

PEPTIDE AND PROTEIN FOLDING AND CONFORMATIONAL EQUILIBRIA: THEORETICAL TREATMENT OF ELECTROSTATICS AND HYDROGEN BONDING WITH IMPLICIT SOLVENT MODELS

By WONPIL IM, JIANHAN CHEN, AND CHARLES L. BROOKS, III

Department of Molecular Biology and Center for Theoretical Biological Physics,
The Scripps Research Institute, La Jolla, California 92037

I. Introduction	174
II. Generalized Born (GB) Models	176
A. GB Electrostatics Theory	176
B. Advances and Achievements	179
C. Remaining Opportunities for Continued Improvement	182
III. Peptide Folding and Conformational Equilibria	184
A. Influence of Backbone H-Bond Strength on Conformational Equilibria	184
B. Influence of Backbone Dihedral Energetics on Conformational Equilibria	189
IV. Concluding Discussion	190
References	192

ABSTRACT

Since biomolecules exist in aqueous and membrane environments, the accurate modeling of solvation, and hydrogen bonding interactions in particular, is essential for the exploration of structure and function in theoretical and computational studies. In this chapter, we focus on alternatives to explicit solvent models and discuss recent advances in generalized Born (GB) implicit solvent theories. We present a brief review of the successes and shortcomings of the application of these theories to biomolecular problems that are strongly linked to backbone H-bonding and electrostatics. This discussion naturally leads us to explore existing areas for improvement in current GB theories and our approach towards addressing a number of the key issues that remain in the refinement of these models. Specifically, the critical importance of balancing solvation forces and intramolecular forces in GB models is illustrated by examining the influence of backbone hydrogen bond strength and backbone dihedral energetics on conformational equilibria of small peptides.

I. INTRODUCTION

Biological function in peptides and proteins is dictated by their conformational equilibria as facilitated through changes of secondary structure and tertiary contacts mediated by hydrophilic and hydrophobic interactions. In particular, hydrogen bonds (H-bonds) formed between backbone carbonyl oxygens and amide protons play a critical role in determining the conformational states of peptide chains. Accurate modeling of such interactions in various environments of interest is therefore a key element in the applications of chemical theory to understanding biological structure and function. For instance, because proteins exist in a predominantly aqueous environment, intramolecular H-bonds compete with protein–water H-bonds. It has been an ongoing interest in the area of molecular modeling to develop theoretical models, or force fields, that capture the delicate balance between solvation forces and intramolecular forces (Cornell *et al.*, 1995; Jorgensen and Tirado-Rives, 1988; MacKerell, 2004; MacKerell *et al.*, 1998; van Gunsteren and Berendsen, 1990). One general difficulty is that the force fields optimized with high-level quantum mechanics in vacuum are not directly transferable to solvent environments (MacKerell, 2004). Furthermore, the paucity of direct experimental measurements of such solvent-mediated interactions has also added to the difficulty in calibrating molecular force fields. It appears that one ultimately needs to examine and improve the quality of the force fields in the context of direct comparison between available experimental observables and simulations (or calculations), through protein–protein or protein–ligand binding thermodynamics, scoring of protein conformations in structure prediction, and peptide and protein folding and unfolding studies.

The most straightforward way of accounting for solvation and its effects on biomolecules is to explicitly include the solvent molecules (primarily water, but sometimes with lipid membranes) (Brooks and Karplus, 1986; Roux, 2002). While this approach yields the most detailed information that can be obtained, the high computational expense, due to (generally uninteresting) solvent molecules, makes it difficult to apply such a methodology routinely to the aforementioned applications. An alternative, which addresses issues of computational efficiency while maintaining physical accuracy, are efforts that have been directed to the development of so-called implicit solvent models in which the mean influence of the solvent molecules around the solute is approximated without having to treat the solvent explicitly (Feig and Brooks, 2004; Lazaridis and Karplus, 2000; Roux and Simonson, 1999; Roux *et al.*, 2000). Depending on the theoretical approaches used to describe the solvation of the solute, implicit solvent models are mainly classified as (effective) empirical solvation models or

continuum electrostatics solvation models (Lazaridis and Karplus, 2000; Roux and Simonson, 1999). In the former approach the solvation-free energy of the solute is generally expressed as a sum of atom or group contributions. Their desolvation energetics due to burial inside the solute is taken into account by weighting the solvent-accessible surface area or solvent-exclusion volume of each atom or group in accord to an empirical-free energy scale. In this way, the solvent-screened charge–charge interactions are included empirically. In the latter approach, continuum electrostatics forms the basis of the theoretical models for solvation. In such theories, the solute interior and the solvent region are described as featureless “low” and “high” dielectric regions, respectively. The dielectric difference leads to the development of surface charges at the dielectric boundary, also called the reaction field potential. Typically, the electrostatic solvation energy of a solute with an arbitrary shape, including the solvent-screened charge–charge interactions, is calculated from numerical solutions of the Poisson–Boltzmann (PB) equation using finite-difference methods (Im *et al.*, 1998; Klapper *et al.*, 1986; Nicholls and Honig, 1991; Warwicker and Watson, 1982). Despite some intrinsic shortcomings arising from the absence of the granularity of solvent molecules, implicit solvent models have been applied quite successfully to the aforementioned research areas (Feig and Brooks, 2004; Lazaridis and Karplus, 2000).

PB continuum electrostatics is the most rigorous and popular method used to estimate the electrostatic solvation energy of a solute with an arbitrary shape, and particular successes in applications to complex biomolecular problems are evident (Honig and Nicholls, 1995; Murray and Honig, 2002; Roux *et al.*, 2000). However, the computational cost of solving the PB equation remains a bottleneck to its application to protein folding and routine dynamics simulations of biomolecules, despite the progress in fast PB computational methodologies (David *et al.*, 2000; Luo *et al.*, 2002; Prabhu *et al.*, 2004). The generalized Born (GB) model, inspired by the Born equation for solvation energies of ions (Born, 1920), has emerged as an efficient alternative for implicit inclusion of the electrostatic solvation energy (Dominy and Brooks, 1999; Ghosh *et al.*, 1998; Hawkins *et al.*, 1996; Im *et al.*, 2003b; Lee *et al.*, 2002, 2003; Onufriev *et al.*, 2000, 2002; Qiu *et al.*, 1997; Scarsi *et al.*, 1997; Schaefer and Karplus, 1996; Spassov *et al.*, 2002; Srinivasan *et al.*, 1999; Still *et al.*, 1990). The GB model is intrinsically based on the same underlying continuum approximation as used in PB theory and thus its accuracy is naturally assessed by comparison with the PB results (Feig and Brooks, 2004; Feig *et al.*, 2004c).

This chapter focuses on the recent developments of GB models and their applications to biomolecular problems that are strongly linked to backbone H-bonds and electrostatics. Brief reviews of successes and

failures of these techniques naturally led us to discuss existing areas for improvement in current GB models and our approach toward addressing a number of the key issues.

II. GENERALIZED BORN (GB) MODELS

In this section, GB electrostatics theory is first briefly described to illustrate the main underlying principles. Then the advances of the models and their achievements in several biological applications are discussed together with issues and opportunities for continued development.

A. GB Electrostatics Theory

The solvation free energy, corresponding to the work required to insert a solute of fixed conformation into a polar solvent, can be approximated as the sum of nonpolar (np) and electrostatic (elec) contributions, that is, $\Delta G_{\text{solv}} = \Delta G_{\text{elec}} + \Delta G_{\text{np}}$ (Roux and Simonson, 1999). The nonpolar solvation energy (ΔG_{np}), which is treated fully empirically, includes the energetic penalty of forming a cavity in the solvent (ΔG_{cav}) and the solvent-solute van der Waals dispersion interactions (ΔG_{vdW}). A popular model is to estimate ΔG_{np} as the product of the solvent-exposed surface area (SA), S , of the solute and a phenomenological surface tension coefficient γ (Gilson *et al.*, 1993; Hermann, 1972; Simonson and Brunger, 1994),

$$\Delta G_{\text{np}} = \Delta G_{\text{vdW}} + \Delta G_{\text{cav}} \quad (1)$$

$$\approx \gamma S \quad (2)$$

However, we note that work by Levy and co-workers has explored an efficient computational method to explicitly include the solvent dispersion term just described, ΔG_{vdW} , in the context of a continuum approach and used Eq. (1) to evaluate ΔG_{np} (Gallicchio and Levy, 2004; Levy *et al.*, 2003).

