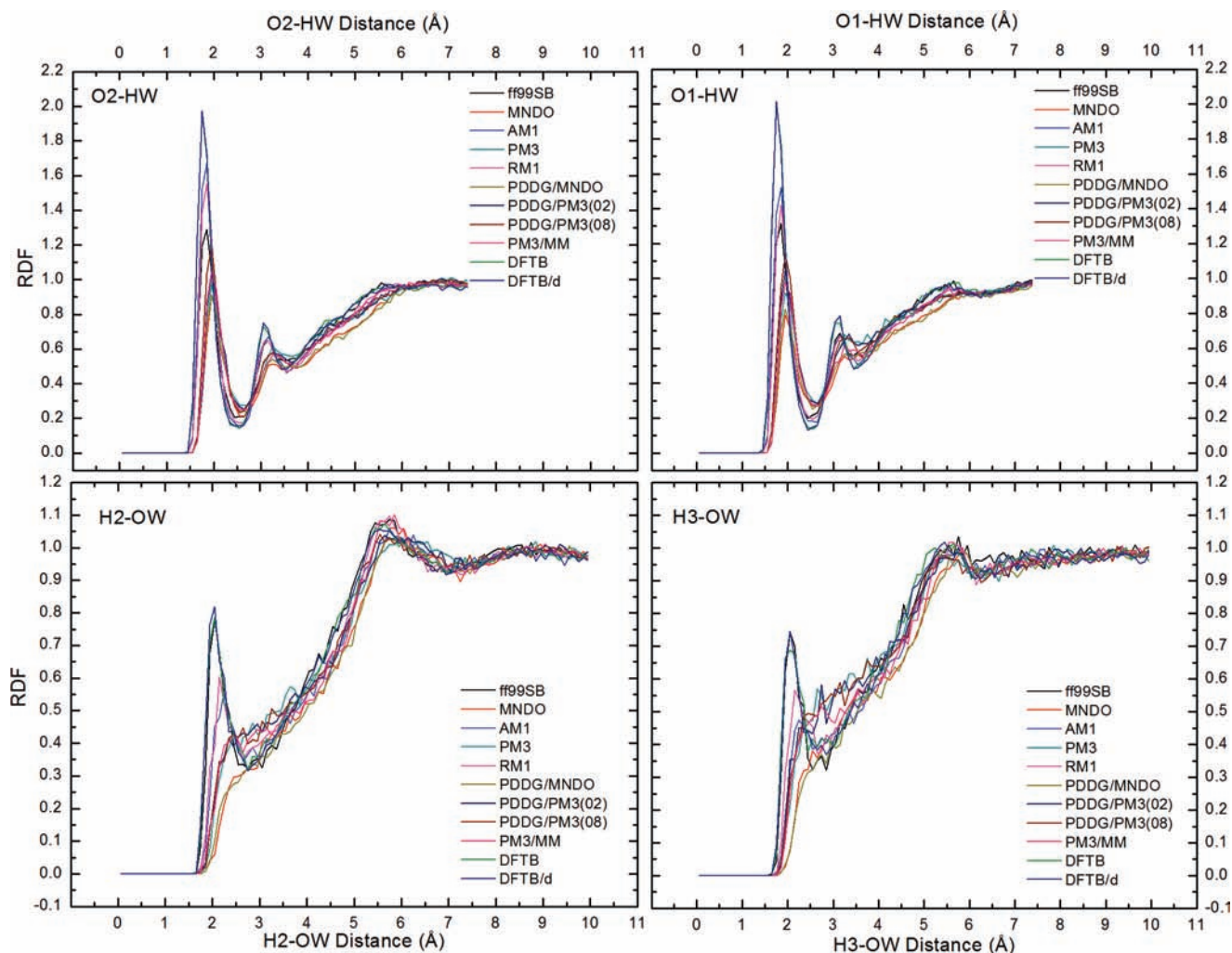


Figure 7. Radial distribution functions for the distribution of water molecules around the central alanine dipeptide, for the different methods studied here. For the atom numbering, refer to Figure 1.

