JCTC Journal of Chemical Theory and Computation

Kinetics from Implicit Solvent Simulations of Biomolecules as a Function of Viscosity

Michael Feig*

Department of Biochemistry and Molecular Biology, and Department of Chemistry, Michigan State University, East Lansing, Michigan 48824-1319

Received March 20, 2007

Abstract: Kinetic properties of alanine dipeptide, the B1 domain of streptococcal protein G, and ubiquitin are compared between explicit solvent and implicit solvent simulations with the generalized Born molecular volume (GBMV) method. The results indicate that kinetics from explicit solvent simulations and experiments can be matched closely when the implicit solvent simulations are combined with Langevin dynamics and a friction coefficient near 10 ps^{-1}. Smaller and larger friction coefficients accelerate and slow down conformational sampling. It is found that local conformational exploration without the crossing of significant barriers can be accelerated by a factor of 4−5; however, the acceleration of barrier crossings is limited to about a factor of 2. The use of a Nosé−Hoover thermostat instead of Langevin dynamics greatly enhances local conformational sampling but slows down the crossing of barriers by at least an order of magnitude because of the lack of solute−solvent stochastic collisions.

Introduction

Computer simulations are often involved in the exploration of how biomolecular structure and dynamics give rise to biological function. Conventional simulation methods employ explicit representations of the solvent environment which can provide a high level of realism but usually at substantial computational costs. As an alternative, implicit solvent models have become increasingly popular in the simulation of biological macromolecules in order to be able to reach larger system sizes and longer time scales.[1,2] Implicit solvent models rely on the assumption that ensemble averages of instantaneous interactions between a solute and explicit solvent molecules may be approximated through a mean-field formalism.[3,4] Explicit solvent molecules can then be omitted from the system, thereby reducing the computational cost of such simulations because of smaller system sizes and the absence of solvent relaxation.

Physically motivated implicit solvent models often decompose the solvation free energy into electrostatic and nonpolar contributions.[5] The electrostatic component is commonly calculated on the basis of a continuum electrostatic model, where the solvent is represented as a homogeneous, high-dielectric medium that surrounds a low-dielectric solute cavity with explicit partial charges at the atomic centers of the solute.[6,7] The Poisson−Boltzmann (PB) equation rigorously describes such a model and can be solved numerically for the electrostatic potential throughout space.[2,8,9] The electrostatic component of the free energy of solvation is readily calculated from the electrostatic potential and the solute partial charges. Direct application of PB theory in biomolecular simulations is possible but hindered by the lack of sufficiently efficient and accurate numerical PB solvers and by difficulties in obtaining continuous first derivatives without altering the solute−solvent dielectric boundary.[2,10−13] Alternatively, generalized Born formalisms provide a convenient and efficient analytical approximation of electrostatic solvation free energies based on the same continuum electrostatic model described by Poisson theory.[14] When recent methodological advances are followed, the latest generation of generalized Born models can reproduce electrostatic solvation energies from Poisson theory accurately at a fraction of the cost that it would take to solve the Poisson equation.[15−21] The nonpolar contribution to the solvation free energy can be approximated with a term proportional to the solvent-accessible surface area (SASA).[5]

* Corresponding author. Phone: (517) 432-7439. Fax: (517) 353-9334. E-mail: feig@msu.edu.

10.1021/ct7000705 CCC: $37.00
Published on Web 07/14/2007

Implicit solvent models address the thermodynamic aspects of solvation but neglect hydrodynamic effects that become relevant in the simulation of kinetic processes. Particularly important are stochastic collisions with solvent molecules and frictional forces which directly impact kinetic rates and the magnitude of conformational fluctuations. Both of these contributions can be included with Langevin dynamics in combination with a given implicit solvent model.[22−24] In Langevin dynamics, a modified equation of motion is applied for a single particle *i*:

$$m_i\mathbf{a}_i(t) = \mathbf{F}_i(\mathbf{r}) - f_i m_i \mathbf{v}_i + \mathbf{F}_{\text{random}}(t) \quad (1)$$

where m_i, $\mathbf{a}_i$, $\mathbf{v}_i$, and $\mathbf{F}_i$ are mass, acceleration, velocity, and the force due to the interactions with the rest of the system. f is the friction coefficient, and $\mathbf{F}_{\text{random}}$ is a stochastic force simulating random collisions with solvent molecules.[22] Langevin dynamics effectively provide a thermostat with a temperature that is controlled through the magnitude of the random forces. It may be compared with other commonly employed thermostats,[24] in particular, the Nosé−Hoover thermostat[25,26] with the following equations of motion:

$$\dot{\mathbf{r}}_i = \mathbf{p}_i/m_i$$

$$\dot{\mathbf{p}}_i = -\nabla V(\mathbf{r}) - \zeta \mathbf{p}_i$$

$$\dot{\zeta} = \frac{1}{q}\left(\sum_i \frac{\mathbf{p}_i^2}{m_i} - gkT\right)$$

where $\mathbf{r}_i$, $\mathbf{p}_i$, and m_i are the position, momentum, and mass of particle *i*. V is the interaction potential according to a given force field; g is the number of degrees of freedom; k is the Boltzmann constant; and T is the temperature of the thermal bath. The extended variable ζ acts like a friction constant that is coupled to the temperature bath according to the coupling constant q. The Nosé−Hoover thermostat resembles Langevin dynamics without stochastic forces, but ζ may assume both positive and negative values and fluctuates during a simulation.

A number of simulation studies with implicit solvent have been reported during recent years.[27−37] The general conclusion from these studies is that is possible to obtain stable trajectories with implicit solvent simulations that exhibit conformational sampling comparable to explicit solvent simulations.[16,21,28,37−40] Some discrepancies, in particular with respect to the stability of salt bridges, have also been reported, but they appear to be resolved at least in part through careful adjustment of the dielectric interface and/or force field reparameterization.[40−43]

A question that has not been fully addressed to date is to what extent the kinetic properties can be reproduced correctly in implicit solvent simulations and to what extent simulations can be accelerated when solvent viscosity is reduced or omitted. Previous studies have found that conformational transitions in the context of protein folding can be accelerated substantially and predictably in Langevin dynamics with reduced friction coefficients,[23] but other evidence also suggests that the choice of the implicit solvent itself may also have a significant impact on the simulated dynamics.[28]

In this study, this question is addressed further with the application of the thermodynamically highly accurate generalized Born molecular volume (GBMV) implicit solvent model[17] to the following two cases: First, kinetic rates of transitions between dominant conformational basins in alanine dipeptide are compared between implicit and explicit solvent simulations. Second, native state conformational sampling of the B1 domain of streptococcal protein G and ubiquitin is compared with experimental data from X-ray crystallography and NMR spectroscopy. The results for both the dipeptide kinetics and protein structure dynamics indicate that kinetic properties can in fact be accurately reproduced with implicit solvent when the effects of solvent friction are included appropriately. Furthermore, conformational sampling can be accelerated by a moderate amount when reduced friction coefficients are applied.

In the following, the simulation methodology is described in more detail before the simulation results are presented and compared to experimental data.

Methods

Explicit and implicit solvent molecular dynamics simulations of blocked alanine dipeptide, the B1 domain of streptococcal protein G, and ubiquitin were performed as summarized in Table 1. The alanine dipeptide simulations were started with peptide torsion angles of $\phi = -65°$ and $\psi = -40°$ corresponding to the α basin. The simulations of protein G and ubiquitin used the experimentally determined structures in aqueous solvent according to NMR spectroscopy (PDB codes: 3GB1[44] and 1D3Z,[45] respectively) as the initial conformations. The all-atom CHARMM force field[46] was used in all of the simulations in conjunction with the CMAP cross-correlation potential[47,48] to improve the sampling of ϕ/ψ backbone torsion angles.

In the explicit solvent simulations, the solute was solvated in a cubic box filled with TIP3P water molecules. Four sodium counterions were added to neutralize the negative charge of protein G. Counterions were not needed for the neutral ubiquitin, and no extra salt was added. The resulting box sizes are 27.23 Å^3 for alanine dipeptide, 61.00 Å^3 for protein G, and 69.30 Å^3 for ubiquitin. Periodic boundary conditions were applied, and particle-mesh Ewald summation with a real-space cutoff of 9 Å was used to calculate electrostatic interactions. The explicit solvent system was simulated in an NVT ensemble at 300 K with a Nosé−Hoover thermostat.[25] A time step of 2 fs was employed in conjunction with SHAKE[49] in order to constrain bonds between heavy atoms and hydrogen atoms. A standard equilibration protocol was applied with initial steepest descent and adopted-basis Newton−Raphson minimization followed by slow heating over 42 ps to the final temperature and slight adjustment of the simulation box size in order to obtain correct bulk water densities corresponding to a temperature of 300 K and a pressure of 1 atm.

Implicit solvent simulations were run with the GBMV variant[20,50] of the generalized Born formalism. The dielectric constant inside the solute cavity was set to $\epsilon = 1$ and to $\epsilon = 80$ for the surrounding medium. GBMV parameters defined in the original references[20,51] were set to $\beta = -12$, $S_0 =$

Table 1. Summary of Simulations Analyzed in This Study

system	solvent	CMAP	length
alanine dipeptide	explicit	original	200 ns
alanine dipeptide	implicit $f = 50, 10, 5, 1, 0.5$ ps^{-1}	original	500 ns
alanine dipeptide	implicit $f = 50, 10, 5, 1, 0.5$ ps^{-1}	modified	500 ns
alanine dipeptide	implicit $f = 5$ ps^{-1} (C), 10 ps^{-1} (N), 20 ps^{-1} (O)	modified	1000 ns
alanine dipeptide	implicit Nosé−Hoover, $q = 1, 10, 100$ kcal/(mol ps^2)	modified	500 ns
protein G	explicit	original	50 ns
protein G	implicit $f = 5$ ps^{-1} (C), 10 ps^{-1} (N), 20 ps^{-1} (O)	original	50 ns
protein G	implicit $f = 50, 5$ ps^{-1}	modified	50 ns
protein G	implicit Nosé-Hoover, $q = 10$	modified	50 ns
protein G	implicit $f = 5$ ps^{-1} (C), 10 ps^{-1} (N), 20 ps^{-1} (O)	modified	50 ns
ubiquitin	explicit	original	22 ns
ubiquitin	implicit $f = 5$ ps^{-1} (C), 10 ps^{-1} (N), 20 ps^{-1} (O)	original	22.5 ns
ubiquitin	implicit $f = 50, 5$ ps^{-1}	modified	22.5 ns
ubiquitin	implicit Nosé-Hoover, $q = 10$ kcal/(mol ps^2)	modified	22.5 ns
ubiquitin	implicit $f = 5$ ps^{-1} (C), 10 ps^{-1} (N), 20 ps^{-1} (O)	modified	22.5 ns

