

Calculation of Protein Heat Capacity from Replica-Exchange Molecular Dynamics Simulations with Different Implicit Solvent Models

In-Chul Yeh,[†] Michael S. Lee,^{†,‡,§} and Mark A. Olson^{*,§}

Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, Maryland 21702, Computational and Information Sciences Directorate, U.S. Army Research Laboratory, Aberdeen Proving Ground, Maryland 21005, and Department of Cell Biology and Biochemistry, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Maryland 21702

Received: March 20, 2008; Revised Manuscript Received: September 22, 2008

The heat capacity has played a major role in relating microscopic and macroscopic properties of proteins and their disorder–order phase transition of folding. Its calculation by atomistic simulation methods remains a significant challenge due to the complex and dynamic nature of protein structures, their solvent environment, and configurational averaging. To better understand these factors on calculating a protein heat capacity, we provide a comparative analysis of simulation models that differ in their implicit solvent description and force-field resolution. Our model protein system is the src Homology 3 (SH3) domain of α -spectrin, and we report a series of 10 ns replica-exchange molecular dynamics simulations performed at temperatures ranging from 298 to 550 K, starting from the SH3 native structure. We apply the all-atom CHARMM22 force field with different modified analytical generalized Born solvent models (GBSW and GBMV2) and compare these simulation models with the distance-dependent dielectric screening of charge–charge interactions. A further comparison is provided with the united-atom CHARMM19 plus a pairwise GB model. Unfolding–folding transition temperatures of SH3 were estimated from the temperature-dependent profiles of the heat capacity, root-mean-square distance from the native structure, and the fraction of native contacts, each calculated from the density of states by using the weighted histogram analysis method. We observed that, for CHARMM22, the unfolding transition and energy probability density were quite sensitive to the implicit solvent description, in particular, the treatment of the protein–solvent dielectric boundary in GB models and their surface-area-based hydrophobic term. Among the solvent models tested, the calculated melting temperature varied in the range 353–438 K and was higher than the experimental value near 340 K. A reformulated GBMV2 model of employing a smoother molecular-volume dielectric interface was the most accurate in reproducing the native conformation and a two-state folding landscape, although the melting transition temperature did not show the smallest deviation from experiment. For the lower-resolution CHARMM19/GB model, the simulations failed to yield a bimodal energy distribution, yet the melting temperature was observed to be a good estimate of higher-resolution simulation models. We also demonstrate that a careful analysis of a relatively long simulation is necessary to avoid trapping in local minima and to find a true thermodynamic transition temperature.

I. Introduction

The prevailing simulation approach for the calculation of thermodynamic folding of proteins is lattice and off-lattice models,^{1–5} while all-atom molecular dynamics (MD) simulations are notoriously difficult.^{6–8} The difficulty arises from computing expectation values of observables from conformational excursions over a large number of degrees of freedom. Many methods have been devised to accelerate the exploration of phase space and improve convergence by smoothing the coarseness in the potential energy landscape while retaining sufficient resolution to accurately estimate the density of states. One of the more notable computational strategies of reducing the manifold of explicit Cartesian coordinates in MD sampling of protein conformational landscapes is the replacement of the microscopic

representation of explicit water interactions with the macroscopic representation of treating implicitly the bulk physical properties of solvation.

Currently, the most popular of implicit electrostatic solvent models for protein dynamics simulations is the generalized Born (GB) approximation. Central to GB models is the treatment of the protein–solvent dielectric boundary. Although there is no unique definition of the dielectric boundary, the more rigorous Poisson-based methods generally apply a Lee–Richards molecular surface.⁹ This surface is considered the *de facto* description for continuum approximations; however, its application in GB models for solvent dynamics is computationally costly. Moreover, because of its abrupt protein–solvent dielectric transition, the molecular surface can yield unstable forces in MD simulations from rapidly changing solute conformations. Alternative smooth dielectric-boundary formulations have been constructed for GB models by applying a superposition of atomic-centered polynomials or Gaussian functions. An example is the smoothing-window dielectric-volume definition imple-

* Corresponding author. E-mail: molson@ncifcrf.gov.

[†] Biotechnology High Performance Computing Software Applications Institute.

[‡] U.S. Army Research Laboratory.

[§] U.S. Army Medical Research Institute of Infectious Diseases.

mented in the GBSW solvent model.¹⁰ The numerical stability of the GBSW model is well suited for MD simulations of proteins, although the dielectric boundary is fundamentally different from the molecular surface in its treatment of the interstitial space between atoms. A truer mimic of the molecular surface is given by the GBMV2 solvent model,^{11,12} where an analytical formulation is used to build a molecular volume based on superposition of spherical functions.

The accuracy of the GBSW and GBMV2 solvent models in their ability to reproduce Poisson solvation energies for single solute conformations is well documented,¹³ yet few studies have been reported on their validation from MD simulations of thermodynamic folding of proteins.^{14–16} A particular issue is the influence of different protein–solvent dielectric-boundary definitions and their smoothing functions on populating conformational states as a function of temperature. Furthermore, the general question arises as to whether the exacting adherence to Poisson theory with a Lee–Richards molecular surface is a true measure of GB models to accurately reproduce calorimetric observables. Given the continued interest to improve upon implicit solvent methods,^{17–19} experimental and theoretical benchmarks are needed to gauge the accuracy of simulation models and their general applicability to a broad range of problems.

In this paper, we report a comparative analysis of GB models for calculating a protein heat capacity and the unfolding–folding phase transition using replica-exchange MD (REMD) simulations.²⁰ Other than the works of Pitera and Swope on modeling the 20-residue miniprotein Trp-cage,²¹ and Duan and co-workers on the 35-residue villin headpiece HP35,²² there is a paucity of studies on computing a protein heat capacity using REMD simulations with implicit solvent models. Our model protein system is the 57-residue src Homology 3 (SH3) domain of α -spectrin, which has been studied extensively with the two-state folding model both experimentally^{23,24} and theoretically.⁴ We applied several different surface-boundary parametrizations of GBSW and GBMV2 solvent models based on the all-atom CHARMM22 force field. For comparison purposes, we also report a simulation with a simple distance-dependent dielectric model to determine whether it is necessary to use the GB models to calculate the heat capacity and melting curve. A further study is provided with the united-atom CHARMM19 force field combined with an earlier GB model based on reparameterization of the original Still formula.²⁵ This later model provides insight into sampling a less rugged energy landscape derived from changing the resolution of the protein representation and solvent description. For each simulation model, we estimated the melting transition temperature and calculated an energy-density profile as a function of the structural deviation from the starting native fold. Our calculations move significantly beyond that of an assessment of implicit solvent models by way of comparison with Poisson theory and offer a much more rigorous benchmark of GB models for continuum solvent dynamics of modeling the thermodynamic stability of proteins.