TABLE 6: Position of the Maxima from the First Peak in the Total Radial Distribution Functions^a

method	O2-HW	H2-OW	O1-HW	H3-OW
ff94	1.85	2.05	1.85	2.05
ff99	1.85	2.05	1.85	2.15
ff99SB	1.85	2.05	1.85	2.05
ff03	1.85	2.05	1.85	2.05
MNDO	1.95	2.95	1.95	2.95
AM1	1.85	2.25	1.85	2.25
PM3	1.95	2.85	1.95	2.65
RM1	1.85	2.15	1.85	2.15
PDDG/MNDO	1.95	2.85	1.95	2.95
PDDG/PM3 (02)	1.95	2.95	1.95	2.75
PDDG/PM3 (08)	1.95	2.65	1.95	2.95
PM3/MM	1.95	2.95	1.95	2.75
DFTB	1.75	2.05	1.75	2.05
DFTB/d	1.75	2.05	1.75	2.05

^a For the atom numbers, refer to Figure 1.

effect is mostly due to changes in the average ψ -angle obtained. On the other hand, the average ϕ -angles sampled change only slightly between PM3 and the two PDDG corrections, to higher values in the α and PP_{II} basins and only slightly lower in the β -basin, and this change is reflected in the calculated $^3J(\text{H}_{\text{N}}, \text{H}_{\alpha})$ coupling constants.

It is also worthwhile to notice that, with the exception of the RM1, the calculated $^3J(\text{H}_{\text{N}}, \text{H}_{\alpha})$ coupling constants from all

MNDO parametrizations fall within a narrow interval from each other (0.58 Hz). The similarity of the $^3J(\text{H}_{\text{N}}, \text{H}_{\alpha})$ coupling constants calculated by the above methods indicate that those methods are sampling similar regions of the ϕ -space, which also becomes evident from Table 4. This similarity is especially noticeable in the PM3-based methods, and seems to point to a specific weakness of the parametrizations which apparently is partly corrected in the RM1 reparameterization. Since RM1 does not introduce any extra term in the Hamiltonian, this weakness is most likely to come from the training set used to parametrize the methods which, in the RM1 case, included a variety of biologically important molecules such as neutral and protonated amino acids, some dipeptides with α -helix or β -sheet conformations, the DNA nitrogen bases including the pairs and phosphates, and some saccharides and disaccharides.²

The SCC-DFTB method has been shown to yield calculated heats of formation mean averaged errors (MAE) somewhat between AM1 and PM3, and much larger than PDDG/PM3, due to an over stabilization of molecules containing S-O bonds.¹ On the other hand, small peptides' relative energies and structures calculated with the SCC-DFTB method are in good agreement with B3LYP/6-31G* density functional theory¹⁰⁶ and ab initio MP2/6-31G*¹⁰ calculations, at a much lower computational cost. In a study of H-bonded systems,¹⁰⁷ it was shown that SCC-DFTB underestimates the H-bond distances in comparison to B3LYP/6-31G*, but those energies are somewhat

TABLE 7: Position of Maxima of the First Peak in the Radial Distribution Functions for Each Conformational Basin, in Å^a

method	basin	O2-HW	H2-OW	O1-HW	H3-OW	method	basin	O2-HW	H2-OW	O1-HW	H3-OW
ff94	α	1.85	1.95	1.85	2.05	ff99	α	1.85	2.05	1.85	2.15
	β	1.85	2.15	1.85	2.15		β	1.75	2.15	1.85	2.05
	PP _{II}	1.85	2.15	1.85	2.15		PP _{II}	1.85	2.05	2.05	2.05
	other	1.75	2.15	1.75	2.15		other	1.75	2.05	1.85	2.05
ff03	α	1.85	2.05	1.85	2.15	ff99sb	α	1.85	1.95	1.85	2.05
	β	1.85	2.05	1.95	2.05		β	1.85	2.05	1.85	2.05
	PP _{II}	1.85	2.05	1.85	1.95		PP _{II}	1.85	2.05	1.85	2.05
	other	1.95	1.95	1.75	2.05		other	1.95	1.95	1.85	2.05
MNDO	α	1.95	2.45	2.15	2.65	AM1	α	1.85	2.25	1.85	2.25
	β	1.95	2.75	1.95	2.95		β	1.85	2.95	1.85	2.85
	PP _{II}	1.95	2.75	1.95	2.95		PP _{II}	1.85	2.15	1.75	2.75
	other	1.95	2.55	2.15	2.75		other	1.85	2.35	1.85	2.05
PM3	α	1.95	2.25	1.95	2.65	RM1	α	1.85	2.15	1.85	2.25
	β	1.95	2.95	1.95	2.65		β	1.85	2.15	1.75	2.25
	PP _{II}	1.95	2.85	1.95	2.95		PP _{II}	1.85	2.15	1.85	2.15
	other	1.95	2.75	2.05	2.15		other	1.85	2.35	1.75	2.35
PM3/ MM	α	1.95	2.85	1.95	2.55	PDDG/ MNDO	α	1.95	2.85	1.95	2.95
	β	1.95	2.95	1.95	2.65		β	1.95	2.95	1.95	2.95
PDDG/ PM3 (02)	PP _{II}	1.95	2.95	1.95	2.75	PDDG/ PM3 (08)	PP _{II}	1.95	2.65	1.95	2.75
	other	1.95	2.55	1.95	2.45		other	1.95	2.25	1.75	2.35
DFTB	α	1.75	2.05	1.75	2.15	DFTB/d	α	1.75	2.05	1.75	2.05
	β	1.75	2.05	1.75	1.95		β	1.75	2.05	1.75	2.05
	PP _{II}	1.75	1.95	1.75	2.05		PP _{II}	1.75	2.05	1.75	1.95
	other	1.75	2.05	1.75	2.05		other	1.75	2.05	1.85	2.05

