Comparative Study of Generalized Born Models: Born Radii and Peptide Folding

Jiang Zhu, Emil Alexov, and Barry Honig*

Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168th Street, New York, New York 10032

Received: August 16, 2004; In Final Form: November 16, 2004

In this study, we have implemented four analytical generalized Born (GB) models and investigated their performance in conjunction with the GROMOS96 force field. The four models include that of Still and co-workers, the HCT model of Cramer, Truhlar, and co-workers, a modified form of the AGB model of Levy and co-workers, and the GBMV2 model of Brooks and co-workers. The models were coded independently and implemented in the GROMOS software package and in TINKER. They were compared in terms of their ability to reproduce the results of Poisson−Boltzmann (PB) calculations and in their performance in the ab initio peptide folding of two peptides, one that forms a β-hairpin in solution and one that forms an α-helix. In agreement with previous work, the GBMV2 model is most successful in reproducing PB results while the other models tend to underestimate the effective Born radii of buried atoms. In contrast, stochastic dynamics simulations on the folding of the two peptides, the C-terminus β-hairpin of the B1 domain of protein G and the alanine-based α-helical peptide 3K(I), suggest that the simpler GB models are more effective in sampling conformational space. Indeed, the Still model used in conjunction with the GROMOS96 force field is able to fold the hairpin peptide to a native-like structure without the benefit of enhanced sampling techniques. This is due in part to the properties of the united-atom GROMOS96 force field which appears to be more flexible, and hence to sample more efficiently, than force fields such as OPLSAA. Our results suggest a general strategy which involves using different combinations of force fields and solvent models in different applications, for example, using GROMOS96 and a simple GB model in sampling and OPLSAA and a more accurate GB model in refinement. The fact that various methods have been implemented in a unified way should facilitate the testing and subsequent use of different methods to evaluate conformational free energies in different applications. Our results also bear on some general issues involved in peptide folding and structure prediction which are addressed in the Discussion.

Introduction

Continuum solvent models are widely used to describe the properties of biological macromolecules.[1] Although many of the early applications were aimed at describing the effect of electrostatic potentials on biological function, there has been increasing emphasis on the use of continuum models in the calculation of the relative conformational free energies.[2−6] These are often obtained through the use of a thermodynamic cycle involving the calculation of gas phase energy differences and differences in the solvation free energies, ΔG_{sol}, between the two conformations.[7,8] The solvation free energy is partitioned into an electrostatic polarization term, ΔG_{pol}, and a nonpolar contribution, ΔG_{np}.[7,8] ΔG_{pol} describes the interaction of charges on the solute with the polarization they induce in the solvent. ΔG_{pol} accounts both for the screening of Coulombic interaction between charges on the solute and for the direct interaction of these charges with the solvent.

ΔG_{np} is the free energy associated with transferring a nonpolar solute with the exact shape of the actual solute molecule from the gas phase to water. It is often written in the form $\Delta G_{np} = \Sigma \sigma_i A_i$, where σ_i is an empirically derived coefficient and A_i is the solvent accessible surface area of atom i. The total solvation free energy can then be written as

$$\Delta G_{sol} = \Delta G_{pol} + \Delta G_{np} = \Delta G_{pol} + \sum_i \sigma_i A_i \qquad (1)$$

In a rigorous continuum model, ΔG_{pol} is calculated by solving the Poisson−Boltzmann (PB) equation. For the case of zero salt, the PB equation reduces to the Poisson equation:

$$\nabla \cdot \epsilon(r) \nabla \phi(r) + 4\pi\rho(r) = 0 \qquad (2)$$

where ϕ is the electrostatic potential in units of kT/q, ϵ is the dielectric constant, and ρ is the charge density. The variables ϕ, ϵ, and ρ are functions of the position vector r. In most applications, the PB equation is solved numerically using a variety of programs that are available.[9−14]

The PB equation provides electrostatic potentials and free energies for a single molecular conformation. Although electrostatic forces on each atom can be calculated, these are not easily incorporated into molecular mechanics algorithms, in part because these forces cannot be simply described in terms of pairwise interactions between solute atoms. Moreover, the time required for a well-converged PB solution is generally significantly longer than the time required for a single calculation of the molecular mechanics energy. A number of strategies have been developed to deal with these problems. It is possible to calculate electrostatic forces directly on individual atoms and use these forces in molecular mechanics algorithms.[12,15−17] With the advent of a new generation of fast PB solvers (DelPhi,[10,11]

* Corresponding author. Phone: 212-305-7970. Fax: 212-305-6926. E-mail: bh6@columbia.edu.

10.1021/jp046307s CCC: $30.25
Published on Web 01/21/2005

ZAP[13]) that can accurately calculate the electrostatic free energy of a medium sized protein in less than a second, the computational cost of incorporating PB methods into molecular mechanics calculations may no longer be prohibitive.[16,17]

An alternative approach is to attempt to capture the physics of the PB equation in a form that is adaptable to molecular mechanics methods.[18] The generalized Born (GB) approximation was developed for just this purpose.[19−26] The GB model provides a method for calculating ΔG_{pol}.

$$\Delta G_{pol}(GB) = -166.0\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{N}\frac{q_i^2}{\alpha_i} - 166.0\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{q_i q_j}{(r_{ij}^2 + \alpha_{ij}^2 e^{-D_{ij}})^{0.5}} \quad (3)$$

where $\alpha_{ij} = (\alpha_i\alpha_j)^{0.5}$ and $D_{ij} = r_{ij}^2/(2\alpha_{ij})^2$, r_{ij} is the distance between atoms i and j that have atomic charges q_i and q_j, respectively, ϵ is the dielectric constant of the solvent, and α_i is the so-called effective Born radius of atom i. The first term in eq 3 accounts for the interaction of every charge on the solute with the polarization charge it induces. If the polarization free energy was obtained from the PB equation, α_i would be given by

$$\alpha_i = -\frac{166.0}{\Delta G'_{pol,i}}\left(1 - \frac{1}{\epsilon}\right) \quad (4)$$

where $\Delta G'_{pol,i}$ is the polarization free energy of the charge on atom i when all other charges on the solute are set to zero. It can be calculated from an integral that extends from the molecular surface to infinity. For an isolated atom, $\alpha_i = R_i$, where R_i is the "Born radius" (which usually is quite close to the atomic radius of that atom).[27] A dielectric offset is often used in GB models to correct atomic or van der Waals radii (see the Materials and Methods section). The second term in eq 3 accounts for the solvent screening of the Coulombic interaction between the charges on atoms i and j. The screening effect of a high dielectric solvent on two charges located in a low dielectric region is accounted for by increasing the effective distance between the two charges by an amount related to their effective Born radii (the denominator in the second term on the right-hand side of eq 3). The ability of GB methods to approximate PB results is critically sensitive on the accuracy of the effective Born radii that are used.[28]

Accurate effective Born radii can be obtained from numerical solutions to the PB equation, but of course, the goal of GB methods is to avoid PB calculations. A number of fast analytical methods have been developed to estimate α_i.[20−26] The underlying idea is that, in the Coulomb field approximation (CFA) upon which GB models are based, the loss of polarization free energy of a charged atom in a solute molecule relative to the isolated atom can be treated as a volume integral of $1/r^4$ over the region of the solvent excluded by all other solute atoms.[18] The effective Born radius of a solute atom i can then be written in terms of an integration over the interior region of the solute molecule, excluding the radius R_i around atom i:[29]

$$\alpha_i^{-1} = R_i^{-1} - \frac{1}{4\pi}\int_{in, r>R_i}\frac{1}{r^4}\,dV \quad (5)$$

Different GB models differ in large part in the method used to evaluate this volume integral (see the discussion by Bashford and Case).[29] The analytical GBSA model of Still and co-workers, the HCT model,[22,30] the ACE model,[21] and the AGBNP model[26] are all based on sums over pairs of atoms where eq 5 is written in the form

$$\alpha_i^{-1} = R_i^{-1} - \frac{1}{4\pi}\sum_j\int_{sphere\ j}\frac{1}{r^4}\,dV \quad (6)$$

There are a number of known shortcomings associated with most GB models. These include intrinsic errors in the CFA and the effects of replacing eq 5 with a discrete sum over atomic volumes as in eq 6. The latter problem includes difficulties in accounting for the effects of overlapping volumes and the treatment of the crevices between atoms. PB methods generally use the molecular surface to define the boundary between the high dielectric solvent and the solute molecule, since, for a given conformation, this surface defines the molecular interior where solvent molecules are excluded.[1] The use of atomic (van der Waals) radii in the pairwise sums in GB models produces a macromolecule that is essentially soaked with water and thus produces an unrealistic physical model for a given solute conformation.[31] This also has the effect of yielding effective Born radii for buried atoms that are too small. The ABGNP model of Levy and co-workers[26] deals with some of the problems associated with the use of pairwise sums, while the GBMV2 model[24,25] of Brooks and co-workers attempts to reduce the errors resulting from the use of both the CFA and pairwise sums.

The goals of this and a second paper in this series (Fan, Mark, Zhu, and Honig, in preparation) are to explore the properties of different GB models in terms of their performance in different applications. Our specific interest is to use continuum solvent models in protein structure prediction and refinement, but in order to do so, we would like to understand the features of different solvent models and force fields. It is well-known that there are differences among them that can lead to different results in some cases,[32−38] especially when an effect of interest results from differences between two large terms. For example, the free energy of salt-bridge formation is a balance between large Coulomb and solvation forces and there are still questions about how to best describe this apparently straightforward electrostatic interaction.[39,40] Our interest in the current work is primarily sampling efficiency, which is an important factor in the ability to carry out structural refinement of protein models, for example, those built by homology to other proteins. We are also interested in determining how well GB models reproduce the predictions of the PB equation so as to have a better sense of the effects of the various approximations they entail.

To carry out as fair a comparison as possible, we have implemented different GB models using a single force field and software package. Specifically, the Still model,[20] the HCT model,[22] a modified version of the AGB model[26] (mAGB), and the latest GBMV2 model[25] have all been implemented in the GROMOS96[41] and GROMACS[42,43] packages (except for GBMV2 which has only been implemented in GROMOS96). They are all used in conjunction with the GROMOS96 force field.[41] We also implemented the mAGB model and the GBMV2 model into TINKER[44] so that they could be used with other force fields (OPLSAA[45−49] in our case). We show below that our implementation produces results that are consistent with those reported by the original authors of the different methods.

The comparisons of the GB models used in this paper include their ability to reproduce Born radii derived from the PB equation and their effect on the ab initio folding of two peptides. The effect of different GB models on protein dynamics is

considered in a second paper. The rationale for evaluating Born radii is that a major difference among GB models is in the methodology used to calculate this parameter, which in turn is a measure of their degree of agreement with standard PB methods. Our focus on ab initio peptide folding is predicated in part on the intrinsic interest in this problem. In addition, we believe that what we learn can be used in various aspects of protein structure prediction, as will be discussed below.

We chose to study two 16-residue peptides: the C-terminus β-hairpin of the B1 domain of protein G and the alanine-based α-helical peptide 3K(I). The β-hairpin has been a focus of extensive experimental[50−53] and theoretical studies[54−64] that are discussed further below. 3K(I) is an alanine-based peptide that is primarily α-helical in solution.[65] Our results indicate that there are characteristic differences in the ability of different GB models to simulate the folding of the two peptides. More generally, our results demonstrate that different GB models and force fields have distinct properties that render them more or less effective in different applications. This issue is considered extensively in the Discussion.

