Conformational Sampling with Implicit Solvent Models: Application to the PHF6 Peptide in Tau Protein

Austin Huang and Collin M. Stultz

Harvard-MIT Division of Health Science and Technology, MIT Department of Electrical Engineering and Computer Science, and MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, Massachusetts

ABSTRACT　Implicit solvent models approximate the effects of solvent through a potential of mean force and therefore make solvated simulations computationally efficient. Yet despite their computational efficiency, the inherent approximations made by implicit solvent models can sometimes lead to inaccurate results. To test the accuracy of a number of popular implicit solvent models, we determined whether implicit solvent simulations can reproduce the set of potential energy minima obtained from explicit solvent simulations. For these studies, we focus on a six-residue amino-acid sequence, referred to as the paired helical filament 6 (PHF6), which may play an important role in the formation of intracellular aggregates in patients with Alzheimer's disease. Several implicit solvent models form the basis of this work—two based on the generalized Born formalism, and one based on a Gaussian solvent-exclusion model. All three implicit solvent models generate minima that are in good agreement with minima obtained from simulations with explicit solvent. Moreover, free-energy profiles generated with each implicit solvent model agree with free-energy profiles obtained with explicit solvent. For the Gaussian solvent-exclusion model, we demonstrate that a straightforward ranking of the relative stability of each minimum suggests that the most stable structure is extended, a result in excellent agreement with the free-energy profiles. Overall, our data demonstrate that for some peptides like PHF6, implicit solvent can accurately reproduce the set of local energy minimum arising from quenched dynamics simulations with explicit solvent. More importantly, all solvent models predict that PHF6 forms extended β-structures in solution, a finding consistent with the notion that PHF6 initiates neurofibrillary tangle formation in patients with Alzheimer's disease.

INTRODUCTION

An appropriate representation of solvent is critical for obtaining physiologically relevant results from biomolecular simulations (1,2,3). The most straightforward approach for modeling solvent is to explicitly include solvent molecules in molecular dynamics (MD) simulations. However, molecular simulations with explicit solvent increase the degrees of freedom in the system and therefore can incur a significant computational cost. Consequently, a number of implicit solvent models have been developed to reduce the computational complexity associated with solvated simulations. Such models modify the potential energy function to reproduce the effects of solvation without explicitly representing solvent atoms (1,2). As simulations with implicit solvent models have led to important insights, these models have gained widespread acceptance in the field of biomolecular simulations (4). As evidence of this, the literature is replete with studies that make conclusions based solely on data obtained from such models (4). Recent studies, however, suggest that implicit solvent models can sometimes lead to results that are at odds with data obtained from explicit solvent simulations and experimental observations (5,6,7). Therefore, it is likely that not all implicit solvent models are appropriate for every application. Moreover, the correct choice of solvent model to use for any given problem likely depends on the system to be studied, whether qualitative or quantitative results are desired, and the degree of accuracy required.

In the study presented here, we explore whether conformational sampling with implicit solvent models can yield results similar to that obtained with explicit solvent simulations. The solvent models that form the basis of this work include: i), an early implementation of the generalized Born (GB) model as described by Brooks and co-workers (8); ii), an alternate implementation of the generalized Born formalism that is based on an integral equation approach and that employs a simple smooth switching function (GBSW) (9); iii), the effective energy function-1 (EEF1) implicit solvent model (10); and iv), the TIP3P model of explicit solvent (11).

The GB model uses a linearized form of Still's equation to estimate the electrostatic component of the solvation free energy (8,12). The equation itself contains six independent parameters that are varied to optimize agreement between GB solvation energies and solvation energies calculated with a finite-difference-Poisson-Boltzmann (FDPB) algorithm (8). As the Born radius is inversely related to the atomic polarization energy, Born radii can be calculated from the GB energies after parameter fitting (8). The model has been widely applied and its utility has been demonstrated in a number of applications (13,14).

The GBSW model, like the GB model, is based on Still's equation; however, GBSW employs a more rigorous integral equation approach to calculate the Born radii. In this method, the electrostatic solvation energy of a given atom is expressed as a sum of two terms—the self-solvation energy in the Coulombic approximation plus a term that accounts for the reaction field (9). Each term is calculated using a surface/volume

integration that employs a smooth switching function at the dielectric boundary to ensure numerical stability during molecular simulations (9). Unlike the GB method, the GBSW model contains two adjustable parameters that dictate the relative importance of the Coulombic field term and the reaction field term (9). As before, the values of these parameters were obtained by minimizing the least-square error between GBSW energies and those calculated with a FDPB approach (9). Once the optimal values of the adjustable parameters are known, the Born radii can be calculated in a straightforward manner. The current implementation of the GBSW algorithm also incorporates a nonpolar contribution to the solvation free energy using the solvent-exposed surface area of the protein of interest, and a user-defined surface tension coefficient. The GBSW model has been used to refine model structures of the C-terminal domain of Hsp33 protein, obtained from sparse NMR data, into native-like folds that matched solved structures (15). In addition, GBSW has been used to examine intermolecular interactions between actin and myosin, leading to new observations regarding a mutation associated with familial hypertrophic cardiomyopathy (16). Overall, the model appears to be applicable to a broad range of problems.

EEF1 estimates the solvation free energy using a Gaussian solvent-exclusion model (10). EEF1 expresses the solvation free energy of a protein as a sum of group contributions, where each contribution is equal to a reference solvation energy (i.e., the solvation energy of the group alone) minus an integral over a solvation free-energy density function. The underlying assumption is that the integral over the free-energy density is well approximated by a sum of Gaussian functions (10). Important aspects of the model are that charged side chains are neutralized and a distance-dependent dielectric is used to further attenuate electrostatic interactions. The model has been used in a number of applications, and interesting results have been obtained. Most notably, EEF1 has been used to calculate unfolding trajectories of proteins (17), discriminate correctly folded from unfolded structures (18), and probe the interactions between regions of α-lytic protease, leading to a better understanding of the relative importance of different interactions in stabilizing the native state (19).

In the study presented here, we address a specific, well-defined problem. We determine whether each of these solvent models can reproduce the set of local energy minima obtained from quenched MD (QMD) simulations with explicit solvent. To this end we perform QMD simulations with each of the aforementioned implicit solvent models and compare these results to those obtained with a TIP3P model of solvent. We note that QMD is a widely used method for locating local energy minima on a given potential surface. The procedure consists of high-temperature MD simulations (typically at 1000 K), followed by minimization of the resulting structures (20). High-temperature simulations ensure that a wide region of conformational space is sampled and the subsequent minimizations assure that only local energy minima are analyzed. Minimization can be performed by cou-

pling the system to a heat bath at 0 K (21,21,22), or by using standard energy minimization algorithms such as steepest descent or conjugate gradients (23). QMD has been used to determine optimal positions and orientations of small functional groups in the binding site of an enzyme (21), estimate the density of states for proteins (24), and study the conformational landscape of peptides and peptide analogs (22,23).

Our studies focus on a six-residue peptide commonly referred to as paired-helical filament 6 (PHF6), which corresponds to the sequence found at the N-terminus of the third microtubule-binding repeat domain of tau protein ([306]VQI-VYK[311]). Tau protein forms intracellular aggregates (also known as neurofibrillary tangles) in patients with Alzheimer's disease (AD), and PHF6 corresponds to the minimal region of tau needed for aggregation to occur in vitro (25,26,27). As the formation of intracellular aggregates may be responsible, in part, for neuronal death in patients with AD, the predominant low energy states of PHF6 are of particular interest (28,29). In performing an analysis of PHF6, the goals of this work are not only to evaluate the ability of several implicit solvent models to reproduce energy minima on a potential surface that explicitly models solvent, but also to determine the most stable conformations of this peptide.