The electrostatic solvation free energy, ΔG_{elec} , of the solute is the work required to assemble the charges, $\{q_\alpha\}$, of the solute in the solvent. Based on continuum electrostatics, in which the solvent is represented as a featureless high dielectric medium, ΔG_{elec} can be expressed in terms of the reaction field potential $\phi_{\text{rf}}(\mathbf{r})$ or the reaction field Green's function $G_{\text{rf}}(\mathbf{r}, \mathbf{r}')$ (Klapper *et al.*, 1986; Sharp and Honig, 1990; Warwicker and Watson, 1982),

$$\begin{aligned}\Delta G_{\text{elec}} &= \frac{1}{2} \sum_{\alpha} q_{\alpha} \phi_{\text{rf}}(\mathbf{r}_{\alpha}) \\ &= \frac{1}{2} \sum_{\alpha\beta} q_{\alpha} G_{\text{rf}}(\mathbf{r}_{\alpha}, \mathbf{r}_{\beta}) q_{\beta},\end{aligned}\quad (3)$$

where $G_{\text{rf}}(\mathbf{r}, \mathbf{r}')$ corresponds to the reaction field potential at \mathbf{r} due to a point charge at \mathbf{r}' . The reaction field potential $\phi_{\text{rf}}(\mathbf{r})$ can be computed by solving the Poisson equation or the PB equation if the influences of salt are included [Eq. (4)], numerically using finite-difference methods (Im *et al.*, 1998; Klapper *et al.*, 1986; Nicholls and Honig, 1991; Warwicker and Watson, 1982)

$$\nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r}) \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}), \quad (4)$$

where $\epsilon(\mathbf{r})$, $\bar{\kappa}(\mathbf{r})$, and $\rho(\mathbf{r})$ are the dielectric constant, the modified Debye–Hückel screening factor (which accounts for nonspecific ionic strength effects of electrostatic interactions), and the fixed charge density of the solute, respectively. Even though it is possible to obtain (numerically) stable electrostatic solvation forces in the context of the finite-difference method (Im *et al.*, 1998; Luo *et al.*, 2002; Prabhu *et al.*, 2004), such calculations are generally too expensive to perform long molecular dynamics (MD) simulations of biomolecules.

An efficient alternative, which uses a simple analytical formula for the reaction field Green’s function $G_{\text{rf}}(\mathbf{r}, \mathbf{r}')$ in Eq. (3), is given by the GB formula, first proposed by Still and co-workers (1990),

$$\Delta G_{\text{elec}} = -\frac{1}{2} \tau \sum_{\alpha\beta} \frac{q_{\alpha} q_{\beta}}{\sqrt{r_{\alpha\beta}^2 + R_{\alpha}^{\text{GB}} R_{\beta}^{\text{GB}} \exp(-r_{\alpha\beta}^2/4R_{\alpha}^{\text{GB}} R_{\beta}^{\text{GB}})}}, \quad (5)$$

where R_{α}^{GB} is the “effective Born radius” of atom α and $\tau = 1/\epsilon_{\text{p}} - 1/\epsilon_{\text{s}}$; ϵ_{p} represents the (low) dielectric response of the interior of the solute and ϵ_{s} the (high) solvent dielectric constant. ΔG_{elec} in Eq. (5) corresponds to the electrostatic free energy of transferring the solute in a medium of ϵ_{p} to a medium of ϵ_{s} . In principle, the “exact” effective Born radii can be calculated by performing PB calculations for one atom at a time by setting all other charges to zero and then inserting the calculated self (or atomic) electrostatic solvation energy $\Delta G_{\text{elec},\alpha}$ into the Born equation (Born, 1920),

$$\Delta G_{\text{elec},\alpha} = -\frac{1}{2} \tau \frac{q_{\alpha}^2}{R_{\alpha}^{\text{GB}}}. \quad (6)$$

$\Delta G_{\text{elec},\alpha}$ or R_{α}^{GB} from PB calculations serve as standard benchmarks for assessing various approximate GB theories.

The effective Born radius represents the distance between a particular atom and a hypothetical spherical dielectric boundary chosen to satisfy Eq. (6). The principal assumption in the GB method is that the solvent-shielded charge–charge interactions in PB can be reproduced by the cross-term in Eq. (5) together with the effective Born radii. It can be shown that Eq. (5) is exact for a pair of atoms, α and β , in the limiting cases of $r_{\alpha\beta} \rightarrow 0$ and $r_{\alpha\beta} \rightarrow \infty$ (Onufriev *et al.*, 2000; Still *et al.*, 1990). Indeed, the GB model of Eq. (5) has been shown to closely reproduce ΔG_{elec} from PB calculations, provided that the effective Born radii are accurate (Feig *et al.*, 2004c). As such, most of the extensive literature on extensions of the GB theory has been focused on efficient and accurate evaluation of the Born radii (Dominy and Brooks, 1999; Gallicchio and Levy, 2004; Ghosh *et al.*, 1998; Hawkins *et al.*, 1996; Im *et al.*, 2003b; Lee *et al.*, 2002, 2003; Onufriev *et al.*, 2000, 2002; Qiu *et al.*, 1997; Scarsi *et al.*, 1997; Schaefer and Karplus, 1996; Srinivasan *et al.*, 1999; Still *et al.*, 1990).

Briefly, the basic idea for the estimation of the effective Born radius is based on the so-called Coulomb field approximation (CFA), which is exact for a charge in the center of a spherical cavity and assumes that the dielectric displacement follows Coulomb’s law, independent of the external dielectric (Bashford and Case, 2000; Scarsi *et al.*, 1997; Still *et al.*, 1990). Expressing $\Delta G_{\text{elec},\alpha}$ in terms of the dielectric displacement results in the following expression that can be used for R_{α}^{GB} through Eq. (6) [for a complete derivation, see Bashford and Case (2000)],

$$\Delta G_{\text{elec},\alpha}^0 = -\frac{1}{2} \tau q_{\alpha}^2 \left(\frac{1}{\eta_{\alpha}} - \frac{1}{4\pi} \int_{r>\eta_{\alpha}} d\mathbf{r} \frac{v(\mathbf{r}; \{\mathbf{r}_{\alpha}\})}{|\mathbf{r} - \mathbf{r}_{\alpha}|^4} \right), \quad (7)$$

where η_{α} is an arbitrarily defined integration starting point, necessary to avoid the singularity at $|\mathbf{r} - \mathbf{r}_{\alpha}| = 0$, and $v(\mathbf{r})$ is a solute volume function, which has a value of one in the interior of a solute and zero in the solvent region. The CFA, Eq. (7), is the basis for most GB implementations and various GB models mainly differ on how the volume integral is evaluated. The integral is usually evaluated by numerical surface/volume integration methods (Ghosh *et al.*, 1998; Im *et al.*, 2003b; Lee *et al.*, 2002, 2003; Scarsi *et al.*, 1997; Still *et al.*, 1990) or pairwise summation approximations (Dominy and Brooks, 1999; Gallicchio and Levy, 2004; Hawkins *et al.*, 1996; Qiu *et al.*, 1997) [for more detailed information, see also Zhu *et al.*, (2005)]. It should be noted that a more elaborate formalism than the CFA is required to accurately estimate the effective Born radii in different dielectric environments, that is, as a function of the dielectric constant of the solvent and the solute (Feig *et al.*, 2004a; Sigalov *et al.*, 2005).