II. Theory

We used three types of GB implicit solvent models in this work: GBSW (generalized Born switching window),¹⁰ GBMV2 (generalized Born molecular volume version 2),^{11,12} and GBORN.²⁶ The first two GB models were developed for the CHARMM22 force field, and GBORN is a reparameterization of an analytical GB formulation²⁵ for CHARMM19. In all of the GB methods, the total generalized Born energy, E_{GB} , is

$$E_{GB} = k \sum_{ij} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp(-r_{ij}^2 / K_s \alpha_i \alpha_j)}} \quad (1)$$

where the summation indices, i and j , run over all atoms, q_i is the charge on atom i , r_{ij} is the distance between atoms i and j , $k = -166.0(\epsilon_{\text{solute}}^{-1} - \epsilon_{\text{solvent}}^{-1})$, $K_s = 8$ for GBMV2, and $K_s = 4$ for GBSW and GBORN. The dielectric constants of the solute, ϵ_{solute} , and solvent, $\epsilon_{\text{solvent}}$, are set to 1 and 80, respectively. Details of the GBORN algorithm can be found elsewhere,²⁶ but we elaborate on the GBSW and GBMV2 methods below. The Born radii, α_i , in both protocols are

$$(\alpha_i)^{-1} = a_0 \left(R_i^{-1} - \frac{1}{4\pi} \iint_{R_i}^{\infty} \frac{V(\vec{r})}{|\vec{r} - \vec{x}_i|^4} dr d\Omega \right) + a_1 \left(\frac{1}{4R_i^4} - \frac{1}{4\pi} \iint_{R_i}^{\infty} \frac{V(\vec{r})}{|\vec{r} - \vec{x}_i|^7} dr d\Omega \right)^{1/4} \quad (2)$$

where R_i is the van der Waals radius of atom i , and \vec{x}_i are the atomic coordinates. In the GBMV2 protocol, $a_0 = (1 - (1/\sqrt{2}))$ and $a_1 = 1$, and $V(\vec{r})$ is described below. In the GBSW method, a_0 and a_1 can be varied to fit Poisson results based on different molecular-surface representations.

In the more rigorous implicit solvent model GBMV2,^{11,12} the solvent excluded volume is

$$V_{\text{GBMV2}}(\vec{r}) = \frac{1}{1 + \exp(\beta_S [S_{\text{MV2}}(\vec{r}) - \lambda])} \quad (3)$$

where the parameter β_S determines the smoothness of the excluded volume and the raw excluded volume, S_{MV2} , is given by

$$S_{\text{MV2}}(\vec{r}) = S_0 \left[\sum_m F_{\text{MV2}}(\|\vec{r} - \vec{x}_m\|) \right] \times \frac{\sum_m \|\vec{r} - \vec{x}_m\|^2 F_{\text{MV2}}^2(\|\vec{r} - \vec{x}_m\|)}{\|\sum_m (\vec{r} - \vec{x}_m) F_{\text{MV2}}(\|\vec{r} - \vec{x}_m\|)\|^2} \quad (4)$$

where S_0 is an adjustable parameter, m indexes another summation over all atoms, and F_{MV2} is an atom-centered function defined in Lee et al.¹²

In this work, we look at a variation of the parameters, β_S and S_0 , suggested by Chocholoušová and Feig²⁷ which provide a smoother excluded volume than the original formulation. The purpose of their reformulation was to enhance energy conservation in the application of GBMV2 in MD simulations where constant energy is required, for example, NVE ensembles. The parameter β_S can be adjusted to make the molecular volume sharper or smoother at the protein–solvent dielectric boundary. The S_0 variable is then adjusted to match Poisson solvation energies. The original parameters for GBMV2 are $\beta_S = -20$ and $S_0 = 0.7$, whereas the reformulated parameters are $\beta_S = -10$ and $S_0 = 0.57$. We denote the original parametrized model as GBMV2- $\beta_S 20$ and the reformulated model as GBMV2- $\beta_S 10$.

The simpler and faster GBSW has an excluded solvent volume of

$$V_{\text{GBSW}}(\vec{r}) = 1 - \prod_i H_i(|\vec{r} - \vec{r}_i|) \quad (5)$$

where the polynomial atomic volume H_i is given by

$$H_i(r) = \begin{cases} 0 & r < R_i - w \\ \frac{1}{2} + \frac{3}{4w}(r - R_i) - \frac{1}{4w^3}(r - R_i)^3 & R_i - w < r < R_i + w \\ 1 & r \geq R_i + w \end{cases} \quad (6)$$

and $2w$ is a smoothing length that confines the region where the smoothing function is applied. In this work, we tested two different parameter sets of the GBSW model.²⁷ These sets are denoted as GBSW1, with $w = 0.3 \text{ \AA}$, $a_0 = -0.1801$, and $a_1 = 1.8174$, and GBSW2, with $w = 0.2 \text{ \AA}$, $a_0 = 1.2045$, and $a_1 = 0.1866$. GBSW1 was designed to mimic the corresponding smoothing-window surface Poisson model, and GBSW2 was parametrized to fit the Lee–Richards molecular-surface Poisson results.

III. Methods

We performed REMD simulations by running the program CHARMM with the MMTSB (Multiscale Modeling Tools for Structural Biology) Tool Set.²⁸ The CHARMM22 force field was used to represent the protein in the GBMV2 and GBSW models, while the CHARMM19 force field was used for the GBORN model. We also ran a simulation with a distance-dependent protein dielectric model ($\epsilon = 4r$) to calculate the thermodynamics of protein folding.

The X-ray crystal structure of the SH3 domain of α -spectrin (PDB ID 1SHG)²⁹ with 57 residues was used as a starting protein conformation. This starting structure was optimized by energy minimization with a distance-dependent dielectric constant of $\epsilon = 4r$, and included 50 steps of initial steepest descent minimization followed by minimization with the corresponding GB or r -dielectric method over 200 steps. During the minimization, C_α/C_β atoms of all residues were restrained to their initial position with a force constant of 0.5 kcal/mol except for the last 100 steps. The resulting minimized conformation was used as a starting structure of the subsequent MD simulations.

For all simulation models, a MD integration time step of 2 fs was used. Distances used for the onset of switching function for nonbonded interaction, the cutoff for nonbonded interactions, and the cutoff for nonbonded list generation were 20, 22, and 25 \AA , respectively. Covalent bonds between the heavy atoms and hydrogens were constrained by the SHAKE algorithm.³⁰

For the nonpolar solvation energy incorporated with the GBMV2 and GBSW models, we applied an energy term expressed as the linear product of the solvent-exposed surface area of the solute and a phenomenological surface tension coefficient (γ) set to 0.03 kcal/(mol $\cdot\text{\AA}^2$) for GBSW1 and 0.005 kcal/(mol $\cdot\text{\AA}^2$) for GBMV2.^{10,28,31} For GBSW2 model calculations, the two values of γ were examined by conducting two separate simulations. Because of the parametrization of the GBORN model with the united-atom CHARMM19 force field, this term was neglected.²⁶

We performed single-temperature MD simulations with both sets of GBMV2 parameters to decide on the temperature range for REMD simulations. Given the temperature for protein-chain unfolding, the REMD simulations for the heat capacity and

melting curves were performed at 32 temperatures exponentially spaced between 298 and 550 K. The exchange of the conformations in neighboring temperatures was attempted at every 100 time steps subject to the Metropolis criterion. The coordinates at each temperature were saved at every 100 time steps for further analysis. All REMD simulations were carried out to 10 ns simulation time per thermal window.