METHODS

Quenched molecular dynamics with explicit solvent

Quenched molecular dynamics consisted of high temperature MD followed by extensive minimization of the structures sampled during the trajectory. A polar hydrogen model of the PHF6 peptide (VQIVYK) was created from the CHARMM19 polar-hydrogen parameter set and initial coordinates for PHF6 were built using the internal coordinate facility, all within CHARMM (30). Both the N- and C-termini of the peptide were patched using NTERM and CTERM patches, as is commonly done, resulting in charged termini (i.e., $-NH_3^+$ and $-COO^-$). The resulting structure was solvated with an equilibrated set of TIP3P water molecules, and waters that overlapped with the peptide or that were outside of a 19 Å radius were removed. A total of 823 water molecules was added to the system. A stochastic boundary setup with a solvent sphere of radius 19 Å was used for these simulations (31). The system was minimized, then heated and equilibrated for 1 ns at 1000 K. Production dynamics were performed for an additional 10 ns at 1000 K. Sampling at this temperature facilitates a broad exploration of the conformational space. The temperature was maintained by weakly coupling (tcoup = 5 ps) the system to a heat bath using the Berendsen method (32). All explicit solvent simulations employed an electrostatic nonbond interaction cutoff of 17 Å, shifted between 14 Å and 16 Å. Switching was used to cut off van der Waals interactions at 16 Å. SHAKE was used to hold hydrogen bond distances close to their equilibrium values and a 2 fs time step was used (33).

Structures were chosen from the trajectory every 10 ps and subsequently minimized, resulting in 1000 distinct minimum energy structures. Minimizations were performed on the entire system consisting of the peptide and all explicit water molecules. In addition, minimizations used the nonbond specifications outlined above and consisted of 2500 steps of steepest descent followed by 2500 steps of conjugate gradient minimization. A root mean-square gradient cutoff of 0.01 kcal/mol/Å was set, such that if the system achieved a root mean-square gradient below this value during the minimization protocol, the minimization was terminated. The procedure for

heating, equilibration, sampling, and minimization was identical for all of the solvent models investigated in this study.

Quenched molecular dynamics in vacuum

Quenched molecular dynamics simulations were performed in vacuum ($\varepsilon = 1$). Comparing the vacuum minima with minima obtained with the different solvent models enabled us to assess the effects of the solvent models on the structure of the peptide. The nonbond cutoffs and the minimization protocol were identical to those used in the explicit solvent simulations.

Quenched molecular dynamics simulations with implicit solvent

We performed a similar procedure for finding local energy minima on the potential energy surface of each implicit solvent model described above. One issue that needs to be resolved is the correct choice of simulation conditions for each implicit solvent model. In general, we rely on prior data to choose simulation conditions that optimize the chance that each implicit solvent simulation would reproduce minima obtained from the explicit solvent simulations. In this regard, we note that some temperature-coupling algorithms may not be appropriate for all implicit solvent simulations (34). In explicit solvent simulations with a Berendsen heat bath, the entire system, consisting of both the solute and the solvent, are coupled to an external heat bath. Implicit solvent simulations that utilize similar thermostats only couple

the peptide to an external bath as a continuum model is used for solvent. It has been noted that some thermostats that couple the solute alone to a heat bath may lead to diminished atomic fluctuations, especially when the peptide itself is tightly coupled (34). Diminished root mean-square (rms) fluctuations would clearly be disadvantageous for an approach that attempts to map local energy minima on a large potential surface.

To determine whether a Berendsen thermostat with a coupling constant of 5 ps would be appropriate for our studies, we conducted MD simulations of PHF6 with each implicit solvent model outlined above and compared these data to simulations conducted with explicit solvent (when both the peptide and solvent are coupled to an external bath). The resulting rms fluctuations were then compared to rms fluctuations arising from the explicit solvent simulations. For PHF6, the rms fluctuations arising from all of the implicit solvent simulations are in reasonable agreement with the rms fluctuations from the explicit solvent simulations (Fig. 1). As we are primarily interested in mapping the local energy minima on the different potential energy surfaces, and not the dynamical properties of PHF6 in different models of solvent, these data suggest that simulations with a Berendsen thermostat would be appropriate for our studies.

Lastly we note that the precise model for the nonbond interactions for each implicit solvent model was chosen based on prior data. The goal here was to optimize the chance that each model would produce data in agreement with the explicit solvent results.

Quenched molecular dynamics with GB: generalized Born simulations utilized the implementation, and Born radii, originally described by
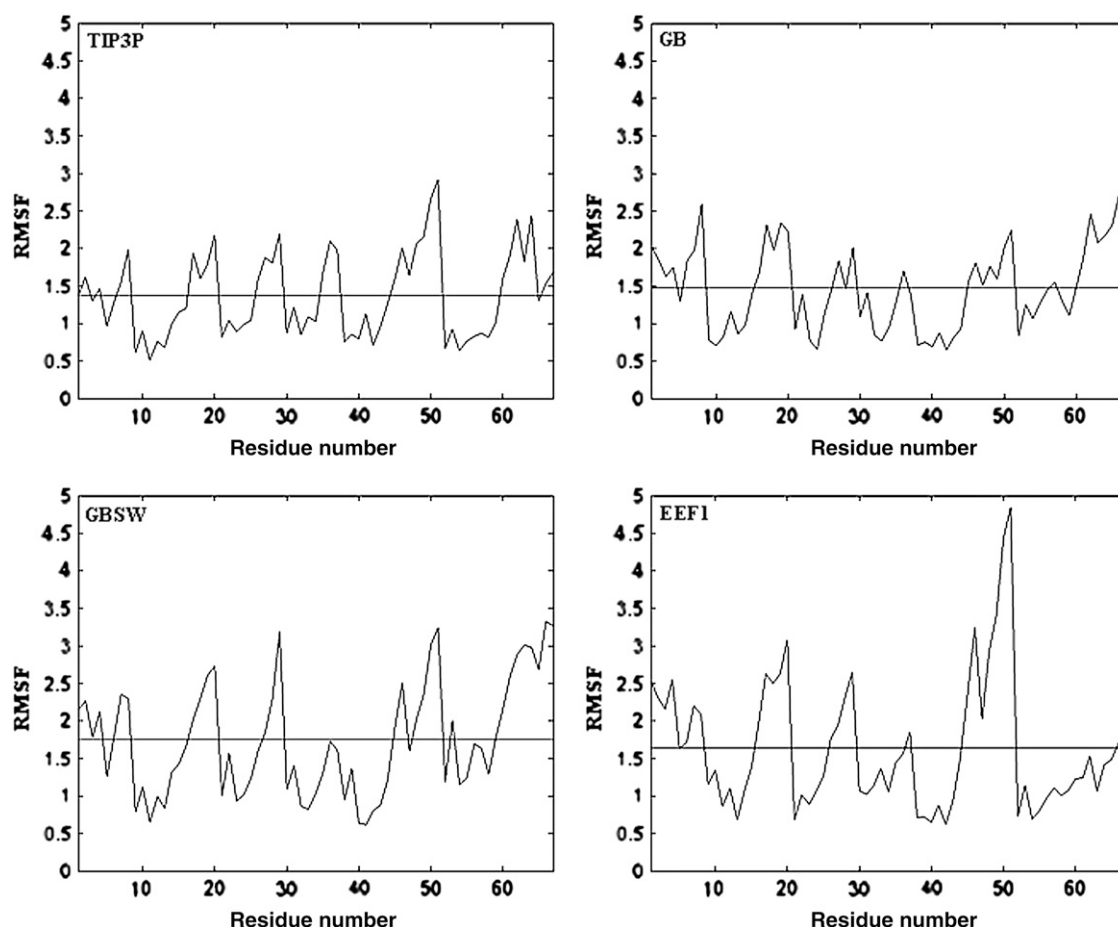

FIGURE 1   RMS fluctuations for 100 ps simulation (100 ps equilibration, 100 ps production dynamics all at 300 K) of PHF6 in different solvent models using a Berendsen heat bath. Simulation parameters, including nonbond cutoffs, are as listed in Methods. The average rms fluctuation over all atoms is denoted with a red line.

Dominy et al. (8). As in our previous study (7), no truncation of nonbond terms was used as this approach yields better results relative to an approach that employs finite nonbond cutoffs (35).

Quenched molecular dynamics with GBSW: GBSW simulations used the implementation previously described by Im et al. with a half smoothing length of 0.3 Å, a nonpolar surface tension coefficient of 0.03 kcal/(mol $\times$ Å^2), and a grid spacing of 1.5 Å (9). Nonbond cutoffs were set to 16 Å using a switching function for both van der Waals and electrostatic interactions. Of the 1000 structures, the minimization protocol described above failed for a single structure, which was excluded from the analysis, yielding 999 distinct structures. The singular failed structure was in a nonphysical conformation corresponding to an energy of 1.3×10^{11} kcal/mol, whereas all other structures had energies that were < -400 kcal/mol.