0.65, $C_0 = -0.1$, and $C_1 = 0.9$ in order to obtain stable trajectories.[38] The GBMV method also provides an estimate of the SASA that was used to calculate the hydrophobic solvation free energy according to γ·SASA with $\gamma = 5.42$ cal/(mol Å^2). Electrostatic interactions were switched to zero from 16 to 18 Å in the simulations of ubiquitin and protein G. No cutoff was applied in the alanine dipeptide simulations. The implicit solvent simulations were run with the same force field and CMAP torsion potential as with the explicit solvent simulations. In addition, a slightly modified CMAP potential was also used in order to improve agreement with the potential of mean force of alanine dipeptide from the explicit solvent simulation (see below). An integration time step of 1.5 fs was used in the protein G and ubiquitin simulations in conjunction with SHAKE[49] to constrain bonds involving hydrogen atoms. Alanine dipeptide was simulated without SHAKE and with a time step of 1 fs. In order to maintain a constant temperature during the simulations, either the Nosé−Hoover algorithm[25] (with coupling constants of 1, 10, or 100 kcal/mol·ps^2 in a single thermostat applied to the entire system) or a Langevin heat bath was applied.[22] Langevin dynamics also include the effects of viscosity. In this study, friction coefficients f of 0.5, 1, 5, 10, and 50 ps^{-1} were applied to non-hydrogen atoms. No memory function was employed in the Langevin algorithm that represents the overdamped limit.

All of the simulations were performed with the CHARMM program.[52] Version c30b2 was used for the explicit solvent simulation of alanine dipeptide, c32a2 for the implicit solvent simulations of alanine dipeptide, and c33a1 for the implicit and explicit solvent simulations of protein G and ubiquitin. The MMTSB Tool Set[53] was used to facilitate and analyze the simulations in conjunction with CHARMM.

Results

Kinetic Transitions in Alanine Dipeptide. The alanine dipeptide system (Figure 1A) is used as a model for peptide backbone dynamics. A potential of mean force map from the 200 ns explicit solvent simulation (Figure 2A) shows that there are essentially four distinct states: α at (−60,−45) to (−100,0), β/PPII at (−60,150) to (−160,160), α_L at (60,50), and $C7_{ax}$ at (50,−150). With the CHARMM force field used here, the α and β basins are practically equi-energetic while the other two minima are slightly higher by a few kilocalories per mole. Implicit solvent simulations with the GBMV model and the same force field (Figure 2B) result in a very similar free energy map, but some differences are noticeable. The main minimum in the β basin is shifted from the poly proline II (PPII) conformations at (−60,150) to fully extended conformations (−160,160), and the second minimum in the α basin at (−100,0) is enhanced. Furthermore, the barriers from the α_L basin to the β and $C7_{ax}$ basins are elevated compared to the explicit solvent simulations. In order to understand the source of these deviations better, it is instructive to examine the energetics of alanine dipeptide conformations with implicit solvent conformations in more detail. Table 2 compares the relative energies for selected conformations between the explicit solvent PMF and adiabatic free energies from the continuum dielectric implicit solvent model. It can be seen that the deviations are relatively small when electrostatic solvation energies are obtained directly from solutions to the Poisson equation. However,

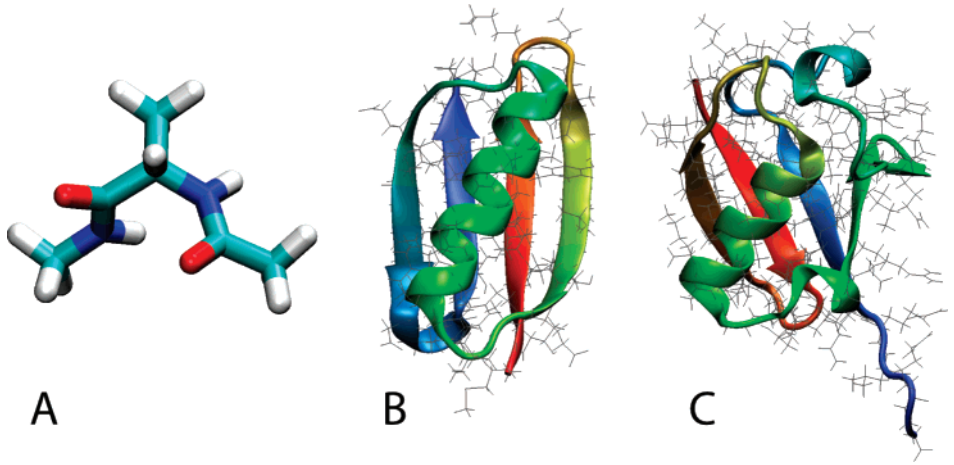

Figure 1. Structures of blocked alanine dipeptide (A), B1 domain of streptococcal protein G (PDB code: 3GB1[44]) (B), and ubiquitin (PDB code: 1D3Z[45]) (C). Graphics were generated with VMD.[69]

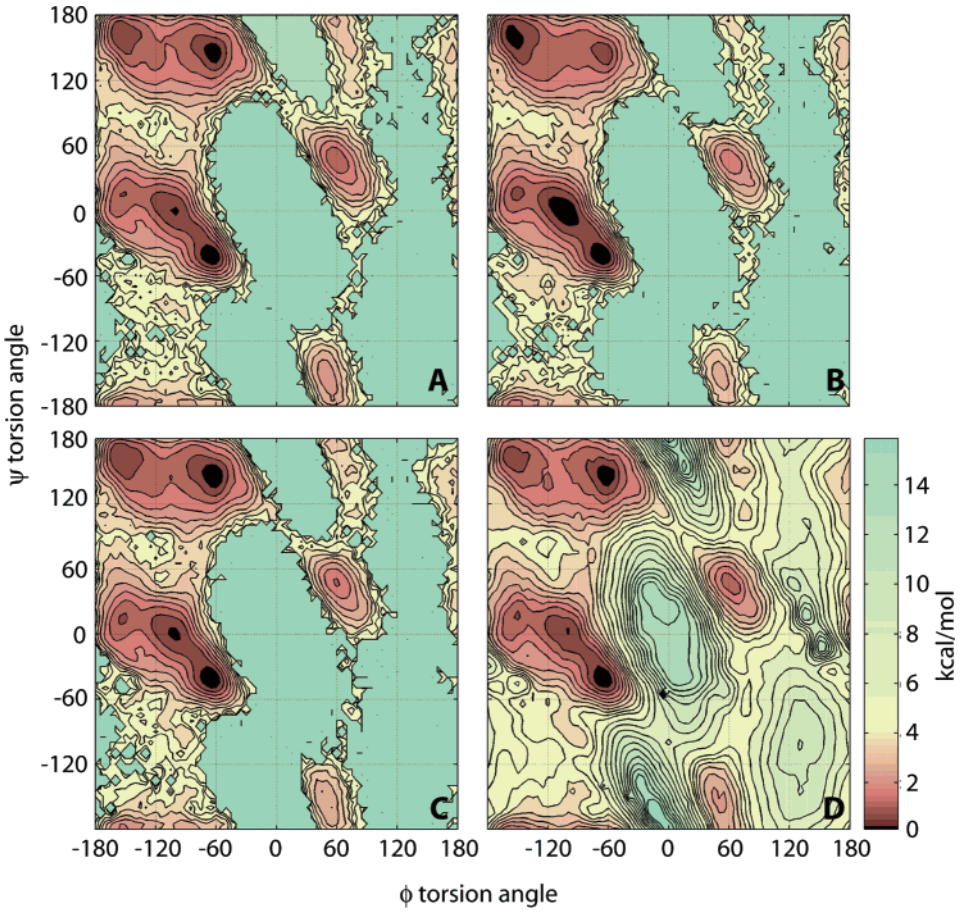

Figure 2. Potential of mean force from molecular dynamics simulations of blocked alanine dipeptide with explicit solvent and original CMAP torsion potential (A), implicit solvent and original CMAP potential (B), and implicit solvent and modified CMAP potential (C). A friction coefficient of 10 ps^{-1} was used in the implicit solvent simulation shown. The results from the simulations are compared with an adiabatic map (D) where implicit solvent energies with the modified CMAP potential are evaluated directly after minimization at different values of ϕ/ψ.

the deviations are amplified when the GBMV formalism is used to approximate the Poisson solutions. Therefore, the (small) deviations between the free energy maps shown in Figure 2A and B represent mostly errors due to the generalized Born approximation rather than the implicit solvent model itself.

In order to focus on a comparison of kinetic properties between implicit and explicit solvent, it is desirable that the underlying thermodynamics is as similar as possible. Most of the differences described above should be resolved in principle with a more accurate generalized Born method. However, further improvements of the already very accurate

Table 2. Free Energies of Selected Conformations of Alanine Dipeptide with Implicit and Explicit Solvent Relative to β Conformation[a]

	f (C−N−Cα−C)	y (N−Cα−C−N)	GBMV [kcal/mol]	PB [kcal/mol]	explicit PMF [kcal/mol]
β	−155	160	0	0	0
PPII	−65	145	0.47	0.12	−0.45
$\beta \rightarrow \alpha_L$	0	105	4.82	4.25	4.02
α_L	60	45	1.34	0.75	0.40

[a] Implicit solvent energies are obtained under the adiabatic approximation by minimizing alanine dipeptide with a harmonic restraint on the ϕ and ψ torsion angles, each with a force constant of 1000 kcal/mol/Å^2, for different values of ϕ and ψ on a grid with a spacing of 15°. Explicit solvent energies are obtained as a potential of mean force from the sampling probabilities in the explicit solvent molecular dynamics simulation. The original CMAP potential was used in all cases. Poisson−Boltzmann (PB) results were obtained with the PBEQ finite difference solver in CHARMM using a grid spacing of 0.15 Å.