We used the weighted histogram analysis method (WHAM)^{32–35} to calculate thermodynamic properties. The density of states $\Omega(U, \xi)$ for a system, where U and ξ are the potential energy and the property of interest, respectively, is given by

$$\Omega(U, \xi) = \frac{\sum_{i=1}^R N_i(U, \xi)}{\sum_{j=1}^R n_j \exp(f_j - \beta_j U)} \quad (7)$$

where n_j is the number of data points in the j th simulation and $\beta_j = 1/(k_B T_j)$, where k_B and T_j are Boltzmann's constant and the temperature of the j th simulation, respectively, and R is the number of thermal windows. The function $N_i(U, \xi)$ is the histogram of U and ξ calculated from the i th simulation. f_j is the scaled free energy obtained by solving the following equations self-consistently

$$P_\beta(U, \xi) = \frac{\sum_{i=1}^R N_i(U, \xi) \exp(-\beta U)}{\sum_{j=1}^R n_j \exp(f_j - \beta_j U)} \quad (8)$$

and

$$\exp(-f_j) = \sum_{U, \xi} \Omega(U, \xi) \exp(-\beta U) \quad (9)$$

where $P_\beta(U, \xi)$ is the probability density at the inverse temperature β . Thermodynamics properties can be determined from the probability density, for example, the heat capacity by the following expression^{34,36}

$$C_v(T) = \frac{\sigma_U^2}{k_B T^2} = \frac{\langle U^2 \rangle - \langle U \rangle^2}{k_B T^2} \quad (10)$$

$$\langle U^n \rangle = \frac{\sum_{U, \xi} U^n P_\beta(U, \xi)}{\sum_{U, \xi} P_\beta(U, \xi)} \quad (11)$$

with $n = 1$ or 2. Similarly, $\langle \xi \rangle(T)$ is the average of the property ξ at the temperature T and can be obtained by the following equation

$$\langle \xi \rangle(T) = \frac{\sum_{U, \xi} \xi P_\beta(U, \xi)}{\sum_{U, \xi} P_\beta(U, \xi)} \quad (12)$$

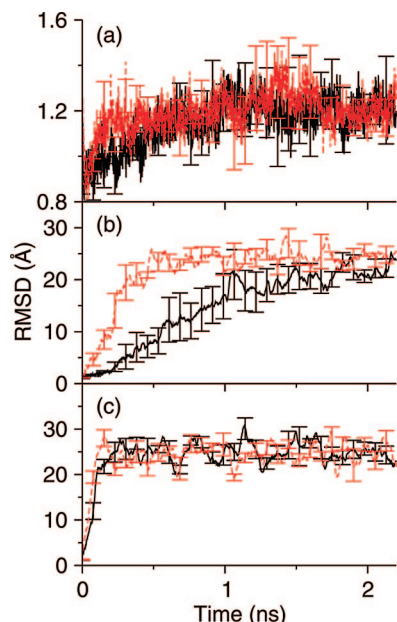


Figure 1. Calculated root-mean-square distance (rmsd) of backbone atoms of the SH3 domain of α -spectrin from MD simulations at temperatures of (a) 298 K, (b) 440 K, and (c) 500 K with respect to its X-ray crystal structure as a function of the simulation time. The black and red lines represent averages of rmsd values obtained from multiple independent simulation runs (4 at 298 and 500 K and 5 at 440 K) calculated by the GBMV2 parameters in the original implementation (GBMV2- β_S20 parameter set, $\beta_S = -20$ and $S_0 = 0.7$) and with a recently reformulated set of parameters (GBMV2- β_S10 parameter set, $\beta_S = -10$ and $S_0 = 0.57$), respectively. Uncertainty is shown as error bars and was estimated by the standard deviation.

Further discussions on theoretical models of the heat capacity in proteins is given in the work by Prabhu and Sharp.³⁷

The root-mean-square distance (rmsd) between two structures is defined as the square-root of the minimum average square distance between respective backbone atoms of the two structures with respect to all rigid-body rotations and translations.³⁸ The fraction of the native contacts, ρ , is calculated by the native contact defined by the residue pair with the distance between side chain centers of geometry less than 6.5 Å.

The REMD simulations were executed on a Linux cluster using an Intel CPU (3.06 GHz Intel IA-32). Application of the CHARMM22/GBMV2 simulation model requires approximately 100 CPU hours per nanosecond of simulation time, while CHARMM19/GBORN is roughly 4 times less demanding.

IV. Results

In the work of Chocholoušová and Feig,²⁷ they reported that modifying the smoothing parameter β_S of the GBMV2 model can enhance energy conservation in MD simulations when using this implicit solvent model. As a preliminary to applying replica-exchange simulations with the GBMV2 model, we tested the effect of modifying β_S on simulation trajectories computed at several different constant temperatures. Figure 1 shows the calculated rmsd between backbone heavy atoms of the X-ray crystal structure of the SH3 domain and structures obtained from single-temperature MD simulations performed at temperatures of 298, 440, and 500 K with the two sets of GBMV2 parameters $\beta_S = -10$ and -20 . The results show that the SH3 domain remained folded at 298 K and it became unfolded quickly at 500 K with both sets of parameters. On the other hand, rmsd changes at an intermediate temperature of 440 K indicate that

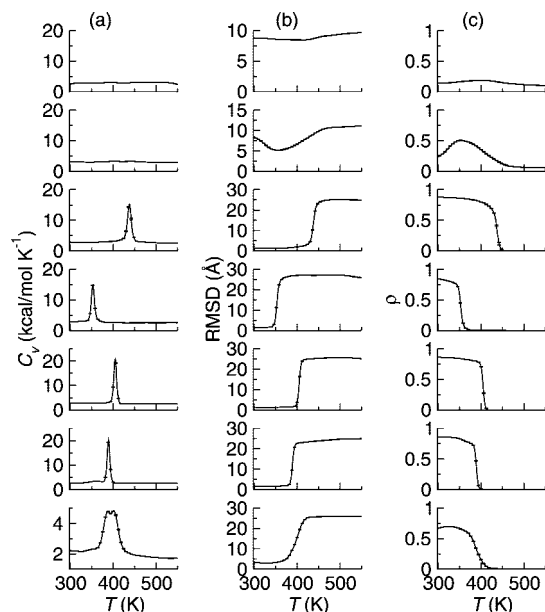


Figure 2. Temperature-dependent profiles of (a) the heat capacity C_v (kcal/mol \cdot K⁻¹), (b) the average rmsd (Å), and (c) the fraction of native contacts ρ of the SH3 domain of α -spectrin calculated from the WHAM analysis of the last 0.5 ns of 10 ns REMD runs with different solvent models. Starting from the top figure and moving down, the CHARMM22 force-field solvent models are (i) the distance-dependent dielectric-screening model, (ii) GBSW1, (iii) GBSW2 ($\gamma = 0.03$ kcal/mol \cdot Å²), (iv) GBSW2 ($\gamma = 0.005$ kcal/mol \cdot Å²), (v) GBMV2- β_S20 , (vi) GBMV2- β_S10 , and the bottom graph (vii) illustrates the CHARMM19 model with GBORN. The error bars represent the statistical uncertainty in the calculated profiles.

the unfolding event progressed faster with the GBMV2- β_S10 model than the original parametrization implemented in GBMV2- β_S20 .

In Figure 2, we compare the temperature-dependent profiles of the heat capacity and the related average values of rmsd and ρ , each calculated with the WHAM analysis of the last 0.5 ns of the REMD simulations with different solvent models. Superimposed on the plots is the statistical uncertainty, which was determined by a Bayesian statistical estimation method developed by Gallicchio et al.³⁵ Other methods of estimating errors from parallel tempering simulations are available (see, e.g., Chodera et al.³⁹). For the computed quantities of Figure 2, the small error bars indicate that the statistical uncertainties in the temperature-based WHAM numerical solutions are negligible. The temperature corresponding to the maximum heat capacity in a temperature-dependent profile is the unfolding–folding critical transition temperature (T_c) or alternatively it can be thought of as an apparent melting temperature. We begin with the CHARMM22 empirical force field and the application of the simplest solvent model based on the distance-dependent scaling of the protein dielectric constant. For this solvent description, the simulation model failed to produce a well-defined folding transition in the calculated heat capacity. Although there is a slight shift in the rmsd and ρ at a temperature above 400 K, the simulation favors a distribution of non-native states.