Quenched molecular dynamics with EEF-1: the EEF-1 implicit solvent model was used as implemented in CHARMM (10,30). As the nonbond cutoff parameters are integral to the model, the previously described nonbond cutoffs were used here.

Generation of Ramachandran plots

Ramachandran density surfaces were created from the minima generated from each of the quenched dynamics simulations. The ϕ/ψ values for residues Gln^2-Tyr^5 were calculated for each of the 1000 minima (999 for GBSW), and a density function was computed using the SCATTERCLOUD function (written by Steve Simon) obtained from the MATLAB central code repository (http://www.mathworks.com/matlabcentral/). The densities were normalized by their maximum values and rendered as surface plots using MATLAB (The MathWorks, Natick, MA). Approximate secondary structure regions as defined in Hovmöller et al. (36) corresponding to α-helical and β-structure are colored.

Generation of minimum pairwise distance plots

Histograms of minimum pairwise backbone rms deviations between minima from different models (a reference model and a comparison model) were computed. These histograms were used to determine whether each minimum in the reference model was adequately represented by a structurally similar minimum in the comparison model. For example, suppose explicit solvent is the reference model and data arising from the EEF1 simulations are the comparison model. The minimum pairwise distance (MPD) plot is used to determine if each explicit solvent minimum is represented in the set of EEF1 minima. For each TIP3P minimum, we find the EEF1 minimum with a backbone conformation closest to the TIP3P minimum in question. This set of rms deviations provides an objective assessment of how well the EEF1 minima reproduce the structures corresponding to the explicit solvent minima. It is also of interest to determine the converse; i.e., whether each EEF1 minimum is well represented by an explicit solvent minimum. The converse is computed by setting EEF1 as the reference model and the explicit solvent results as the comparison model. If EEF1 generated many spurious minima that did not correspond to explicit solvent minima, then the resulting histogram of rms deviations would contain many large values. Therefore, two sets of MPD plots were computed for each of the implicit solvent models. In one set of calculations, the explicit solvent minima formed the reference set, and in the other set of calculations, the implicit solvent model served as the reference. Histograms were computed using MATLAB and plots of aligned structures were constructed with Visual Molecular Dynamics (37).

Potential of mean-force calculations for PHF6

Free-energy profiles for PHF6 were computed for each solvent model. The reaction coordinate for these simulations was the radius of gyration of the peptide main-chain atoms. The simulations began by restraining the backbone to adopt an extended conformation with a radius of gyration of 5.5 Å using a harmonic constraining potential with a force constant of 25 kcal/mol/Å^2. The system was then equilibrated at 300 K for 1 ns. The potential of mean force (pmf) for a given solvent model was calculated by running a series of simulations (windows), where the peptide is restrained to a different radius of gyration using a harmonic force constant of 25 kcal/mol/Å^2. The first window was centered at 5.5 Å and subsequent windows began with the final state from the preceding window. The radius of gyration was decreased by 0.1 Å for each new window. Restrained molecular dynamics for each window involved 20 ps of equilibration followed by 80 ps of production dynamics. Additional dynamics were performed to extend the pmf boundaries and improve sampling for regions of the pmf that exhibited discontinuities. Specifically, windows for extended states of the peptide were run at 0.1 Å intervals for radius of gyration (rgyr) constraints ranging between 5.6 Å and 6.6 Å to extend the boundaries of the pmf.

To compute the potential of mean force, the radius of gyration was computed every 20 fs for each window of dynamics. From these data, a biased probability density, ρ_i^* is computed and the potential of mean force, $W_i(\xi)$, is computed using the relation (31)

$$W_i(\xi) = -k_B T \ln \rho_i^*(\xi) - V_i(\xi) + C_i, \qquad (1)$$

where k_B is Boltzmann's constant, T is the temperature, V_i is the restraining potential for window i, and C_i is a constant. To construct one continuous potential of mean force, the different pmf from each window need to be linked together—a process performed by the program SPLICE (38).

To determine that our pmf had converged, we performed additional simulations for new windows constrained at rgyr that were offset 0.05 Å from the original window constraints and determined that this convergence criterion was satisfied. Our metric for convergence of the pmf was based on the location of the pmf minimum, since this is the primary quantity of interest for this study. Specifically, we required that the location of the pmf minimum changed by < 0.25 Å as the window step size was halved.

Representative structures from the global energy minimum in each pmf were generated by first averaging the structures sampled at the window corresponding to the global energy minimum followed by minimization to the nearest local energy minimum. All molecular figures were constructed with Visual Molecular Dynamics (37).

Calculating vibrational entropies

Vibrational entropies were calculated from the 1000 distinct minima obtained from EEF1 simulations. To ensure that only nonnegative eigenvalues would be generated from the normal mode calculations, each minimum was further minimized using 1000 steps of steepest descent minimization followed by 2500 steps of adopted-basis Newton Rhapson minimization. The corresponding Hessian matrix was then diagonalized to yield the normal modes and their corresponding frequencies. The vibrational entropy for a given minimum was computed as follows (39,40):

$$-TS_{\text{vib}} = \sum_{i=1}^{3N-6} \left(k_B T \ln\left(1 - e^{-h\nu_i/(k_B T)}\right) - \frac{h\nu_i}{e^{h\nu_i/(k_B T)} - 1} \right), \quad (2)$$

where N is the number of atoms in the system, h is Planck's constant, k_B is Boltzmann's constant, and $\{\nu_i\}_{i=1}^{3N-6}$ are the normal mode frequencies. CHARMM was used to create the Hessian matrix from minimized structures and MATLAB was used to calculate vibrational entropies from the Hessian matrix, yielding 1000 vibrational entropy measures; i.e., one for each minimum (30).

We note that a harmonic analysis could only be performed on minima arising from the EEF1 simulations, as second derivative calculations with GB are not supported in CHARMMv32a2, and despite the extensive additional minimization, Hessian matrices for GBSW structures had negative eigenvalues, thereby preventing a normal mode analysis.

TABLE 1   Statistics of minima obtained from quenched molecular dynamics simulations with different solvent models

Solvent model	Average rgyr (Å) $\pm$ SD	Minimum rgyr (Å)	Maximum rgyr (Å)
TIP3P	4.1 $\pm$ 0.63	3.0	5.6
Vacuum	3.1 $\pm$ 0.06	3.0	3.5
GB	4.7 $\pm$ 0.46	3.2	5.7
GBSW	4.5 $\pm$ 0.55	3.1	5.7
EEF1	4.7 $\pm$ 0.47	3.4	5.9

RESULTS

Minimum energy conformations with explicit solvent

Minima on the potential energy surface of PHF6 were obtained from high temperature molecular dynamics simulations with explicit solvent followed by extensive minimization (i.e., quenched dynamics). After 10 ns of molecular dynamics at 1000 K, a range of conformations was sampled and subsequent energy minimization yielded 1000 distinct structures corresponding to different local energy minima. These structures span a range of conformations from the compact, with a rgyr near 3 Å, to a rgyr of almost 5.6 Å (Table 1). By contrast, minima arising from quenched molecular simulations in vacuum are relatively homogeneous and have radii of gyration that are distributed over a narrow range—between 3.0 Å and 3.5 Å, suggesting that compact states are overwhelmingly favored in the vacuum simulations (Table 1).

To quantify the diversity among the different minimum energy structures, we computed the backbone rms deviation between all pairs of minima (Fig. 2). As we are interested in distinguishing extended structures from compact structures, in addition to secondary structural motifs sampled by the peptide, we focus on comparisons of the backbone rms deviation between different pairs of conformers. These data confirm that the explicit solvent minima are considerably more diverse than minima arising from the vacuum simulations. In particular, the most extended structure from the

vacuum simulations has a radius of gyration of only 3.5 Å and contains a salt bridge between the N- and C-termini (Fig. 3). In vacuum, this salt bridge is exceptionally stable in that it has an interaction energy near −90 kcal/mol and remains intact even at 1000 K. Hence virtually all minima have this salt bridge and the resulting vacuum structures are all compact.