B. Advances and Achievements

1. Performance of GB Compared to PB

Because the GB model is intrinsically based on the same underlying continuum representation as in PB theory, the accuracy of various GB models has been primarily examined by quantitative comparisons between ΔG_{elec} from PB and ΔG_{elec} from GB for small molecules as well as folded and misfolded protein conformations. To overcome the well-known issue that the CFA underestimates $\Delta G_{\text{elec},\alpha}$ and thus overestimates R_{α}^{GB} compared to PB results, Lee *et al.* (2002, 2003) introduced higher order terms to Eq. (7) as heuristic corrections beyond the CFA, resulting in effective Born radii very close to PB results and, consequently, more accurate estimates of total solvation energies; less than 1% error on average was achieved for absolute electrostatic solvation energies of a large set of proteins and for relative solvation energies of protein conformations, compared to the corresponding PB solvation energies. Feig *et al.* (2004c) summarized the performance of various GB implementations against PB calculations. By using the same smoothed dielectric boundary in both PB and GB, Im *et al.* (2003b) showed that GB could reproduce corresponding PB electrostatic solvation forces as well as energies. Over the past decade, methodological developments in GB formalisms have reached a mature stage in which the accuracy of the GB models is almost identical to the PB method. Furthermore, recent PB implicit solvent MD simulations may provide an opportunity for direct comparisons between PB and GB models for dynamic properties of proteins (Prabhu *et al.*, 2004).

2. Scoring of Protein Conformations

Distinguishing native and near-native structures from nonnative decoys is an important task in protein structure prediction. The nature of these problems reinforces the point that the scoring function must be fast and accurate enough for good scoring performance, which provides motivation, challenge, and validation to the physics-based energy functions (e.g., modern molecular mechanics plus continuum dielectric solvent models). While there exists some evidence that a GB model might not help for this purpose (Morozov *et al.*, 2003), a number of other studies have shown that the physics-based energy functions perform well in distinguishing native and near-native folds (Dominy and Brooks, 2001; Feig and Brooks, 2002; Felts *et al.*, 2002; Zhu *et al.*, 2003) and loop conformations (Fiser *et al.*, 2002; Forrest and Woolf, 2003) from a large set of nonnative decoys. Recent examination of GB and PB implicit solvation models in the context of computational protein design revealed a systematic bias that the burial of

polar amino acids in the protein interior is more favored than that of nonpolar ones, which can be problematic in protein design (Jaramillo and Wodak, 2005). This study suggests that further improvements of the GB implicit solvent models can be achieved (see Section II, C for more detailed discussions).

3. *Protein-Protein and Protein-Ligand Binding*

Molecular recognition such as occurs in protein-protein, protein-ligand, and protein-DNA interactions is a key process for many biological functions. Proper evaluation of protein-protein interactions will be very helpful in understanding cellular processes (Wodak and Mendez, 2004), and insights into protein-ligand interactions are necessary for drug discovery (Ferrara *et al.*, 2004). However, it is this area where successful applications of GB implicit solvent models have appeared to be limited (Ferrara *et al.*, 2004; Wang and Wade, 2003), although some success has been documented (Gohlke and Case, 2003). Therefore, further optimization of implicit solvent force fields for this purpose is anticipated.

4. *Protein Folding/Unfolding*

One of the most exciting aspects of developing efficient implicit solvent models is the potential they provide for studies of protein folding/unfolding in atomic detail (Bursulaya and Brooks, 2000). Considerable efforts have focused on examination of the implicit solvent models by close comparison with explicit solvent simulations (Bursulaya and Brooks, 2000; Nymeyer and Garcia, 2003; Zhou, 2003; Zhou and Berne, 2002) and with experimental data (Lin *et al.*, 2003; Pitera and Swope, 2003; Steinbach, 2004; Zagrovic *et al.*, 2001). Discrepancies in both global minima of free energy surfaces and detailed local structures, such as salt bridge formation and helical content, have been observed, indicating that room exists for further improvement of the GB implicit solvent models (see Section II, C for more detailed discussions). Nevertheless, successful applications of implicit solvent models to address more specific problems in protein folding have been reported (Karanicolas and Brooks, 2004; Ohkubo and Brooks, 2003). In addition, recent years have seen quite a few successful applications of various GB implicit solvent models to *ab initio* structure prediction for a number of miniproteins, such as protein A (Jang *et al.*, 2003a), villin headpiece (Jang *et al.*, 2003a; Liu and Beveridge, 2002), Trp-cage (Pitera and Swope, 2003), Trp-zip (Okur *et al.*, 2003; Steinbach, 2004; Yang *et al.*, 2004), $\beta\beta\alpha$ motifs (Jang *et al.*, 2003b), and the fd Coat transmembrane protein (Im and Brooks, 2004). These calculations have demonstrated the feasibility of general structure prediction procedures based on efficient

physics-based force fields, especially when combined with advanced sampling techniques such as the replica-exchange (REX) method (Feig *et al.*, 2004b; Hansmann, 1997; Sugita and Okamoto, 1999).

5. Nuclear Magnetic Resonance (NMR) and X-Ray Structure Refinement

Electrostatic interactions are often oversimplified or ignored in the energy functions for NMR and X-ray structure calculations because it is difficult to evaluate them reliably without proper description of the dielectric screening by solvent. In light of recent improvements in implicit solvent models, Xia *et al.* (2002) showed that simulated annealing refinement in a GB implicit solvent could lead to noticeable improvement in the final protein NMR structures in terms of the backbone dihedral angle distributions and hydrogen bond patterns. However, the impact of an implicit solvent is rather small when a sufficient number of experimental restraints exist (such as in the final stage of NMR structure determination). In contrast, it was demonstrated that replica exchange molecular dynamics (REX-MD) refinement in a GB implicit solvent model could significantly improve the quality of structures and the radius of convergence when experimental data are limited (Chen *et al.*, 2004). For example, while conventional structure calculations using an initial set of sparse NOE restraints were unable to identify a unique topology for a protein domain, high-quality native-like initial folds were generated through REX-MD refinement of the initial structures with a GB implicit solvent (Chen *et al.*, 2005). These models could be then used to make further assignments of ambiguous NOEs and speed up the structure determination process. Moulinier *et al.* (2003) and Korostelev *et al.* (2004) documented the use of continuum electrostatics PB and GB approaches for refining X-ray structures.

6. Constant pH Molecular Dynamics

The stability and function of proteins, as well as many biological processes, are dependent on the environmental pH; examples include fibril formation of amyloid peptides and prion proteins (Clippingdale *et al.*, 2001; Kelly, 1997), membrane fusion of influenza virus (Bullough *et al.*, 1994), and proton gradient-driven ATP synthesis (Rastogi and Girvin, 1999). Conventional MD simulations use predefined protonation states of the protein, making it difficult to explore such pH-coupled biological phenomena. To overcome these difficulties and achieve efficient conformational sampling, pH-coupled MD simulation techniques based on GB implicit solvent models have been developed (Dlugosz and Antosiewicz,

2004; Khandogin and Brooks, 2005; Lee *et al.*, 2004; Mongan and Case, 2005; Mongan *et al.*, 2004). The protonation states of titratable residues are adjusted on the fly based on their pK_a values during the simulations either discontinuously using a Monte Carlo (MC) technique (Dlugosz and Antosiewicz, 2004; Mongan *et al.*, 2004) or continuously using a λ -dynamics technique (Khandogin and Brooks, 2005; Lee *et al.*, 2004).

7. *Implicit Membrane Environment*

PB continuum electrostatics has proven useful in many aspects of membrane modeling (Im and Roux, 2002; Murray and Honig, 2002; Roux *et al.*, 2000); however, repeated numerical solutions to the PB equation remain too cumbersome for applications requiring extensive conformational sampling. To circumvent these difficulties, a number of methodologies for implicit membrane modeling based on GB electrostatics theory have appeared in the past few years (Im *et al.*, 2003a; Spassov *et al.*, 2002; Tanizaki and Feig, 2005). These methods provide a rapid means of evaluating the energies and forces of proteins in a membrane environment and may serve as platforms for exploring dynamics, insertion, folding, and assembly of membrane proteins and peptides. For example, by combining a membrane GB model with the REX-MD method, it has been shown that it is possible to fold and assemble simple helical membrane peptides (Im *et al.*, 2003a), predict the structure of a small membrane-bound protein and reproduce its solid-state NMR properties reasonably well (Im and Brooks, 2004), and explore the membrane insertion mechanism and interfacial folding of designed peptides (Im and Brooks, 2005).