^a For the atom names, refer to Figure 1.

improved by considering the waters as classical (TIP3P) in a QM/MM scheme. SCC-DFTB has also been shown to reproduce the B3LYP/6-31G* and basis set superposition error corrected MP4/cc-pVTZ(-f)/MP2/6-31G* geometries and energetics of various small peptide models in the gas phase, including alanine dipeptide, in cases where AM1 and PM3 present difficulties, but still underestimates some steric interactions and dipole moments on the C^α conformer of the alanine dipeptide.⁸⁷ In a comparison with different classical force fields for calculations of alanine and glycine dipeptides in explicit water, an SCC-DFTB/MM scheme using SPC or TIP3P waters yielded a better agreement with a distribution of angles extracted from high-resolution PDB structures than any of the classical force fields employed.³ An in-house test with our own set of 54 different alanine tetrapeptide conformers (to be published) has shown that the SCC-DFTB relative energies agree with MP2/6-311G* by as much as other higher level methods with the same basis set, while AM1 and PM3 completely fail to reproduce the energy ordering obtained with higher levels of theory. Those results motivated us to implement the SCC-DFTB method into the Amber package allowing its use in a more consistent QM/MM scheme including long-range electrostatics.⁴⁸

Our results show that SCC-DFTB overpopulates the α and β -basins (Table 3). As shown by Figure 5, there is no minimum in the PP_{II}-basin, and the population in this area seems to be mostly due to very low rotational barriers around the (ϕ , ψ)-dihedral angles. Although α is the most populated basin, the population is centered at a ϕ -angle around 106° (Table 4) which, together with the high β -population in relation to PP_{II}, leads to prediction of a high dipolar coupling constant (Table 2). The addition of dispersion interactions only worsens this scenario, increasing the population of the α -basin while decreasing the β and PP_{II} ones. However, it must be noted that the dispersion parameters used were originally derived for nucleic acid base

pairs,⁶² although they have been used before for the simulation of peptides.^{108,109}

The QM/MM Interaction

Differences in the interaction energies with the classical water molecules may also contribute to the differences seen among the different semiempirical methods. In consideration of this interaction, we calculated the interaction energy at each saved snapshot as

$$IE = E_{\text{complex}} - (E_{\text{water}} + E_{\text{peptide}}) \quad (2)$$

where E_{complex} is the energy of the complex formed by the peptide and only the closest water molecule, E_{peptide} is the energy of the isolated peptide at that same configuration, and E_{water} is the energy for this one water molecule. This analysis was performed independently for each peptide conformational basin, and the final results, calculated as an average of the instantaneous IEs, are shown in Figure 6. All the methods predict higher average interaction energy with the β -basin, although the difference as compared to other basins is small and, in many cases, the energies overlap when the error bars are taken into consideration. With the exception of AM1 and RM1, all MNDO-based methods predict very low interaction energies as compared to the full classical calculation. Interestingly, AM1 and RM1 predict interaction energies that are very close to the classical ones, and the strongest interaction energies are predicted by the DFTB methods. However, there seems to be no correlation between those results and the basin populations and dipolar coupling constant calculated, an indication that the differences in the results provided by the different methods more likely come from the intrinsic differences among the methods, and not from the way each one interacts with the classical waters.