Minimum energy conformations with implicit solvent

Minima arising from QMD simulations with implicit solvent sample a range of radii of gyration that is similar to that found in the set of explicit solvent minima (Table 1). A comparison between representative minima from the different solvent models further illustrates the close correspondence between the implicit solvent results and the explicit solvent results (Fig. 4); i.e., the backbone conformations of the implicit solvent minima are similar to that arising from the explicit solvent simulations.

The degree of similarity between the implicit solvent minima and the TIP3P minima was quantified by computing MPD plots. Each MPD plot is a histogram of the minimum pairwise backbone rms deviations between minima from two different models: a reference model and a comparison model. For each minimum in the reference model, the closest minimum in the comparison model is found and used to generate a histogram of rms deviations. For example, in Fig. 5 A, the TIP3P minima is the reference model and the GB minima is the comparison model. These data demonstrate that every explicit solvent minimum is within 1.5 Å of a GB minimum (Fig. 5 A). Fig. 5 A also shows an overlay of the explicit solvent minimum that is farthest away from a GB minimum; even for this worst case, the two minima have very similar backbone conformations. MPD plots for the other implicit solvent models reveal the same trend, i.e., each explicit solvent minimum is within 1.3 Å of a GBSW minimum (Fig. 5 B) and 1.5 Å from an EEF1 minimum (Fig. 5 C).
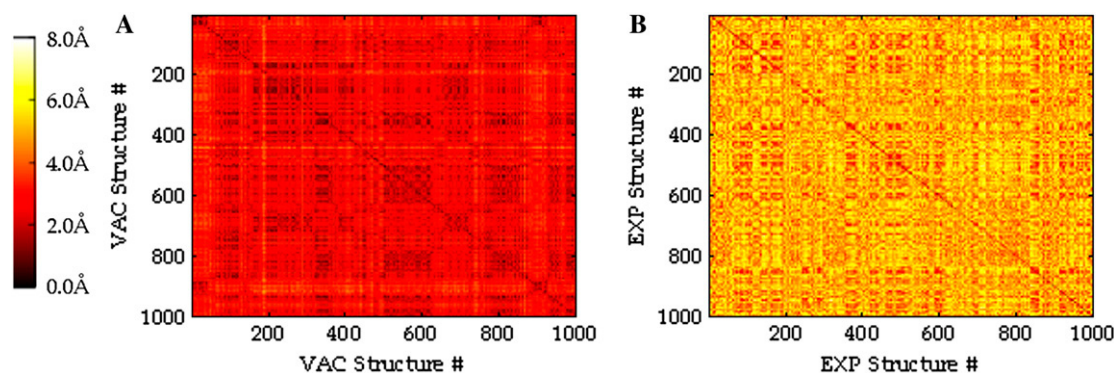
FIGURE 2   Pairwise distance matrices between minimized structures from the (A) vacuum simulations and (B) explicit solvent simulations. Each pixel color corresponds to a pairwise backbone RMS distance. The color scale is shown at the left of the figure.
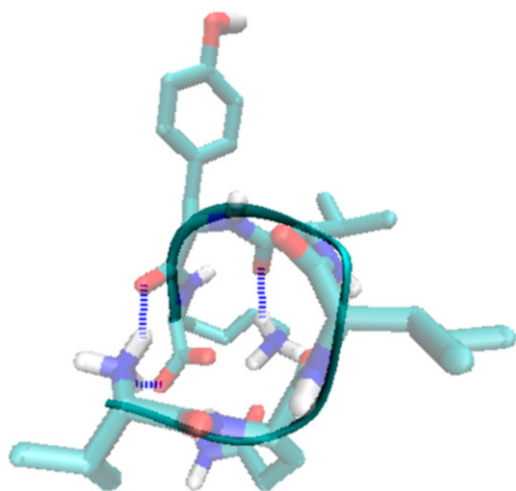
FIGURE 3   Structure of the most extended PHF6 minimum arising from the vacuum simulations (rgyr = 3.5 Å).

done, i.e., MPDs were computed with each implicit solvent minima serving as the reference model and TIP3P serving as the comparison model (Fig. 5, D–F). These data verify that the implicit solvent simulations do not produce many extraneous minima—that is, each implicit solvent minimum is close to an explicit solvent minimum.

A conformational analysis of the TIP3P minima suggests that the four residues in PHF6 with defined ϕ/ψ angles (residues 2–5) preferentially sample regions of conformational space corresponding to β-structure (Fig. 6). Gln^2, in particular, is most likely to adopt ϕ/ψ angles belonging to the β-strand region of conformational space. The ϕ/ψ densities of the GB, GBSW, and EEF1 minima are similar to that obtained from the TIP3P simulations in that β-strand configurations are also favored (Fig. 6). By contrast, the vacuum simulations yield minima where residues 2, 4, and 5 adopt ϕ/ψ angles that belong to the α-helical region of conformational space (Fig. 6).

Potential of mean-force calculations

Free-energy profiles were calculated for PHF6 in explicit solvent to determine the predominant conformation of the peptide in solution (Fig. 7). The reaction coordinate for these simulations was the radius of gyration of the peptide. The

Although every explicit solvent minimum is close to an implicit solvent minimum, it may be that the implicit solvent simulations produce extraneous minima that do not correspond to any minimum arising from the TIP3P simulations. To determine whether the implicit solvent simulations produced such superfluous minima, the reverse comparison was
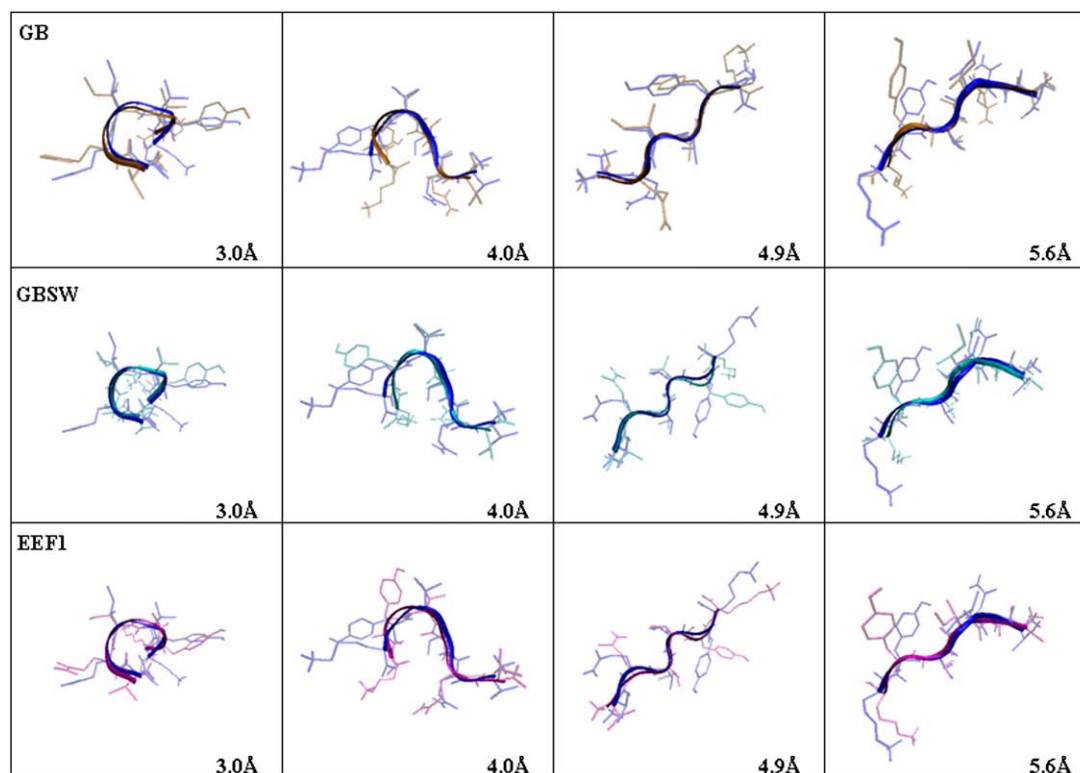