In summary, it is evident that GB formalisms have reached a mature stage and that their accuracy is essentially identical to the PB method (Feig *et al.*, 2004c). Many successful applications to various biological problems are encouraging, demonstrating the great potential of the GB implicit solvent models for studies of biomolecular structure and function. Nonetheless, the limited successes in protein folding, and protein-protein and protein-ligand binding point to remaining opportunities for continued improvement of GB implicit solvent force fields.

C. *Remaining Opportunities for Continued Improvement*

Continuum dielectric solvent models may yield considerable disagreement with explicit water simulations, especially when the detailed interplay of a few water molecules (which are significantly distinct from the bulk water) in solvent-mediated intramolecular interactions exists. Despite this limitation, the previous section demonstrates that there are many

biological problems for which GB-based continuum dielectric solvent models can provide insights that are very difficult to gain from explicit solvent models. The successes and failures of various implicit solvent models in applications to the aforementioned biological problems arise in principle from their ability in balancing delicate energetics between sets of the competing interactions, that is, the solvation preference of each side chain and the peptide backbone in aqueous bulk solution versus the strength of solvent-mediated interactions between these moieties in a complex protein environment. In terms of electrostatic interactions, the intramolecular Coulombic interaction energy in the protein is known to be strongly anticorrelated with the electrostatic solvation energy. Similarly, the intramolecular van der Waals dispersion interaction energy in the protein also strongly anticorrelates with the nonpolar solvation energy. These competing, opposing forces mostly cancel each other, and a shift in the balance, depending on the extent of specific interactions in a given protein conformation and environment, can lead to a bias in conformational equilibria. To what extent a GB force field can capture this delicate balance appears to be a key in the success of its applications. Achieving sufficient balance of these competing interactions in a force field for complex biological systems is a very challenging task. In addition to the general difficulty that force fields optimized with high-level quantum mechanics are not directly transferable to solvent environments, the lack of direct experimental data on solvation energies of proteins, as well as the pairwise interactions between polar groups in solvent environments, has also added to difficulties in improving the quality of implicit solvent force fields. As such, it appears that one has to resort to thermodynamic data from explicit water simulations and available experimental observables such as protein stability and conformation equilibria.

Toward this end, it might be a reasonable starting point to examine the solvent-mediated interactions between polar groups and further optimize the implicit solvent force field based on potentials of mean force (PMF) calculated using explicit solvent MD simulations. Masunov and Lazaridis (2003) carefully examined explicit water PMFs between all possible ionizable amino acid side chain pairs in various protonation states. Comparisons with various implicit solvent models revealed that stronger salt bridges are formed in implicit solvent models than in explicit water, which has been reported as well in other studies (Nymeyer and Garcia, 2003; Zhou, 2003; Zhou and Berne, 2002). This overstabilization might be amplified even more in the low dielectric protein interior, which appears to be problematic in many applications, such as protein design (Jaramillo and Wodak, 2005). In continuum dielectric solvent models, the extent of solvent exposure of each atom at the dielectric boundary dictates all of the electrostatic

and most of the nonpolar solvation energetics. Thus, it is physically appropriate to optimize the input radii, by which the low dielectric region and the high dielectric region are divided, not only based on solvation of individual sidechains, but also in consideration of solvent-mediated interactions. In principle, partial charges, Lennard–Jones parameters, and backbone dihedral energetics in current molecular mechanics force fields may also need to be adjusted for a specific implicit solvent model to achieve the sufficient balance. However, considering that the current force fields have been calibrated extensively over the past decades to achieve proper solvent–solute and solute–solute interactions in explicit solvent models, at this moment, adjusting the input radii in implicit dielectric solvent models to reproduce explicit-solvent PMFs is a reasonable “first pass” approach. As an illustration and first step toward this aim, in the next section the backbone input radii for a GB model are optimized to reproduce solvent-mediated backbone H-bond strength in an explicit water model and then the influence of backbone H-bond strength on peptide conformational equilibria is examined and verified by folding simulations of small peptides. In addition, we also illustrate the influence of backbone dihedral energetics on conformation equilibria.

III. PEPTIDE FOLDING AND CONFORMATIONAL EQUILIBRIA

A. *Influence of Backbone H-Bond Strength on Conformational Equilibria*

Backbone H-bonds play an important role in determining the conformational states of peptide chains. Their inherent stability is therefore a key element in the calibration of theoretical models. For example, overstabilization of backbone H-bonds cannot only lead to excessive helical content, but also hinders efficient sampling of the protein conformation space by creating energetic traps. Destabilization of backbone H-bonds in a force field, however, can greatly reduce the overall stability of the protein structure and result in its rapid unfolding. Strong dependence of the H-bond strength on its environment makes it very difficult to choose representative model systems to explore these issues. For example, Kelly and co-workers employed an amide-to-ester perturbation to estimate the position-dependent backbone H-bond contributions to protein stability (Deechongkit *et al.*, 2004a,b); backbone H-bonds were estimated to stabilize a three-stranded β protein by 1.5 to 5.0 kcal/mol, depending on the position in the protein. Furthermore, the lack of direct experimental data has also added to the difficulty in calibrating force fields for such

interactions. It seems that one ultimately needs to examine the quality of the force field in the context of actual peptide and protein folding and unfolding simulations to achieve the proper balance of interactions. An ideal force field should mimic the delicate balance among the competing interactions and be capable of providing correct conformational equilibria.

Explicit solvent simulations arguably provide the highest accuracy in describing the conformational equilibria of peptides and proteins. The paucity of direct experimental data of the H-bond strength makes these simulations a useful guide for calibrating implicit solvent models, with the caveat that these force fields are also not perfect. In the present work, to calibrate the backbone parameters of a GB implicit solvent model, we have utilized the potential of mean force (PMF) of H-bond formation between a modified alanine dipeptide dimer, shown in Fig. 1. The approach is then verified by folding simulations of a synthetic peptide with the sequence of $(AAQAA)_3$ and a small peptide from residues 101–111 of α -lactalbumin (α -lac) whose conformational equilibria have been determined previously (Demarest *et al.*, 1999; Shalongo *et al.*, 1994). We note that these efforts represent only preliminary steps in the development of a fully consistent implicit (GB) solvent force field, but the present studies illustrate the general approach one should use in developing such a force field.

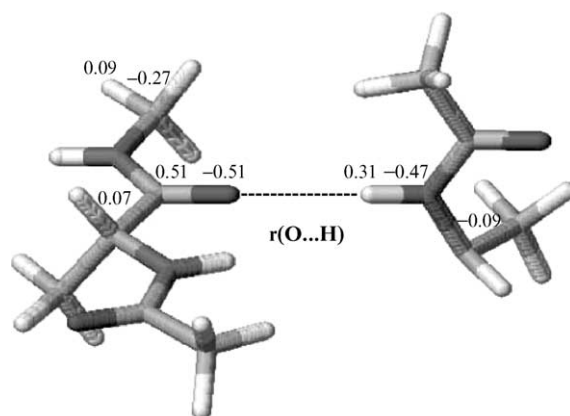


FIG. 1. The modified alanine dipeptide dimer with the partial charges and the reaction coordinate [$r(O \dots H)$ for hydrogen bond formation] indicated. The dimer is constrained to move along a straight line, and the relative orientation of two molecules, defined by planes of heavy atoms, are also constrained to be perpendicular to each other. Note that the H-bond partner is a model system with only the backbone H-bonding functionality included.