Materials and Methods

Calculating Born Radii. DelPhi V.4[10,11] was used to solve the PB equation with a grid spacing of 1.0 Å and a convergence criterion of 10^{-4} kT/e. The percentage of the lattice occupied by the solute is 70%. To maintain consistency with GB calculations, the internal solute dielectric constant was set to 1.0 and the external solvent dielectric constant was set to 78.3. The probe radius used to define the molecular surface is 1.4 Å, and the salt concentration is zero. The Born radius of a solute atom i was obtained by inverting the solvation energy of a reference system, in which atom i bears a unit charge and all other solute atoms are neutral. For GB calculations, the Born radius is directly calculated from the formulae given below.

Defining Effective Born Radii. As summarized above, three of the GB models we tested (Still, HCT, and mAGB) replace the integral in eq 5 with a pairwise sum of interactions of atom i with other atoms, j, in the solute (eq 6). (It is worth noting here that the GBMV2 model uses eq 5 directly.) However, the volume integral in eq 6 is over the full volume of atom j that does not overlap with atom i but may overlap the volume of other atoms in the summation. The different GB models account for this overcounting problem in different ways. In the following, we summarize the four methods and describe our implementation of each.

The Still Model. In this model, the integral over the volume of each atom is replaced by terms of the form V_j/r_{ij}^4; that is, there is no explicit integration over the volume of each atom. The problem of overlapping spheres is accounted for through the use of a series of different scaling factors for bonded and nonbonded atoms. In this work, we have used the original Still formula[20] without modification, which can be rewritten in a form similar to eq 6:

$$\alpha_i^{-1} = \frac{1}{R_{\text{vdW-}i} + \phi + P_1} - \frac{1}{4\pi}\left(\sum^{\text{stretch}} \frac{P_2 V_j}{r_{ij}^4} + \sum^{\text{bend}} \frac{P_3 V_j}{r_{ij}^4} + \sum^{\text{nonbonded}} \frac{P_4 V_j \text{CCF}}{r_{ij}^4}\right) \quad (7)$$

where r_{ij} is the distance between atoms i and j, $R_{\text{vdW-}i}$ is the van der Waals radius of atom i, ϕ is the dielectric offset (−0.09), and V_j is the van der Waals volume of atom j. P_1 is a single-atom scaling factor, while P_2-P_4 empirically scale the contribution from surrounding atoms to account for the overlap in their atomic volumes. CCF is a switching function that reduces this overlap when nonbonded atoms approach within a certain fraction of the sum of their van der Waals radii:

$$\text{CCF} = 1.0 \quad \text{if} \quad \left(\frac{r_{ij}}{R_{\text{vdW-}i} + R_{\text{vdW-}j}}\right)^2 > \frac{1}{P_5}$$

otherwise

$$\text{CCF} = \left\{0.5\left[1.0 - \cos\left\{\left(\frac{r_{ij}}{R_{\text{vdW-}i} + R_{\text{vdW-}j}}\right)^2 P_5 \pi\right\}\right]\right\}^2 \quad (8)$$

P_5 is a cutoff parameter. The first three terms in eq 7 are treated as constants and only calculated once at the beginning of the energy minimization or dynamics simulation. The parameters P_1-P_5 are taken from the original paper[20] without modification.

The HCT Model. In this model, eq 6 is interpreted in such a way that the integral over the volume of atom j outside atom i is calculated with a two-sphere integral and a set of atom type dependent parameters, S_i, that reduce the overcounting induced by the pairwise sums. The effective Born radius in this model is given by the following expression:

$$\alpha_i^{-1} = R_i^{-1} - \frac{1}{2}\sum_{j\neq i}\left[\frac{1}{L_{ij}} - \frac{1}{U_{ij}} + \frac{r_{ij}}{4}\left(\frac{1}{U_{ij}^2} - \frac{1}{L_{ij}^2}\right) + \frac{1}{2r_{ij}}\ln\frac{L_{ij}}{U_{ij}} + \frac{S_i^2 R_j^2}{4r_{ij}}\left(\frac{1}{L_{ij}^2} - \frac{1}{U_{ij}^2}\right)\right] \quad (9)$$

where

$$\begin{aligned}
L_{ij} &= 1 && \text{if} \quad r_{ij} + S_i R_j \le R_i \\
L_{ij} &= R_i && \text{if} \quad r_{ij} - S_i R_j \le R_i \le r_{ij} + S_i R_j \\
L_{ij} &= r_{ij} - S_i R_j && \text{if} \quad R_i \le r_{ij} - S_i R_j \\
U_{ij} &= 1 && \text{if} \quad r_{ij} + S_i R_j \le R_i \\
U_{ij} &= r_{ij} + S_i R_j && \text{if} \quad R_i \le r_{ij} + S_i R_j
\end{aligned}$$

where r_{ij} is the distance between atoms i and j and R_i is a modified atomic radius given by $(R_{\text{vdW-}i} + \phi)$, where $R_{\text{vdW-}i}$ is the van der Waals radius of atom i and ϕ is the dielectric offset (−0.09). The key variable, S_i, is the screening parameter of atom i and depends on the force field and atom type. We used $\{S_i\}$ as implemented in TINKER[44] and slightly adapted the S_i values of N-terminal nitrogens (−NH2 and −NH3) and arginine nitrogens (N_ϵ and N_η) from 0.79 to 0.65 for the GROMOS96 force field[41] so as to improve the correlation of Born radii between the HCT model and Delphi.[10,11]

The mAGB Model. The AGBNP model of Levy and coworkers[26] is a combination of an analytical GB (AGB) model and a new nonpolar (NP) term. The AGB model is based on a different interpretation of eq 6 than that used in the HCT model. The integral over the volume of atom j outside atom i is calculated with a two-sphere integral and then multiplied by a scaling factor, S_{ji}, to reduce the overcounting effect with other atoms. Here, S_{ji} is calculated separately and depends on the local geometry and not the atom type. The effective Born radius in the AGB model is expressed as

$$\alpha_i^{-1} = R_i^{-1} - \frac{1}{4\pi}\sum_{j\neq i} S_{ji}Q_{ji} \quad (10)$$

where R_i is a modified atomic radius of atom i defined as in the HCT model. Q_{ji} is the integral over the volume of atom j that lies outside atom i.[26] The full expression for Q_{ji} at different distance conditions between atom i and atom j can be found in Appendix B of ref 26. S_{ji} is a scaling factor, which can be approximated by

$$S_{ji} \approx S_j + \frac{1}{2}\frac{V_{ij}}{V_j} \quad (11)$$

where V_j is the van der Waals (vdW) volume of atom j and V_{ij} is the two-body overlap volume between atoms i and j. $S_j = V_j'/V_j$, V_j' is the self-volume of atom j, which is the region of atom j that can be assigned only to that atom (in such a way that the sum of the self-volumes gives the total solute volume excluding internal voids). In the AGB model, the volume overlap formed by n spheres is calculated with a Gaussian integration described by Grant and Pickup.[66] The exact calculation of the self-volumes, $\{V_i'\}$, requires that various combinations of overlapping atoms be explored and is time-consuming.

To circumvent this problem, we have developed a modified AGB (mAGB) model where we replaced S_{ji} with S_j in eq 10 and defined V_j' as the vdW volume of atom j minus those subvolumes that lie inside directly bonded atoms. A simple volume algorithm as implemented in the Still model[20] is used to calculate V_j'. Another change we made is to use R_i instead of $R_{\text{vdW-}i}$ so that we can adjust the model behavior by tuning the ϕ value. We have a used a value of -0.09 here so as to remain consistent with the original AGB model. However, it is possible to tune this parameter to improve the correlation with PB results.

We have carried out a number of tests on the structure of the B1 domain of protein G (PDB code 2gb1) (which is used in the folding studies described below) to verify that the mAGB model retains the features of the original AGB model. As shown in Figure 1a, self-volumes of atoms, V_j', calculated with the two methods are very well correlated; they are practically identical for polar atoms, which have relatively small radii, while the mAGB model assigns relatively larger volumes to nonpolar atoms, which usually have larger radii (e.g., carbons). This reflects the fact that the Gaussian algorithm takes into account all volume overlaps while the simple algorithm only considers overlaps with bonded atoms. A further test involved comparing Born radii calculated with the original AGB model and with the mAGB model. As shown in Figure 1b, the two sets of Born radii are almost perfectly correlated with a slope near 1. A minor difference between the two sets is that the original AGB model gives slightly larger Born radii than the mAGB model (see the offset of -0.123 in Figure 1b). However, both models assign smaller Born radii to buried protein atoms with respect to the Delphi results.

As a further test of the two methods, we carried out 2 ns stochastic dynamics (SD) simulations on the C-terminal domain of the ribosomal protein (PDB code 1ctf) using an in-house SD program which uses TINKER subroutines for calculating internal molecular mechanics terms (bonded and nonbonded) and the mAGB and Still GB models. To compare to results reported previously for the AGBNP model, the OPLSAA force field[45−49] was used. As shown in Figure 1c, the root-mean-square deviations (rmsds) of the C_α atoms from the native structure remain below 1.5 Å and are ~0.5−1.0 Å lower than the fluctuations reported by Gallicchio et al.[26] using the same force field (Figure 6 in ref 26). Figure 1d reports a similar simulation to that reported in Figure 1c except that the Still model rather than the mAGB model was used. Here again, the observed fluctuations remain below 1.5 Å. It is possible that the differences between our results and those in ref 26 are due to differences in the nonpolar model. We have used a value of σ of 0.005 kcal/(mol Å^2), whereas Levy and co-workers used a novel and more complex model that is described in detail in their paper.[26]

The GBMV2 Model. This model was designed by Brooks and co-workers[24,25] so as to maximize consistency with PB methods. First, the GBMV2 model is based on the volume integral in eq 5 instead of the pairwise summation of eq 6. Second, to reproduce the solute volume defined in PB methods, which is the volume enclosed by the molecular surface, the GBMV2 model employs a set of complex, vector-based functions that detect and fill internal voids. The volume they obtain is termed the standard molecular volume (SMV)[24,25] Third, the GBMV2 model uses an additional term to correct the CFA. Specifically, the Born radius is calculated from the following empirical formula:

$$\alpha_{\text{GBMV2}} = C_1\alpha_i + C_0 \quad (12)$$

where

$$\alpha_i^{-1} = \left(1 - \frac{1}{\sqrt{2}}\right)\left(\frac{1}{R} - \frac{1}{4\pi}\int\!\!\int\!\!\int_{|\vec{r}-\vec{x}_i|>R} \frac{V(\vec{r})}{|\vec{r}-\vec{x}_i|^4}\,dx\,dy\,dz\right) + \left(\frac{1}{4R^4} - \frac{1}{4\pi}\int\!\!\int\!\!\int_{|\vec{r}-\vec{x}_i|>R} \frac{V(\vec{r})}{|\vec{r}-\vec{x}_i|^7}\,dx\,dy\,dz\right)^{1/4}$$

where α_{GBMV2} is a linear transformation of α_i, the Born radius of atom i. α_{GBMV2} is used instead of α_i in eq 3 to calculate ΔG_{pol}. The volume integral is calculated with quadrature techniques as described in Lee et al.'s paper.[25] Our implementation of the GBMV2 model differs from the original model only in the method used to generate a lookup table; specifically, we generate an atomic neighbor list for each grid point instead of for the center of each cell. This adjustment was made for convenience and should have no influence on the performance. The theoretical details and model parameters can be found in the original paper.[25]