GBMV method compared to other generalized Born formalisms are not easily achievable. On the other hand, direct solutions of the Poisson equation at a very high level of accuracy are not practical in simulation applications. A pragmatic solution is the use of a slightly modified CMAP term for the ϕ/ψ torsion angles. The CMAP potential function is obtained from a spline-interpolated grid-based two-dimensional function of ϕ and ψ.[47,48,54] The deviations between the explicit and implicit solvent simulations are easily incorporated by adding the difference between the respective free energy maps to the original CMAP grid data. The modified CMAP is available from the author by request. The use of different ϕ/ψ torsion maps for implicit and explicit solvent has been suggested previously to correct for deficiencies of the implicit solvent model.[40] As a result, it is possible to reproduce the explicit solvent free energy map nearly exactly (see Figure 2C).

Implicit solvent allows the calculation of the complete free energy map under the adiabatic approximation (shown in Figure 2D). As would be expected, the resulting map agrees closely with the potentials of mean force obtained from the simulations, but the adiabatic map also offers insight into higher-energy regions that are not sampled during sub-microsecond simulations. In particular, a clearer view of possible transition pathways between the four main minima is given. Transitions between α and β basins may progress either through the C7eq conformation near (−60,60) or through the second transition state at (−100,−120). Transitions between β and α_L basins proceed through the transition state at (0,90), while transitions between the α_L and $C7_{ax}$ can also follow two routes along $\psi = 80$ either through positive or negative ϕ angles. Additional transitions may occur directly between β and $C7_{ax}$ across the transition state at (120,120) and from α to $C7_{ax}$ across the transition state at (0,−100) or to α_L across (130,−30).

The sub-microsecond simulations described here sample only the α/β, β/α_L, and $\alpha_L/C7_{ax}$ transitions sufficiently to obtain meaningful statistical averages. Table 3 shows the kinetic rates for these transitions in alanine dipeptide (following either pathway in the case of α/β and $\alpha_L/C7_{ax}$ transitions) obtained from explicit and implicit solvent simulations. It can be seen that the implicit solvent simulations are in good qualitative agreement with the explicit solvent results when the modified CMAP is used to match the sampling of ϕ/ψ angles to the explicit solvent and when a friction coefficient between 10 and 50 ps^{-1} is used. With a friction coefficient of 10 ps^{-1}, most transitions are faster by about a factor of 2, except for the transition from β to α_L, which is too slow. On the other hand, the rates are significantly reduced to about half the value found with explicit solvent when a friction coefficient of 50 ps^{-1} is used. If the friction coefficient is reduced further from 10 ps^{-1}, some but not all of the rates are accelerated. Most noteworthy, the α_L-to-β rate appears to be largely unaffected and actually slightly decreases as the friction is reduced. Previous studies have suggested that a single friction coefficient may not be optimal.[34] After trying a number of combinations (data not shown), it was found that a friction coefficient of 10

Table 3. Kinetic Rates in ns^{-1} for Conformational Transitions in Blocked Alanine Dipeptide from Implicit and Explicit Solvent Simulations[a]

solvent	CMAP	friction	α/β	β/α_L	$\alpha_L/C7_{ax}$
expl.	orig.		2.5 (*225*)	0.3 (*26*)	3.8 (*35*)
			2.6 (*224*)	2.7 (*25*)	14.8 (*33*)
impl.	orig.	50 ps^{-1}	1.1 (*316*)	0.02 (*4*)	1.4 (*9*)
			1.6 (*317*)	0.8 (*5*)	8.8 (*8*)
	orig.	10 ps^{-1}	3.8 (*1000*)	0.06 (*12*)	5.5 (*47*)
			5.0 (*1000*)	1.4 (*12*)	25.7 (*47*)
	orig.	5 ps^{-1}	4.8 (*1273*)	0.09 (*19*)	6.7 (*75*)
			6.3 (*1273*)	1.4 (*16*)	38.3 (*71*)
	orig.	1 ps^{-1}	5.4 (*1391*)	0.1 (*25*)	6.5 (*92*)
			6.9 (*1388*)	1.6 (*22*)	38.6 (*89*)
	orig.	0.5 ps^{-1}	5.4 (*1422*)	0.1 (*21*)	8.7 (*79*)
			7.0 (*1421*)	2.9 (*26*)	55.6 (*8.5*)
	mod.	50 ps^{-1}	1.4 (*312*)	0.09 (*21*)	1.7 (*20*)
			1.4 (*311*)	1.6 (*19*)	7.5 (*19*)
	mod.	10 ps^{-1}	4.6 (*893*)	0.2 (*45*)	6.4 (*62*)
			4.3 (*897*)	4.9 (*48*)	22.9 (*64*)
	mod.	5 ps^{-1}	5.7 (*1299*)	0.2 (*53*)	6.5 (*77*)
			5.9 (*1302*)	4.8 (*57*)	29.5 (*80*)
	mod.	1 ps^{-1}	6.5 (*1427*)	0.3 (*58*)	7.0 (*99*)
			6.3 (*1425*)	4.6 (*65*)	27.2 (*107*)
	mod.	0.5 ps^{-1}	6.0 (*1340*)	0.3 (*75*)	7.1 (*121*)
			6.0 (*1346*)	4.4 (*75*)	33.3 (*117*)
	mod.	mixed[b]	4.1 (*1940*)	0.2 (*110*)	4.2 (*146*)
			4.0 (*1944*)	3.4 (*118*)	18.5 (*154*)

[a] Transitions are counted when one of the four basins is reached from one of the other basins. No transitions are recorded if a basin is left and re-entered without visiting another basin. Forward and backward rates are give in the first and second row, respectively. The number of observed transitions is given in parentheses. [b] Friction coefficient of 10 ps^{-1} for nitrogen, 20 ps^{-1} for oxygen, and 5 ps^{-1} for carbon.

Table 4. Mean First Passage Time in ps of Conformational Transitions in Blocked Alanine Dipeptide from Implicit Solvent Simulations in Comparison with Results from Explicit Solvent Simulation[a]

solvent	CMAP	friction	α/β	β/α_L	$\alpha_L/C7_{ax}$	$\alpha/C7_{ax}$
expl.	orig.		**426** (*27*)	**6257** (*1130*)	**4441** (*1181*)	**8968** (*1748*)
			459 (*35*)	**430** (*78*)	**602** (*376*)	**744** (*97*)
impl.	orig.	50 ps^{-1}	929 (*56*)	94 617 (*15420*)	53 305 (*29034*)	157 676 (*41942*)
			660 (*40*)	1278 (*599*)	77 (*30*)	2088 (*737*)
	orig.	10 ps^{-1}	282 (*9*)	25 562 (*4537*)	**5710** (*1901*)	25465 (*4536*)
			223 (*8*)	**686** (*188*)	**2216** (*1378*)	**693** (*168*)
	orig.	5 ps^{-1}	220 (*6*)	21 029 (*4522*)	**4115** (*1601*)	24 193 (*5579*)
			178 (*6*)	**608** (*107*)	2219 (*1076*)	693 (*147*)
	orig.	1 ps^{-1}	195 (*6*)	14 487 (*2557*)	**2897** (*851*)	17 714 (*3359*)
			166 (*6*)	**545** (*88*)	2134 (*923*)	**658** (*94*)
	orig.	0.5 ps^{-1}	195 (*5*)	16 350 (*4741*)	**4682** (*1651*)	18 778 (*5473*)
			159 (*5*)	**372** (*59*)	**1176** (*1027*)	518 (*70*)
	mod.	50 ps^{-1}	762 (*42*)	20 901 (*4017*)	14 357 (*4976*)	27 691 (*6156*)
			832 (*46*)	721 (*200*)	4621 (*2288*)	1083 (*235*)
	mod.	10 ps^{-1}	234 (*8*)	**8089** (*1033*)	**5328** (*1161*)	**12473** (*1871*)
			261 (*9*)	247 (*30*)	**950** (*485*)	534 (*67*)
	mod.	5 ps^{-1}	192 (*5*)	**7549** (*750*)	**5160** (*970*)	12526 (*1748*)
			190 (*6*)	254 (*32*)	**670** (*362*)	495 (*53*)
	mod.	1 ps^{-1}	166 (*4*)	**6637** (*665*)	**3708** (*624*)	**9116** (*1013*)
			183 (*6*)	267 (*32*)	**514** (*192*)	451 (*50*)
	mod.	0.5 ps^{-1}	178 (*5*)	**4998** (*605*)	**3121** (*618*)	**7776** (*1230*)
			192 (*6*)	238 (*25*)	**476** (*137*)	437 (*41*)
	mod.	mixed[b]	265 (*6*)	7943 (*691*)	**6203** (*797*)	12667 (*1241*)
			287 (*7*)	**359** (*29*)	**345** (*149*)	**616** (*46*)

[a] Passage times in forward and backward direction are given for each transition. Estimated Statistical errors calculated as $\sigma/\sqrt{N}$ from the standard deviation σ and the number of transitions N are given in parentheses. Values that agree with the explicit solvent simulation within the error intervals are shown in boldface. [b] Friction coefficient of 10 ps^{-1} for nitrogen, 20 ps^{-1} for oxygen, and 5 ps^{-1} for carbon.

ps^{-1} applied to the amide nitrogen, 20 ps^{-1} applied to the carbonyl oxygen, and 5 ps^{-1} applied to carbon atoms gave the best results. In that case, the rates are in good quantitative agreement with the explicit solvent simulations (see Table 3), although the transition rates between the α and β basins are still significantly overestimated.

Transition rates obtained with implicit solvent and the original CMAP torsion term do not agree as well with the explicit solvent simulations. As would be expected from the differences in the free energy maps, the rates between α and β basins are more asymmetric with faster rates from β to α reflecting a more favorable free energy of the α basin versus the poly proline II conformation. As a result of the increased barrier height at (0,100), transitions between β and α_L are significantly slowed down. However, despite the differences in detail, the implicit solvent simulations with the original CMAP term also provide an overall reasonable qualitative description of the kinetics of alanine dipeptide if a friction coefficient of 10 ps^{-1} is chosen.