The next two simulation models illustrated in Figure 2 are different parametrizations of the GBSW solvent model. The first is the original implementation given by GBSW1, and the calculated results show a heat capacity that lacks a single sharp spike representing an unfolding–folding transition. The plot of rmsd versus temperature indicates significant density of native-like states at 300 K; however, at higher temperatures, the protein

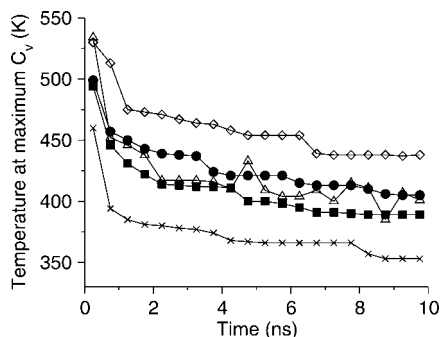


Figure 3. Melting transition temperatures estimated from the temperature corresponding to the maximum in the heat capacity profiles calculated by the WHAM analysis of 10 ns REMD runs with different GB solvent models at 0.5 ns intervals. Filled circles and squares represent REMD runs with CHARMM22 simulation models GBMV2- β_{s20} and GBMV2- β_{s10} , respectively. Open diamonds and crosses represent REMD runs with CHARMM22/GBSW2 models with $\gamma = 0.03$ and 0.005 kcal/mol $\cdot\text{\AA}^2$, respectively. The triangles represent the CHARMM19/GBORN simulation model.

chain fails to undergo unfolding to an rmsd greater than 10 \AA . In contrast, the GBSW2 model simulations with the surface tension value set at either $\gamma = 0.03$ kcal/mol $\cdot\text{\AA}^2$ or $\gamma = 0.005$ kcal/mol $\cdot\text{\AA}^2$ produced heat capacities that exhibit values for T_c of 438 and 353 K, respectively. In addition, both calculations show unfolding to an rmsd of approximately 25 \AA and a transition in the fraction of native contacts from roughly 0.85 at 300 K to negligible contacts at higher temperatures, indicating significant loss of structural similarity with the native state.

Comparable to GBSW2, both GBMV2- β_{s20} and GBMV2- β_{s10} models produced sharp unfolding–folding transitions corresponding to maximum positions in the heat capacity at T_c 's of 405 and 389 K, respectively. The GBSW2 and GBMV2 simulation models at 300 K show very similar heat capacity values of approximately $2.0\text{--}3.5$ kcal $\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$, all in excellent agreement with the experimental determination of $2.4\text{--}3.6$ kcal $\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$.^{23,24} However, as a collective group of models, the simulations produced profiles much narrower than the experimental heat capacity, an effect that can be attributed to starting the simulations from the native conformation rather than combining simulations starting from the unfolded state. Given the size of SH3, ab initio folding of this protein from an extended conformation would be computationally beyond most available computer resources.

The final simulation model in Figure 2 is the application of CHARMM19 with the GBORN solvent model. The simulation yields a broad heat capacity of multiple peaks with a maximum value at T_c of 401 K. The observed transitions in rmsd and ρ clearly demonstrate protein-chain unfolding, although the shifts are less sharp than those calculated with the two GBMV2 models and GBSW2.

The transition temperatures were observed to change during the course of REMD simulations. Figure 3 shows the T_c determined from the temperature corresponding to the maximum in the heat capacity profiles calculated by the WHAM analysis of 10 ns REMD runs with different GB models at 0.5 ns time intervals. At the end of the 10 ns run, we report the final T_c values. The T_c changed more rapidly from REMD runs with the GBMV2- β_{s10} model, which introduces a smoother molecular volume, reaching 389 K at the end of a 10 ns simulation. With the GBSW2 models, the transition temperature changed slowly to a much higher value of 438 K for $\gamma = 0.03$ kcal/mol $\cdot\text{\AA}^2$, and for $\gamma = 0.005$ kcal/mol $\cdot\text{\AA}^2$, the final temperature is 353 K. Estimated transition temperatures from the CHARMM19/GBORN simulation model

were comparable to those obtained with the GBMV2 models, even though the estimated transition temperatures fluctuated during the course of the simulation because of the complex peak shapes in the heat capacity profiles (Figure 2).

Normalized probability-density distributions of potential energies of SH3 conformations at T_c are illustrated in Figure 4 for simulation models that displayed a sharp transition in their respective heat capacity. Also shown are two-dimensional contour maps of potential energy versus rmsd and ρ at or near the transition temperature. We define the potential energy as the protein internal energy plus the GB solvent energy. Histograms for the CHARMM22 simulation models near their T_c values are observed to be bimodal distributions. The left peak of the histogram corresponds to the folded state, and the right peak corresponds to the unfolded state. At the transition temperature, SH3 exists in both states with equal probability. The difference in the maxima of the two peaks indicates the existence of a free-energy barrier that separates the transition between the SH3 folded and unfolded states. For the united-atom CHARMM19/GBORN model, the potential energies demonstrate a broad probability distribution which lacks a bimodal shape.

The contour map from the GBSW2 model with $\gamma = 0.03$ kcal/mol $\cdot\text{\AA}^2$ shows that, near T_c , the folded state is distributed between two major clusters of SH3 conformations. A near-native cluster is observed at $\rho \sim 0.8$; however, it is less densely populated than the second cluster positioned at $\rho \sim 0.5$. The more populated cluster is higher in energy and is presumably favored by entropic contributions. For the GBSW2 model with $\gamma = 0.005$ kcal/mol $\cdot\text{\AA}^2$, there is less bifurcation of the folded state, yet the ensemble of unfolded states now exhibits two disconnected clusters, one displaying significant population with native contacts at $\rho \sim 0.15$.

Of the two GBMV2 simulation models, application of GBMV2- β_{s10} produced the most compacted near-native cluster with the lowest average energy. Similar observations are made for the rmsd contour. In contrast with the results with the CHARMM22 force field, the CHARMM19/GBORN simulation model yields contour maps near T_c that show considerable reduction in the separation among the conformational states that contribute to the energy-density profile. The native or near-native state is missing in the contour, while the most probable cluster is $\rho \sim 0.2$, containing partially folded structures. Similarly, there are multiple clusters that span the rmsd range, indicating a significant reduction in the energy barrier to unfolding.

We constructed a histogram of SH3 conformations based on an rmsd measure from the native and selected a conformation representative of the most populated rmsd value for each cluster at T_c . Figure 5 illustrates the selected structures. For comparison purposes, the X-ray structure for SH3 (panel f) is an all- β protein with a β -barrel topology, consisting of two hydrophobic sheets. The first sheet is dominated by a three-stranded β arrangement, and the second is a short two-stranded sheet near the two terminal ends. The GBSW2 model with $\gamma = 0.03$ kcal/mol $\cdot\text{\AA}^2$ shows the most populated state at a rmsd of 5.84 \AA with a conformation that breaks the short β -sheet near the terminus region. Reducing γ to 0.005 kcal/mol $\cdot\text{\AA}^2$, the GBSW2 model yields a rmsd of 1.75 \AA for the most populated native state and shows a structure exhibiting minor disruptions of the major β -sheet arrangement.