In the explicit solvent simulations the modified alanine dipeptide dimer is constrained to move along a straight line in a specific orientation using the MMFP module in CHARMM (Brooks *et al.*, 1983). The system was solvated by about 750 water molecules in a rectangular box with periodic boundary conditions. The all-atom parameter set PARAM22 for the solute (MacKerell *et al.*, 1998) and a modified TIP3P water model (Jorgensen *et al.*, 1983) were used. To remove the artifacts associated with truncation of electrostatic forces, electrostatic interactions were calculated using the particle mesh Ewald method (PME) (Essmann *et al.*, 1995). The van der Waals energy was smoothly switched off at 10–12 Å by use of a switching function (Brooks *et al.*, 1985; Steinbach and Brooks, 1994). Biased sampling along the reaction coordinate was carried out using the umbrella sampling technique (Torrie and Valleau, 1977), and the final PMF was calculated using the weighted histogram analysis method (WHAM) (Boczko and Brooks, 1993; Kumar *et al.*, 1992; Roux, 1995). For each window, equilibration simulations of 60 ps at constant pressure and temperature (NPT) were followed by 1 ns of production sampling at constant volume and temperature (NVT). The SHAKE algorithm (Ryckaert *et al.*, 1977) was applied to fix lengths of all bonds involving hydrogen atoms and a time step of 2 fs was used. Corresponding PMFs in implicit solvent were computed directly by translating the molecules away from each other along the reaction coordinate. Note that PMFs computed with the aforementioned setup do not include the contribution of solute conformational entropy. However, this contribution is assumed to be similar in both explicit and implicit solvent models and thus omitting it in both cases should not affect the GB optimization. Figure 2 shows the PMF of H-bond formation in the dipeptide dimer in TIP3P water. The H-bond

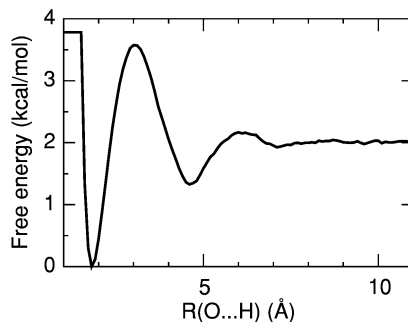


FIG. 2. Free energy profile for hydrogen bond formation of the modified alanine dipeptide dimer in TIP3P water.

stability appears to be about 2.0 kcal/mol, which falls within previously determined theoretical estimates of H-bond strengths, ranging from less than 0.5 kcal/mol for the well-exposed formamide dimer (Sneddon *et al.*, 1989) to about 2.7 kcal/mol in a β -sheet environment (Tobias *et al.*, 1992).

As mentioned in Section II,C, one of the key parameters in GB implicit solvent models is the input atomic radii that are used to define the solvent-solute dielectric boundaries. The choice of optimal radii is coupled intrinsically with the choice of the boundary surfaces. In the present study, we focus on optimization of a GB model with a simple switching function (GBSW) (Im *et al.*, 2003b) and ϵ_p is set to 1 to be consistent with the molecular mechanics force field. The input radii consistent with such a surface definition have been optimized previously based on the radial solvent charge distribution as well as the charging free energies calculated by MD free energy simulations (MD/FES) for 20 standard amino acids (Nina *et al.*, 1997, 1999) (hereinafter referred to as the Nina's radii). However, even though the Nina's radii set has been shown to work well in several applications from folding to NMR structure refinement (see Section II,B), optimizing the overall electrostatic solvation free energy does not explicitly balance the interactions between polar groups in the solvent, which might be more important in studying large conformational changes, such as in folding simulations. For example, the modified alanine dipeptide dimer is overstabilized by about 0.4 kcal/mol in the GBSW implicit solvent with Nina's radii (see the black curve in Fig. 3a). Note that with the smoothed surface definition, the solvation peak around 3.0 Å is reduced greatly (see Fig. 2). While this may alter the kinetics of H-bond formation, the lack of kinetic barriers can speed up conformational sampling and is thus advantageous when one is interested primarily in identifying the most stable thermodynamic states, such as in *ab initio* protein structure prediction. As demonstrated clearly in Fig. 3, small changes in the underlying backbone interactions can translate into substantial changes in conformational equilibria of the peptides. The average helicity of (AAQAA)₃ has been measured by NMR to be around 50% at 274 K. This information can be used as a guide, in addition to the explicit solvent simulations, to identify and verify a set of optimal input radii. Due to the uncertainties in simulations as well as experiments, multiple sets of atomic radii, which appear to provide comparable results for a single system, should be then examined further using folding simulations of mini and small proteins. For example, Fig. 4 compares the residue helicity of peptide α -lac using Nina's radii and an optimized radii set in which the amide nitrogen was set to 1.95 Å (the purple curves in Fig. 3a). It also demonstrates that modified radii lead to satisfactory agreement with the experimental results that residues 108–111 are largely unstructured

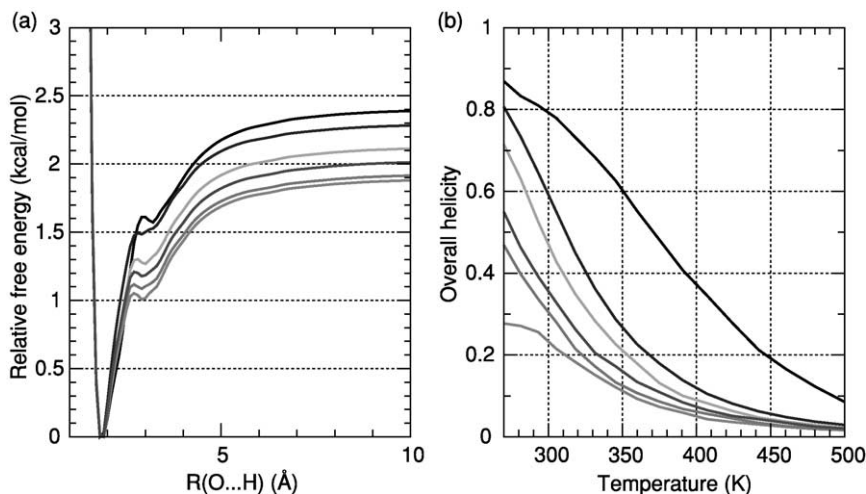


FIG. 3. (a) Free energy profiles for H-bond formation in the modified alanine dipeptide dimer in GBSW implicit solvent, and (b) corresponding simulated helicity of $(AAQAA)_3$ as a function of temperature. The same colored curves in both panels were obtained using the same input radii. Only the input radius of the amide nitrogen was adjusted from 2.23 Å (black) in original Nina's radii to 2.15 Å (blue), 2.0 Å (yellow), 1.95 Å (purple), 1.9 Å (green), and 1.85 Å (red). Replica exchange molecular dynamics (REX-MD) simulations of 10 ns for each curve were carried out using 16 replicas from 270 to 500 K to obtain the conformational equilibria (Feig *et al.*, 2004b; Sugita and Okamoto, 1999). Conformations from the last 8 ns were included in the helicity calculation. Longer simulations, as well as repeated simulations, appear to indicate that 10 ns is sufficient for the convergence of computed helicity (data not shown). The helicity was computed from average 1–4 H-bond frequency defined by $d_{O_i,HN_{i+4}} \leq 2.6$ Å, where $d_{O_i,HN_{i+4}}$ is the distance between the carbonyl oxygen of residue i , O_i , and the amide hydrogen of residue $i + 4$, HN_{i+4} . Note that using backbone dihedral criteria resulted in similar but shifted helicity curves (data not shown). Original Nina's radii (black curves) clearly overstabilize the dimer and give rise to helicity larger than experimental values. (See Color Insert.)

(Demarest *et al.*, 1999). Note that while such small adjustments of input radii can greatly improve the backbone interactions as reflected in better agreement of simulated conformational equilibria with experimental results, they do not significantly alter the overall electrostatic solvation-free energy. For example, the electrostatic solvation-free energy of alanine dipeptide changes only by about 5% from -14.8 kcal/mol (Nina's radii) to -15.6 kcal/mol (the optimized set) in the GBSW implicit solvent. Multiple nanosecond REX-MD simulations starting from the native structures for several small- to medium-sized proteins verified that these proteins are still sufficiently stable with optimized radii (data not shown). Extensive control

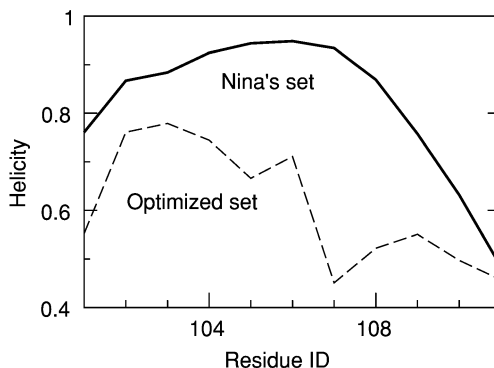


FIG. 4. Residue helicity of a small peptide from residues 101–111 of α -lactalbumin IIDYWLAHKALA, calculated with Nina's radii and an optimized radii set where the amide nitrogen input radius was adjusted from original 2.23 to 1.95 Å. Conformations at 270 K extracted from the last 8 ns of 10-ns REX-MD simulations using 16 replicas were used to compute the helicity. A residue is considered helical if it belongs to a segment of at least three residues whose backbone dihedral angles are within 30° from $(\phi, \psi) = (-57^\circ, -47^\circ)$. Note that the short length of the peptide made it very difficult to use the same hydrogen bond definition used in Fig. 3, i.e., $d_{O_r,HN_{r+4}} \leq 2.6$ Å.

and folding simulations of more complex systems, as well as similar calibration of the interactions between polar side chains, are necessary to more rigorously test and calibrate the implicit solvent force field.