Peptide Folding

Peptide Systems. Two peptides were studied: the C-terminus β-hairpin of the B1 domain of protein G (residues 41−56, GEWTYDDATKTFTVTE; PDB code 2gb1) and the alanine-based peptide 3K(I) (AAAAKAAAAKAAAAKA). The β-hairpin peptide is capped with an Ace group at the N-terminal and an Nme group at the C-terminal, while 3K(I) is capped with an Ace group at the N-terminal and an NH_2 group at the C-terminal. As described below, the simulations on both peptides were carried out starting from an extended conformation defined here as having backbone dihedral angles ψ, ϕ, and ω of $-135°$, $135°$, and $180°$. All side chain dihedral angles were set to $180°$ with the exception of χ_1 and χ_2 of aromatic residues that were set to $60°$ and $90°$, respectively, to avoid steric clashes. We define the "native structure" of the β-hairpin as corresponding to the structure of residues 41−56 in 2gb1 (residues 41−56), which is characterized by two β-pleated strands and hydrophobic contacts between Y45 and F52 and between W43 and V54. For 3K(I), an ideal α-helical conformation was constructed with the TINKER[44] package with ψ, ϕ, and ω set to $-57°$, $-47°$, and

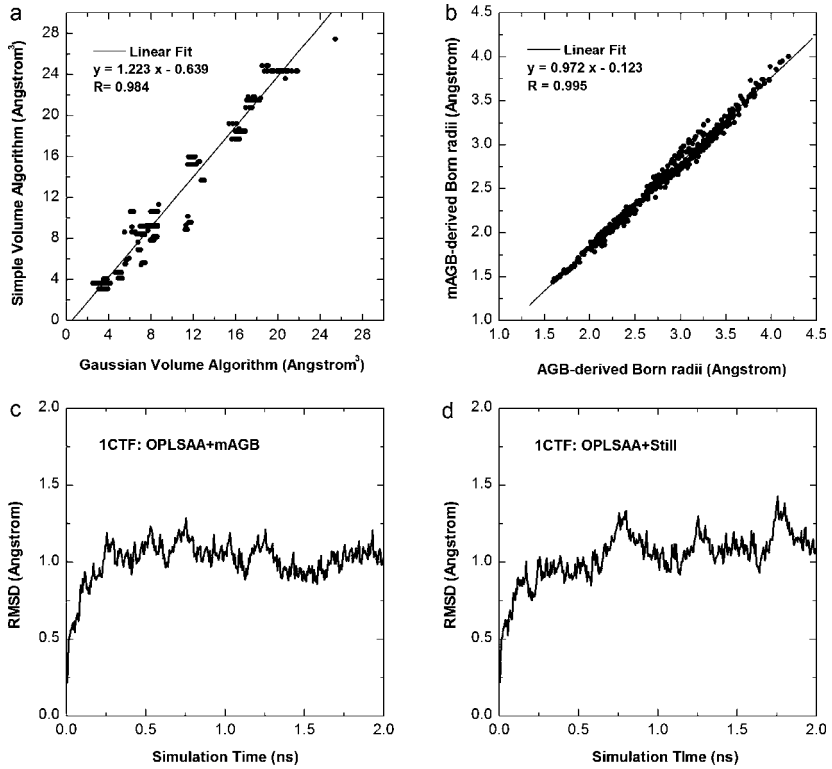

Figure 1. Comparison of the AGB and modified mAGB models: (a) self-volumes of atoms in the protein 2gb1; (b) effective Born radii for atoms in 2gb1; (c) rmsds from the native conformation of 1ctf obtained from SD simulation using the OPLSAA force field and the mAGB model; (d) rmsds from the native conformation of 1ctf obtained from SD simulations using the OPLSAA force field and the Still GB model.

180° and $\chi_1-\chi_4$ set to 180°. The molecular topologies of the peptides were built with PROGMT in the GROMOS96 package.[41] Polar and aromatic hydrogens were built with PROGCH in the same package.

Simulation Details. All folding simulations were performed with the GROMOS96[41] package. The four GB models were introduced into this package, and expressions for the derivatives were obtained from the GB equation (eq 3) and from the formulae for different Born radii given in the Materials and Methods section. The GROMOS96 43A1 force field[41] was used with the protonation states of both peptides corresponding to those expected at neutral pH. Standard stochastic dynamics (SD)[67] was combined with different GB models. A time step of 2 fs was used in the simulations with the covalent bond lengths constrained by SHAKE[68] with a tolerance of 10^{-4}. The temperature was kept constant by weak coupling[69] to an external heat bath at 300 K, with the relaxation times being 0.1 ps. The nonbonded interactions were calculated without cutoffs. The friction coefficient of water, γ_w, is 91 ps^{-1} and was weighted by an atomic accessibility factor.[70] Coordinates and energies were recorded every 0.5 ps. The nonpolar solvation energy and derivatives were updated every 10 time steps so as to enhance computational efficiency.[71] σ was set to 0.005 kcal/(mol Å^2) for all atom types except for sulfur (-0.005 kcal/(mol Å^2)) and polar hydrogens (0.0 kcal/(mol Å^2)).[71]

The initial structures were relaxed with a 200-step steepest descent energy minimization using the GB model being tested. The minimized structures were then adapted to implicit solvent with 20 ps restrained dynamics during which the positions of non-hydrogen atoms were restrained with a harmonic potential. Then, ten 20 ns SD simulations for the β-hairpin and five 20 ns SD simulations for the α-helix were performed without restraints. Each folding simulation uses a different random number to assign initial atomic velocities. For purposes of comparison, folding simulations using the OPLSAA force field[45−49] and the Still GB model were performed with an in-house SD program based on TINKER subroutines.[44] The SD protocols and parameters used with this program were essentially the same as those used in the GROMOS96 simulations.

Trajectory Analysis. Modified GROMOS96[41] programs were used to analyze the SD trajectories in peptide folding. A clustering algorithm in DycoBlock[72,73] was redesigned to analyze the structural distributions along folding pathways and to characterize the sampling features of different GB models. The algorithmic details are as follows: A structure, *S*, extracted from a trajectory is compared with "representative" structures of existing clusters, $\{R_i\}$, according to C_α rmsds and energies. If the rmsd between structure *S* and the representative structure, R_i, is smaller than a cutoff, the energy E_S is compared with E_R. If $E_S > E_R$, *S* is assigned to this cluster; if $E_S < E_R$, *S* replaces R_i and becomes the new representative of this cluster. If the rmsd between *S* and more than one representative structure is less than the cutoff, the clusters are merged into one and *S* becomes the representative of the new cluster. If the rmsd between structure *S* and any representative structure, R_i, is larger than the cutoff, *S* is added to the tail of the list of representative structures. This process is repeated until $\{R_i\}$ does not change between the current clustering cycle (the current iteration) and the previous cycle or the number of iterations exceeds a user-specified number. Here, the rmsd cutoff is set to 1.5 Å and the maximum number of cycles is set to 10.

Results

Born Radii

Folded Proteins. We compared Born radii calculated with the four GB models to those obtained with DelPhi using atomic radii taken from the GROMOS96 force field.[71] The results

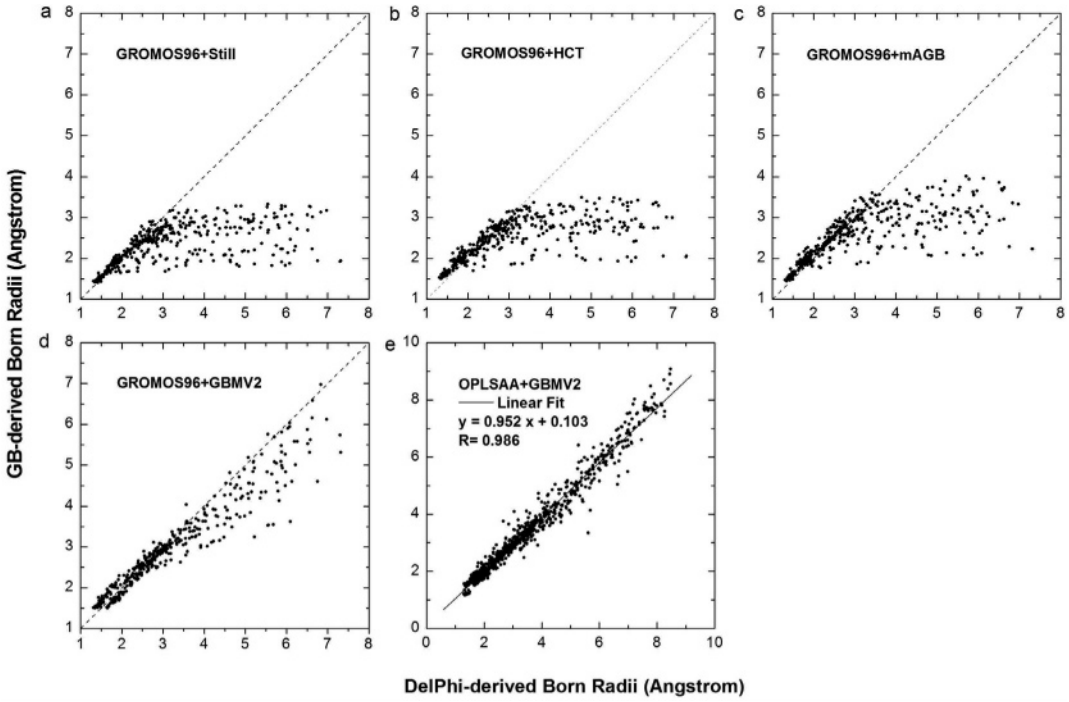

Figure 2. Effective Born radii obtained from the PB equation and various GB models for atoms in 2gb1: (a) Still GB model with the GROMOS96 force field; (b) HCT model with the GROMOS96 force field; (c) mAGB model with the GROMOS96 force field; (d) GBMV2 model with the GROMOS96 force field; (e) GBMV2 model with the OPLSSA force field. Atomic radii of all-atom force fields as defined in TINKER were used here.

obtained using the B1 domain of protein G (2gb1) are shown in Figure 2. Similar results have been obtained for other proteins (data not shown). The Still model, the HCT model, and the mAGB model clearly underestimate the Born radii of buried atoms (points on the right side and below the diagonal), while the GBMV2 model shows a substantial improvement but still underestimates the Born radii of some atoms by ∼30%. Lee et al.[25] have obtained better agreement with PB calculations than those shown in Figure 2, but the original GBMV2 parameters were tuned for the CHARMM22 all-atom force field.[74] To test whether the use of the united-atom GROMOS96 force field affected the correlation, we recalculated the Born radii with the OPLSAA all-atom force field[45−49] (Figure 2e). The atomic radii for the all-atom force fields contained in TINKER[44] were used in all PB and GB calculations. The correlation reported in Figure 2e is comparable to that reported by Lee et al.,[25] suggesting that GBMV2 parameters can be used with different all-atom force fields and are not sensitive to atomic radii. There is some degradation of performance with GROMOS96, but GBMV2 still gives the best agreement with PB calculations of Born radii.

To further characterize the properties of the different models, we calculated the average Born radii of backbone and aromatic hydrogens. The average Born radius of backbone hydrogens is 2.01 Å for the Still model, 2.43 Å for the HCT model, 2.49 Å for the mAGB model, 3.22 Å for the GBMV2 model, and 3.49 Å for DelPhi. The corresponding values for aromatic hydrogens are 1.81 Å for the Still model, 1.96 Å for the HCT model, 2.08 Å for the mAGB model, 3.53 Å for GBMV2 model, and 4.00 Å for DelPhi. It is clear from these numbers, and from Figure 2, that the GBMV2 model is most successful in reproducing PB results. The "ranking" of the four models in terms of their ability to reproduce PB-derived Born radii is GBMV2 > mAGB > HCT > Still.

Overall, the results reflect the fact discussed above that GB models of proteins are, in effect, more highly solvated than PB models, since the latter use the molecular surface of the entire protein to define the molecular interior while GB models, through their use of vdW volumes for individual atoms, effectively allow more water to penetrate the protein interior, even where there is no space to place individual water molecules. One important consequence will be to reduce the strength of pairwise electrostatic (including hydrogen bonding) interactions for buried atoms which, in molecular dynamics (MD) simulations, could lead to larger structural fluctuations than are warranted.