FIGURE 4   Representative explicit solvent structures (*blue*) aligned with their closest implicit solvent structures. The first row depicts the alignment of GB (*orche*) minima to TIP3P minima; the second row shows the alignment of GBSW (*cyan*) minima to TIP3P; and the last row shows the alignment of EEF1 (*purple*) minima to TIP3P minima.
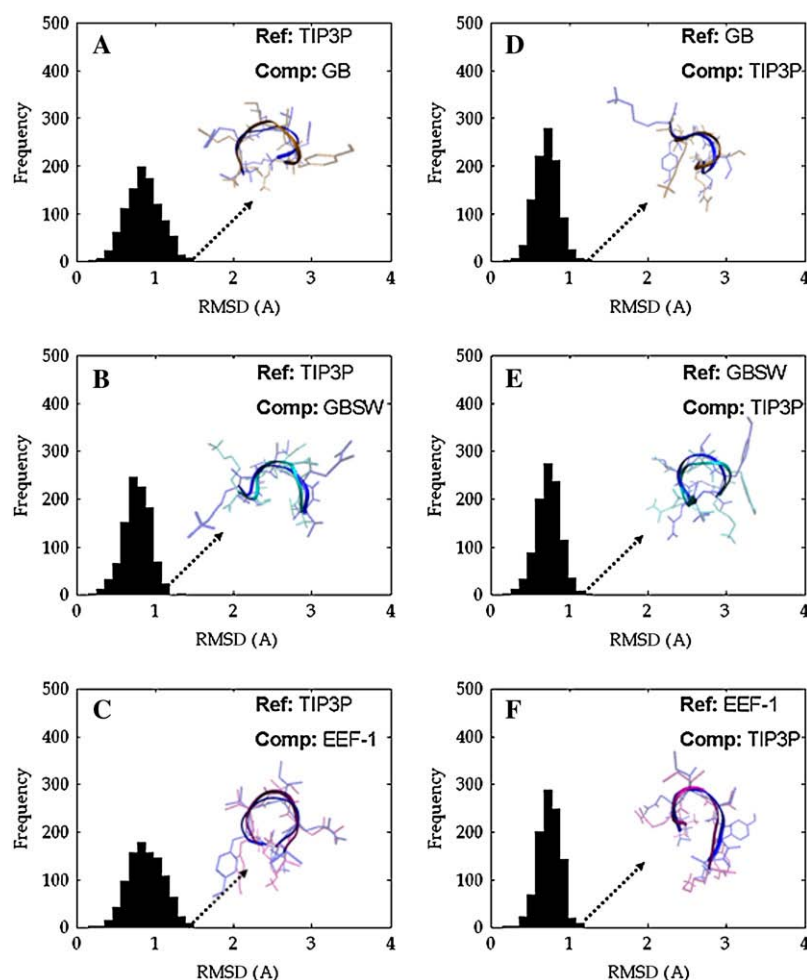
FIGURE 5   MPD plots (see text). The reference and comparison sets are labeled. In each case, the two structures having the greatest RMS difference are overlaid.

global free-energy minimum of the peptide in explicit solvent occurs at ~5.2 Å, corresponding to a relatively extended conformation of the peptide (Fig. 7)—a finding consistent with the ϕ/ψ densities of explicit solvent minima.

The free-energy profiles calculated with each of the implicit solvent models are similar to the pmfs calculated with explicit solvent; i.e., each has a global minimum located between 5 Å and 5.5 Å (Fig. 7). Average structures from windows corresponding to the pmf minima confirm that these low energy structures are relatively extended (Fig. 8). In addition, residues 2–5 from the average structure arising from the explicit solvent pmf minimum have ϕ/ψ angles that fall within a region of conformational space consistent with β-structure. The GBSW average structure, however, is least similar to the average structure from the explicit solvent pmf minimum (Fig. 8). The backbone rms deviation between the GBSW pmf minimum and the TIP3P pmf minimum is ~2.7 Å, whereas the GB and EEF1 structures are within 1 Å of the TIP3P pmf minimum structure (Fig. 8). Hence, whereas all of the implicit solvent models show qualitative agreement with the explicit solvent pmf, the average structure arising from EEF1 simulations at the global free-energy minimum is

most similar to the average structure obtained from corresponding simulations with explicit solvent.

Ranking minima from the implicit solvent models

Ideally, any sampling protocol designed to find low energy states on a potential surface should not only discover local energy minimum, but it should also deduce which of the resulting low energy structures are the most stable. In this regard, we note that EEF1 and potentials based on the generalized born formalism have been shown to correctly identify the most stable protein conformation from sets consisting of native and misfolded structures (18,41,42,43). Moreover, a number of these studies suggest that the most stable state can be deduced from static energy calculations on energy-minimized structures (18,41,43). Given these observations, we explored whether static energy calculations on the different implicit solvent minima could provide enough information for identifying the most stable conformation.

A comparison of the relative energies of the different minima is shown in Fig. 9. Both the GB and GBSW minima have a number of low energy states that are within 2 kcal/mol
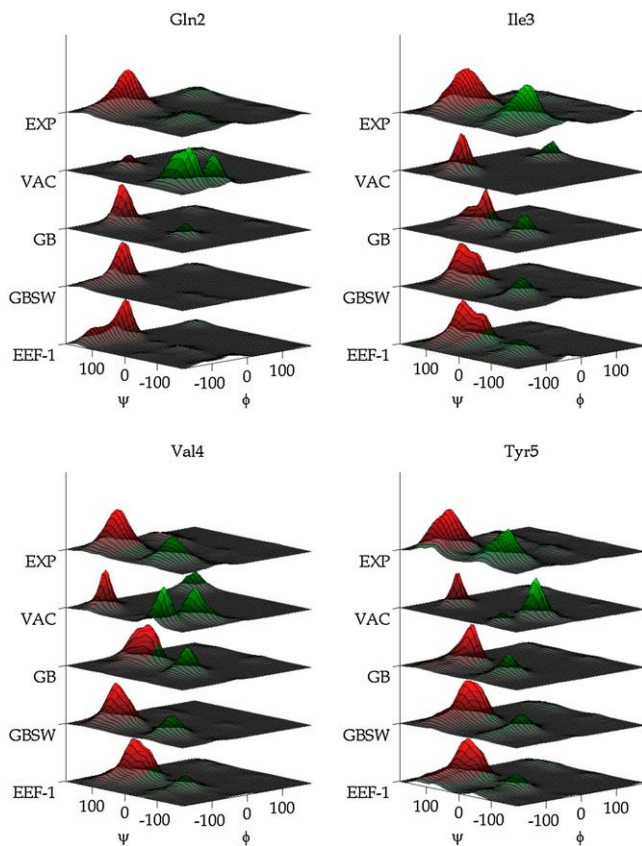
FIGURE 6 Comparison of normalized ϕ/ψ densities of minima obtained by quenched molecular dynamics for residues Gln[2]-Tyr[5]. The region corresponding to the β-structure peak is colored red and the region corresponding to the α-helix peak is colored green. Following the secondary structural definitions used in Hovmöller et al. (36), the region of β-sheet conformations consists of ϕ/ψ angles within the range of $\phi = [-180°, 45°]$ and $\psi = [45°, 225°]$ and the region of α-helix conformations consists of ϕ/ψ angles within the range of $\phi = [-180°, 0°]$ and $\psi = [-100°, 45°]$.

of the lowest energy structure, and all of these conformations are relatively compact with radii of gyration near 3.5 Å (Fig. 9). By contrast, the set of EEF1 minima contains a prominent minimum with a radius of gyration of 5.08 Å, a value close to the global free energy minimum in the EEF1 and TIP3P free energy profiles (Fig. 9). Hence the most stable conformation of PHF6 can be identified from an analysis of the EEF1 energies alone.