B. Influence of Backbone Dihedral Energetics on Conformational Equilibria

In addition to backbone H-bonds, accurate treatment of the peptide backbone dihedral energetics is crucial to achieve correct conformational distributions in simulation studies (Feig *et al.*, 2003; MacKerell *et al.*, 2004a,b). For example, correct α - and π -helical contents were obtained by modifying the CHARMM22 backbone ϕ/ψ potential surface to reproduce high-level quantum mechanical (QM) calculations using CMAP dihedral cross terms (Feig *et al.*, 2003). Further modification was later introduced empirically based on explicit solvent MD simulations of several proteins in both crystal and aqueous environment (MacKerell *et al.*, 2004a), where agreement of simulated backbone ϕ, ψ distributions with experimental crystallographic data was the target. However, these empirical adjustments may not be transferable from one solvent model (namely TIP3P) to another (e.g., GB implicit solvent models). Therefore, it is necessary to examine the influence of backbone dihedral energetics

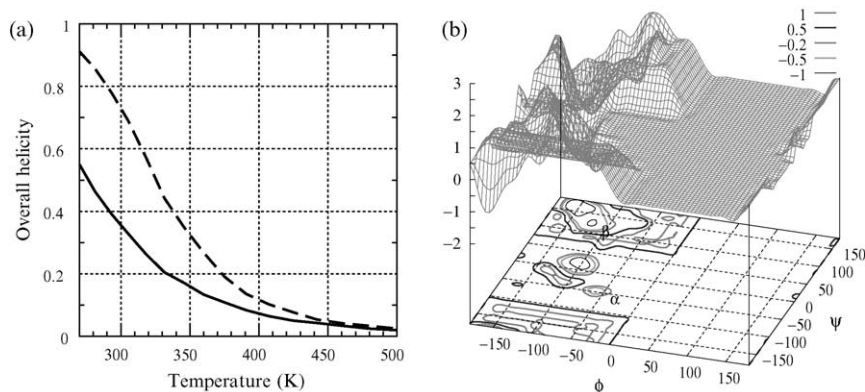


FIG. 5. (a) Simulated helicity of $(AAQAA)_3$ as a function of temperature with the QM-based (solid line) and further modified (dashed line) CMAP potentials. The same setup for REX-MD simulations and H-bond calculations as in Fig. 3 was used. The same optimized input radii as those in Fig. 4 were used. (b) Surface and contour plots of the difference between QM-based and modified CMAP potentials, i.e., $E_{\text{CMAP}}^{\text{mod}} - E_{\text{CMAP}}^{\text{QM}}$. (See Color Insert.)

on conformational equilibria in the context of specific implicit solvent models. As shown in Fig. 5a, the simulated helicity of $(AAQAA)_3$ in GBSW implicit solvent increases substantially with the newest CMAP potential, which is consistent with the difference between two CMAP potentials shown in Fig. 5b. It should be stressed that this does not necessarily mean that the new CMAP potential is problematic. However, it does raise a warning flag that the modifications possibly overstabilize specific secondary structures. As mentioned earlier, folding simulations on peptides and small proteins might provide an ultimate check on the quality of the force field.

IV. CONCLUDING DISCUSSION

In computational biology, the environment specific to biomolecules of interest must be treated properly for meaningful studies of their structural and dynamic features. Solvation exerts competing, opposing forces on intramolecular interactions in peptides and proteins and plays a critical role in determining their conformational states that, in turn, dictates the biological functions of peptides and proteins. As an efficient and physical model to account for solvation, we have introduced GB continuum electrostatic theory and described its underlying principles and recent developments. Methodological developments in GB formalism have reached a

mature stage in which the accuracy of the GB models is almost identical to the PB method. To illustrate how one can gain useful insights into various biological problems using these efficient implicit solvent models, we have briefly reviewed their recent applications to protein structure prediction, molecular recognition problems, protein folding and unfolding, NMR and X-ray structure refinement, pH-coupled MD simulations, and membrane protein modeling. This also provides us an opportunity to discuss existing areas for improvement in current GB implicit solvent force fields.

In principle, one may need to optimize partial charges, Lennard–Jones parameters, and backbone dihedral energetics in current molecular mechanics force fields for a specific implicit solvent model to achieve sufficient balance between solvent–solute and solute–solute interactions. Given the fact that the current force fields have been calibrated extensively to meet such a proper balance, however, we illustrated another possible route to optimize a GB implicit solvent force field; in the present study, the backbone H-bond strength was calibrated by adjusting the backbone input radii for a GB model. Because the dielectric boundary in continuum dielectric solvent models dictates all of the electrostatic and most of nonpolar solvation energetics, it is physically appropriate to optimize input radii not only based on solvation of individual side chains, but also in consideration of solvent-mediated interactions. Our study of conformational equilibria of two peptides, (AAQAA)₃ and α -lac, demonstrated that the helical content can be changed significantly with small modification of backbone input radii based on explicit-solvent H-bond PMFs. For instance, a small change of the amide nitrogen input radius from 2.23 Å (the Nina’s radii) to 1.95 Å, which corresponds to 0.4 kcal/mol H-bond strength difference in an alanine dipeptide model, alters the average helicity of (AAQAA)₃ from 83 to 55%, closer to the 50% value observed in experiments. In the case of α -lac, the same modification results in about a 25% reduction in the helicity of residues 108–111, which is in accord with the experimental results that the peptide is largely unstructured in water. In addition, we illustrated that the backbone dihedral energetics can have a large impact on conformation equilibria.

Because of the paucity of experimental measurements of direct solvent-mediated interactions, comparison with other experimental observables (e.g., those reflecting conformational equilibria) should be used as a guide, in addition to the explicit solvent simulations, to identify and verify optimal input radii in continuum models such as PB and GB. Due to the uncertainties in simulations as well as experiments, multiple sets of atomic radii, which appear to provide comparable results for a single system, should be then examined further in terms of folding and

unfolding simulations of mini and small proteins and by monitoring stability of various medium and large proteins. It is hoped that proper parameterization of GB implicit solvent force fields will enable us to gain more useful insights into various biological problems, which are often very difficult to obtain from explicit solvent models or through experimental study.

ACKNOWLEDGMENTS

We thank Jana Khandogin for helpful discussions. We are grateful to Michael Feig for his efforts to develop the MMTSB Tool Set (<http://mmtsb.scripps.edu>). This work was supported by the National Institutes of Health (RR12255) and by the Center for Theoretical Biological Physics (CTBP) through funding from the National Science Foundation (PHY0216576).

REFERENCES

- Bashford, D., and Case, D. A. (2000). Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**, 129–152.
- Boczko, E. M., and Brooks, C. L., III (1993). Constant-temperature free energy surfaces for physical and chemical process. *J. Phys. Chem.* **97**, 4509–4513.
- Born, M. (1920). Volumen und hydrationswärme der ionen. *Z. Phys.* **1**, 45–48.
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
- Brooks, C. L., III, and Karplus, M. (1986). Theoretical approaches to solvation of biopolymers. *Methods Enzymol.* **127**, 369–400.
- Brooks, C. L., III, Pettitt, B. M., and Karplus, M. (1985). Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids. *J. Chem. Phys.* **83**, 5897–5908.
- Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43.
- Bursulaya, B. D., and Brooks, C. L., III (2000). 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.
- Chen, J., Im, W., and Brooks, C. L., III (2004). NMR structure refinement by replica exchange. *J. Am. Chem. Soc.* **126**, 16038–16047.
- Chen, J., Won, H., Im, W., Dyson, H. J., and Brooks, C. L., III (2005). Generation of native-like models from limited NMR data. *J. Biomol. NMR* **31**, 59–64.
- Clippingdale, A. B., Wade, J. D., and Barrow, C. J. (2001). The amyloid-peptide and its role in Alzheimer's disease. *J. Peptide Sci.* **7**, 227–249.
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
- David, L., Luo, R., and Gilson, M. K. (2000). Comparison of generalized Born and Poisson models: Energetics and dynamics of HIV protease. *J. Comput. Chem.* **21**, 295–309.