Peptide Folding. As part of studying the effect of different GB models on peptide folding, we compared GB- and PB-derived Born radii for the unfolded and folded conformations of the β-hairpin (Figure 3). The extended conformation was taken as a model of unfolded state, and we assume that it is representative of extended conformations of the polypeptide chain. The native structure was taken form the crystal structure as described above. It is evident that in the unfolded state all GB models yield good correlations between GB and PB radii. Specifically, the underestimate of Born radii for buried atoms that is seen for folded structures is not an issue. This is due to the fact that in the unfolded state most atoms are exposed to water so that their electrostatic solvation energy can be well approximated by the Born formula. The HCT and mAGB models tend to overestimate the Born radii in the unfolded state, while the GBMV2 model, though yielding a good correlation, produces radii that are either too large or too small. This may be due to the fact, as mentioned above, that GBMV2 was tuned to the CHARMM22 force field. Overall, the performance of all GB models for the folded state of the β-hairpin is similar to that seen for the folded state of small proteins (Figure 2).

The net effect of GB "errors" on folding simulations is hard to predict, since both Coulomb and solvation interactions are affected by the Born term. It has been shown that when GB-derived Born radii are in agreement with PB results then solvent screened Coulomb energies are in agreement as well.[28] When GB and PB radii are not in agreement, there will also be errors

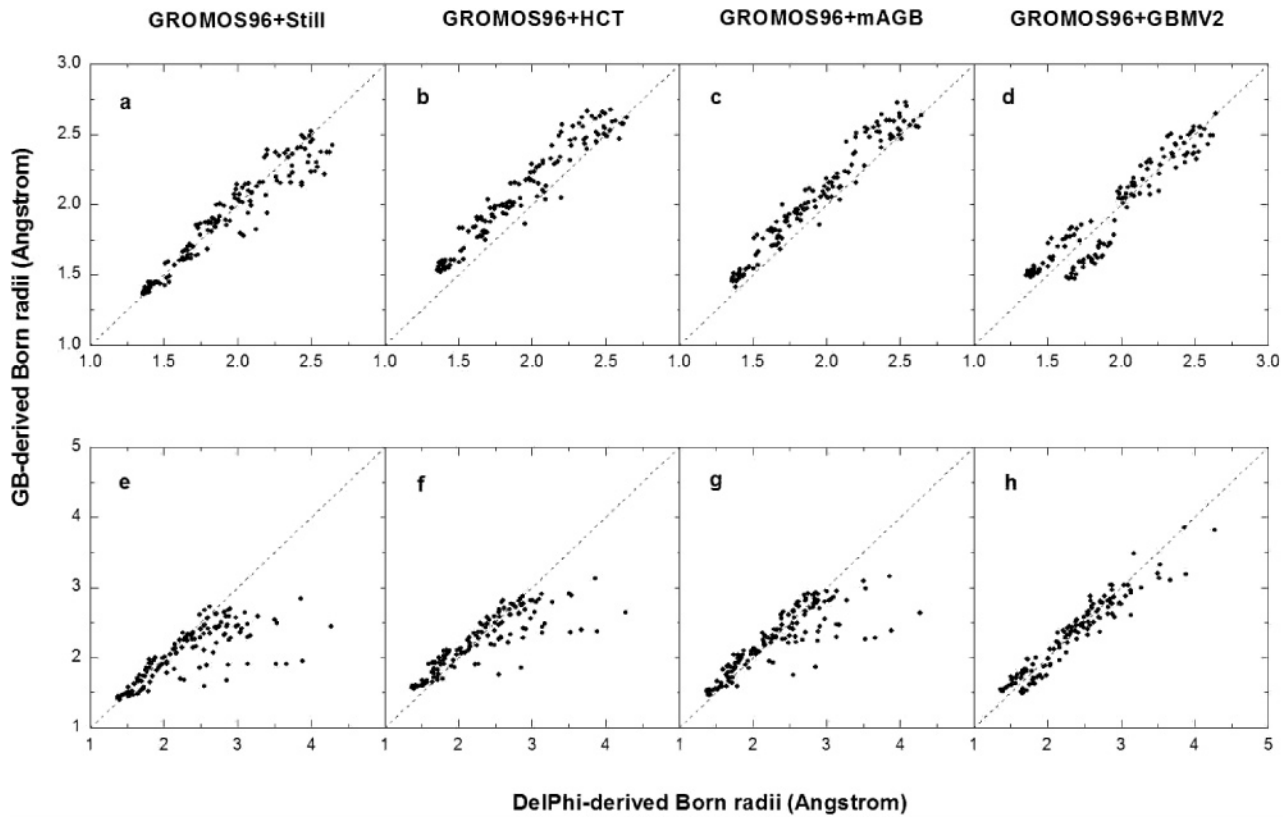

Figure 3. Effective Born radii obtained from the PB equation and various GB models for atoms in the hairpin peptide in the folded state (native conformation) and unfolded state (extended conformation). The upper panel is the folded state, and the lower panel is the unfolded state: (a and e) Still GB model with the GROMOS96 force field; (b and f) HCT model with the GROMOS96 force field; (c and g) mAGB model with the GROMOS96 force field; (d and h) GBMV2 model with the GROMOS96 force field.

in the Coulomb energies so that forces and individual atoms can be inaccurate. On the other hand, the good correlation of GB and PB solvation energies suggests that the total energies tend to be much more insensitive to the effective Born radii than the individual terms.[20,71,75−77] Nevertheless, differences in radii calculated with different models are expected to affect local forces that drive peptide folding and perhaps the free energy differences between different conformations as well.

Ab Initio Peptide Folding

C-Terminal β-Hairpin of the B1 Domain of Protein G

Lowest-rmsd Structures. The lowest-rmsd structures are identified by comparing to the native structure taken from the protein 2gb1 (41−56). The results from different folding simulations are summarized in Table 1. The top part of the table describes some of the structural properties of the most native-like structures (defined by C_α rmsd relative to the native structure) obtained in each of the 10 folding simulations that were carried out. The bottom part of the table characterizes the sampling features of the four GB models from the results of a clustering analysis. We first point out that the Still model produces a low-rmsd structure (simulation 7), which has six correct native hydrogen bonds formed (out of a total of 8, as defined by DSSP[78]). This, to our knowledge, is the first time that a native-like structure has been obtained with a room temperature simulation of this peptide without the use of enhanced sampling techniques. As is evident from the table, the Still model produces a number of other low-rmsd structures (four below 2 Å rmsd) and none worse than 3.5 Å. As can be seen from Figure 4a−d, the four best structures all form hydrophobic clusters, although the specific contacts differ in each case. Other than the best structure (structure 7), none of these lowest-rmsd structures have the correct native backbone hydrogen bonds, but all form hairpins, which appear driven by the aggregation of the aromatic residues. None of the other GB models produce structures with rmsds below 2 Å with the exception of the mAGB model, which produces two (from simulations 4 and 9). As seen in Figure 4e, a hairpin structure is formed with a hydrophobic cluster, although V54 faces in the wrong direction and does not participate. Column 4 in Table 2 shows that there are, on average, 1.6 native hydrogen bonds in the structural ensemble corresponding to this lowest-rmsd structure. In Figure 4f, a β-hairpin is formed with a number of non-native hydrogen bonds. A hydrophobic cluster cannot form because Y45 and W43 face one side of the sheet while F52 and V54 face the other.

In terms of the average lowest-rmsd over 10 simulations, the Still model gives the best results followed by the mAGB and GBMV2 models which are equivalent by the rmsd criterion. The HCT model gives the worst results. With the exception of the HCT model, which produces overly compact structures, all other GB models yielded radii of gyration that are in good agreement with experiment. None of the lowest-rmsd structures shown in the table had more than one native hydrogen bond (with the exception of structure 7 with the Still model). On the other hand, many of the structures had some hydrogen bonds formed and many of these were involved in hairpin-like structures. The averages reported here are probably not statistically meaningful, since they are based on only 10 simulations per model. To determine whether there is a clear basis for the observed behavior of the four GB models, we used clustering tools to characterize the effect of each model on the sampling of conformational space.

Conformational Sampling. We applied a number of novel cluster analysis tools, as described in the Materials and Methods

TABLE 1: Analysis of the Results of Folding Simulations on the C-Terminal β-Hairpin Peptide of the B1 Domain of Protein G

		1	2	3	4	5	6	7	8	9	10	av
Structural Data for the Lowest-rmsd Structures												
rmsd[a]	Still	1.49	3.45	1.88	3.39	2.87	1.50	0.67	3.66	2.53	3.02	**2.45**
	HCT	2.28	3.08	3.22	3.58	3.69	3.80	4.14	2.77	3.27	3.44	**3.33**
	mAGB	2.42	3.07	2.72	2.04	2.29	5.00	2.61	4.60	1.82	2.78	**2.94**
	GBMV2	3.63	3.30	2.46	3.07	3.04	2.84	2.73	3.05	2.35	2.92	**2.94**
RG[b]	exptl	8.26										
	Still	8.13	7.67	8.20	7.07	8.05	8.19	8.03	7.71	7.91	7.73	**7.87**
	HCT	7.75	7.35	7.33	7.92	7.19	6.93	7.19	7.25	7.15	7.24	**7.33**
	mAGB	7.72	7.99	7.82	7.79	8.17	7.05	8.01	7.51	7.86	7.88	**7.78**
	GBMV2	7.84	7.80	7.95	8.02	8.05	8.32	8.27	7.94	8.36	8.14	**8.07**
HB[c]	exptl	8										
	Still	1/0	5/1	3/0	2/1	3/1	0/0	6/6	3/0	5/0	2/0	
	HCT	1/0	2/0	4/1	1/0	1/0	6/0	4/0	3/1	5/1	4/0	
	mAGB	3/0	3/0	1/0	5/1	4/0	4/0	4/0	3/0	4/0	4/1	
	GBMV2	6/0	4/0	3/1	0/0	3/0	0/0	0/0	4/1	0/0	2/0	
Clustering Analysis												
NC[d]	Still	5/17	4/43	6/22	6/46	6/50	9/42	8/45	5/35	7/45	8/47	**6.6/44**
	HCT	5/31	8/23	12/35	7/24	10/36	5/42	12/43	9/35	11/30	12/34	**9.5/36**
	mAGB	7/37	11/41	6/40	9/35	11/47	8/23	10/37	8/32	9/22	12/33	**9.3/38**
	GBMV2	8/28	4/42	8/23	7/27	3/40	10/26	7/27	10/40	8/23	9/27	**7.4/30**
PERC[e] (%)	Still	69/84	25/67	70/86	40/74	20/66	33/63	35/78	36/70	33/71	31/70	**39/73**
	HCT	54/82	74/89	64/87	75/90	60/89	30/74	53/84	52/84	57/83	58/84	**58/85**
	mAGB	44/78	42/74	36/76	50/82	42/81	70/88	48/80	54/81	71/85	65/86	**52/81**
	GBMV2	58/80	16/53	49/61	61/79	15/51	61/80	47/71	49/74	61/77	67/87	**48/71**

[a] rmsd: the lowest root-mean-square deviation of all C_α atoms with respect to the native structure, which is taken from 2gb1. [b] RG: radius of gyration of the lowest-rmsd structure. [c] HB: the number of hydrogen bonds in the lowest-rmsd structure. [d] NC: number of large clusters with more than 1000 structures and number of meaningful clusters with more than 200 structures. [e] PERC: percentage of population in all large clusters and in all meaningful clusters.

section. Of interest was the extent to which various simulations explored conformational space and whether the simulations were trapped in local minima. We defined clusters that had greater than 200 structures (lifetime ~0.1 ns) as *meaningful regions* in conformational space and clusters with greater than 1000 structures (lifetime ~0.5 ns) as *dominant regions*, which might correspond to energetically favorable local minima. The top half of the Clustering Analysis section of Table 1 indicates the number of dominant and meaningful clusters formed in each simulation. The bottom half of the table indicates the percentage of all the structures generated in a simulation that belong to one of the two types of clusters. As is evident from the averages given in the last column, the Still model produces the largest number of meaningful clusters and the smallest number of dominant clusters. This suggests that the Still model is most effective in sampling conformational space; that is, it is not easily trapped in local minima and explores the largest number of structurally distinct conformations. This feature is responsible for the fact that it is most effective in producing low-rmsd structures.