A comparison of mean first passage times is given in Table 4. The mean first passage time between two basins measures the time it takes to reach the second basin after the first basin has been entered. According to this definition, a transition from β to α may involve a simple barrier crossing from β to α or a more complicated path such as β to α_L, α_L to β, and finally β to α. Therefore, transitions between the α and $C7_{ax}$ basins are also considered along with transitions between α/β, β/α_L, and $\alpha_L/C7_{ax}$. An analysis of the mean first passage times provides similar conclusions as for the kinetic rates of single barrier crossings described above. Overall good agreement with the explicit solvent simulation is found when the modified CMAP potential is used in conjunction with the mixed friction coefficients as described above. When a single friction coefficient is used for all non-hydrogen atoms, the data from explicit solvent are reproduced best with friction coefficients between 1 and 10 ps^{-1}. However, larger discrepancies remain for transitions between α and β basins that are too fast with implicit solvent unless a friction coefficient of 50 ps^{-1} is employed.

The distribution of mean first passage times for transitions between α and β shown in Figure 3 provides a more detailed view of how kinetics from implicit and explicit solvent compare. It can be seen that the relative frequency of long passage times becomes more similar to the explicit solvent simulations beyond 500 ps to 1 ns while the differences are largest for the distribution of short-time transition events. Furthermore, it is apparent that the explicit solvent distribution of mean first passage times at short times is represented best with the mixed friction coefficients around 10 ps^{-1}. However, a friction coefficient of 50 ps^{-1} provides significantly better agreement with explicit solvent for the distribution of passage times beyond 500 ps (β-to-α transition) and beyond 800 ps (α-to-β transition). This finding suggests that Langevin dynamics with any choice of friction coefficient do not fully capture the hydrodynamic effects of water throughout the entire range of relevant time scales if compared to explicit solvent.

As discussed briefly above, transitions are accelerated over explicit solvent with small friction coefficients and slowed down with large friction coefficients. Figure 4 shows the

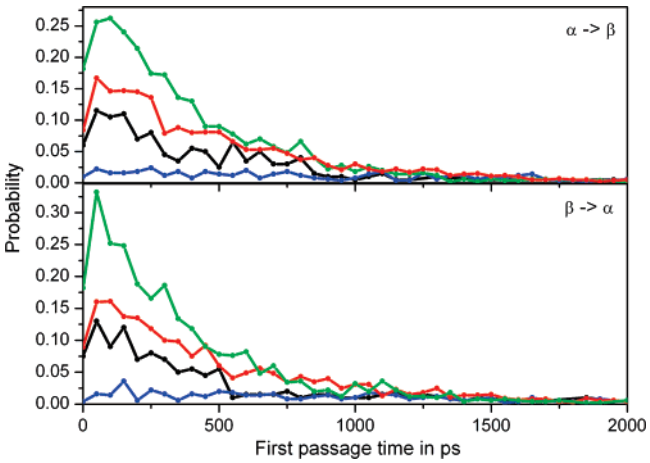

Figure 3. Distribution of first passage times for $\alpha \rightarrow \beta$ (top) and $\beta \rightarrow \alpha$ (bottom) transitions in blocked alanine dipeptide. Results from explicit solvent simulation are shown in black, from implicit solvent simulations with modified CMAP potential in blue ($f = 50$ ps^{-1}), red (mixed friction, see text), and green ($f = 5$ ps^{-1}), respectively.

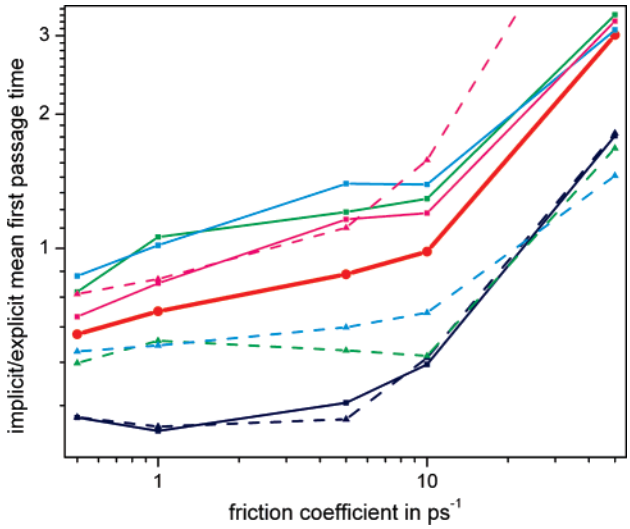

Figure 4. Mean first passage times from implicit solvent simulations with modified CMAP potential relative to explicit solvent values for conformational transitions in blocked alanine dipeptide as a function of friction coefficient. Both axes are shown in logarithmic scale. Individual transitions are indicated by color as follows: α/β (blue), β/α_L (green), $\alpha_L/C7_{ax}$ (magenta), and $\alpha/C7_{ax}$ (cyan). Solid (dashed) lines indicate forward (backward) transitions. The red line indicates the average relative mean first passage times from all transitions.

mean first passage times relative to explicit solvent as a function of the friction coefficient for the observed transitions in the alanine dipeptide system. The graph readily identifies two distinct regimes above and below a friction coefficient of 10 ps^{-1}. At 10 ps^{-1}, the mean first passage times are roughly equal between implicit and explicit solvent. Transitions below 10 ps^{-1} are accelerated with decreasing friction according to $r = f \times 0.034 + 0.67$, where r is the speed relative to explicit solvent and f is the implicit solvent friction coefficient in picoseconds^{-1}. The extrapolation of this function to zero friction gives a speedup by less than a factor of 2. A slightly better fit is found with $\log(r) = \log(f) \times 0.137 - 0.34$; however, this function does not extrapolate to zero friction. Solute friction is dominant in this regime, as pointed out previously, and consequently the transitions are only accelerated moderately. On the other hand, friction coefficients larger than 10 ps^{-1} slow down the kinetics much more rapidly, indicative of a solvent friction-dominated regime. Because only two points (10 and 50 ps^{-1}) were simulated, a functional form for this regime cannot be given with confidence.

Table 5. Mean First Passage Time in ps of Conformational Transitions in Blocked Alanine Dipeptide from Implicit Solvent Simulations with Modified CMAP Torsion Potential (see Text) and Nosé−Hoover Thermostat as a Function of the Coupling Constant q[a]

	$q = 1$	$q = 10$	$q = 100$	explicit
$\alpha \rightarrow \beta$	11 367 (*2357*)	85 948 (26 652)	5683 (982)	426 (*27*)
$\beta \rightarrow \alpha$	15 040 (*18*)	44 753 (24 621)	7048 (888)	459 (*35*)

[a] Statistical errors are given in parentheses. Results from the explicit solvent simulations are shown for comparison.

So far, only the implicit solvent simulations with Langevin dynamics have been discussed. Table 5 shows the mean first passage times between α and β basins when a Nosé−Hoover thermostat is used instead of Langevin dynamics. It can be seen that with any of coupling constants, the first passage times are 1−2 orders of magnitude longer than with the explicit solvent. Transitions to the right side of the Ramachandran plot were observed, but they occurred so rarely that significant statistics could not be obtained from 500 ns simulations. Although simulations with a Nosé−Hoover thermostat represent the limit of zero friction, the absence of stochastic collisions with water molecules greatly diminishes the effectiveness of overcoming transitional barriers, which explains the very slow kinetics in the alanine dipeptide system when Langevin dynamics are not used.

Native State Dynamics of Protein G and Ubiquitin. Protein G and ubiquitin were simulated with explicit and implicit solvent over 50 and 22.5 ns, respectively, in order to examine how the conclusions from the alanine dipeptide system apply to native state simulations of proteins.

Average Structural and Dynamic Properties. First, average structural and dynamic properties are compared between implicit and explicit solvent simulations and with experimental data to establish that the implicit solvent model results in a thermodynamically sufficiently accurate description of native state dynamics. Figures 5 and 6 show root-mean-square deviations (rmsd) as a function of time for both systems. Overall, the data demonstrate that the simulated structures remain very close to the experimental structure with both explicit and implicit solvent. The most notable exception is an excursion as far away as about 4 Å for about 5 ns in one of the implicit simulations of protein G (see Figure 5D). Other significant deviations from the experimental structure are observed around 22 ns with explicit solvent and at different times past 20 ns in the implicit solvent simulation with a Nosé−Hoover thermostat. On the basis of the time evolutions, the simulations were, somewhat arbitrarily, considered fully equilibrated after 30ns (protein G) and 10ns (ubiquitin), respectively. The protein G simulations from 30 to 50 ns and the ubiquitin simulations from

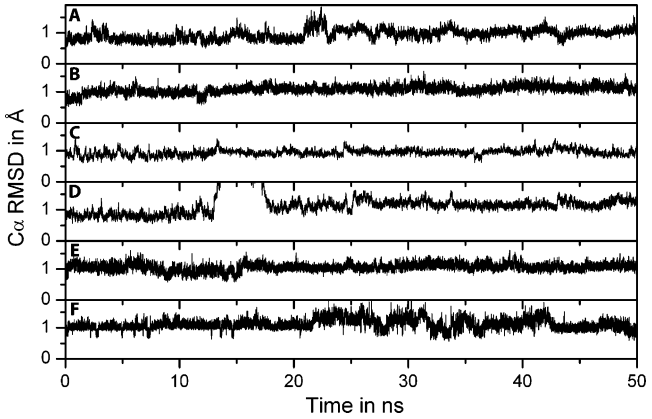

Figure 5. Cα root-mean-square deviations from NMR conformation (3GB1) during molecular dynamics simulations of protein G with explicit solvent (A) or implicit solvent: mixed friction, original CMAP (B); mixed friction, modified CMAP (C); $f = 50$ ps^{-1}, modified CMAP (D); $f = 5$ ps^{-1}, modified CMAP (E); Nosé−Hoover thermostat, modified CMAP (F).

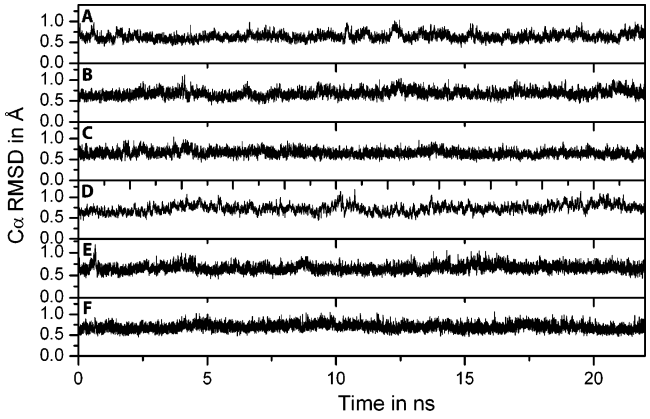

Figure 6. Cα root-mean-square deviations from NMR conformation (1D3Z) during molecular dynamics simulations of ubiquitin (residues 1−72) with explicit solvent (A) or implicit solvent: mixed friction, original CMAP (B); mixed friction, modified CMAP (C); $f = 50$ ps^{-1}, modified CMAP (D); $f = 5$ ps^{-1}, modified CMAP (E); Nosé−Hoover thermostat, modified CMAP (F).