- Deechongkit, S., Dawson, P. E., and Kelly, J. W. (2004a). Toward assessing the position-dependent contributions of backbone hydrogen bonding to beta-sheet folding thermodynamics employing amide-to-ester perturbations. *J. Am. Chem. Soc.* **126**, 16762–16771.
- Deechongkit, S., Nguyen, H., Powers, E. T., Dawson, P. E., Gruebele, M., and Kelly, J. W. (2004b). Context-dependent contributions of backbone hydrogen bonding to beta-sheet folding energetics. *Nature* **430**, 101–105.
- Demarest, S. J., Hua, Y. X., and Raleigh, D. P. (1999). Local interactions drive the formation of nonnative structure in the denatured state of human alpha-lactalbumin: A high resolution structural characterization of a peptide model in aqueous solution. *Biochemistry* **38**, 7380–7387.
- Dlugosz, M., and Antosiewicz, J. M. (2004). Constant-ph molecular dynamics simulations: A test case of succinic acid. *Chem. Phys.* **302**, 161–170.
- Dominy, B. N., and Brooks, C. L., III (1999). Development of a generalized Born model parametrization for proteins and nucleic acids. *J. Phys. Chem. B* **103**, 3765–3773.
- Dominy, B. N., and Brooks, C. L., III (2001). Identifying native-like protein structures using physics-based potentials. *J. Comput. Chem.* **23**, 147–160.
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593.
- Feig, M., and Brooks, C. L., III (2002). Evaluating CASP4 predictions with physical energy functions. *Proteins* **49**, 232–245.
- Feig, M., and Brooks, C. L., III (2004). Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **14**, 217–224.
- Feig, M., Im, W., and Brooks, C. L., III (2004a). Implicit solvation based on generalized Born theory in different dielectric environment. *J. Chem. Phys.* **120**, 903–911.
- Feig, M., Karanicola, J., and Brooks, C. L., III (2004b). MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Comp. Graph and Modl.* **22**, 3777–3795.
- Feig, M., MacKerell, A. D., Jr., and Brooks, C. L., III (2003). Force field influence on the observation of π -helical protein structures in molecular dynamics simulations. *J. Phys. Chem. B* **107**, 2831–2836.
- Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A., and Brooks, C. L., III (2004c). Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **25**, 265–284.
- Felts, A. K., Gallicchio, E., Wallqvist, A., and Levy, R. M. (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.
- Ferrara, P., Gohlke, H., Price, D. J., Klebe, G., and Brooks, C. L., III (2004). Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **47**, 3032–3047.
- Fiser, A., Feig, M., Brooks, C. L., III, and Sali, A. (2002). Evolution and physics in comparative protein structure modeling. *Acc. Chem. Res.* **35**, 413–421.
- Forrest, L. R., and Woolf, T. B. (2003). Discrimination of native loop conformations in membrane proteins: Decoy library design and evaluation of effective energy scoring functions. *Proteins* **52**, 492–509.
- Gallicchio, E., and Levy, R. M. (2004). AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* **25**, 479–499.

- Ghosh, A., Rapp, C. S., and Friesner, R. A. (1998). Generalized Born model based on a surface integral formulation. *J. Phys. Chem. B* **102**, 10983–10990.
- Gilson, M. K., Davis, M. E., Luty, B. A., and McCammon, J. A. (1993). Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. *J. Phys. Chem.* **97**, 3591–3600.
- Gohlke, H., and Case, D. A. (2003). Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **25**, 238–250.
- Hansmann, U. H. E. (1997). Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140–150.
- Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996). Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **100**, 19824–19839.
- Hermann, R. B. (1972). Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem.* **76**, 2754–2759.
- Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149.
- Im, W., Beglov, D., and Roux, B. (1998). Continuum solvation models: Electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Comm.* **111**, 59–75.
- Im, W., and Brooks, C. L., III (2004). *De novo* folding of membrane proteins: An exploration of the structure and NMR properties of the fd coat protein. *J. Mol. Biol.* **337**, 513–519.
- Im, W., and Brooks, C. L., III (2005). Interfacial folding and membrane insertion: A molecular dynamics study with synthetic peptides. *Proc. Natl. Acad. Sci. USA* **102**, 6771–6776.
- Im, W., Feig, M., and Brooks, C. L., III (2003a). An implicit membrane generalized Born theory for the study of structures, stability, and interactions of membrane proteins. *Biophys. J.* **85**, 2900–2918.
- Im, W., Lee, M. S., and Brooks, C. L., III (2003b). Generalized Born model with a simple smoothing function. *J. Comput. Chem.* **24**, 1691–1702.
- Im, W., and Roux, B. (2002). Ion permeation and selectivity of OmpF porin: A theoretical study based on molecular dynamics, Brownian dynamics, and continuum electrodiffusion theory. *J. Mol. Biol.* **322**, 851–869.
- Jang, S., Kim, E., Shin, S., and Pak, Y. (2003a). Ab initio folding of helix bundle proteins using molecular dynamics simulations. *J. Am. Chem. Soc.* **125**, 14841–14846.
- Jang, S., Shin, S., and Pak, Y. (2003b). Molecular dynamics study of peptides in implicit water: Ab initio folding of beta-hairpin, beta-sheet, and beta beta alpha-motif. *J. Am. Chem. Soc.* **124**, 4976–4977.
- Jaramillo, A., and Wodak, S. J. (2005). Computational protein design is a challenge for implicit solvation models. *Biophys. J.* **88**, 156–171.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
- Jorgensen, W. L., and Tirado-Rives, J. (1988). The OPLS force field for proteins: Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666.
- Karanicolas, J., and Brooks, C. L., III (2004). Integrating folding kinetics and protein function: Biphasic kinetics and dual binding specificity in a WW domain. *Proc. Natl. Acad. Sci. USA* **101**, 3432–3437.

- Kelly, J. W. (1997). Amyloid fibril formation and protein misassembly: A structural quest for insights into amyloid and prion diseases. *Structure* **5**, 595–600.
- Khandogin, J., and Brooks, C. L., III (2005). A two-dimensional λ -dynamics method to include proton tautomerism in molecular dynamics with continuous titration coordinates. *Biophys. J.* **89**, 141–157.
- Klapper, I., Hagstrom, R., Fine, R., Sharp, K., and Honig, B. (1986). Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins* **1**, 47–59.
- Korostelev, A., Fenley, M. O., and Chapman, M. S. (2004). Impact of a Poisson-Boltzmann electrostatic restraint on protein structures refined at medium resolution. *Acta Cryst.* **D60**, 1786–1794.
- Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules I. The method. *J. Comput. Chem.* **13**, 1011–1021.
- Lazaridis, T., and Karplus, M. (2000). Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* **10**, 139–145.
- Lee, M. S., Feig, M., Salsbury, F. R., Jr., and Brooks, C. L., III (2003). A new analytical approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **24**, 1348–1356.
- Lee, M. S., Salsbury, F. R., Jr., and Brooks, C. L., III (2002). Novel generalized Born methods. *J. Chem. Phys.* **116**, 10606–10614.
- Lee, M. S., Salsbury, F. R., Jr., and Brooks, C. L., III (2004). Constant-pH molecular dynamics using continuous titration coordinates. *Proteins* **56**, 738–752.
- Levy, R. M., Zhang, L. Y., Gallicchio, E., and Felts, A. K. (2003). On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute-solvent interaction energy. *J. Am. Chem. Soc.* **125**, 9523–9530.
- Lin, C.-Y., Hu, C.-K., and Hansmann, U. H. E. (2003). Parallel tempering simulations of hp-36. *Proteins* **52**, 436–445.
- Liu, Y., and Beveridge, D. L. (2002). Exploratory studies of *ab initio* protein structure prediction: Multiple copy simulated annealing, amber energy functions, and a generalized Born/solvent accessibility solvation model. *Proteins* **46**, 128–146.
- Luo, R., David, L., and Gilson, M. K. (2002). Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.* **23**, 1244–1253.
- MacKerell, A. D., Jr. (2004). Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **25**, 1584–1604.
- MacKerell, A. D., Jr., Bashford, D., Bellot, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Joseph-McCarthy, D., Ha, S., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., III, Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., and Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616.
- MacKerell, A. D., Jr., Feig, M., and Brooks, C. L., III (2004a). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415.
- MacKerell, A. D., Jr., Feig, M., and Brooks, C. L., III (2004b). Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **126**, 698–699.
- Masunov, A., and Lazaridis, T. (2003). Potentials of mean force between ionizable amino acid side chains in water. *J. Am. Chem. Soc.* **125**, 1722–1730.