The interaction energy between two TIP3P water molecules has been calculated before as 6.83 kcal/mol, overestimating the experimental value of 5.4 ± 0.7 .¹⁰⁷ We scanned the interaction energies between one TIP3P and one QM water, where the QM water has the oxygen involved in the H-bond, and the minima obtained are shown in Table 5. In our calculations, the interaction energy between two TIP3P water molecules was slightly lower, but it is still higher than any of the peptide–water interaction energies shown in Figure 6.

Although the QM/MM interaction energies' effect on the basin distribution is small, they do influence the distribution of water molecules around the dipeptide, as can be seen in the radial distribution functions plotted in Figure 7. As could be predicted from the interaction energies, the stronger the IE, the closer to the peptide the waters get, as can be seen in the position of the first RDF peak, listed in Table 6. Also, in accordance to the conclusion that the IE has little influence on the basin distributions, there is little difference between the RDFs for different basins, as shown in the breakdown in Table 7.

Conclusions

It is a general dogma in the field of Computational Chemistry that the ability of theoretical methods to reproduce experimental properties is roughly proportional to computational cost; i.e., it should increase as the methods change from force fields to semiempiricals and then to ab initio and density functional theory methods. The present results challenge this dogma, at least from the viewpoint of using classical force fields or semiempirical models to obtain structural ensemble information for biological molecules.

It has been shown here that the specific QM/MM interaction energies between the QM peptide and classical waters have only a very weak influence on the relative populations of the different basins. On the other hand, the extensive sampling necessary for biological calculations amplifies the intrinsic issues of the SE methods. Obviously, with parametrizations, isolated cases may happen where the older or more primitive of the methods occasionally shows better agreement with experimental data due to some fortunate side effect, and this was the case with the population distributions calculated with MNDO, and the dipolar coupling constant calculated from the ff94 coordinates.

Aside from those fortunate side effects, none of the methods employed here could accurately reproduce experimental data such as dipolar coupling constants and population distributions for the simplest possible peptide model. However, one must also consider that, taking for example the population distributions obtained by Grdadolnik of 60–76% PP_{II},¹⁰³ a quick estimate of the free energy difference required to obtain this level of precision, using $\Delta G = -RT \ln(P_I/P_{II})$, indicates that the free energy of the PP_{II} basin should lie only about 0.24–0.67 kcal/mol below other minima, an accuracy level hard to reach even for high-level QM methods.

Comparison with the latest generation of classical force fields shows that results provided by the semiempirical Hamiltonians are not especially closer to experiment than the classical ones. Indeed, results from the classical ff99SB force field are generally in better agreement with experiment than most of the quantum methods. The exception was the RM1 parametrization, which was the only method to show consistently better results, although it still did not fully agree with experimental numbers. This improved performance likely originates from the explicit inclusion of biological molecules in its training set.

We note that the present results do not invalidate current QM/MM studies: QM methods are still necessary for calculation of

processes that involve bond breaking and forming, charge redistribution, etc., while the current force fields may be better suited for structural ensemble property studies. Additionally, it should be noted that most current usage of semiempirical QM/MM methods restrict the QM region to just an organic ligand molecule and in some cases a few water molecules or side chains, while the backbone energetics are handled by force fields. On the other hand, the partial success of the RM1 method may also point the way for further improvement of the semiempirical methods, with the inclusion of more realistic and biologically relevant molecules in the training sets.

Finally, it is worthwhile to notice that the present work does not include the latest OMx series of semiempirical methods from Prof. Thiel's group,⁷ which improve on the existing semiempirical models by the inclusion of orthogonalization corrections. Those methods have been shown to considerably improve on the peptide conformations as compared to AM1 and PM3,¹¹⁰ to better reproduce the water–water interaction energies in a QM/MM environment,¹¹¹ and have recently been expanded with a dispersion correction term similar to the one used for SCC-DFTB/d.¹¹²

Acknowledgment. R.C.W. acknowledges DoE SciDAC grant DE-AC36-99GO10337 for partial funding of this work. We also additionally thank the San Diego Supercomputer Center and the NSF advanced User Support Program for supporting the development of software used in this work and the NSF TeraGrid program for providing supercomputer time under grants TG-MCB090110 to R.C.W. and TG-MCA05S010 to A.E.R. We also acknowledge the University of Florida High-Performance Computing Center for providing computational resources. Work at University of Florida was funded by the National Science Foundation grant number CHE-0822-935.