10 to 22.5 ns were subjected to further analysis. Average root-mean-square deviations for Cα and Cβ atoms from both NMR and crystallographic structures are given in Table 6. Relatively few significant differences can be discerned suggesting that implicit and explicit solvent simulations essentially sample very similar, nativelike conformations.

A more rigorous comparison with experimental structures should be based on average structures from the simulations. The corresponding root-mean-square deviations shown in Table 7 are in fact smaller than the average root-mean-square deviations from Table 6, as close as 0.4 Å for ubiquitin Cα atoms and as far as 1 Å for protein G Cβ atoms. This excellent agreement is generally matched with implicit solvent.

Average dynamic structural fluctuations in protein G and ubiquitin can be compared to experimental B factors from crystallography[55] and S^2 order parameters from NMR spectroscopy.[56] B factors are related to atomic mean-square displacements σ_i under the assumption that fluctuations are isotropic according to $B_i = 8\pi^2\sigma_i/3$. The comparison of B factors from simulations and experiments is shown in Figures 7 and 8 for protein G and ubiquitin, respectively. B factors for potein G are available from two different crystal structures (1PGA[57] and 1PGB[57]). The two sets of B factors are substantially different, indicating the degree of uncertainty for this type of data. The data extracted from the explicit solvent simulation agree best with the B factors from 1PGA,[57] although the loop between the first two strands of the β sheet near residue 10 is significantly more flexible in the simulation. A closer look at the crystal structure reveals that these residues are involved in crystal packing interactions to a greater extent in 1PGA than in 1PGB. This corresponds with larger crystallographic B factors from 1PGB in this region. Significant differences from explicit solvent are found in most of the implicit solvent simulations. However, the implicit solvent simulations also do not fully agree with each other in particular with respect to the flexibility around residues 10, 20, and 40. B factors for ubiquitin are also available from two different sources (1UBQ[58] and 1UBI[59]). The differences between these two data sets are much smaller than for protein G. Furthermore, B factors from the explicit

Table 6. Time-Averaged Coordinate Root-Mean-Square Deviations in Å from Explicit and Implicit Solvent Simulations of Protein G and Ubiquitin in Comparison with NMR (3gb1, 1d3z) and X-ray Structures (1pga, 1ubq)[a]

		protein G (30-50ns)		ubiquitin (10−22.5 ns)	
		Cα	Cβ	Cα	Cβ
explicit solvent	NMR	1.01 (*0.11*)	1.20 (*0.13*)	0.65 (*0.09*)	0.77 (*0.10*)
	X-ray	0.83 (*0.11*)	1.00 (*0.14*)	0.68 (*0.09*)	0.79 (*0.09*)
GBMV, orig. CMAP mixed friction	NMR	1.13 (*0.10*)	1.37 (*0.11*)	0.70 (*0.09*)	0.82 (*0.10*)
	X-ray	0.97 (*0.13*)	1.18 (*0.14*)	0.69 (*0.09*)	0.79 (*0.09*)
GBMV, mod. CMAP mixed friction	NMR	1.25 (*0.14*)	1.63 (*0.24*)	0.64 (*0.07*)	0.76 (*0.08*)
	X-ray	1.10 (*0.16*)	1.45 (*0.26*)	0.66 (*0.07*)	0.77 (*0.07*)
GBMV, mod. CMAP $f = 50$ ps^{-1}	NMR	1.20 (*0.11*)	1.39 (*0.12*)	0.76 (*0.10*)	0.87 (*0.11*)
	X-ray	1.12 (*0.15*)	1.29 (*0.16*)	0.73 (*0.10*)	0.84 (*0.10*)
GBMV, mod. CMAP $f = 5$ ps^{-1}	NMR	1.09 (*0.10*)	1.31 (*0.11*)	0.67 (*0.08*)	0.77 (*0.08*)
	X-ray	0.90 (+*0.12*)	1.08 (*0.24*)	0.69 (*0.09*)	0.80 (*0.08*)
GBMV, mod. CMAP Nosé−Hoover	NMR	1.08 (*0.19*)	1.32 (*0.23*)	0.69 (*0.08*)	0.80 (*0.09*)
	X-ray	1.00 (*0.20*)	1.23 (*0.24*)	0.68 (*0.08*)	0.79 (*0.09*)

[a] Standard deviations are given in parentheses. Only residues 1−72 were considered for ubiquitin because of the highly flexible C terminus.

Table 7. Coordinate Root-Mean-Square Deviations in Å of Average Structures from Explicit and Implicit Solvent Simulations of Protein G and Ubiquitin in Comparison with NMR (3gb1, 1d3z) and X-ray Structures (1pga, 1ubq)[a]

		protein G (30-50ns)		ubiquitin (10−22.5 ns)	
		Cα	Cβ	Cα	Cβ
explicit solvent	NMR	0.83	1.00	0.41	0.49
	X-ray	0.60	0.74	0.44	0.52
GBMV, orig. CMAP mixed friction	NMR	0.97	1.16	0.47	0.53
	X-ray	0.77	0.94	0.45	0.49
GBMV, mod. CMAP mixed friction	NMR	1.10	1.42	0.43	0.51
	X-ray	0.92	1.22	0.46	0.52
GBMV, mod. CMAP $f = 50$ ps^{-1}	NMR	1.01	1.18	0.57	0.63
	X-ray	0.93	1.06	0.54	0.60
GBMV, mod. CMAP $f = 5$ ps^{-1}	NMR	0.94	1.12	0.45	0.51
	X-ray	0.71	0.84	0.48	0.54
GBMV, mod. CMAP Nosé−Hoover	NMR	0.83	0.99	0.46	0.52
	X-ray	0.72	0.87	0.45	0.49

[a] Only residues 1−72 were considered for ubiquitin because of the highly flexible C terminus.

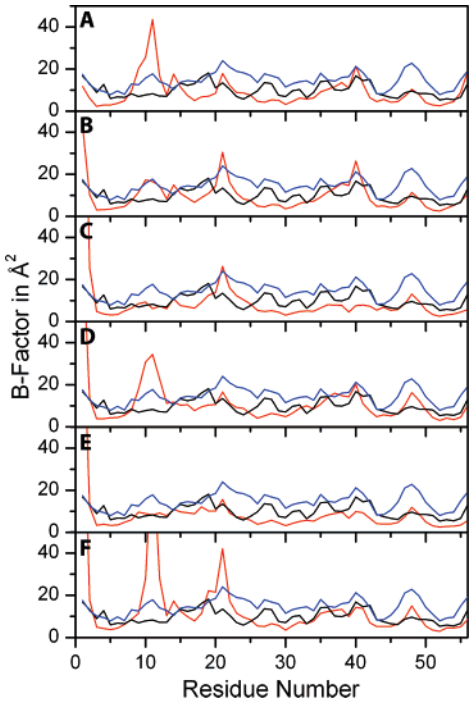

Figure 7. Cα B factors calculated from root-mean-square fluctuations during molecular dynamics simulations of protein G (30-50ns) with explicit solvent (A) or implicit solvent: mixed friction, original CMAP (B); mixed friction, modified CMAP (C); $f = 50$ ps^{-1}, modified CMAP (D); $f = 5$ ps^{-1}, modified CMAP (E); Nosé−Hoover thermostat, modified CMAP (F). Simulation results (red) are compared with crystallographic data from 1PGA (black) and 1PGB (blue).

solvent simulation of ubiquitin agree well with the experimental data. The flexibility between residues 50 and 65 is only slightly underestimated, and flexibility around residue 47 is overestimated. Contrary to the results for protein G, the implicit simulations of ubiquitin show no significant differences from the explicit solvent results.

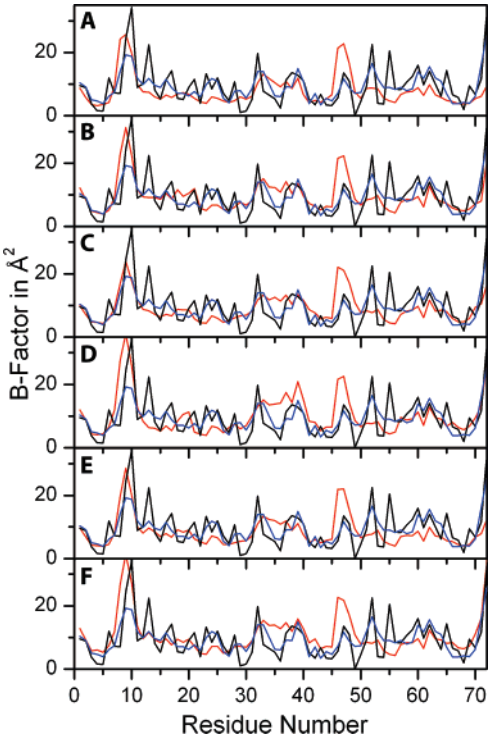

Figure 8. Cα B factors calculated from root-mean-square fluctuations during molecular dynamics simulations of ubiquitin (10−22.5ns) with explicit solvent (A) or implicit solvent: mixed friction, original CMAP (B); mixed friction, modified CMAP (C); $f = 50$ ps^{-1}, modified CMAP (D); $f = 5$ ps^{-1}, modified CMAP (E); Nosé−Hoover thermostat, modified CMAP (F). Simulation results (red) are compared with crystallographic data from 1UBI (black) and 1UBQ (blue).