The HCT and mAGB models are approximately equivalent based on the clustering criteria: they form fewer meaningful clusters and more dominant clusters than the Still model, both with high populations. This suggests that both models produce more stable local minima than the Still model. The GBMV2 model forms the fewest meaningful clusters and appears to be the least effective method in terms of sampling. This is likely due in part to the complex formula it requires for the calculation of the Born radius which makes the solvation energy more sensitive to small conformational changes, which in turn leads to a more rugged energy surface. As a consequence, the ability to fold is more sensitive to initial conditions rather than to the global shape of the energy surface.

The clustering analysis indicates that some of the lowest-rmsd structures reported in the Structural Data for the Lowest-rmsd Structures section of Table 1 are transient conformations while others belong to the meaningful or dominant clusters. For the Still model, 6 of the 10 lowest-rmsd structures are transient, 3 are found in meaningful clusters, and 1 (simulation 7, 0.67 Å) is in a dominant cluster which occupies 4.3% of the total population of conformations and has an average rmsd of 1.21 Å. Similar patterns are found for the HCT and mAGB models. However, none of the lowest-rmsd structures in Table 1 belong to the largest and second-largest clusters shown in Table 2 (see below). These results point to problems in the standard practice of using lowest-rmsd structures as an indicator of the success of folding simulations. If such simulations were to be used for predictions, one could use a conformational energy criterion to select native-like conformations, but such energies do not necessarily pick out the lowest-rmsd structures (see below). Analysis of clusters may provide a more effective approach, but this requires that the cluster provides an adequate sampling of conformational space.

Conformational Ensembles. Table 2 lists a number of structural characteristics, average energies, and the percentage of the total population for the two largest clusters formed in each simulation. The precise characterization of each structural ensemble can be inferred on the basis of the information from the table. The average rmsds are of course significantly higher than the lowest rmsds in Table 1, but the two numbers are correlated. The average number of native hydrogen bonds is zero for most clusters with only five exceptions (marked in bold). An average effective energy (defined as the sum of the internal force field energy and GBSA solvation free energy) is used to characterize each cluster. This energy is not strongly correlated with the rmsd from the native structure. For example, in the Still model, the lowest average energy is found for the largest cluster in simulation 1, which has a partially formed β-hairpin structure and an average energy of -513.8 kcal mol^{-1}. The second-largest cluster in simulation 7 corresponds to a distribution of native-like structures, which have an average of 4.6 native hydrogen bonds and an average energy of -513.3 kcal mol^{-1}. The largest cluster in simulation 2 has a comparable energy, -513.4 kcal mol^{-1}, compared to the other two but no

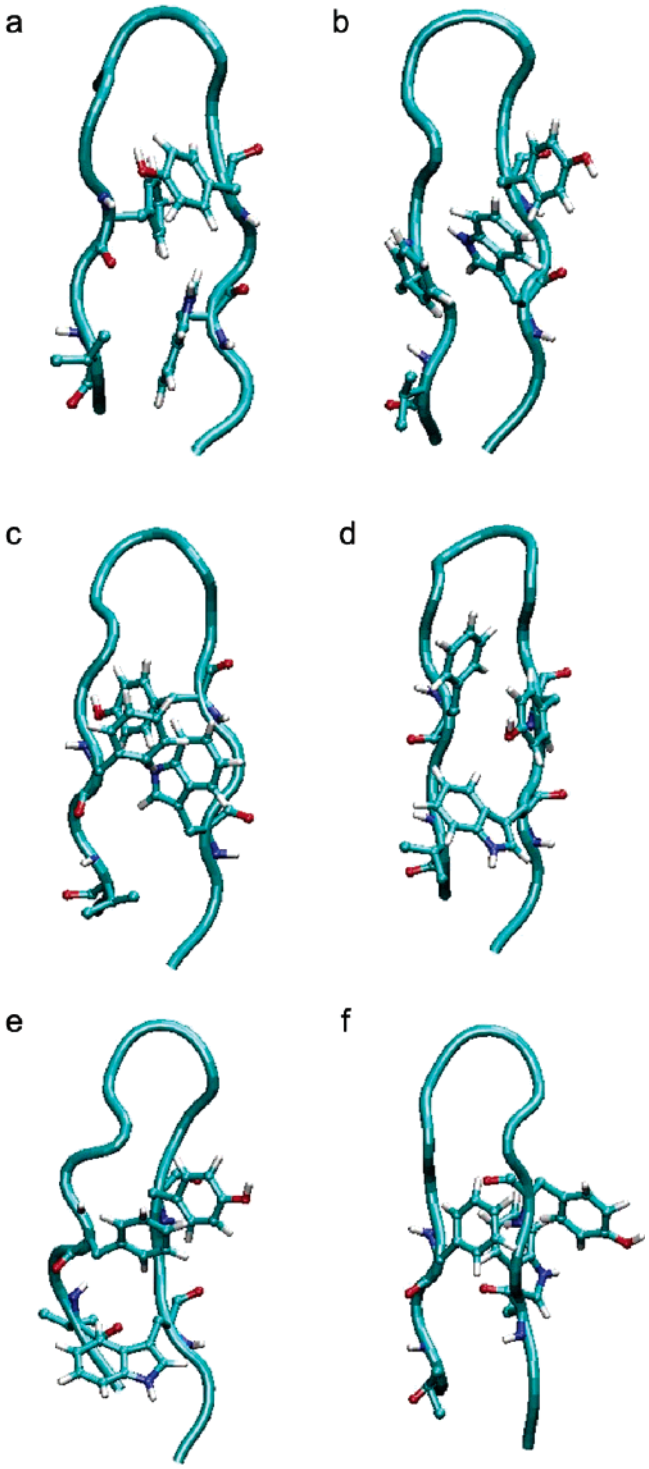

Figure 4. Selected low-rmsd structures obtained from folding simulations on the hairpin peptide (see text and Table 1 for details). The peptide backbone is shown as a tube, and the four residues (Y45−F52, W43−V54) forming a hydrophobic cluster in the protein 2gb1 are shown in ball-and-stick: (a) simulation 1 with the Still GB model; (b) simulation 3 with the Still GB model; (c) simulation 6 with the Still GB model; (d) simulation 7 with the Still GB model; (e) simulation 4 with the mAGB model; (f) simulation 9 with the mAGB model.

native hydrogen bonds. Similar results are found for other GB models.

An important question in structure prediction is whether native-like structures can be detected as those that have the lowest conformational energies. To address this issue within the context of this work, we carried a 2 ns SD simulation starting from the native structure with each of the four GB models. All structures within 1.5 Å rmsd to the native structure (the same threshold used in the clustering analysis) were defined as belonging to the native ensemble. The average energy in each ensemble was defined as the "native energy". For the Still model, the native energy is −514.5 kcal mol^{-1}, which is ∼1 kcal mol^{-1} lower than the lowest energies found for any cluster (the largest cluster for simulation 1 and the second largest for simulation 7). For the HCT model, the native energy is −491.0 kcal mol^{-1}, which is 1 kcal mol^{-1} higher than the lowest energy of any cluster (the largest cluster in simulation 6). For the mAGB model, the native energy is −497.6 kcal mol^{-1}, which is ∼3 kcalmol^{-1} lower than the lowest energy (the second-largest cluster in simulation 1). Finally, for the GBMV2 model, the native energy is −561.0 kcal mol^{-1}, which is ∼2 kcal mol^{-1} higher than the energy of the largest cluster in simulation 6.

It is somewhat surprising that the native structure using the GBMV2 model does not have the lowest energy, even though this model produces the best Born radii for the folded state, as judged by the agreement with PB results (Figures 2 and 3). The problem may well be due to our use of the GBMV2 model with GROMOS96, as discussed above. On the other hand, the native structure for the GBMV2 model was significantly lower in energy than the average energy for all of the other clusters that were formed; indeed, the energy gap for GBMV2 between the lowest energy and highest energy clusters reported in Table 2 was larger than that for the other GB models.

OPLSAA Simulation. The surprisingly good sampling efficiency found here for β-hairpin folding has not, to our knowledge, been reported for other force fields. To directly compare our results with published work, we performed five 20 ns SD simulations using the OPLSAA force field[45−49] and the Still GB model. The results are shown in Figure 5. Five GROMOS96 trajectories and five OPLSAA trajectories are described by plotting the C_α rmsd to the native structure as a function of simulation time. In the GROMOS96 simulations (Figure 5a), all five trajectories reach ∼4 Å rmsd within ∼4 ns and two reach ∼2 Å rmsd. In contrast, only two OPLSAA trajectories reach ∼4 Å at 16 ns, none reach 2 Å, and three OPLSAA trajectories fluctuate around 8−9 Å for the entire duration of the simulation. It is evident from the figure that GROMOS96 is a more "flexible" force field than OPLSAA, an observation that has been reported previously by van der Spoel et al. using explicit solvent simulation.[35] As discussed by Hu et al.,[36] this difference between the two force fields may be due to local geometric terms, such as those that describe torsional barriers.

Alanine-Based α-Helix 3K(I)

Lowest-rmsd Structures. Given the simplicity of the α-helical structure, we performed only five 20 ns simulations with each of the four GB models. Table 3 summarizes the characteristics of the lowest-rmsd structures and is analogous to Table 1 in the type of information it contains. The Still, HCT, and mAGB models all produce at least one very low rmsd structure, with the HCT model producing the best average results of the three. All three models generate at least one conformation with many native hydrogen bonds. It is of interest that, in contrast to the results for the helical peptide, the HCT model yielded the worst results for β-hairpin folding where it generated compact structures characterized by a rmsd of ∼5 Å and a small RG of 6−7 Å. Some of the misfolded structures had some helical content, suggesting that the HCT model tends to form more local contacts than do the other models. Clearly, this feature

TABLE 2: Properties of the Largest and Second-Largest Clusters Obtained in the Folding of the β-Hairpin Peptide