References and Notes

- (1) Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. A* **2006**, *110*, 13551–13559.
- (2) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. *J. Comput. Chem.* **2006**, *27*, 1101–1111.
- (3) Hu, H.; Elstner, M.; Hermans, J. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 451–463.
- (4) Zhou, H. Y.; Tajkhorshid, E.; Frauenheim, T.; Suhai, S.; Elstner, M. *Chem. Phys.* **2002**, *277*, 91–103.
- (5) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 498–510.
- (6) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601–1622.
- (7) Weber, W.; Thiel, W. *Theor. Chem. Acc.: Theory, Comput. Mod. (Theor. Chim. Acta)* **2000**, *103*, 495–506.
- (8) Bräuer, M.; Kunert, M.; Dinjus, E.; Klußmann, M.; Döring, M.; Görls, H.; Anders, E. *J. Mol. Struct. THEOCHEM* **2000**, *505*, 289–301.
- (9) Bernal-Uruchurtu, M. I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-López, M. F. *J. Comput. Chem.* **2000**, *21*, 572–581.
- (10) Bohr, H. G.; Jalkanen, K. J.; Elstner, M.; Frimand, K.; Suhai, S. *Chem. Phys.* **1999**, *246*, 13–36.
- (11) Morpurgo, S.; Bossa, M.; Morpurgo, G. O. *J. Mol. Struct. THEOCHEM* **1998**, *429*, 71–80.
- (12) Anh, N. T.; Frison, G.; Solladié-Cavallo, A.; Metzner, P. *Tetrahedron* **1998**, *54*, 12841–12852.
- (13) Dannenberg, J. J. *J. Mol. Struct. THEOCHEM* **1997**, *401*, 279–286.
- (14) Csonka, G. I.; Ángyán, J. G. *J. Mol. Struct. THEOCHEM* **1997**, *393*, 31–38.
- (15) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616–626.
- (16) Thiel, W. In *Advances in Chemical Physics*; John Wiley & Sons Inc: New York, 1996; Vol. 93, pp 703–757.
- (17) Ludwig, O.; Schinke, H.; Brandt, W. *J. Mol. Model.* **1996**, *2*, 341–350.
- (18) Dos Santos, H. F.; De Almeida, W. B. *J. Mol. Struct. THEOCHEM* **1995**, *335*, 129–139.
- (19) Kolb, M.; Thiel, W. *J. Comput. Chem.* **1993**, *14*, 775–789.
- (20) Dewar, M. J. S.; Jie, C.; Yu, J. *Tetrahedron* **1993**, *49*, 5003–5038.