Generalized S^2 order parameters were calculated for the dynamics of backbone N−H vectors $\mathbf{u}_i$ according to

$$S^2 = \frac{1}{2}[3\sum_{\alpha=1}^{3}\sum_{\beta=1}^{3}\langle \mathbf{u}_{i,\alpha}\mathbf{u}_{i,\beta}\rangle^2 - 1]$$

where α and β are the Cartesian coordinates of $\mathbf{u}_i$. The results from the simulations are compared with experimental data in Figures 9 and 10 for protein G and ubiquitin, respectively. The data from the explicit solvent simulation of protein G are in fairly good agreement with the experimental S^2 values according to the Lipari−Szabo analysis,[60,61] only slightly overestimating flexibility near residues 10, 17, 22, and 41 (see Figure 9A). The generally better agreement of the explicit solvent simulation with the NMR data than with the crystallographic data near residue 10 further supports the view that the reduced flexibility in this region observed in the crystals is an artifact of crystallization while this region is actually more flexible in solution. Again, larger deviations between implicit and explicit solvent are observed for protein G while the differences between simulated and experimental S^2 values for ubiquitin are negligible (see Figure 10).

Sampling Efficiency. Short-time sampling efficiency with different implicit solvent methods was examined in the context of protein simulations to examine kinetic properties due to differences between thermostats and Langevin friction coefficients. Figures 11 and 12 show how the rmsd difference

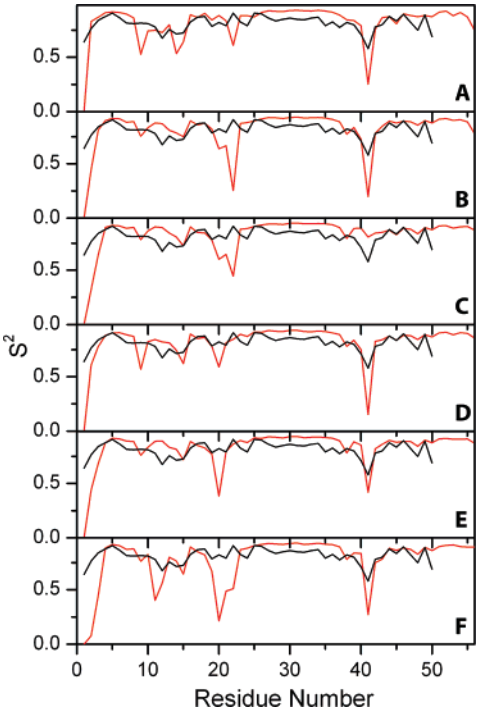

Figure 9. Backbone N−H order parameters S^2 during molecular dynamics simulations of protein G (30−50 ns) with explicit solvent (A) or implicit solvent: mixed friction, original CMAP (B); mixed friction, modified CMAP (C); $f = 50$ ps^{-1}, modified CMAP (D); $f = 5$ ps^{-1}, modified CMAP (E); Nosé−Hoover thermostat, modified CMAP (F). Simulation results (red) are compared with data from NMR spectroscopy (black).[70,71]

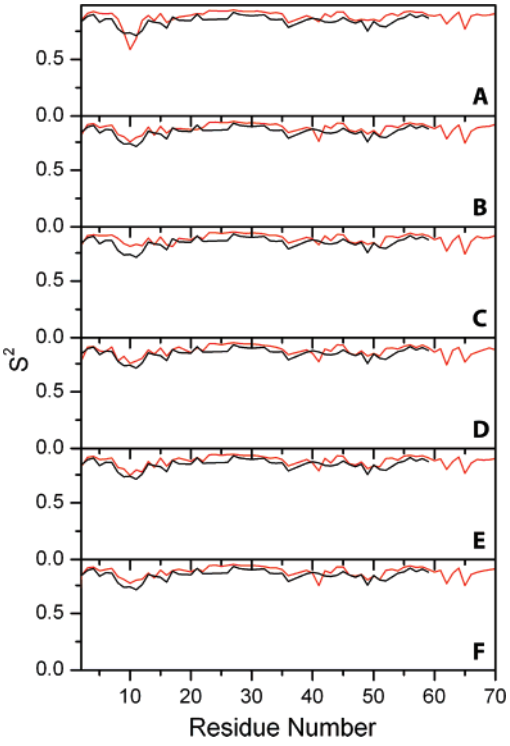

Figure 10. Backbone N−H order parameters S^2 during molecular dynamics simulations of ubiquitin (10−22.5 ns) with explicit solvent (A) or implicit solvent: mixed friction, original CMAP (B); mixed friction, modified CMAP (C); $f = 50$ ps^{-1}, modified CMAP (D); $f = 5$ ps^{-1}, modified CMAP (E); Nosé−Hoover thermostat, modified CMAP (F). Simulation results (red) are compared with data from NMR spectroscopy (black).[72]

between simulated conformations at times t and $t + \delta t$ varies as a function of δt in the simulations of protein G and ubiquitin, respectively. A rapid increase in rmsd as a function of the time interval means that less time is spent in a given basin and that conformational sampling of different conformations is more efficient. A slow rise of rmsd, on the other hand, means that it takes a long time before a given conformational basin is left and that conformational sampling is inefficient. Similar conclusions are found for both, protein G and ubiquitin. Conformational sampling with implicit solvent is less efficient than with explicit solvent only if Langevin dynamics with a friction coefficient of 50 ps^{-1} are employed. When a friction coefficient of 5 ps^{-1} is used, sampling becomes more efficient by about a factor of 4−5 when comparing how much time it takes with implicit solvent to reach the average rmsd at $\delta t = 50$ ps with explicit solvent. The sampling appears to be much more efficient when a Nosé−Hoover thermostat is used instead of Langevin dynamics. In that case, conformations diverge rapidly after a very short time. A curious oscillatory behavior is seen, where conformations initially diverge further but return slightly after about 5 ps before diverging again.

Discussion

This study was carried out in order to examine in detail how the application of implicit solvent affects kinetic properties in simulations of peptides and proteins. Because very long simulations are needed to provide meaningful comparisons, only a limited number of systems and only one implicit solvent model, the GBMV method, with different viscosity parameters could be tested. Nevertheless, interesting observations could be made that are expected to be relevant at least qualitatively in a more general context for other systems and implicit solvent methods.

The first part of this study focuses on kinetic transitions in the well-studied alanine dipeptide system. It is found that implicit solvent combined with Langevin dynamics can reproduce the kinetic behavior seen in explicit solvent simulations quite well when a suitable friction coefficient of around 10 ps^{-1} is chosen. Physical insight suggests that different friction coefficients should be used for different atom types rather than a single friction coefficient for all non-hydrogen atoms.[34] Results from this study provide some evidence that different friction coefficients may provide results closer to those for the explicit solvent, but it is unclear whether the rather modest improvements seen here would warrant substantial parametrization efforts for all protein side chains.

Not surprisingly, the agreement between implicit and explicit solvent is quite sensitive to the underlying free energy surface. Small deviations in the conformational preferences between the GBMV model and explicit solvent are manifest in significant differences in some of the kinetic rates. Sampling differences in the alanine dipeptide system are

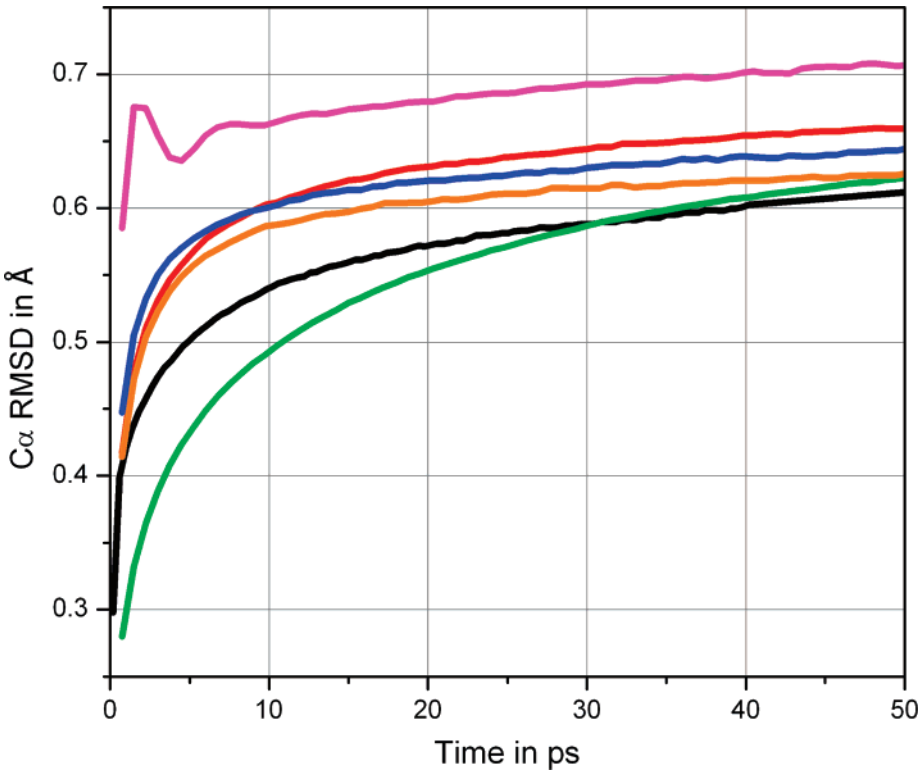

Figure 11. Average Cα rmsd between structures at times t and $t + \Delta t$ during simulations of protein G (30−50 ns) with explicit solvent (black) or implicit solvent: mixed friction, original CMAP (red); mixed friction, modified CMAP (orange); $f = 50$ ps^{-1}, modified CMAP (green); $f = 5$ ps^{-1}, modified CMAP (blue); Nosé−Hoover thermostat, modified CMAP (magenta).