- Mongan, J., and Case, D. A. (2005). Biomolecular simulations at constant pH. *Curr. Opin. Struct. Biol.* **15**, 157–163.
- Mongan, J., Case, D. A., and McCammon, J. A. (2004). Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **25**, 2038–2048.
- Morozov, A. V., Kortemme, T., and Baker, D. (2003). Evaluation of models of electrostatic interactions in proteins. *J. Phys. Chem. B* **107**, 2075–2090.
- Moulinier, L., Case, D. A., and Simonson, T. (2003). Reintroducing electrostatics into protein x-ray structure refinement: Bulk solvent treated as a dielectric continuum. *Acta Cryst.* **D59**, 2094–2103.
- Murray, D., and Honig, B. (2002). Electrostatic control of the membrane targeting of C2 domains. *Mol. Cell.* **9**, 145–154.
- Nicholls, A., and Honig, B. (1991). A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* **12**, 435–445.
- Nina, M., Beglov, D., and Roux, B. (1997). Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B* **101**, 5239–5248.
- Nina, M., Im, W., and Roux, B. (1999). Optimized atomic radii for protein continuum electrostatics solvation forces. *Biophys. Chem.* **78**, 89–96.
- Nymeyer, H., and Garcia, A. E. (2003). Simulation of the folding equilibrium of alpha-helical peptides: A comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. USA* **100**, 13934–13939.
- Ohkubo, Y. Z., and Brooks, C. L., III (2003). Exploring Flory's isolated-pair hypothesis: Statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNAse A. *Proc. Natl. Acad. Sci. USA* **100**, 13916–13921.
- Okur, A., Strockbine, B., Hlornak, V., and Simmerling, C. (2003). Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. *J. Comput. Chem.* **24**, 21–31.
- Onufriev, A., Bashford, D., and Case, D. A. (2000). Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **104**, 3712–3720.
- Onufriev, A., Bashford, D., and Case, D. A. (2002). Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comput. Chem.* **23**, 1297–1304.
- Pitera, J. W., and Swope, W. (2003). Understanding folding and design: Replica-exchange simulations of “Trpcage” fly miniproteins. *Proc. Nat. Acad. Sci. USA* **100**, 7587–7592.
- Prabhu, N. V., Zhu, P., and Sharp, K. A. (2004). Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *J. Comput. Chem.* **25**, 2049–2064.
- Qiu, D., Shenkin, P. S., Hollinger, F. P., and Still, W. C. (1997). The gb/sa continuum model for solvation: A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **101**, 3005–3014.
- Rastogi, V. K., and Girvin, M. E. (1999). Structural changes linked to proton translocation by subunit *c* of the ATP synthase. *Nature* **402**, 263–268.
- Roux, B. (1995). The calculation of the potential of mean force using computer simulations. *Comput. Phys. Comm.* **91**, 275–282.
- Roux, B. (2002). Theoretical and computational models of ion channels. *Curr. Opin. Struct. Biol.* **12**, 182–189.
- Roux, B., Bernèche, S., and Im, W. (2000). Ion channels, permeation, and electrostatics: Insight into the function of KcsA. *Biochemistry* **39**, 13295–13306.

- Roux, B., and Simonson, T. (1999). Implicit solvent models. *Biophys. Chem.* **78**, 1–20.
- Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equation of motions of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Chem.* **23**, 327–341.
- Scarsi, M., Apostolakis, J., and Caffisch, A. (1997). Continuum electrostatic energies of macromolecules in aqueous solutions. *J. Phys. Chem. A* **101**, 8098–8106.
- Schaefer, M., and Karplus, M. (1996). A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* **100**, 1578–1599.
- Shalongo, W., Dugad, L., and Stellwagen, E. (1994). Distribution of helicity within the model peptide acetyl(AAQAA)₃ amide. *J. Am. Chem. Soc.* **116**, 8288–8293.
- Sharp, K. A., and Honig, B. (1990). Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Chem.* **19**, 301–332.
- Sigalov, G., Scheffel, P., and Onufriev, A. (2005). Incorporating variable dielectric environments into the generalized Born model. *J. Chem. Phys.* **122**, 94511–94525.
- Simonson, T., and Brunger, A. (1994). Solvation free energies estimated from macroscopic continuum theory: An accuracy assessment. *J. Phys. Chem.* **98**, 4683–4694.
- Sneddon, S. F., Tobias, D. J., and Brooks, C. L., III (1989). Thermodynamics of amide hydrogen bond formation in polar and apolar solvents. *J. Mol. Biol.* **209**, 817–820.
- Spassov, V. Z., Yan, L., and Szalma, S. (2002). Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models. *J. Phys. Chem. B* **106**, 8726–8738.
- Srinivasan, J., Trevathan, M. W., Beroza, P., and Case, D. A. (1999). Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **101**, 426–434.
- Steinbach, P. J., and Brooks, B. R. (1994). New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **15**, 667–683.
- Steinbach, P. J. (2004). Exploring peptide energy landscapes: A test of force fields and implicit solvent models. *Proteins* **57**, 665–677.
- Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990). Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127–6129.
- Sugita, Y., and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151.
- Tanizaki, S., and Feig, M. (2005). A generalized Born formalism for heterogeneous dielectric environments: Application to the implicit modeling of biological membranes. *J. Chem. Phys.* **122**, 124706.
- Tobias, D. J., Sneddon, S. F., and Brooks, C. L., III (1992). Stability of a model beta-sheet in water. *J. Mol. Biol.* **227**, 1244–1252.
- Torrie, G. M., and Valleau, J. P. (1977). Non-physical sampling distributions in monte-carlo free-energy estimation - umbrella sampling. *J. Comp. Phys.* **23**, 187–199.
- van Gunsteren, W. F., and Berendsen, H. J. C. (1990). Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.* **29**, 992–1023.
- Wang, T., and Wade, R. C. (2003). Implicit solvent models for flexible protein-protein docking by molecular dynamics simulations. *Proteins* **50**, 158–169.
- Warwicker, J., and Watson, H. C. (1982). Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *J. Mol. Biol.* **157**, 671–679.
- Wodak, S. J., and Mendez, R. (2004). Prediction of protein-protein interactions: The capri experiment, its evaluation and implications. *Curr. Opin. Struct. Biol.* **14**, 242–249.

- Xia, B., Tsui, V., Case, D. A., Dysonand, H. J., and Wright, P. E. (2002). Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J. Biomol. NMR* **22**, 317–331.
- Yang, W. Y., Pitera, J. W., Swope, W. C., and Gruebele, M. (2004). Heterogeneous folding of the trpzip hairpin: Full atom simulation and experiment. *J. Mol. Biol.* **336**, 241–251.
- Zagrovic, B., Sorin, E. J., and Pande, V. (2001). β -hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **313**, 151–169.
- Zhou, R. (2003). Free energy landscape of protein folding in water: Explicit vs. implicit solvent. *Proteins* **53**, 148–161.
- Zhou, R., and Berne, B. J. (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.
- Zhu, J., Alexov, E., and Honig, B. (2005). Comparative study of generalized Born models: Born radii and peptide folding. *J. Phys. Chem. B* **109**, 3008–3022.
- Zhu, J., Zhu, Q., Shi, Y., and Liu, H. (2003). How well can we predict native contacts in proteins based on decoy structures and their energies? *Proteins* **52**, 598–608.