The GBMV2- β_{s20} model at T_c populates conformations with a distorted β -sheet arrangement, and the representative structure shows a coordinate rmsd of 1.87 \AA and an energy value of -1271 kcal/mol. For GBMV2- β_{s10} , a similar rmsd is populated at a lower

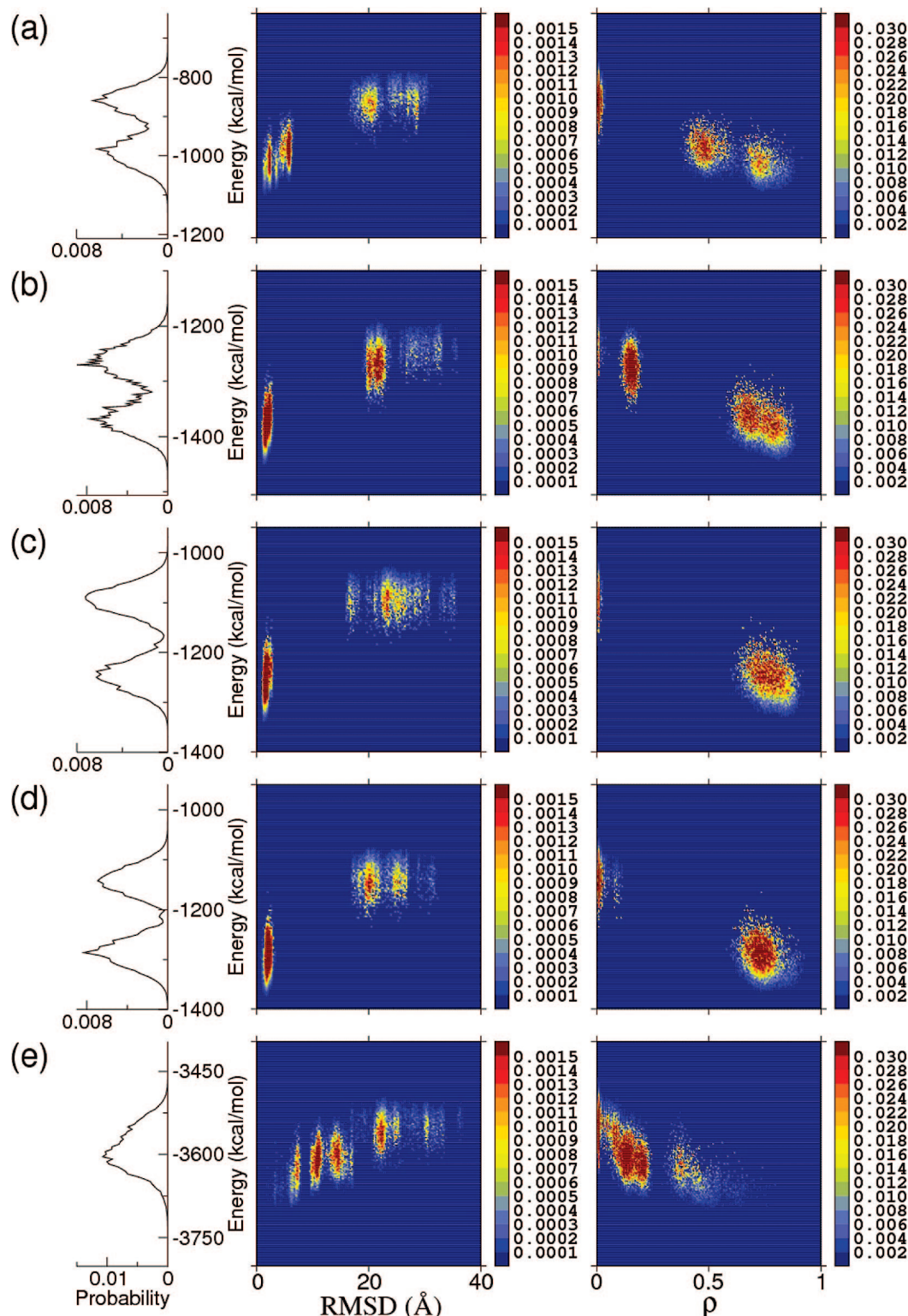


Figure 4. Two-dimensional probability-density distributions with respect to the potential energy and rmsd (left) and also with respect to the potential energy and the fraction of native contacts ρ (right) at the unfolding–folding transition temperatures calculated by the WHAM analysis of REMD runs with (a) the GBSW2 ($\gamma = 0.03 \text{ kcal/mol} \cdot \text{\AA}^2$) model, (b) the GBSW2 ($\gamma = 0.005 \text{ kcal/mol} \cdot \text{\AA}^2$) model, (c) the GBMV2- β_520 model, (d) the GBMV2- β_510 model, and (e) the GBORN model. The leftmost graphs show the projections of two-dimensional probability-density distributions to the probability distributions with respect to the potential energy alone.

energy of -1293 kcal/mol . More importantly, the GBMV2- β_510 simulation model produces a SH3 domain topology that maintained the highest fraction of native contacts of the X-ray structure. The least favorable structure compared to the starting X-ray structure is obtained from the CHARMM19/GBORN simulation and shows a partially folded conformation containing only a β -hairpin ar-

range with an rmsd of 11.02 \AA . This particular structure from the united-atom force field was not observed in any of the other contours using CHARMM22 simulation models.

Figure 6 illustrates the Metropolis acceptance ratio as a function of the replica-exchange pair for the solvent models tested. The data used for each calculation was taken from the

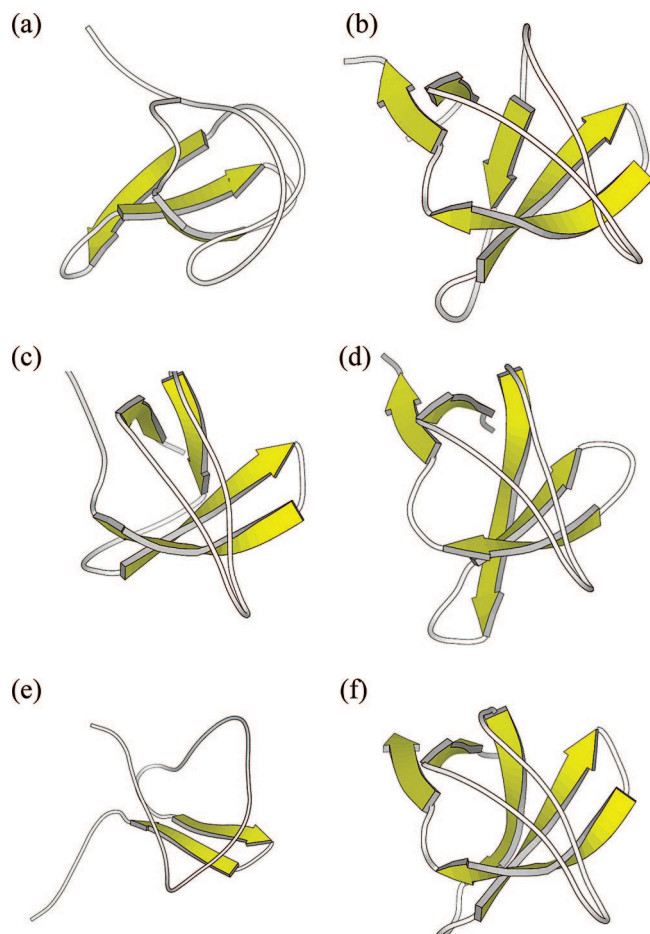


Figure 5. Representative structures of SH3 conformations selected from the most populated rmsd clusters depicted in Figure 4 at the unfolding–folding transition temperature: (a) Structure obtained from the GBSW2 model ($\gamma = 0.03$ kcal/mol $\cdot\text{\AA}^2$) at an rmsd of 5.84 \AA from the X-ray structure; (b) GBSW2 ($\gamma = 0.005$ kcal/mol $\cdot\text{\AA}^2$) modeled structure at an rmsd of 1.75 \AA ; (c) GBMV2- β_s20 modeled structure at an rmsd of 1.87 \AA ; (d) GBMV2- β_s10 modeled structure at an rmsd of 1.87 \AA ; (e) CHARMM19/GBORN modeled structure at an rmsd of 11.02 \AA ; (f) X-ray crystal structure.