We note that methods that identify the most stable conformation of a protein from static energy evaluations on distinct energy-minimized conformers typically assume that the solute entropy at each local energy minimum is roughly the same, and therefore can be ignored (42,44,45). Such approximations may be valid for a number of proteins, but it is not clear whether such a premise is valid for small peptides like PHF6 (44,45). Although static energy calculations with EEF1 lead to results that agree with calculated free energy profiles, this does not necessarily imply that the solute entropy is the same at each minimum. Therefore, to explore

the role that the solute entropy has in determining the relative stability of the PHF6 minima, we computed the vibrational entropy of each EEF1 minimum within the context of a harmonic approximation (40). The relative free energy of each minimum was then estimated using the sum of the internal energy (i.e., the EEF1 energy) and the vibrational entropy (Table 2). Ranking the EEF1 minima using this new measure leads to conclusions that are identical to what was obtained from an analysis of the EEF1 energies alone. In particular, the lowest energy conformations are extended, and the lowest energy structure is the same (Table 2). However, as is clear from Table 2, the vibrational entropy spans a range of more than 10 kcal/mol, a somewhat larger range than was noted in prior studies on proteins (44,45). Including the vibrational entropy also leads to a change in the ranking of the PHF6 minima. Consequently, even though our results are similar to those seen when the vibrational entropy is explicitly included, it is clear that it can play a role in determining the relative ordering of different minima.

DISCUSSION

Given their considerable computational efficiency, a number of problems can be approached with the aid of implicit solvent models that would be intractable if only explicit solvent models were available (13–19). However, not all implicit solvent models are created equal, and some may be more appropriate for particular problems. As such, studies, such as the work presented here, which aim to delineate the limitations as well as the advantages of different implicit solvent models, may help to decide which model to use for any given application.

This study was designed to address a specific question–namely, could selected implicit solvent models adequately reproduce the set of local energy minima found on a potential surface that explicitly includes solvent. Toward this end, we mapped local energy minima on different potential surfaces and compared these minima to minima obtained from simulations with explicit solvent. We found that GB, GBSW, and EEF1 performed quite admirably in that they were able to successfully reproduce the set of minima obtained from explicit solvent simulations. Ramachandran plots of the resulting structures confirm that all solvent models sampled similar regions of conformational space. Furthermore, free-energy profiles obtained from all three implicit solvent models were in good agreement with free-energy profiles obtained with explicit solvent. However, visual inspection of the structures suggests that EEF1 provides a slightly more accurate representation of the most favored conformations on the peptide's free energy surface.

All of the implicit solvent simulations generate pmfs that are in good agreement with the explicit solvent simulations in a fraction of the central processing unit (CPU) time required for the explicit solvent simulations (Fig. 10 A). Of the different implicit solvent simulations, EEF1 required
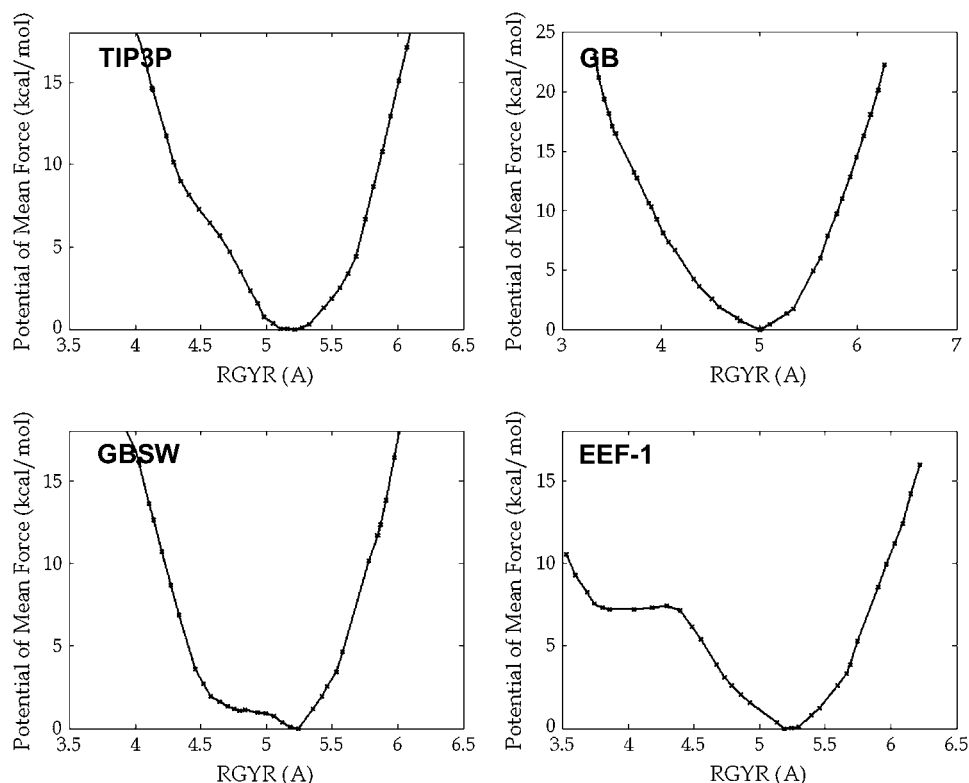
FIGURE 7 Potential of mean force plots for the different solvent models analyzed in this study.
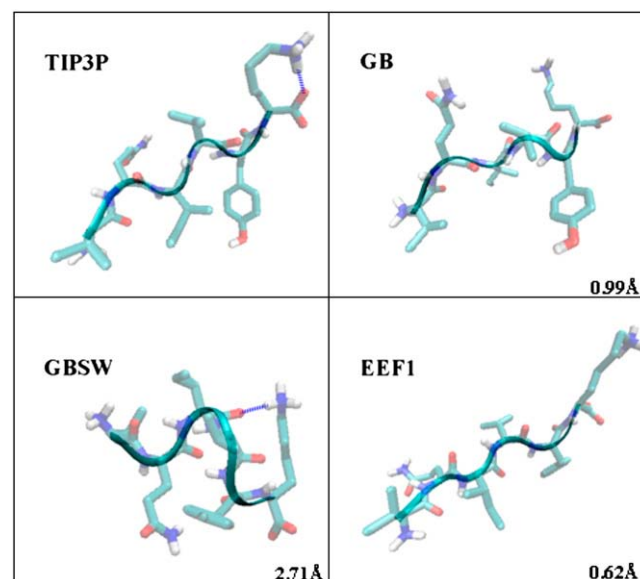

FIGURE 8 Representative structures from the simulation windows corresponding to the global free energy minimum in each pmf. The backbone rms deviation from the TIP3P structure is explicit shown for each of the implicit solvent structures.

the least CPU time (Fig. 10 *B*). This is due, in part, to the different nonbond cutoffs in each model. As the nonbond specifications in EEF1 are part of the model, all EEF1 simulations employ a relatively short cutoff of 9 Å (10). The nonbond cutoffs for the GB and GBSW models were considerably larger. The GB simulations employed an infinite cutoff because it has been shown that this cutoff scheme yields results that are in good agreement with explicit solvent for some systems (35). The GBSW simulations used a finite nonbond cutoff of 16 Å because this value leads to reasonable computation times with relatively small errors in the calculated forces (9). Nevertheless, a 16 Å cutoff for a small peptide like PHF6 leads to almost no truncation of the nonbond terms. As a result, the nonbond lists for the GB and GBSW simulations are quite similar. The longer simulation time for GBSW is due to the fact that, unlike GB, GBSW employs a relatively expensive surface/volume integration to calculate the electrostatic contribution to the solvation energy (8,9).