		1	2	3	4	5	6	7	8	9	10	av[a]
						Average over the Largest Cluster						
rmsd[b]	Still	2.38	5.41	3.03	4.81	5.43	4.81	3.50	6.01	3.38	5.24	**4.40**
	HCT	7.19	3.94	5.53	6.53	4.68	6.20	6.78	4.17	4.76	6.25	**5.60**
	mAGB	4.40	3.64	4.24	3.29	3.76	7.30	4.48	5.48	3.75	3.70	**4.40**
	GBMV2	4.44	4.34	3.97	4.56	4.41	5.53	3.92	4.64	3.99	4.42	**4.42**
RG[c]	Still	7.51	7.05	8.20	6.7	7.21	7.01	8.16	7.27	7.48	6.74	**7.33**
	HCT	6.67	6.93	6.83	6.66	6.50	7.10	6.83	6.97	6.58	6.73	**6.78**
	mAGB	7.27	7.27	7.55	7.49	7.21	6.80	7.03	6.95	7.35	7.26	**7.22**
	GBMV2	7.91	7.43	7.78	7.16	8.55	6.82	7.73	7.09	7.70	7.08	**7.53**
HB[d]	Still	3.5/2.1	3.1/0	5.2/0.6	3.2/0	3.6/0	3.6/0.2	4.3/0	2.9/0	3.7/0	3.7/0	
	HCT	3.3/0	3.8/0	3.7/0.4	5.0/0	4.5/0.5	6.1/0.3	3.7/0.4	4.5/0.1	4.2/0.6	3.9/0	
	mAGB	3.1/0	2.9/0	3.3/0	5.0/1.6	4.8/0.7	2.7/0	4.6/0	4.3/0	4.3/0	4.7/0.1	
	GBMV2	4.6/0.2	4.1/0	4.8/0	5.0/0	3.7/0	4.6/0	4.5/0.1	3.8/0.2	5.3/0	4.8/0	
ENE[e]	Still	−513.8	−513.4	−509.0	−510.4	−512.1	−511.4	−502.6	−510.7	−507.7	−513.5	
	HCT	−488.6	−484.7	−490.1	−487.3	−488.4	−492.0	−481.9	−483.7	−489.3	−482.6	
	mAGB	−493.8	−491.5	−491.7	−494.3	−492.3	−488.8	−494.9	−490.7	−489.0	−490.2	
	GBMV2	−553.2	−556.7	−554.5	−558.8	−543.5	−563.3	−551.3	−553.6	−555.6	−555.5	
						Average over the Second-Largest Cluster						
rmsd	Still	2.25	4.67	2.99	3.80	4.40	2.43	1.32	5.27	5.30	4.60	**3.70**
	HCT	7.77	5.67	5.11	6.55	5.84	6.51	6.41	6.04	4.70	4.11	**5.87**
	mAGB	6.02	4.75	5.77	3.04	4.39	5.75	4.32	5.61	2.48	3.52	**4.57**
	GBMV2	4.73	5.63	4.58	3.88	6.12	5.52	4.36	3.92	5.53	4.16	**4.84**
RG	Still	7.71	6.81	8.29	7.53	7.08	7.49	8.00	6.98	6.98	7.12	**7.40**
	HCT	6.78	6.58	6.90	6.60	7.03	7.20	6.73	6.64	6.65	7.17	**6.83**
	mAGB	6.73	7.28	7.39	7.31	7.14	7.16	7.21	7.06	8.04	7.54	**7.29**
	GBMV2	7.18	6.79	7.12	7.72	7.23	7.08	7.25	7.63	7.26	7.47	**7.27**
HB	Still	3.0/2.0	4.5/0.5	4.9/0.6	3.5/0	2.8/0	2.5/1.6	5.2/4.6	1.6/0	2.0/0	3.8/0	
	HCT	3.1/0.1	4.5/0	3.9/0.1	2.6/0	3.7/0.5	7.3/0	3.8/0.7	4.7/0.7	3.6/0	4.3/0	
	mAGB	4.1/0	1.7/0	2.9/0	3.9/0.6	2.5/0.3	5.7/0	4.8/0	4.7/0	5.0/0.1	4.6/0.1	
	GBMV2	4.6/0.3	4.2/0	3.7/0	4.5/0	2.4/0	5.0/0	4.9/0.1	4.2/0.2	3.9/0	4.8/0	
ENE	Still	−511.4	−512.8	−507.8	−506.8	−504.1	−503.0	−513.3	−510.6	−505.8	−512.2	
	HCT	−485.3	−482.3	−485.7	−475.8	−476.5	−489.9	−482.6	−488.1	−486.2	−480.1	
	mAGB	−494.1	−487.3	−492.4	−492.4	−486.3	−492.7	−490.8	−490.3	−484.5	−487.6	
	GBMV2	−559.5	−551.8	−555.4	−548.9	−544.4	−559.4	−554.6	−550.8	−553.6	−555.1	

[a] av: averages calculated for rmsd and RG over 10 simulations. [b] rmsd: average C_α root-mean-square deviation (rmsd) with respect to the native structure, averaged over all member structures. [c] RG: average radius of gyration (RG) with respect to the native structure, averaged over all member structures. [d] HB: average number of hydrogen bonds (the total number/number of native hydrogen bonds). [e] ENE: average "effective energy", which is defined as the sum of GROMOS96 internal energy and the GBSA solvation free energy, averaged over all member structures.

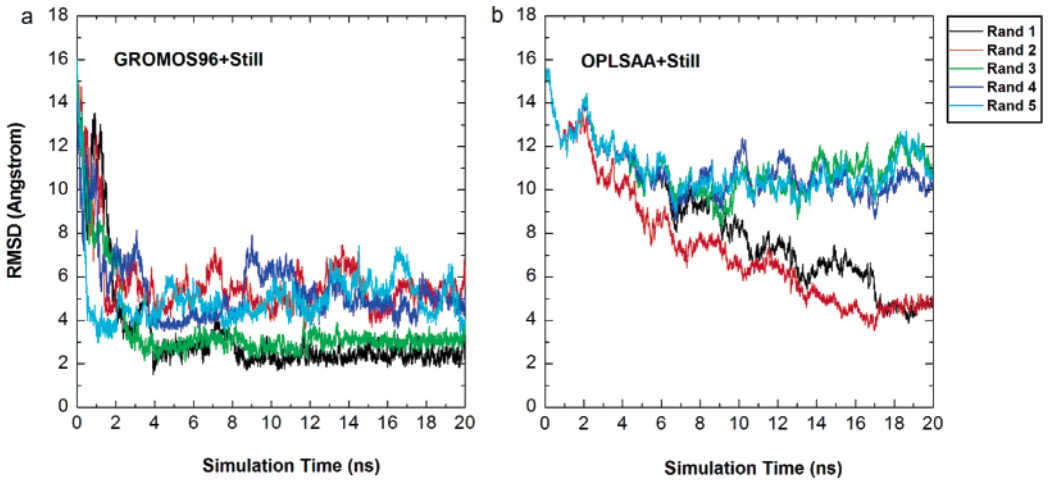

Figure 5. rmsds from the native structure as functions of simulation time for folding simulations on the hairpin peptide. Different colors denote different random numbers used to assign the initial atomic velocities: (a) GROMOS96 force field and Still solvation model; (b) OPLSAA force field and Still solvation model.

will facilitate the folding of helical peptides. Figure 6a−c shows the three lowest-rmsd structures from simulations: the Still model (simulation 5), the HCT model (simulation 3), and the mAGB model (simulation 5), respectively. They all correspond to an almost perfect α-helix with a slightly unraveled N-terminal.

Conformational Sampling. On the basis of the clustering analysis in the Clustering Analysis section of Table 3, the Still and mAGB models form the largest number of meaningful clusters while the HCT model forms more dominant clusters and, as was the case for the β-hairpin, appears to sample less effectively. All three models formed fewer dominant clusters than they did in hairpin folding, perhaps because helical structures are more restrictive. The GBMV2 failed to fold 3K(I) to near-native structures in any of the five simulations, and based on the large average value of RG, the peptide failed

TABLE 3: Analysis of the Results of Folding Simulations on the Alanine-Based α-Helix 3K(II)

		1	2	3	4	5	av
Structural Data for the Lowest-rmsd Structures							
rmsd[a]	Still	2.44	4.58	4.33	2.55	0.32	**2.84**
	HCT	4.59	0.43	0.26	3.12	3.52	**2.38**
	mAGB	3.23	4.82	4.91	3.41	0.29	**3.33**
	GBMV2	3.85	3.64	4.36	4.08	4.24	**4.03**
RG[b]	exptl	8.12					
	Still	9.34	10.08	8.22	8.88	8.03	**8.91**
	HCT	9.46	8.13	8.05	8.33	7.49	**8.29**
	mAGB	8.16	8.04	9.33	9.45	8.01	**8.60**
	GBMV2	8.48	10.57	11.12	9.17	9.64	**9.80**
HB[c]	exptl	14					
	Still	8/8	3/0	5/0	6/5	11/9	
	HCT	3/0	12/12	13/13	4/2	8/7	
	mAGB	5/4	2/0	3/0	4/3	12/12	
	GBMV2	2/1	3/1	2/0	2/0	2/1	
Clustering Analysis							
NC[d]	Still	3/28	5/32	8/40	1/35	6/47	**4.6/36**
	HCT	8/29	8/31	6/22	7/27	6/36	**7.0/29**
	mAGB	7/41	6/40	5/42	6/40	8/37	**6.4/40**
	GBMV2	3/15	0/0	0/7	0/1	0/30	
PERC[e] (%)	Still	11/35	53/80	47/79	7/43	26/65	**28/60**
	HCT	57/83	45/75	68/86	58/82	45/80	**55/81**
	mAGB	38/73	37/71	28/71	27/65	45/73	**35/71**
	GBMV2	12/24	0/0	0/5	0/1	0/28	

[a] rmsd: the lowest root-mean-square deviation of all C_α atoms with respect to the native structure, which here is the ideal α-helix. [b] RG: radius of gyration of the lowest-rmsd structure. [c] HB: the number of hydrogen bonds in the lowest-rmsd structure. [d] NC: number of large clusters with more than 1000 structures and number of meaningful clusters with more than 200 structures. [e] PERC: percentage of population in all large clusters and in all meaningful clusters.

to form compact structures. The GBMV2 model also failed to generate meaningful clusters.

Conformational Ensembles. As shown in Table 4, effective energies using any of the three simple GB models can generally distinguish native-like and non-native structures. As an example, in the Still model, the largest cluster in simulation 5 has 9 native hydrogen bonds and an average energy of −344.6 kcal mol^{-1}. This value is 5.4 kcal mol^{-1} lower than that of the largest cluster formed in the simulation 4 which has 6.3 native hydrogen bonds and 7.5 kcal mol^{-1} lower than the largest cluster formed in simulation 1 which has 2.8 native hydrogen bonds. The two structures without native hydrogen bonds have much higher energies. The same general trends are observed for the other GB models. Figure 6d−f shows three examples of misfolded structures from the Still model (simulation 2), the HCT model (simulation 5), and the mAGB model (simulation 2). They are representative structures of the largest cluster in each simulation. The HCT model forms a well-packed helix with three turns, as shown in Figure 6d. This structure appears to be energetically favorable due to the well-formed helix and tightly bound terminal residues. In contrast, the misfolded structures obtained with both the Still and mAGB models both contain a three-strand hairpinlike structure. It appears that the HCT model favors helix formation but, as seen above, had difficulty in generating sheet-like structures. The Still and mAGB models appear able to form both types of structures with no strong bias in one direction or another. We did not observe any cases where 3K(I) folded into a π-helix, as reported by Feig et al. using the CHARMM force field[34] and GROMOS96 force field[41] in the gas phase.