- (21) Csonka, G. I. *J. Comput. Chem.* **1993**, *14*, 895–898.
- (22) Burk, P.; Herodes, K.; Koppel, I.; Koppel, I. *Int. J. Quantum Chem.* **1993**, *48*, 633–641.
- (23) Zerner, M. C. In *Reviews in Computational Chemistry*; Kenny B., Lipkowitz, D. B. B., Eds.; VCH Publishers: New York, 1991; Vol. 2, pp 313–365.
- (24) Stewart, J. J. P. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1990; Vol. 1, pp 45–81.
- (25) Stewart, J. J. P. *J. Comput.-Aided Mol. Design* **1990**, *4*, 1–103.
- (26) Dewar, M. J. S.; Liotard, D. A. *J. Mol. Struct. THEOCHEM* **1990**, *206*, 123–133.
- (27) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221–264.
- (28) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209–220.
- (29) Dewar, M. J. S.; Jie, C. *J. Mol. Struct. THEOCHEM* **1989**, *187*, 1–13.
- (30) Thiel, W. *Tetrahedron* **1988**, *44*, 7393–7408.
- (31) Dewar, M. J. S.; Dieter, K. M. *J. Am. Chem. Soc.* **1986**, *108*, 8075–8086.
- (32) Dewar, M. J. S.; Zebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (33) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4907–4917.
- (34) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.
- (35) Caramella, P.; Houk, K. N.; Domelsmith, L. N. *J. Am. Chem. Soc.* **1977**, *99*, 4511–4514.
- (36) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.
- (37) Yang, Y.; Yu, H.; York, D.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2008**, *4*, 2067–2084.
- (38) Nam, K.; Cui, Q.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2007**, *3*, 486–504.
- (39) Hu, H.; Lu, Z.; Yang, W. *J. Chem. Theory Comput.* **2007**, *3*, 390–406.
- (40) Acevedo, O.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2006**, *128*, 6141–6146.
- (41) Naray-Szabo, G.; Berente, I. *J. Mol. Struct. THEOCHEM* **2003**, *666–667*, 637–644.
- (42) Torras, J.; Seabra, G. M.; Deumens, E.; Trickey, S. B.; Roitberg, A. E. *J. Comput. Chem.* **2008**, *29*, 1564–1573.
- (43) Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct. THEOCHEM* **2003**, *632*, 29–41.
- (44) Nemukhin, A. V.; Grigorenko, B. L.; Bochenkova, A. V.; Topol, I. A.; Burt, S. K. *J. Mol. Struct. THEOCHEM* **2002**, *581*, 167–175.
- (45) Stewart, J. J. P. *J. Mol. Model.* **2009**, *15*, 765–805.
- (46) Case, D. A.; Darden, T. A.; Cheatham, L.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; K.F.Wong; Paesani, F.; Vanicek, J.; Xu, Wu; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California: San Francisco, 2008.
- (47) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019–1031.
- (48) Seabra, G. M.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E. *J. Phys. Chem. A* **2007**, *111*, 5655–5664.
- (49) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
- (50) Brooks, B. R.; Brucoleri, R. E.; Olafson, D. J.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (51) Hammes-Schiffer, S. *Acc. Chem. Res.* **2006**, *39*, 93–100.
- (52) Claeysens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schutz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. *Angew. Chem., Int. Ed.* **2006**, *45*, 6856–6859.
- (53) Truhlar, D. G.; Gao, J. L.; Garcia-Viloca, M.; Alhambra, C.; Corchado, J.; Sanchez, M. L.; Poulsen, T. D. *Int. J. Quantum Chem.* **2004**, *100*, 1136–1152.
- (54) Truhlar, D. G.; Gao, J.; Alhambra, C.; Garcia-Viloca, M.; Corchado, J.; Sanchez, M. L.; Villa, J. *Acc. Chem. Res.* **2002**, *35*, 341–349.
- (55) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467–505.
- (56) Cui, Q.; Elstner, M.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 2721–2740.
- (57) Frauenheim, T.; Porezag, D.; Elstner, M.; Jungnickel, G.; Elsner, J.; Haugk, M.; Sieck, A.; Seifert, G. *Mater. Res. Soc. Symp. Proc.* **1998**, *491*, 91–104.
- (58) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B: Condens. Matter* **1998**, *58*, 7260–7268.
- (59) Avbelj, F.; Grdadolnik, S. G.; Grdadolnik, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1272–1277.
- (60) Shi, Z.; Chen, K.; Liu, Z.; Ng, A.; Bracken, W. C.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17964–17968.
- (61) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.
- (62) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149–5155.
- (63) Stewart, J. P., Frank J. *MOPAC 6.0*; Seiler Research Laboratory, United States Air Force Academy: Colorado Springs, CO, 1990.