To determine whether the most stable state of PHF6 could be identified from an analysis of the minima alone without additional umbrella sampling, we examined the relative energies of minima arising from each implicit solvent simulation. The lowest energy structure from the set of EEF1 minima is extended and has a radius of gyration near that found in the free-energy profiles. By contrast, the lowest energy structures from the GB and GBSW simulations are relatively compact. Hence, for PHF6, one could correctly deduce that extended structures are most stable from an analysis of the EEF1 energies alone. These data are encouraging as they suggest that an analysis of minima obtained from simulations with EEF1 may provide insights that are
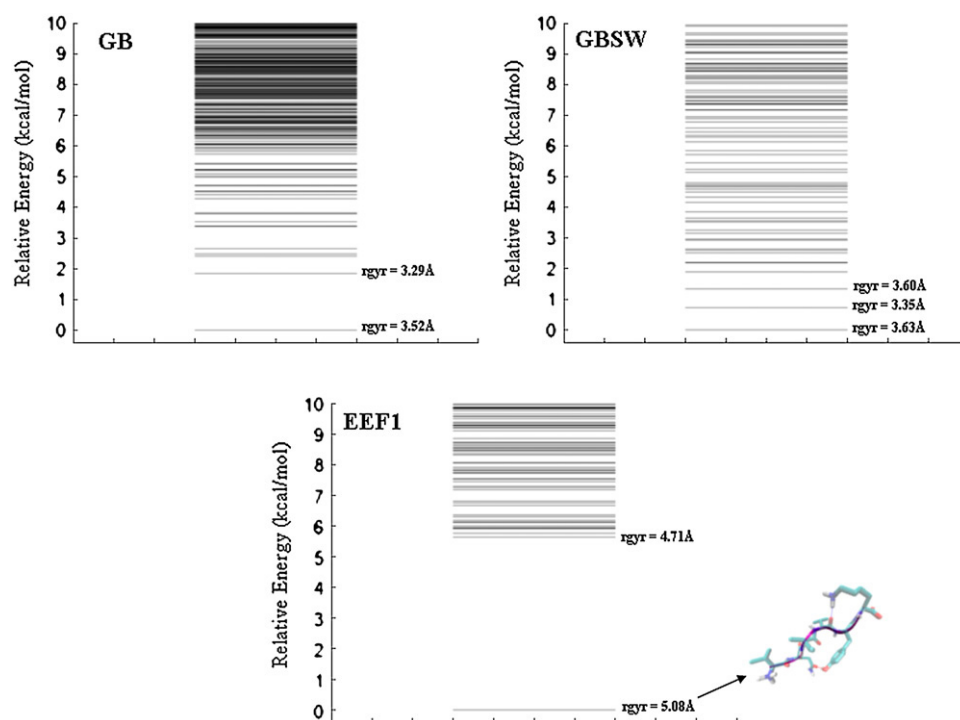
FIGURE 9 Relative energies of minima from each implicit solvent simulation. The radii of gyration of the low energy structures in each solvent model are explicitly shown. The structure of the lowest energy minimum arising from the EEF1 simulations is explicitly shown.

comparable to what one would obtain from umbrella sampling calculations with explicit solvent—a considerably more taxing approach.

It should be noted that this conclusion may not be generally applicable. Ranking EEF1 minima based solely on static EEF1 energies assumes that the solute entropy at each minimum can be safely ignored. However, estimates of the vibrational entropy reveal that the solute entropy can vary significantly at each minimum. Although our conclusions are the same when the vibrational entropy of each minimum is explicitly calculated, the ranking of the EEF1 minima is somewhat altered when this is done. Therefore, we cannot rule out that estimates of the solute entropy are needed to accurately identify the most stable conformation of other peptides. In this regard, we note that static energy evalua-

tions of GB and GBSW minima lead to conclusions that differ from that obtained from the pmf calculations in explicit solvent. As normal mode analyses could not be performed on GB and GBSW minima, it may be that more accurate results could be obtained if a vibrational analysis was performed on these minima.

In our previous study, we found that both EEF1 and GB were unable to reproduce the free-energy profile obtained from simulations with explicit solvent using a different peptide system (7). In that work, we used umbrella sampling calculations with explicit solvent to calculate a peptide's potential of mean force as a function of its radius of gyration (7). The FRET efficiency for this peptide, which was calculated from the pmf, was in excellent agreement with experiment. Central to the success of the explicit solvent simulations was the formation of a stable salt bridge between glutamate 5 and arginine 11. By contrast, in both the GB and EEF1 simulations, the formation of a glutamate-arginine salt bridge was unfavorable, and consequently simulations with these implicit solvent models lead to calculated FRET efficiencies that disagreed with the explicit solvent results (7). Although the solvation energy of individual side chains is likely well modeled by these implicit solvation models, it is not clear that energetics of salt-bridge formation is appropriately modeled by these approaches (7,46). This may be particularly true for salt-bridges that involve arginine residues (46). As such, the absence of multiple charged side chains in the sequence of PHF6 likely explains the difference between the results presented in this work and those of our prior work. For PHF6, representative structures from the

TABLE 2 EEF1/vibrational energies of selected EEF1 minima; minima are ranked in order of increasing energy

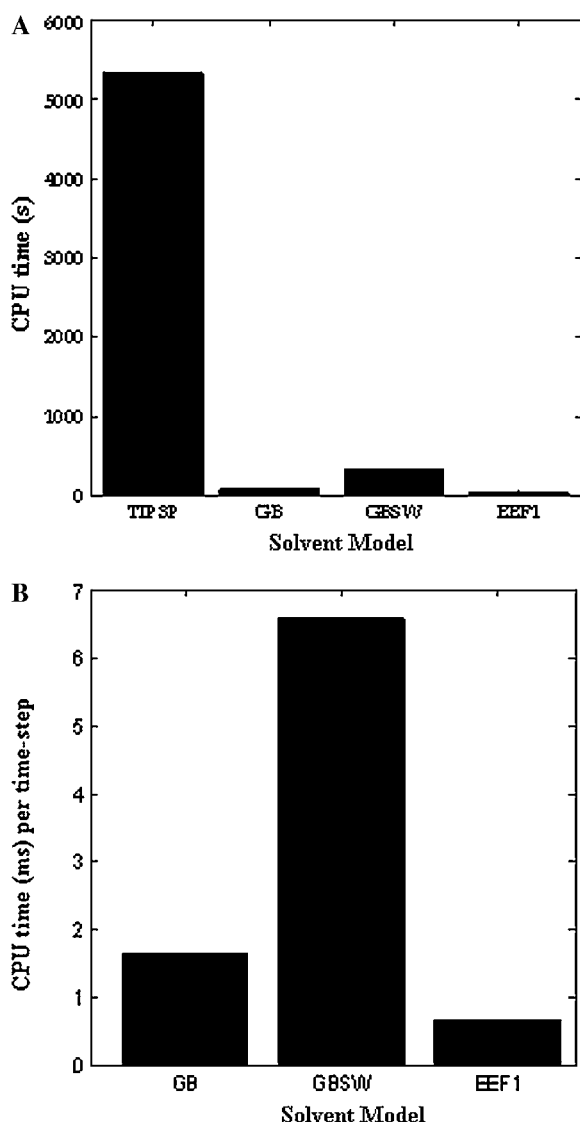| Ranking	rgyr (Å)	E (kcal/mol)	$-TS_{vib}$ (kcal/mol)	$A = E - TS_{vib}$ (kcal/mol)
1	5.08	−191.93	−23.22	−215.15
2	4.90	−177.75	−33.31	−211.06
3	5.16	−185.81	−24.92	−210.73
4	5.03	−186.17	−24.53	−210.70
5	5.10	−185.98	−23.82	−209.80
15	5.04	−186.03	−22.63	−208.66
38	5.01	−177.09	−30.60	−207.69
64	4.45	−185.74	−20.99	−206.73
141	4.46	−172.93	−32.45	−205.37
843	3.99	−175.72	−20.53	−196.25

FIGURE 10 (*A*) CPU time for running one window of pmf simulations in each solvent model. (*B*) Close up of CPU requirements for the various implicit solvent models. All calculations were performed on one XEON 2.8GHz processor running Linux.

lowest energy state within the explicit solvent pmf contain one salt-bridge between the side chain of lysine 6 and the C-terminal carboxyl of the same residue (Fig. 8). Therefore, the explicit solvent pmf suggests that the lowest energy state is extended without any salt bridges or hydrogen bonds between moieties that are separated in the sequence. This simple extended state that lacks salt bridges or hydrogen bonds between distant residues is well modeled by the implicit solvent models investigated in this work.

All of the solvent models predict that PHF6 preferentially adopts extended structures in solution, and a conformational analysis of amino-acids in PHF6 argues that residues 2–5 adopt ϕ/ψ values corresponding to the β-strands. These findings have important implications for the pathogenesis

of neurofibrillary tangle formation in patients with AD. In particular, there is growing consensus that the ability of amyloidogenic proteins like tau to aggregate stems from properties of the protein backbone. In many instances, protein aggregation requires the formation of intermolecular backbone hydrogen bonds yielding a cross β-structure (i.e., the β-strands are perpendicular to the axis of the fibril), and for tau this process is likely important for the initiation of neurofibrillary tangle formation (47,48,49).