last 0.5 ns of the simulation and averaged over the data set. The i th exchange denotes the exchange between i th and $(i+1)$ th conditions. In general, the acceptance is near 0.6, except for windows containing an unfolding–folding transition. In the latter, the low acceptances indicate the lack of thermodynamic coexistence between the unfolded and folded states, although it should be noted that potential energies are typically applied in the exchanges rather than free energies that include entropic contributions, thus making coexistence difficult to achieve.

V. Discussion and Conclusions

Computational studies aimed at understanding the protein-folding process involve investigation beyond whether the simulations can sample the native and unfolded conformations correctly but must address issues of folding thermodynamics, for example, changes in the heat capacity on protein folding or unfolding and the corresponding melting curve of the protein. There are few reported studies on computing a protein heat capacity using all-atom MD simulations with implicit solvent models. Without sufficient benchmarks to gauge the accuracy of different simulation models, questions remain as to the influence of GB models on the simulation results and whether

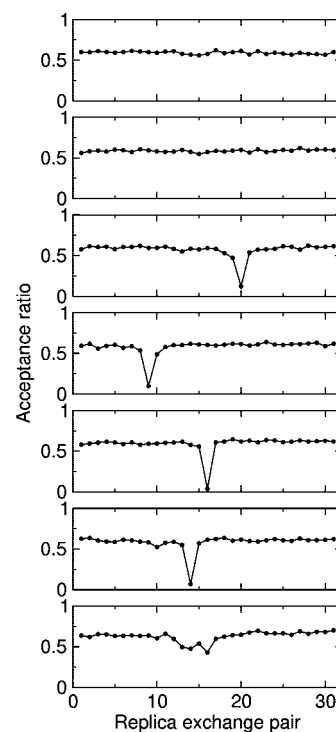


Figure 6. Replica-exchange acceptance ratios as a function of the nearest-neighbor exchange clients for each solvent model. Exchanges were computed using 32 replicas. Starting from the top plot and moving downward, the solvent models are (i) the distance-dependent dielectric screening, (ii) GBSW1, (iii) GBSW2 with $\gamma = 0.03$ kcal/mol $\cdot\text{\AA}^2$, (iv) GBSW2 with $\gamma = 0.005$ kcal/mol $\cdot\text{\AA}^2$, (v) GBMV2- β_s20 , (vi) GBMV2- β_s10 , and (vii) the CHARMM19/GBORN model. Each plot was calculated from averaging data culled from the last 0.5 ns simulation.

the most simplistic solvent models are adequate in capturing the correct thermodynamics of protein folding.

While there are a number of different GB model implementations in MD simulations of proteins, the primary difference among the models is the definition and calculation of the Born radii. To evaluate the Born radii, two approximations are invoked. The first is the Coulomb field approximation (CFA) of estimating the atomic self-energy. The CFA in its most rigorous form is expressed as a surface/volume integration of a monopole-induced dipole energy term, and depending on the particular GB formulation, higher-order non-Coulomb correction terms may be added to the Born radii to account for the reaction field. The second approximation is the description of the molecular volume or surface. It is particularly this latter approximation that our comparative study of GB models has significant implications on how to properly treat the solute–solvent dielectric boundary.

For the CHARMM22 force field, we investigated the GBSW and GBMV2 solvent models and their parametrizations. The two surface-boundary parametrizations of GBSW were found to yield strikingly different results. Even though there is a small difference in the smoothing length between the two GBSW models ($w = 0.3$ \AA for GBSW1 and $w = 0.2$ \AA for GBSW2), the main difference is scaling of the CFA and its higher-order correction term.^{10,13,27} The GBSW1 model was parametrized from fitting Born radii to a van der Waals-based surface with a smooth dielectric boundary and gives less weight to the CFA and more to the correction term than the GBSW2 parametrization, which attempts to mimic a sharp molecular-surface boundary. The results for the GBSW2 model simulations

produced an SH3 unfolding–folding transition that exhibited a sharp spike in the heat capacity at T_c , while the application of GBSW1 failed to undergo a complete unfolding–folding transition. Although GBSW1 has been successfully applied to many different applications,^{40–45} our results reveal shortcomings of this solvent model and its applicability for modeling the thermodynamic folding of polypeptides.

Our study of the GBSW2 model and the effect of modifying the surface tension of the hydrophobic effect showed several anomalies in the probability-density contour maps computed from the 10 ns simulations. Two values were used, $\gamma = 0.03$ kcal/mol·Å², which is the frequently applied value for GBSW models, and $\gamma = 0.005$ kcal/mol·Å², which is the default value for GBMV2. Application of both values of γ produced bifurcation of the native state, creating multiple conformational clusters, although less pronounced for the lower surface tension. On the other hand, reducing γ led to a much less greasy protein that created a cluster of unfolded states that retained residual native contacts. This latter GBSW2 simulation model yielded the lowest T_c among the solvent models; nevertheless, because of an imperfect balance between the hydrophobic term and electrostatic interactions, artifacts were produced in the unfolding–folding landscape.

Unlike GBSW2, the GBMV2 models generated trajectories that were more consistent with a two-state folding model used in fitting the experimental data of the protein heat capacity of SH3.^{23,24} The GBMV2 approximation models a molecular-volume dielectric interface and applies a switching function through the parameter β_s that diminishes the steep surface boundary. The original parametrization determined for the GBMV2- β_s20 model maximizes the agreement with Poisson theory.^{11,12} Of the two GBMV2 models tested, GBMV2- β_s10 exhibited the lowest T_c of 389 K. Nevertheless, the predicted T_c is higher than experimental observations. Depending on the sequence variant of the SH3 domain and experimental pH conditions, the reported melting temperature is roughly 340 K.²⁴ Although the deviation between the experimental temperature and our calculation is less drastic than that observed in the work of Pitera and Swope on modeling the Trp-cage,²¹ where they predicted a T_c value nearly 85 K higher than the experimental determination, both studies indicate that REMD simulations with GB models tend to overstabilize native structures. The work of Duan and co-workers on modeling HP35, however, shows a T_c value that is in close agreement with the experimentally derived value.²²