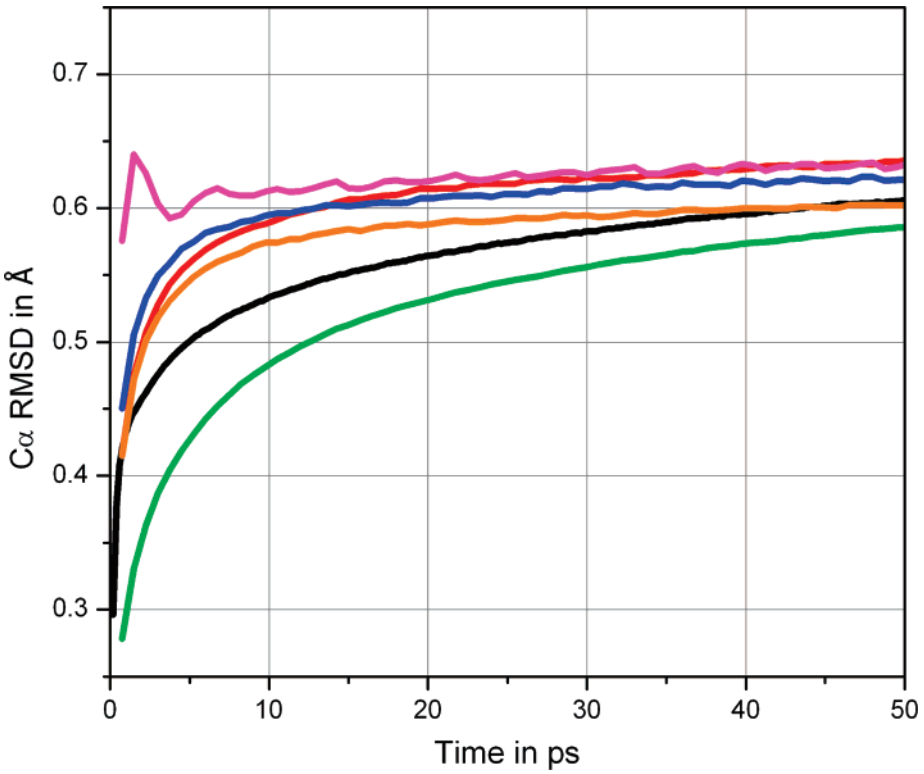

Figure 12. Average Cα rmsd between structures at times t and $t + \Delta t$ during simulations of ubiquitin (10−22.5 ns, residues 1−72) with explicit solvent (black) or implicit solvent: mixed friction, original CMAP (red); mixed friction, modified CMAP (orange); $f = 50$ ps^{-1}, modified CMAP (green); $f = 5$ ps^{-1}, modified CMAP (blue); Nosé−Hoover thermostat, modified CMAP (magenta).

easily remedied by using a modified CMAP torsion potential that accounts for the difference between the implicit and explicit solvent energetics. Previous work has also suggested the use of a modified force field for implicit solvent.[40] However, on the basis of the data presented here, it is impossible to tell whether the modified CMAP potential that

improves sampling of the alanine dipeptide surface would also be able to compensate implicit solvent inadequacies in a wider range of systems, and it is noted that application of the modified CMAP term had little effect on the sampling of protein G or ubiquitin compared to the original CMAP.

A number of previous studies have examined how viscosity affects the barrier crossing kinetic rates in isomerization transitions[62] and during protein folding.[23,63−65] An interesting conclusion from some of these studies is the identification of two regimes where either solute−solute or solute−solvent friction is dominant. Similar conclusions can also be made from the data presented here: Small friction coefficients accelerate alanine dipeptide kinetics by a relatively modest amount in the solute−solute friction-dominated regime.[23,63] On the other hand, the kinetic rates are affected much more dramatically with larger friction coefficients in the regime where solvent−solute friction dominates. The ability to speed up kinetics through the use of implicit solvent depends on the characteristics of the solute−solute friction-dominated regime. On the basis of the results presented here, barrier crossings can be accelerated only by a maximum factor of about 2 with the GBMV method. This finding may be compared with results from Pande and Zagrovic[23] that suggest that protein folding kinetics could be accelerated up to about a factor of 20 with low friction coefficients when a different, thermodynamically less-accurate[17,39] generalized Born model is used.

A comparison of how individual kinetic rates are affected at reduced solvent viscosity indicates that different barrier crossings may not be accelerated uniformly. This finding could have implications for transition pathways extracted from such simulations. A recent study of protein folding that looked for such an effect did not find evidence that different pathways may be sampled as a function of solvent viscosity.[65] However, it is possible that folding funnels are sufficiently robust to accommodate subtle changes in individual barrier crossing rates. Further exploration of this aspect is needed to determine more conclusively to what extent reduced friction implicit solvent simulations might affect sampled kinetic pathways.

Temperature control of the solute with a Nosé−Hoover thermostat without any account of solvent viscosity greatly diminishes the ability to cross barriers in the alanine dipeptide system. This can be understood from a complete absence of stochastic collisions with the solvent that typically provide the kinetic energy for crossing a given barrier. From a statistical mechanics point of view, the coupling of only the solute with a Nosé−Hoover thermostat in the absence of explicit solvent also creates a different ensemble than a Nosé−Hoover thermostat applied to a solute plus a surrounding box of explicit water. In the latter case, much larger temperature fluctuations of the solute itself are allowed as kinetic energy is free to transfer between the solute and solvent through atomic collisions. Temperature control of the solute alone suppresses most of these temperature fluctuations in the implicit solvent systems. The result is a dramatic slowdown of the barrier crossing rates.

The kinetic rates and mean first passage times obtained in this study are also interesting in themselves, as there are only very few studies that explore alanine dipeptide kinetics on 100 ns time scales. The results can be compared to other theoretical studies where kinetic rates were obtained with Brownian dynamics methods on an implicit solvent surface[66] or through the extraction of state-to-state transition functions.[67,68] Although differences due to the choice of other force fields are expected, the agreement is reasonable. In the study of Chekmarev et al.,[66] mean first passage times between α and β basins are estimated to be 29 and 293 ps for forward and backward transitions, respectively, compared to 426 and 459 ps with explicit solvent and about half that with implicit solvent and a friction coefficient of 10 ps^{-1} in this study. Transitions between α and $C7_{ax}$, on the other hand, are found to take about 11 ns and 224 ps in the forward and backward directions compared with 9 ns and 744 ps with explicit solvent and 12 ns and 534 ps with implicit solvent in this study, respectively. The simulations presented here also support the preference for indirect transitions between $C7_{ax}$ and α that follow a path through α_L and β rather than a direct transition between $C7_{ax}$ and α due to the significant barrier height near $\phi = 0$ and $\psi = -100$.[66]

The detailed analysis of kinetic rates in alanine dipeptide is contrasted with native state dynamics of two well-studied proteins: protein G and ubiquitin. Essentially, it is found that implicit solvent simulations closely reproduce results from explicit solvent and experiments. To the extent that differences exist, in particular in the protein G simulations, they likely indicate incomplete convergence of conformational sampling of a more flexible system compared to ubiquitin. The choice of different thermostats and Langevin friction coefficients could affect convergence rates of average structural and dynamic properties. However, no systematic differences can be found from the simulations presented here, suggesting that the long-time convergence of native state dynamics in simulations over tens of nanoseconds is not significantly affected by the thermostats tested here.

In contrast, differences in the exploration of conformational space during tens of picoseconds are clearly present. Conformational sampling in this time regime can be more efficient by a factor of 4−5 when a reduced friction coefficient is used in Langevin dynamics. With a Nosé−Hoover thermostat, the decorrelation of conformational sampling is even more rapid. This can be understood from the complete lack of friction and a continuous adjustment of the atomic velocities in the Nosé−Hoover method to maintain coupling with the thermal bath. As a result, it is expected that sampling can be accelerated significantly with the Nosé−Hoover thermostat as long as no significant kinetic barriers are present. This is the presumed case of conformational dynamics within the native basin of most macromolecules. However, if conformational changes over kinetic barriers that are large compared to kT are involved, such as in loop rearrangements or protein folding/unfolding, it is expected that significant differences as a function of the thermostat are manifest in a similar way as in the alanine dipeptide system. The practical question of which thermostat would be most suitable for a given system then depends on the particular application and expected dynamic properties. While the present study provides the first insight into the

effect of different thermostats on the sampling of biological systems with implicit solvent, further studies are clearly needed to understand the effect of such a methodology on the sampling of larger conformational changes in biological macromolecules where implicit solvent is expected to be most beneficial.

Previous studies have indicated that different implicit solvent methods may significantly contribute to the solute−solute friction.[28] While only one implicit solvent method, the GBMV method, was considered here, it is likely that the general conclusions would apply in a similar fashion for other implicit solvent methods as well. However, there is some evidence that the optimal choice of solvent friction coefficients may be different in those cases.[28] Further studies will be needed to address this point more systematically.

Summary

This paper compares the kinetic properties of biomolecules in simulations with implicit and explicit solvent. The first important conclusion is that it is possible to closely match kinetic properties from explicit solvent by combining an implicit solvent with Langevin dynamics. Friction coefficients near 10 ps^{-1} appear to be optimal in conjunction with the GBMV implicit solvent model studied here. The second conclusion is that conformational sampling can be accelerated with reduced friction coefficients although the degree of acceleration depends on the circumstances. Local conformational exploration without the crossing of significant barriers can be accelerated substantially on the basis of native state simulations of proteins. However, when the crossing of significant kinetic barriers is involved, solvent viscosity appears to play a less important role. Consequently, reduced friction coefficients affect the kinetics only to a limited extent in agreement with previous studies of protein folding kinetics.[23] The use of Nosé−Hoover thermostats instead of Langevin dynamics in conjunction with an implicit solvent has quite dramatic consequences. Local conformational sampling is greatly enhanced over Langevin dynamics because of the lack of friction. However, the crossing of barriers is slowed down by at least an order of magnitude because of tight solute temperature control and the lack of stochastic collisions to provide the necessary kinetic energy to overcome barriers. For practical applications that may involve both the exploration of local conformational minima and the crossing of significant barriers, the results from this study indicate that Langevin dynamics with small friction coefficients of less than 5 ps^{-1} may offer substantially improved sampling over explicit solvent simulations while maintaining a thermodynamically accurate description of the simulated system.

Acknowledgment. Financial support from NSF CAREER grant 0447799 and the Alfred P. Sloan Foundation and access to computational resources through the Michigan Center for Biological Information at Michigan State University is acknowledged.

References

(1) Feig, M.; Brooks, C. L., III. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217−224.

(2) Baker, N. A. Improving implicit solvent simulations: a Poisson-centric view. *Curr. Opin. Struct. Biol.* **2005**, *15* (2), 137−143.

(3) Roux, B.; Simonson, T. Implicit Solvent, Models. *Biophys. Chem.* **1999**, *78*, 1−20.

(4) Cramer, C. J.; Truhlar, D. G. Implicit solvation models: Equilibria, structure spectra and dynamics. *Chem. Rev.* **1999**, *99*, 2161−2200.

(5) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(6) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144−1149.

(7) Sharp, K. A.; Honig, B. Electrostatic Interactions in Macromolecules - Theory and Applications. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301−332.