- (64) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (65) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
- (66) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (67) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (68) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinform.* **2006**, *65*, 712–725.
- (69) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (70) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (71) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (72) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32*, 523–535.
- (73) Sindhikara, D.; Meng, Y.; Roitberg, A. E. *J. Chem. Phys.* **2008**, *128*, 024103–10.
- (74) Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2008**, *4*, 297–306.
- (75) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11–15.
- (76) Hu, J. S.; Bax, A. *J. Am. Chem. Soc.* **1997**, *119*, 6360–6368.
- (77) Kang, Y. K. *J. Phys. Chem. B* **2006**, *110*, 21338–21348.
- (78) Kim, Y. S.; Wang, J.; Hochstrasser, R. M. *J. Phys. Chem. B* **2005**, *109*, 7511–7521.
- (79) Lavrich, R. J.; Plusquellic, D. F.; Suenram, R. D.; Fraser, G. T.; Walker, A. R. H.; Tubergen, M. J. *J. Chem. Phys.* **2003**, *118*, 1253–1265.
- (80) Poon, C. D.; Samulski, E. T.; Weise, C. F.; Weisshaar, J. C. *J. Am. Chem. Soc.* **2000**, *122*, 5642–5643.
- (81) Smith, P. E. *J. Chem. Phys.* **1999**, *111*, 5568–5579.
- (82) Aleman, C.; Leon, S. *J. Mol. Struct. THEOCHEM* **2000**, *505*, 211–219.
- (83) Apostolakis, J.; Ferrara, P.; Caffisch, A. *J. Chem. Phys.* **1999**, *110*, 2099–2108.
- (84) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908–5920.
- (85) Boreesch, S.; Willensdorfer, M.; Steinhauser, O. *J. Chem. Phys.* **2004**, *120*, 3333–3347.
- (86) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569–585.
- (87) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2001**, *263*, 203–219.
- (88) Han, W.-G.; Jalkanen, K. J.; Elstner, M.; Suhai, S. *J. Phys. Chem. B* **1998**, *102*, 2587–2602.
- (89) Improta, R.; Barone, V. *J. Comput. Chem.* **2004**, *25*, 1333–1341.
- (90) Iwaoaka, M.; Okada, M.; Tomoda, S. *J. Mol. Struct. THEOCHEM* **2002**, *586*, 111–124.
- (91) Kalko, S. G.; Guardia, E.; Padro, J. A. *J. Phys. Chem. B* **1999**, *103*, 3935–3941.
- (92) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, 13749–13754.
- (93) Marrone, T. J.; Gilson, M. K.; McCammon, J. A. *J. Phys. Chem.* **1996**, *100*, 1439–1441.
- (94) Tobias, D. J.; Brooks, C. L. *J. Chem. Phys.* **1992**, *96*, 3864–3870.
- (95) Vargas, R.; Garza, J.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* **2002**, *106*, 3213–3218.
- (96) Wei, D.; Guo, H.; Salahub, D. R. *Phys. Rev. E: Stat., Nonlin., Soft Matter Phys.* **2001**, *64*, 011907/1–011907/4.
- (97) Wang, Z.-X.; Duan, Y. *J. Comput. Chem.* **2004**, *25*, 1699–1716.
- (98) Poon, C.-D.; Samulski, E. T.; Weise, C. F.; Weisshaar, J. C. *J. Am. Chem. Soc.* **2000**, *122*, 5642–5643.
- (99) Avbelj, F.; Kocjan, D.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17394–17397.
- (100) Avbelj, F.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5742–5747.
- (101) Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K. *J. Am. Chem. Soc.* **2001**, *123*, 9628–9633.
- (102) Schweitzer-Stenner, R.; Measey, T.; Kakalis, L.; Jordan, F.; Pizzanelli, S.; Forte, C.; Griebenow, K. *Biochemistry* **2007**, *46*, 1587–1596.
- (103) Grdadolnik, J.; Grdadolnik, S. G.; Avbelj, F. *J. Phys. Chem. B* **2008**, *112*, 2712–2718.

(104) Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C. *J. Comput. Chem.* **2003**, *24*, 21–31.

(105) García, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782–2787.

(106) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2000**, *256*, 15–27.

(107) Han, W.-G.; Elstner, M.; Jalkanen, K. J.; Frauenheim, T.; Suhai, S. *Int. J. Quantum Chem.* **2000**, *78*, 459–479.

(108) Valdes, H.; Reha, D.; Hobza, P. *J. Phys. Chem. B* **2006**, *110*, 6385–6396.

(109) Reha, D.; Valdés, H.; Vondrásek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S. *Chem.—Eur. J.* **2005**, *11*, 6803–6817.

(110) Möhle, K.; Hofmann, H.-J.; Thiel, W. *J. Comput. Chem.* **2001**, *22*, 509–520.

(111) Geerke, D. P.; Thiel, S.; Thiel, W.; Gunsteren, W. F. v. *Phys. Chem. Chem. Phys.* **2008**, *10*, 297–302.

(112) Tuttle, T.; Thiel, W. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159–2166.

JP903474V