Our findings imply that PHF6 exhibits a strong preference for extended β-structures in solution—a finding that suggests that PHF6 promotes neurofibrillary tangle formation by facilitating the formation of cross β-structure between tau monomers. This premise is consistent with recent data suggesting that the sequence of PHF6 is the minimal region of tau required for tau aggregation into cross β-filaments and hence neurofibrillary tangles (25). As neurofibrillary tangle formation may play a role in neurodegeneration (28), therapies directed at modifying the structural preference for PHF6 may lead to new treatments for dementias like AD and the tauopathies (50).

REFERENCES

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

 2. Roux, B., and T. Simonson. 1999. Implicit solvent models. *Biophys. Chem.* 78:1–20.

 3. Brooks, C. L., and M. Karplus. 1989. Solvent effects on protein motion and protein effects on solvent motion. *J. Mol. Biol.* 208:159–181.

 4. Feig, M., and C. L. Brooks. 2004. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* 14:217–224.

 5. Jaramillo, A., and S. J. Wodak. 2005. Computational protein design is a challenge for implicit solvent models. *Biophys. J.* 88:156–171.

 6. Zhou, R., and B. J. Berne. 2002. Can a continuum solvent model reproduce the free energy landscape of a β-hairpin folding in water? *Proc. Natl. Acad. Sci. USA.* 99:12777–12782.

 7. Stultz, C. 2004. An assessment of potential of mean force calculations with implicit solvent models. *J. Phys. Chem. B.* 108:16525–16532.

 8. Dominy, B. N., and C. L. Brooks. 1999. Development of a generalized Born model parameterization for proteins and nucleic acids. *J. Phys. Chem.* 103:3765–3773.

 9. Im, W., S. Lee Michael, and C. L. Brooks. 2003. Generalized Born with a simple smoothing function. *J. Comput. Chem.* 24:1691–1702.

10. Lazaridis, T., and M. Karplus. 1999. Effective energy function for proteins in solution. *Prot. Struct. Func. Gen.* 35:133–152.

11. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating water. *J. Chem. Phys.* 79:926–935.

12. Still, W. C., A. Tempczyk, R. C. Hawley, and T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.

13. Dominy, B. N., and C. L. Brooks. 2001. Identifying native-like protein structures using physics-based potentials. *J. Comput. Chem.* 23:147–160.

14. Rod, T. H., J. L. Radkiewicz, and C. L. Brooks. 2003. Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA.* 100:6980–6985.

15. Chen, J., H. Won, W. Im, H. J. Dyson, and C. L. Brooks. 2005. Generation of native-like protein structures from limited NMR data, modern force fields and advanced conformational sampling. *J. Biomol. NMR.* 31:59–64.

16. Liu, Y., M. Scolari, W. Im, and H. Woo. 2006. Protein-protein interactions in actin-myosin binding and structural effects of R405Q mutation: a molecular dynamics study. *Prot. Struct. Func. Bioinf.* 64: 156–166.

17. Lazaridis, T., and M. Karplus. 1997. ''New view'' of protein folding reconciled with the old through multiple unfolding simulations. *Science.* 278:1928–1930.

18. Lazaridis, T., and M. Karplus. 1998. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J. Mol. Biol.* 288:477–487.

19. Izunuka, Y., and T. Lazaridis. 2000. On the unfolding of α-lytic protease and the role of the pro region. *Prot. Struct. Func. Gen.* 41: 21–32.

20. Bruccoleri, R. E., and M. Karplus. 1990. Conformational Sampling using high-temperature molecular dynamics. *Biopolymers.* 29:1847–1862.

21. Stultz, C. M., and M. Karplus. 1999. MCSS functionality maps for a flexible protein. *Proteins.* 37:512–529.

22. Kisore, A., and R. Kishore. 2002. Folded conformation of an immuno-stimulating tetrapeptide rigin: high temperature molecular dynamics simulation study. *Bioorg. Med. Chem.* 10:4083–4090.

23. O'Connor, S. D., P. E. Smith, F. Al-Obeidi, and B. M. Pettitt. 1991. Quenched molecular dynamics simulations of tuftsin and proposed cyclic analogues. *J. Med. Chem.* 35:2870–2881.

24. Sullivan, D. C., and C. Lim. 2006. Toward absolute density of states calculations for proteins. *J. Phys. Chem. B.* 110:12125–12128.

25. Kumar, V., R. S. Cotran, and S. L. Robbins. 2003. Robbins Pathology, 7th ed. W.B. Saunders, Philadelphia.

26. Von Bergen, M., P. Friedhoff, J. Biernat, J. Heberle, E. M. Mandelkow, and E. Mandelkow. 2000. Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif (306VQIVYK311) forming beta structure. *Proc. Natl. Acad. Sci. USA.* 97:5129–5134.

27. Von Bergen, M., S. Barghorn, L. Li, A. Marx, J. Biernat, E. M. Mandelkow, and E. Mandelkow. 2001. Mutations of tau protein in frontotemporal dementia promote aggregation of paired helical filaments by enhancing local β-structure. *J. Biol. Chem.* 276:48165–48174.

28. Iqbal, K., A. C. Alonso, S. Chen, O. Chohon, E. El-Akkad, C. Gong, S. Khatoon, B. Li, F. Liu, A. Rahman, H. Tanimukai, and I. Grundke-Iqbal. 2005. Tau pathology in Alzheimer's disease and other tauopathies. *Biochim. Biophys. Acta.* 1739:198–210.

29. Gamblin, T. C. 2005. Potential structure/function relationships of predicted secondary structural elements of tau. *Biochim. Biophys. Acta.* 1739:40–149.

30. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.

31. Brooks, C. L., M. Karplus, and B. M. Petitt. 1988. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. John Wiley & Sons, New York.

32. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

33. Van Gunsteren, W. F., and H. J. C. Berendsen. 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* 34: 1311–1327.

34. Shen, M., and K. F. Freed. 2002. Long time dynamics of met-enkephalin: comparison of explicit and implicit solvent models. *Biophys. J.* 82:1791–1808.

35. Bursulaya, B. D., and C. L. Brooks. 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.* 104:12378–12383.

36. Hovmöller, S., T. Zhou, and T. Ohlson. 2002. Conformations of amino acids in proteins. *Acta Crystallogr. D. Biol. Crystallogr.* 58:768–776.

37. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 14:33–38.

38. Stultz, C. 2002. Localized unfolding of collagen explains collagenase cleavage near imino-poor sites. *J. Mol. Biol.* 319:997–1003.

39. Tidor, B., and M. Karplus. 1994. The contribution of vibrational entropy to molecular association. *J. Mol. Biol.* 238:405–414.

40. Mcquarrie, D. 2000. Statistical Mechanics. University Science Books, Sausalito, CA.

41. Dominy, B. N., and C. L. Brooks. 2002. Identifying native-like protein structures using physics-based potentials. *J. Comput. Chem.* 23:147–160.

42. Lee, M. C., and Y. Duan. 2004. Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized Born solvent model. *Proteins.* 55:620–634.

43. Felts, A. K., E. Gallicchio, A. Wallqvist, and R. M. Levy. 2002. Distinguishing native conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the surface generalized Born solvent model. *Proteins.* 48:404–422.

44. Lee, M. R., Y. Duan, and P. A. Kollman. 2000. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins.* 39:309–316.

45. Vorobjev, Y. N., J. C. Almagro, and J. Hermans. 1998. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. *Proteins.* 32:399–413.

46. Masunov, A. M., and T. Lazaridis. 2003. Potentials of mean force between ionizable amino acid side chains in water. *J. Am. Chem. Soc.* 125:1722–1730.

47. Dobson, C. M. 2004. Principles of protein folding, misfolding and aggregation. *Sem. Cell Devel. Biol.* 15:3–16.

48. Margittai, M., and F. Langen. 2004. Templated-assisted filament growth by parallel stacking of tau. *Proc. Natl. Acad. Sci. USA.* 101: 10278–10283.

49. Berriman, J., L. C. Serpell, K. A. Oberth, A. L. Fink, M. Goedert, and R. A. Crowther. 2003. Tau Filaments from human and from in vitro assembly of recombinant protein show cross-β structure. *Proc. Natl. Acad. Sci. USA.* 100:9034–9038.

50. Tolnay, M., and A. Probst. 1999. Tau protein pathology in Alzheimer's disease and related disorders. *Neuropath. Appl. Neuro.* 25:171–187.