(8) Warwicker, J.; Watson, H. C. Calculation of the Electric Potentialin the Active Site Cleft due to α-Helix Dipoles. *J. Mol. Biol.* **1982**, *157*, 671−679.

(9) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.* **2002**, *15*, 377−392.

(10) Im, W.; Beglov, D.; Roux, B. Continuum Solvation Model: Computation of Electrostatic Forces from Numerical Solutions to the Poisson-Boltzmann Equation. *Comput. Phys. Commun.* **1998**, *111*, 59−75.

(11) Friedrichs, M.; Zhou, R. H.; Edinger, S. R.; Friesner, R. A. Poisson-Boltzmann analytical gradients for molecular modeling calculations. *J. Phys. Chem. B* **1999**, *103*, 3057−3061.

(12) Lu, B. Z.; Chen, W. Z.; Wang, C. X.; Xu, X.-j. Protein Molecular Dynamics With Electrostatic Force Entirely Determined by a Single Poisson-Boltzmann Calculation. *Proteins* **2002**, *48*, 497−504.

(13) Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann Calculations for Static and Dynamic Systems. *J. Comput. Chem.* **2002**, *23*, 1244−1253.

(14) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(15) Yu, Z. Y.; Jacobson, M. P.; Friesner, R. A. What role do surfaces play in GB models? A new-generation of surface-generalized Born model based on a novel Gaussian surface for biomolecules. *J. Comput. Chem.* **2006**, *27* (1), 72−89.

(16) 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* (2), 383−394.

(17) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A. Brooks, C. L., III. Performance Comparison of Generalized Born and Poisson Methods in the Calculation of Electrostatic Solvation Energies for Protein Structures. *J. Comput. Chem.* **2004**, *25*, 265−284.

(18) Grycuk, T. Deficiency of the Coulomb-field approximation in the generalized Born model: An improved formula for Born radii evaluation. *J. Chem. Phys.* **2003**, *119*, 4817−4826.

(19) Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2003**, *25*, 238−250.

(20) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. Novel generalized Born methods. *J. Chem. Phys.* **2002**, *116* (24), 10606−10614.

(21) Gallicchio, E.; Levy, R. M. AGBNP: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-Resolution Modeling. *J. Comput. Chem.* **2004**, *25*, 479−499.

(22) Brooks, C. L.; Berkowitz, M.; Adelman, S. A. Generalized Langevin Theory for Many-Body Problems in Chemical-Dynamics - Gas-Surface Collisions Vibrational-Energy Relaxation in Solids and Recombination Reactions in Liquids. *J. Chem. Phys.* **1980**, *73* (9), 4353−4364.

(23) Zagrovic, B.; Pande, V. Solvent Viscosity Dependence of the Folding Rate of a Small Protein: Distributed Computing Study. *J. Comput. Chem.* **2003**, *24*, 1432−1436.

(24) Hoover, W. G.; Aoki, K.; Hoover, C. G.; De Groot, S. V. Time-reversible deterministic thermostats. *Physica D* **2004**, *187* (1−4), 253−267.

(25) Nose, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255−268.

(26) Hoover, W. G. Constant-Pressure Equations of Motion. *Phys. Rev. A: At., Mol., Opt. Phys.* **1986**, *34* (3), 2499−2500.

(27) Zagrovic, B.; Sorin, E. J.; Pande, V. β-Hairpin Folding Simulations in Atomistic Detail Using an Implicit Solvent Model. *J. Mol. Biol.* **2001**, *313*, 151−169.

(28) Fan, H.; Mark, A. E.; Zhu, J.; Honig, B. Comparative study of generalized Born models: Protein dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (19), 6760−6764.

(29) David, L.; Luo, R.; Gilson, M. K. Comparison of Generalized Born and Poisson Models: Energetics and Dynamics of HIV Protease. *J. Comput. Chem.* **2000**, *21*, 295−309.

(30) Tsui, V.; Case, D. A. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.

(31) Calimet, N.; Schaefer, M.; Simonson, T. Protein Molecular Dynamics With the Generalized Born/ACE Solvent Model. *Proteins* **2001**, *45*, 144−158.

(32) Ferrara, P.; Apostolakis, J.; Caflish, A. Evaluation of a Fast Implicit Solvent Model for Molecular Dynamics Simulations. *Proteins* **2002**, *46*, 24−33.

(33) Suenaga, A. Replica-exchange, molecular dynamics simulations for a small-sized protein folding with implicit solvent. *THEOCHEM* **2003**, *634*, 235−241.

(34) Shen, M.-y.; Freed, K. F. Long Time Dynamics of Met-Enkephalin: Comparison of Explicit and Implicit Solvent Models. *Biophys. J.* **2002**, *82*, 1791−1808.

(35) Krol, M. Comparison of Various Implicit Solvent Models in Molecular Dynamics Simulations of Immunoglobulin G Light Chain Dimer. *J. Comput. Chem.* **2003**, *24*, 531−546.

(36) Bursulaya, B. D.; Brooks, C. L., III. Comparative Study of the Folding Free Energy Landscape of a Three-Stranded β-Sheet Protein with Explicit and Implicit Solvent Models. *J. Phys. Chem. B* **2000**, *104* (51), 12378−12383.

(37) Chocholousova, J.; Feig, M. Implicit solvent simulations of DNA and DNA-protein complexes: Agreement with explicit solvent vs experiment. *J. Phys. Chem. B* **2006**, *110* (34), 17240−17251.

(38) Chocholousova, J.; Feig, M. Balancing an Accurate Representation of the Molecular Surface in Generalized Born Formalisms with Integrator Stability in Molecular Dynamics Simulations. *J. Comput. Chem.* **2006**, *27*, 719−729.

(39) Zhu, J.; Alexov, E.; Honig, B. Comparativestudy of generalized Born models: Born radii and peptide folding. *J. Phys. Chem. B* **2005**, *109* (7), 3008−3022.

(40) Chen, J. H.; Im, W. P.; Brooks, C. L. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *J. Am. Chem. Soc.* **2006**, *128* (11), 3728−3736.

(41) Geney, R.; Layten, M.; Gomperts, R.; Hornak, V.; Simmerling, C. Investigation of salt bridge stability in a generalized born solvent model. *J. Chem. Theory Comput.* **2006**, *2* (1), 115−127.

(42) Zhu, J.; Shi, Y.; Liu, H. Parametrization of a Generalized Born/Solvent-Accessible Surface Area Model and Applications to the Simulation of Protein Dynamics. *J. Phys. Chem. B* **2002**, *106*, 4844−4853.

(43) 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.

(44) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. Improving the Packing and Accuracy of NMR Structures with a Pseudopotential for the Radius of Gyration. *J. Am. Chem. Soc.* **1999**, *121*, 2337−2338.

(45) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837.

(46) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, J. D.; Evanseck, M. J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(47) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(48) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.

(49) 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* (3), 327−341.

(50) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. New Analytical Approximation to the Standard Molecular Volume Definition And Its Application to Generalized BornCalculations. *J. Comput. Chem.* **2003**, *24*, 1348−1356.

(51) Im, W.; Lee, M. S.; Brooks, C. L., III. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* **2003**, *24*, 1691−1702.

(52) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy Minimization and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(53) Feig, M.; Karanicolas, J.; Brooks, C. L., III. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graphics Modell.* **2004**, *22*, 377−395.

(54) Feig, M.; MacKerell, A. D.; Brooks, C. L. Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations. *J. Phys. Chem. B* **2003**, *107* (12), 2831−2836.

(55) Drenth, J. *Principles of Protein X-ray Crystallography*; Springer: New York, 1994; pp 89−91.

(56) Henry, E. R.; Szabo, A. Influence of Vibrational Motion on Solid-State Line-Shapes and Nmr Relaxation. *J. Chem. Phys.* **1985**, *82* (11), 4753−4761.

(57) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. 2 Crystal-Structures of the B1 Immunoglobulin-Binding Domain of Streptococcal Protein-G and Comparison with Nmr. *Biochemistry* **1994**, *33* (15), 4721−4729.

(58) Vijaykumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8 a Resolution. *J. Mol. Biol.* **1987**, *194* (3), 531−544.

(59) Alexeev, D.; Bury, S. M.; Turner, M. A.; Ogunjobi, O. M.; Muir, T. W.; Ramage, R.; Sawyer, L. Synthetic Structural and Biological Studies of the Ubiquitin System - Chemically Synthesized and Native Ubiquitin Fold into Identical 3-Dimensional Structures. *Biochem. J.* **1994**, *299*, 159−163.

(60) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, *104* (17), 4546−4559.

(61) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **1982**, *104* (17), 4559−4570.

(62) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides - the Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N′-Methylamide. *Biopolymers* **1992**, *32* (5), 523−535.

(63) Hagen, S. J.; Qiu, L. L.; Pabit, S. A. Diffusional limits to the speed of protein folding: fact or friction? *J. Phys.: Condens. Matter.* **2005**, *17* (18), S1503−S1514.

(64) Gee, P. J.; van Gunsteren, W. F. Numerical simulation of the effect of solvent viscosity on the motions of a beta-peptide heptamer. *Chem.—Eur. J.* **2005**, *12* (1), 72−75.

(65) Jagielska, A.; Scheraga, H. A. Influence of Temperature Friction and Random Forces on Folding of the B-Domain of Staphylococcal Protein A: All-Atom Molecular Dynamics in Implicit Solvent. *J. Comput. Chem.* **2007**, *28*, 1068−1082.

(66) Chekmarev, D. S.; Ishida, T.; Levy, R. M. Long-time conformational transitions of alanine dipeptide in aqueous solution: Continuous, and discrete-state kinetic models. *J. Phys. Chem. B* **2004**, *108* (50), 19487−19495.

(67) Swope, W. C.; Pitera, J. W.; Suits, F. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B* **2004**, *108* (21), 6571−6581.

(68) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and beta-hairpin peptide. *J. Phys. Chem. B* **2004**, *108* (21), 6582−6594.

(69) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33.

(70) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.* **1999**, *121* (10), 2337−2338.

(71) Idiyatullin, D.; Daragan, V. A.; Mayo, K. H. (NH)-N-15 backbone dynamics of protein GB1: Comparison of order parameters and correlation times derived using various "model-free" approaches. *J. Phys. Chem. B* **2003**, *107* (11), 2602−2609.

(72) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **1998**, *120* (27), 6836−6837.

CT7000705