The accuracy obtained by any simulation study is strongly influenced by simulation protocols, for example, force-field description, short versus long cutoffs for nonbonded interactions, unfolding REMD versus folding REMD, conformational sampling, and the simulation temperature range. Of the latter, the range of 298–550 K used in the simulations was determined from the GBMV2 model to unfold the protein chain and suggests an approximate upper bound temperature for the other solvent models. For models that exhibit greater resistance to protein unfolding, the application of higher temperatures is needed to overcome the activation enthalpy. Because of computational efficiency issues, the maximum temperature should be only slightly above the transition temperature where the enthalpy vanishes⁴⁶ and our temperature range is suitable for most of the tested GB models. It is also important to note that different sampling protocols may influence the simulation results and alternative methods might include, for example, adaptive REMD schemes that incorporate feedback iterations to identify an optimal set of temperatures for exchanges near conformational

transitions,⁴⁷ Hamiltonian REMD using soft-core interactions,⁴⁸ and REMD with global energy reassignment.⁴⁹

A significant source of error in GB simulations is overstabilization of solvent-exposed salt bridges.^{50–54} Recent calculations of potentials of mean force for ion-pair formation in native protein structures have consistently demonstrated that GB models overpredict thermodynamic stability in comparison with explicit solvent simulations.^{50,51} More generally, Poisson-based implicit solvent models have difficulty in accurately reproducing explicit solvent simulations of electrostatic charging free energies of charged residues computed for protein structures.⁵⁵

Our comparative analysis of GB solvent models extends beyond that of simply scoring static structures and their ability to reproduce Poisson theory. While the Lee–Richards molecular-surface representation is considered the *de facto* surface for Poisson-based implicit solvent models, there is limited direct evidence from MD simulations over a wide range of temperatures that this surface is indeed the best dielectric-boundary definition. Although many GB models can be parametrized to stabilize the native state at 298 K, it is the generation of conformational ensembles and their density of states over unfolding–folding temperatures that clearly reveal the distinction among solvent models.^{52,53,56–59} The REMD simulations presented here demonstrated that a smooth molecular-surface representation (GBSW2 or GBMV2) is far superior to a surface based on a smooth van der Waals representation (GBSW1) for modeling the folded state. Moreover, a molecular-volume dielectric surface built from a superposition of spherical functions with a Gaussian-type smoothing function as in GBMV2 performed better than a smoothing-window dielectric-volume definition built from atomic functions used by GBSW2. The magnitude of the difference between the original parametrization of the GBMV2- β_s20 model and GBSW2 with the default value of γ is surprisingly large, as the simulations show dissimilarity in the basins containing the folded state and the computed T_c at the end of 10 ns REMD runs.

The molecular-volume representation of the GBMV2- β_s20 model has, nevertheless, computational problems due to the sharp discontinuous dielectric boundary. One outstanding difficulty is obtaining forces directly from Poisson theory, and several GB models, including GBSW1 and GBSW2, were developed to circumvent these problems in the application of MD simulations. Recent work by Chocholoušová and Feig has shown that GBMV2- β_s20 results in energy drift and numerical artifacts in MD simulations of biological macromolecules.²⁷ They proposed the GBMV2- β_s10 model parametrization, which smooths the surface boundary by readjusting the value of β_s , and showed that this model is capable of maintaining very good agreement with Poisson theory while providing energy conservation and stable MD simulations. We did not observe any numerical instability in the MD simulations of SH3 with either of the GBMV2 solvent models. However, in general, the numerical instability of MD simulations with any GB solvent model can be sensitive to the size of the protein and its structural properties of hydration.

Although the GBMV2- β_s10 model did not produce a T_c value that exhibited the smallest deviation from the experimental value, our REMD simulation with this model produced the most favorable overall outcome. In addition to yielding a two-state energy landscape without significant distortion, GBMV2- β_s10 showed a densely populated native basin that maintained SH3 conformations with the lowest energy and rmsd clustering from the starting X-ray crystal structure. This result is notable given that secondary-structure elements of β -sheets are characteristi-

cally problematic for GB solvent models due to their specific hydrogen-bonding pattern.^{52,53} Furthermore, modifying the β_S smoothing parameter produced faster convergence of GBMV2 in obtaining T_c . The latter suggests that a less-abrupt molecular surface can constructively yield a less-rugged potential energy landscape without diminishing the resolution of conformational states and their disorder–order transition. The end result is much more efficient sampling, yet we observed that, even for a relatively small protein like SH3, long simulation times were still required. Note that, unlike the standard protocol of computing ensemble averages of structural properties and their naturally occurring fluctuations over the entire equilibrated simulation trajectory, the unfolding–folding transition temperatures were slowly decaying over time and configurational averaging of the density of states was restricted to the last 2500 conformations (0.5 ns) per temperature replica of the 10 ns simulation. This correctly avoided overweighing the trapping of excursions into local energy minima found sampling the energy landscape and yielded a better measure of a true thermodynamic temperature. As illustrated by other workers,³⁵ the application of WHAM significantly reduces the statistical uncertainty in temperature-biased simulation data, and we were able to obtain very smooth heat capacities with sharp unfolding–folding transitions.

Although Chocholoušová and Feig report that the GBMV2- β_S10 solvent model produces greater deviation from Poisson theory than the original β_S20 parametrization,²⁷ our calculations strongly indicate that matching a discontinuous molecular-surface dielectric boundary should not be the gold standard of evaluating the accuracy of implicit solvent models. An alternative and complementary choice is matching dynamics with extensive thermodynamic quantities, rather than scoring single protein chains. A more stringent benchmark is a comparison with explicit solvent calculations of sampling energy landscapes, as well as the determination of electrostatic charging free energies. It is conceivable that the GBMV2- β_S10 model would do much better than other parametrizations with both comparisons and perhaps minimize problems of distorting conformational ensembles that plague many current GB solvent models.^{51–54} For the work presented here, a rigorous measure of accuracy would be an explicit solvent calculation of the heat capacity using REMD, although the computational cost using standard protocols is likely prohibitive for modeling proteins of the size of SH3. To improve efficiency of explicit solvent REMD simulations, Simmerling and co-workers developed a novel approach of using a hybrid explicit/implicit solvation model.⁵¹ Their method involves simulations performed with fully explicit solvent, while the replica-exchange probability is computed with a hybrid model of keeping a small solvation shell around the protein and applying GB to treat the bulk dielectric boundary. Preliminary studies indicate a reduction in the system size and a dramatic decrease in computational cost of REMD simulations. For applications of their approach using the CHARMM22 force field, we propose that GBMV2- β_S10 be the default choice for modeling implicit solvent.

Because of the large computational cost of applying CHARMM22 with GB models in MD calculations, we investigated two cost-effective alternatives. The simulation results for the distance-dependent dielectric-screening model showed that, for modeling thermodynamic folding and unfolding of proteins, this computational strategy of treating solvent effects was incapable of producing the correct density of states. Despite reports that dielectric-screening models in MD simulations can cause significant protein structural distortions,^{60,61} they are still

popular among many applications in the literature. Other than for structural refinement of near-native conformations using very short REMD simulations, our results strongly support the notion that the application of these models should be abandoned in large-scale sampling of energy landscapes.

The second alternative to ease the computational demand is to reduce the resolution of the force field and its parametrized implicit solvent model. The question is, can these models provide sufficient resolution to detect the unfolding–folding transition? The simulation results showed that the united-atom CHARMM19/GBORN model produces a heat capacity and melting curve that exhibit thermodynamic transitions, although the energy distribution profile indicated the loss of a bimodal energy distribution that was observed with the all-atom models. Unexpectedly, the T_c computed with a reduced protein representation was found to be in good agreement with the higher-resolution models. This encouraging outcome achieved by the CHARMM19/GBORN model was obtained using nearly 4 times less CPU hours than simulations with CHARMM22/GBMV2. Despite the computational burden of the GBMV2 model, the overall accuracy is significantly enhanced over the faster method. Notably, inspection of the computed density of states and the SH3 conformations taken at the transition temperature clearly indicated that smoothing the coarseness in the potential energy landscape by the CHARMM19/GBORN model predicts folded structures that poorly resemble the native state. These results are, nevertheless, still more realistic than most coarse-grained models in producing tertiary chain conformations that contain protein-like characteristics. It is plausible that some improvement in accuracy of the generated conformations may be gained by replacing the pairwise GB solvent model with GBMV2- β_S10 in a united-atom representation, albeit this hybrid approach has never been tested for numerical stability and may require reparameterization of the molecular-volume smoothing function.