Discussion

Force Field and Solvation Model. The widespread use of molecular mechanics simulation techniques and continuum solvent models attests to the major impact these methods have had in chemical and biological applications. However, the results that are obtained can be quite sensitive to the force field and solvent model that are used even though standard methods are based on similar physical models and attempt to capture the same underlying physical interactions. The sensitivity to parameters requires that the extent to which a particular result is truly robust needs to be carefully considered in different

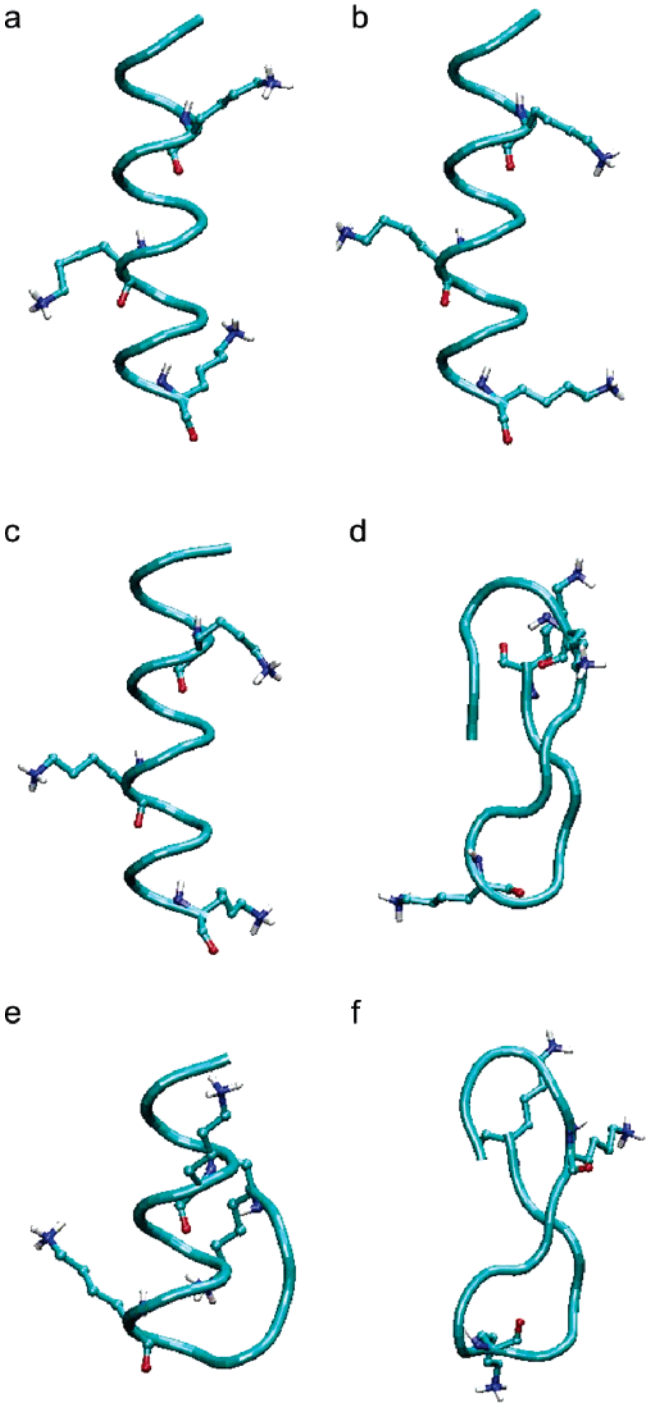

Figure 6. Selected low-rmsd structures obtained from folding simulations on the 3K(I) peptide (see text and Tables 3 and 4 for details). The peptide backbone is shown as a tube, and the lysines are shown in ball-in-stick: (a) simulation 5 with the Still GB model; (b) simulation 3 with the HCT GB model; (c) simulation 5 with the mAGB model; (d) a representative structure of the largest cluster in simulation 2 with the Still GB model; (e) a representative structure of the largest cluster in simulation 5 with the HCT model; (f) a representative structure of the largest cluster in simulation 2 with the mAGB model.

TABLE 4: Properties of the Largest and Second-Largest Clusters Obtained in the Folding of the α-Helix[a]

		1	2	3	4	5	av[b]
Average over the Largest Cluster							
rmsd[c]	Still	5.24	7.45	7.87	4.16	0.86	**5.12**
	HCT	6.55	1.82	0.76	5.02	6.13	**4.06**
	mAGB	7.47	7.50	7.44	7.34	4.42	**6.83**
RG[d]	Still	6.51	7.25	7.41	7.54	8.22	**7.39**
	HCT	6.25	8.08	8.13	6.63	6.54	**7.13**
	mAGB	7.21	6.96	6.63	6.65	7.51	**6.99**
HB[e]	Still	5.3/2.8	4.0/0	5.3/0	7.4/6.3	9.7/9.0	
	HCT	5.5/0.7	9.3/8.7	10.9/10.3	7.0/2.1	7.5/6.5	
	mAGB	5.6/1.5	5.1/0	4.2/0.2	4.3/0	7.7/6.5	
ENE[f]	Still	−337.1	−333.7	−331.1	−339.2	−344.6	
	HCT	−315.3	−320.1	−326.0	−317.3	−321.6	
	mAGB	−318.3	−320.8	−314.8	−318.1	−325.4	
Average over the Second-Largest Cluster							
rmsd	Still	4.60	7.55	7.93	6.64	5.44	**6.43**
	HCT	6.35	5.33	1.79	5.06	6.70	**5.05**
	mAGB	6.96	7.42	6.38	6.41	0.85	**5.60**
RG	Still	6.64	7.55	7.28	6.96	6.74	**7.03**
	HCT	6.55	6.94	8.24	6.07	6.92	**6.94**
	mAGB	6.78	7.24	6.52	6.26	8.19	**7.00**
HB	Still	7.0/3.1	4.0/0.1	5.0/0	5.1/2.8	6.3/5.1	
	HCT	5.1/0.4	7.4/6.4	9.8/9.0	7.1/2.4	5.3/3.1	
	mAGB	5.8/3.2	4.9/0.2	4.7/0.6	5.5/1.1	10.4/9.8	
ENE	Still	−337.3	−330.1	−329.4	−333.6	−336.2	
	HCT	−312.1	−320.0	−321.2	−319.2	−314.8	
	mAGB	−321.1	−319.4	−319.2	−320.7	−332.1	

[a] The GBMV2 model failed to produce meaningful clusters; thus, only results from the Still model, the HCT model, and the mAGB model are given in this table. [b] av: averages calculated for rmsd and RG over five simulations. [c] rmsd: average C_α root-mean-square deviation (rmsd) with respect to the native structure, averaged over all member structures. [d] RG: average radius of gyration (RG) with respect to the native structure, averaged over all member structures. [e] HB: average number of hydrogen bonds (the total number/number of native hydrogen bonds) [f] ENE: average effective energy, which is defined as the sum of GROMOS96 internal energy and the GBSA solvation free energy, averaged over all member structures.

applications. This is particularly true as applications increasingly turn to the description of rather subtle phenomena, for example, the role of a particular interaction on binding specificity or on the details of a folding pathway. While it is not generally feasible to test every force field and solvent model in every application, it is possible to consider the effects of different parameters and models so as to develop a basis for deciding whether a particular result is robust. This type of understanding also leads to the possibility of using different energy descriptions in different applications, not randomly but rather based on quite specific goals. A number of recent studies have described the performance of different force fields.[32,35−38] The major focus of this study is the properties of different GB solvent models. We have studied the ability of four GB methods to sample conformational space and to reproduce native-like structures in the ab initio folding of simple peptides. A second paper in this series (Fan, Mark, Zhu and Honig, in preparation) describes the effect of different GB models on MD simulations of native proteins.

To obtain results that are primarily sensitive to the solvent model, we used only a single force field (GROMOS96[41]) and a single sampling technique (standard SD simulation). In addition, to compare different models in an unbiased fashion, all four GB models studied here were implemented in a single program (GROMOS96), although TINKER[44] was used as well in order to test force fields other than GROMOS96. The mAGB model is a simplified version of the AGB model of Levy and co-workers[26] and is reported here for the first time. In the accompanying paper, the Still, HCT, and mAGB models were implemented into the GROMACS MD and applied to protein dynamics. The availability of the different models within the computationally efficient GROMACS program may enable the use of more than one solvent model in future applications.

Born Radii. We first studied the correlation of Born radii produced by GB models with those obtained from PB calculations, which tests the ability of each model to reproduce the predictions of the continuum model upon which they are based. As expected, the GBMV2 model produces by far the best correlations while the other models always underestimate the Born radii of buried atoms and thus overestimate their solvation energies. However, the success of the GBMV2 model comes with a high computational price; it is about 12 times slower than the other models.

***β*-Hairpin Folding.** The first peptide studied in this work contains 16 residues that correspond to residues 41−56 in the B1 domain of protein G. This region forms a *β*-hairpin in the native protein, and a similar structure was observed in the seminal NMR study of Blanco et al.[52,53] However, the structure of the peptide is not identical to that of the corresponding region in the native protein and, in addition, the *β*-hairpin is only ∼40% populated in the peptide. Most theoretical studies have compared results to the NMR structure observed for the entire domain which contains eight backbone−backbone hydrogen bonds as defined by DSSP[78] within the hairpin structure formed by residues 41−56 (six are between the antiparallel strands). It is not entirely clear that all these hydrogen bonds exist within the entire ensemble of native-like structures or that they all exist in a single conformation. It is worth pointing out that others[32,54,55,79] have used a somewhat different definition of hydrogen bonds in their folding studies, and this leads to five hydrogen bonds that they use to define the native state.

Previous studies of peptide folding have included MD folding simulations and calculations of potentials of mean force (PMFs). The former can explore folding pathways directly, while PMF calculations explore the conformational free energy landscape from which information about folding mechanisms can be inferred. All previous studies used enhanced sampling techniques such as ensemble dynamics,[57] generalized ensemble algorithms,[32,54,55,80] Tsallis statistics methods,[81−84] and multiple-copy simulated annealing[85] to improve sampling efficiency. As Jang et al.[81] have pointed out, enhanced sampling technique limits the extent to which the description of folding pathways may be viewed as realistic. In addition, sampling methods may obscure the properties of individual force fields by moving past energy barriers that are intrinsic to a particular set of local-geometry parameters.[35,36] The folding dynamics studies of Pande and co-workers were based on ensemble dynamics, the OPLSAA force field, and the Still GB model.[57] The peptide was found to fold to conformations within ∼3 Å rmsd relative to the native structure and with a radii of gyration fluctuating around 7.0 Å. Jang et al.[86] used MD simulations with the CHARMM19 force field[74] and a modified version[76] of the Still GB model to study the folding of this peptide. Although hairpin-like structures were formed, rmsd values were not reported. Liu and Beveridge[85] used multiple-copy simulated annealing dynamics, the AMBER force field,[87] and the HCT model.[22,30] A 4.5 Å rmsd structure was reported.

The folding simulations described in this work were carried out without the use of enhanced sampling methods. Nevertheless, low-rmsd conformations are generated with every GB model and a very low rmsd conformation with a significant

number of native hydrogen bonds was generated using the Still solvation model within the 20 ns simulation time. Moreover, this structure belongs to a large cluster, indicating that sampling native-like conformations is not a problem with the GROMOS96 force field and the Still GB model. Since previous studies on the same peptide have also used the Still model, this suggests that our results were due to our use of the GROMOS96 force field. This was confirmed with our folding simulations using the OPLSAA force field and the Still GB model which indicated that OPLSAA is less effective than GROMOS96 in sampling conformational space (an effect attributed to the different inherent rigidity of the two force fields[36]).

As can be deduced from Table 2, there is no clear correlation between rmsd and energy for any of the solvent models. Some of the large clusters that are formed using the Still or mAGB models contain a significant number of hydrogen bonds and radii of gyration that are consistent with experiment (see Table 2, results marked in bold), and for these two models (as opposed to HCT and GBMV2), the native structure is found to be lower in energy than any of the low-rmsd structures: $\sim$1 kcal mol^{-1} lower for the Still model and $\sim$3 kcal mol^{-1} lower for the mAGB model. On the basis of the figures (Figure 4a−e), the formation of the hairpin-like structures appears to be driven in large part by hydrophobic interactions, although there are obviously alternate ways of simultaneously forming hydrophobic clusters and backbone−backbone hydrogen bonds.

In a series of papers, Zhou and Berne[32,54,55] investigated the folding free energy landscape of the hairpin peptide produced with an explicit water model and the OPLSAA force field, the SGB model[23] with OPLSAA, and different versions of AMBER used with the Still GB model. They found that the native-like structure corresponds to the global minimum in the explicit water simulations, but the native-like structure was not the lowest with some combinations of force fields and GB models. Levy and co-workers[79] carried out related calculations using the AGBNP model and the OPLSAA force field and found that the free energy basin is very flat for this peptide. Given experimental uncertainties as to the details of the structure of the isolated peptide in solution, it is difficult to draw strong conclusions as to the accuracy of a particular solvent model and force field. Nevertheless, the fact that we can generate a significant number of clusters with low rmsds, a number of native hydrogen bonds, and a native-like radius of gyration suggests that the GROMOS96 force field used with different solvent models does provide a reasonably accurate description of the conformational energetics of the hairpin peptide.

It should be emphasized that the failure to fold does not imply any particular limitations in the accuracy of a given model. Indeed, as shown above (Figure 2d and e), the GBMV2 model is most effective in reproducing the more rigorous PB results and, in this sense, is probably the most accurate of the four models that were tested. It is of interest that in the 2 ns SD simulations of the hairpin peptide that start from the native structure, the GBMV2 model keeps 78% of the structures below 1.5 Å as compared to 51% for the Still model, 36% for the mAGB model, and 34% for the HCT model. This is probably due to its less effective sampling efficiency, but it may also be related to model accuracy in that GBMV2 appears to have somewhat larger energy gaps from misfolded structures than do the other models. If so, GBMV2 may be particularly suitable for the terminal refinement once a structure is close to the native state. Preliminary results from our lab suggest that this is indeed the case, although the ability to draw firm conclusions will require additional work.

α-Helix Folding. Our simulations on the 3K(I) peptide are consistent with the conclusions drawn from the β-hairpin peptide. For the case of 3K(I), it is clearly possible to fold from an extended structure into correct α-helical conformations without the use of enhanced sampling techniques. All solvation models with the exception of GBMV2 are successful in this regard. It is of interest that standard SD simulations using GROMOS96 and either the Still or mAGB model are able to sample conformational space for the helical peptide as effectively as found by Pak et al.[81] using the *q*-jumping MD methods together with the CHARMM22 force field and the Still GB model. (It is also of interest that we do not observe any π-helices in our simulations despite evidence that GROMOS96 favors π-helical conformations in the gas phase.[34]) Consistent with the results of Levy and co-workers on another helical peptide,[79] there is a deep energy well in the α-helical region that has a large energy gap relative to alternate conformations. This is very different from what is observed for the hairpin peptide.

Perspective. The fact that a single force field combined with different solvent models can fold both a helical and β-hairpin suggests that our simulations capture many of the important elements of peptide folding. GROMOS96 is clearly very effective for this problem and has the additional benefit that it is a united-atom force field, which reduces computational costs. Our results clearly indicate that GROMOS96 used together with the Still or mAGB solvent model is an efficient and effective means of sampling conformational space. The HCT model seems less effective than the Still model in terms of sampling efficiency and moreover, based on its preference for overly dense conformations, seems less accurate than both the Still and mAGB models in describing the conformational energetics of peptides in solution. GBMV2 is clearly the least effective solvent model in terms of sampling, but given its ability to reproduce PB-based effective Born radii and keep the native structure stable, it may well prove to be the most effective of the four models in terms of detailed energy refinement.

It is important to emphasize that, strictly speaking, our comparison of different solvent models is valid only for the GROMOS96 force field. However, although the features we describe do appear to be quite general, a number of caveats should be pointed out. First, although we do report folding simulations using with OPLSAA force field, we have not carried out a systematic study of how different force fields affect the performance of other GB models. Second, we used the atomic radii from the work of Zhu et al.[71] without reparametrization. Since both PB and GB calculations are sensitive to the atomic radii, it may be possible to obtain improved performance though an adjustment of parameters that affects the balance of the internal energy and solvation energy for a particular force field. Third, we used a surface-area-based nonpolar term and used the same set of surface-area coefficients for all GB models. It is impossible that improved performance could be obtained through the use of model-specific nonpolar terms or with other physical models such as that suggested by Gallicchio et al.[88,89] Finally, where possible, we have used the GB parameters reported in these original papers, although in principle these could be rederived for a particular application and force field.

We believe that our study is the first of its kind that compares different GB models in a consistent fashion. Moreover, all models have been implemented within GROMOS96 and all but GBMV2 have been implemented in GROMACS, and thus should now be easily accessible to other researchers. In addition to the conclusions about the properties of different solvent

models, our study also suggests a general strategy for using the strengths and weaknesses of each model in different applications. For example, it might be reasonable to explore strategies that use GROMOS96 and the Still model for sampling and then OPLSAA or CHARMM, perhaps in conjunction with GBMV2, for refinement. The approach we have described here should facilitate more flexible strategies that have been possible in the past which can be based on an improved understanding of the properties of different solvent models and force fields.

Acknowledgment. We thank Professor Wilfred van Gunsteren for providing us with the GROMOS96 package. J.Z. gratefully thanks Dr. Michael S. Lee for his generous help in implementing the GBMV2 model. J.Z. gratefully thanks Prof. Haiyan Liu, School of Life Sciences, USTC, China, for valuable discussions and suggestions. The work was supported by NIH grant GM-30518.

References and Notes

(1) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
(2) Gilson, M. K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 7.
(3) Felts, A. K.; Gallicchio, E.; Wallqvist, A.; Levy, R. M. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 404.
(4) Wallqvist, A.; Gallicchio, E.; Felts, A. K.; Levy, R. M. Detecting native protein folds among large decoy sets with the OPLS all-atom potential and the surface generalized born solvent model. *Computational Methods for Protein Folding*; 2002; Vol. 120, p 459.
(5) Zhu, J.; Zhu, Q. Q.; Shi, Y. Y.; Liu, H. Y. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 598.
(6) Feig, M.; Brooks, C. L. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 232.
(7) Honig, B.; Sharp, K.; Yang, A. S. *J. Phys. Chem.* **1993**, *97*, 1101.
(8) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
(9) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435.
(10) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128.
(11) Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2001**, *105*, 6507.
(12) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. *Comput. Phys. Commun.* **1995**, *91*, 57.
(13) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608.
(14) Bashford, D. An object-oriented programming suite for electrostatic effects in biological molecules. In *Scientific Computing in Object-Oriented Parallel Environments*, volume 1343 of Lecture Notes in Computer Science; Ishikawa, Y., Oldehoeft, R. R., Reynders, J. V. W., Tholburn, M., Eds.; Springer: Berlin, 1997; pp 233−240.
(15) Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591.
(16) Prabhu, N. V.; Zhu, P.; Sharp, K. A. *J. Comput. Chem.* **2004**, *25*, 2049.
(17) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244.
(18) Gilson, M. K.; Honig, B. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 5.
(19) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(20) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
(21) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578.
(22) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.
(23) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983.
(24) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *J. Chem. Phys.* **2002**, *116*, 10606.
(25) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1348.
(26) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479.
(27) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588.
(28) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297.
(29) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.
(30) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122.
(31) Alexov, E. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 94.
(32) Zhou, R. H. *Proteins: Struct., Funct., Genet.* **2003**, *53*, 148.
(33) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 265.
(34) Feig, M.; MacKerell, A. D.; Brooks, C. L. *J. Phys. Chem. B* **2003**, *107*, 2831.
(35) van der Spoel, D.; Lindahl, E. *J. Phys. Chem. B* **2003**, *107*, 11178.
(36) Hu, H.; Elstner, M.; Hermans, J. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 451.
(37) Mu, Y. G.; Kosov, D. S.; Stock, G. *J. Phys. Chem. B* **2003**, *107*, 5064.
(38) Price, D. J.; Brooks, C. L. *J. Comput. Chem.* **2002**, *23*, 1045.
(39) Xiao, L.; Honig, B. *J. Mol. Biol.* **1999**, *289*, 1435.
(40) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994**, *3*, 211.
(41) van Gunsteren, W. F. B., S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Groningen Molecular Simulation (GROMOS) System*; University of Groningen: The Netherlands, ETH Zurich, Switzerland, 1996.
(42) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43.
(43) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.
(44) Ponder, J. W. *TINKER-software tools for molecular design*, version 3.7; Washington University: St. Louis, MO, 1999.
(45) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(46) Jorgensen, W. L.; McDonald, N. A. *THEOCHEM* **1998**, *424*, 145.
(47) McDonald, N. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 8049.
(48) Price, M. L. P.; Ostrovsky, D.; Jorgensen, W. L. *J. Comput. Chem.* **2001**, *22*, 1340.
(49) Jorgensen, W. L.; TiradoRives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1666.
(50) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
(51) Munoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5872.
(52) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584.
(53) Blanco, F. J.; Serrano, L. *Eur. J. Biochem.* **1995**, *34*, 634.
(54) Zhou, R. H.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931.
(55) Zhou, R. H.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777.
(56) Zhou, Y. Q.; Linhananta, A. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 154.
(57) Zagrovic, B.; Sorin, E. J.; Pande, V. *J. Mol. Biol.* **2001**, *313*, 151.
(58) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 345.
(59) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2544.
(60) Bryant, Z.; Pande, V. S.; Rokhsar, D. S. *Biophys. J.* **2000**, *78*, 584.
(61) Kolinski, A.; Ilkowski, B.; Skolnick, J. *Biophys. J.* **1999**, *77*, 2942.
(62) Roccatano, D.; Amadei, A.; Di Nola, A.; Berendsen, H. J. C. *Protein Sci.* **1999**, *8*, 2130.
(63) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062.
(64) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068.
(65) Smythe, M. L.; Nakaie, C. R.; Marshall, G. R. *J. Am. Chem. Soc.* **1995**, *117*, 10555.
(66) Grant, J. A.; Pickup, B. T. *J. Phys. Chem.* **1995**, *99*, 3503.
(67) van Gunsteren, W. F.; Berendsen, H. J. C.; Rullmann, J. A. C. *Mol. Phys.* **1981**, *44*, 69.
(68) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(69) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(70) Shi, Y. Y.; Wang, L.; van Gunsteren, W. F. *Mol. Simul.* **1988**, *2*, 217.
(71) Zhu, J. A.; Shi, Y. Y.; Liu, H. Y. *J. Phys. Chem. B* **2002**, *106*, 4844.
(72) Zhu, J.; Fan, H.; Liu, H. Y.; Shi, Y. Y. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 979.
(73) Zhu, J.; Yu, H. B.; Fan, H.; Liu, H. Y.; Shi, Y. Y. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 447.
(74) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. *CHARMM: The Encyclopedia of Computational Chemistry*; John Wiley & Sons: New York, 1998.
(75) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. *J. Phys. Chem. B* **1997**, *101*, 1190.
(76) Dominy, B. N.; Brooks, C. L. *J. Phys. Chem. B* **1999**, *103*, 3765.

(77) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401.
(78) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(79) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins: Struct., Funct., Bioinformatics* **2004**, *56*, 310.
(80) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96.
(81) Pak, Y.; Jang, S.; Shin, S. *J. Chem. Phys.* **2002**, *116*, 6831.
(82) Pak, Y.; Wang, S. M. *J. Chem. Phys.* **1999**, *111*, 4359.
(83) Tsallis, C. *J. Stat. Phys.* **1988**, *52*, 479.
(84) Curado, E. M. F.; Tsallis, C. *J. Phys. A: Math. Gen.* **1991**, *24*, L69.
(85) Liu, Y. X.; Beveridge, D. L. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 128.
(86) Jang, S.; Shin, S.; Pak, Y. *J. Am. Chem. Soc.* **2002**, *124*, 4976.
(87) Pearlman, D. A. C.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; Fergusen, D. M.; Siebel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 4.1*; University of San Francisco: San Francisco, CA, 1995.
(88) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517.
(89) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem. B* **2000**, *104*, 6271.