Acknowledgment. We thank Dr. M. Feig and Dr. A. Wallqvist for helpful discussions, and also Dr. E. Gallicchio and Dr. R. Levy for making their software code available to us for estimating statistical uncertainties. Funding support for this work came from the Department of Defense High Performance Computing (HPC) Modernization Program Office, under the HPC Software Application Institute initiative, and the Army Medical Research and Materiel Command (Project Nos. 21248 and 132259). Computational time was provided by the U.S. Army Research Laboratory Major Shared Resource Center and the Advanced Biomedical Computing Center at the National Cancer Institute. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army.

References and Notes

- (1) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10.
- (2) Onuchic, J. N.; Nymeyer, H.; García, A.; Chahine, J.; Socci, N. *Adv. Protein Chem.* **2000**, *53*, 88.
- (3) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619.
- (4) Borreguero, J. M.; Nikolay, V.; Dokholyan, N. V.; Buldyrev, S. V.; Shakhnovich, E. I.; Stanley, H. E. *J. Mol. Biol.* **2002**, *318*, 863.
- (5) Cellmer, T.; Bratko, D.; Prausnitz, J. M.; Blanch, H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11692.
- (6) Jayachandran, G.; Vishal, V.; Pande, V. S. *J. Chem. Phys.* **2006**, *124*, 164902.
- (7) Snow, C. D.; Sorin, E. J.; Rhee, Y. M.; Pande, V. S. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 43.
- (8) Shea, J. E.; Brooks, C. L., III. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499.
- (9) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 151.

- (10) Im, W.; Lee, M. S.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1691.
- (11) Lee, M. S.; Salsbury, F. R.; Brooks, C. L., III. *J. Chem. Phys.* **2002**, *116*, 10606.
- (12) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1348.
- (13) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *25*, 265.
- (14) Zhu, J.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2005**, *109*, 3008.
- (15) Fan, H.; Mark, A. E.; Zhu, J.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6760.
- (16) Huang, A.; Stultz, C. M. *Biophys. J.* **2007**, *92*, 34.
- (17) Tjong, H.; Zhou, H. X. *J. Phys. Chem. B* **2007**, *111*, 3055.
- (18) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J. Chem. Theory Comput.* **2007**, *3*, 156.
- (19) Swanson, J. M. J.; Wagoner, J. A.; Baker, N. A.; McCammon, J. A. *J. Chem. Theory Comput.* **2007**, *3*, 170.
- (20) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
- (21) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587.
- (22) Lei, H.; Wu, C.; Liu, H.; Duan, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 4925.
- (23) Martinez, J. C.; Viguera, A. R.; Berisio, R.; Wilmanns, M.; Mateo, P. L.; Filimonov, V. V.; Serrano, L. *Biochemistry* **1999**, *38*, 549.
- (24) Cobos, E. S.; Filimonov, V. V.; Vega, M. C.; Mateo, P. L.; Serrano, L.; Martinez, J. C. *J. Mol. Biol.* **2003**, *328*, 221.
- (25) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
- (26) Dominy, B. N.; Brooks, C. L., III. *J. Phys. Chem. B* **1999**, *103*, 3765.
- (27) Chocholoušová, J.; Feig, M. *J. Comput. Chem.* **2006**, *27*, 719.
- (28) Feig, M.; Karanicolas, J.; Brooks, C. L., III. *J. Mol. Graphics Model.* **2004**, *22*, 377.
- (29) Musacchio, A.; Noble, M.; Pauptit, R.; Wierenga, R.; Saraste, M. *Nature* **1992**, *359*, 851.
- (30) Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- (31) Feig, M.; Brooks, C. L., III. *Proteins* **2002**, *49*, 232.
- (32) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195.
- (33) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
- (34) Shea, J. E.; Nochomovitz, Y. D.; Guo, Z. Y.; Brooks, C. L., III. *J. Chem. Phys.* **1998**, *109*, 2895.
- (35) Gallicchio, E.; rec, M.; Felts, A. K.; Levy, R. M. *J. Phys. Chem. B* **2005**, *109*, 6722.
- (36) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976.
- (37) Prabhu, N. V.; Sharp, K. A. *Annu. Rev. Phys. Chem.* **2005**, *56*, 521.
- (38) Hummer, G.; García, A. E.; Garde, S. *Phys. Rev. Lett.* **2000**, *85*, 2637.
- (39) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Sech, C.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3*, 26.
- (40) Formanek, M. S.; Cui, Q. *J. Comput. Chem.* **2006**, *27*, 1923.
- (41) Khandogin, J.; Brooks, C. L., III. *Biophys. J.* **2005**, *89*, 141.
- (42) Chen, J.; Won, H.-S.; Im, W.; Dyson, H. J.; Brooks, C. L., III. *J. Biomol. NMR* **2005**, *31*, 59.
- (43) Im, W.; Brooks, C. L., III. *J. Mol. Biol.* **2004**, *337*, 513.
- (44) Chen, J.; Im, W.; Brooks, C. L., III. *J. Am. Chem. Soc.* **2004**, *126*, 16038.
- (45) Chen, J.; Im, W.; Brooks, C. L., III. *J. Am. Chem. Soc.* **2006**, *128*, 3728.
- (46) Nymeyer, H. *J. Chem. Theory Comput.* **2008**, *4*, 626.
- (47) Trebst, S.; Troyer, M.; Hansmann, U. H. J. *J. Chem. Phys.* **2006**, *124*, 174903.
- (48) Hritz, J.; Oostenbrink, C. *J. Chem. Phys.* **2008**, *128*, 144121.
- (49) Li, X.; O'Brien, C. P.; Collier, G.; Vellere, N. A.; Wang, F.; Latour, R. A.; Bruce, D. A.; Stuart, S. D. *J. Chem. Phys.* **2007**, *127*, 164116.
- (50) Swanson, J. M. J.; Mongan, J.; McCammon, J. A. *J. Phys. Chem. B* **2005**, *109*, 14769.
- (51) Okur, A.; Roe, D. R.; Cui, G.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 420.
- (52) Zhou, R.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777.
- (53) Zhou, R. *Proteins* **2003**, *53*, 148.
- (54) Simmerling, C.; Strockbine, B.; Roitberg, A. *J. Am. Chem. Soc.* **2002**, *124*, 11258.
- (55) Lee, M. S.; Olson, M. A. *J. Phys. Chem. B* **2005**, *109*, 5223.
- (56) Nymeyer, H.; García, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934.
- (57) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins* **2004**, *56*, 310.
- (58) Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 1846.
- (59) Strodel, B.; Wales, D. J. *J. Chem. Theory Comput.* **2008**, *4*, 657.
- (60) Olson, M. A. *Proteins* **2004**, *57*, 645.
- (61) Calimet, N.; Schaefer, M.; Simonson, T. *Proteins* **2001**, *45*, 144.

JP802469G