

Exploring Protein Native States and Large-Scale Conformational Changes With a Modified Generalized Born Model

Alexey Onufriev,¹ Donald Bashford,² and David A. Case^{2*}

¹Department of Computer Science, Virginia Tech, Blacksburg, Virginia

²Department of Molecular Biology, The Scripps Research Institute, La Jolla, California

ABSTRACT Implicit solvation models provide, for many applications, a reasonably accurate and computationally effective way to describe the electrostatics of aqueous solvation. Here, a popular analytical Generalized Born (GB) solvation model is modified to improve its accuracy in calculating the solvent polarization part of free energy changes in large-scale conformational transitions, such as protein folding. In contrast to an earlier GB model (implemented in the AMBER-6 program), the improved version does not overstabilize the native structures relative to the finite-difference Poisson–Boltzmann continuum treatment. In addition to improving the energy balance between folded and unfolded conformers, the algorithm (available in the AMBER-7 and NAB molecular modeling packages) is shown to perform well in more than 50 ns of native-state molecular dynamics (MD) simulations of thioredoxin, protein-A, and ubiquitin, as well as in a simulation of Barnase/Barstar complex formation. For thioredoxin, various combinations of input parameters have been explored, such as the underlying gas-phase force fields and the atomic radii. The best performance is achieved with a previously proposed modification to the torsional potential in the Amber ff99 force field, which yields stable native trajectories for all of the tested proteins, with backbone root-mean-square deviations from the native structures being ~ 1.5 Å after 6 ns of simulation time. The structure of Barnase/Barstar complex is regenerated, starting from an unbound state, to within 1.9 Å relative to the crystal structure of the complex. *Proteins* 2004;55:383–394. © 2004 Wiley-Liss, Inc.

Key words: Generalized Born approximation; molecular dynamics; macromolecules

INTRODUCTION

An accurate description of the aqueous environment is essential for realistic biomolecular simulations, but may become very expensive computationally. For example, an adequate representation of the solvation of a medium-size protein typically requires thousands of discrete water molecules to be placed around it. Alternatively, one can use an electrostatic model, which replaces discrete water molecules by an infinite continuum medium with the dielectric properties of water. Although this model repre-

sents an approximation at a fundamental level, it has in many cases been successful in calculating various macromolecular properties.^{1–7} The continuum models have several advantages over the explicit water representation, especially in molecular dynamics (MD) simulations:

1. The computational cost associated with the use of these models in simulations is generally considerably smaller than the cost of representing water explicitly.
2. The models describe an instantaneous solvent dielectric response, which eliminates the need for the lengthy equilibration of water that is typically necessary in explicit water simulations.
3. Due to the absence of viscosity associated with the explicit water environment, the molecule can more quickly explore the available conformational space.
4. The continuum model corresponds to solvation in an infinite volume of solvent, thereby avoiding possible artifacts of the replica interactions that occur in the periodic systems typically used for explicit water calculations.
5. Since solvent degrees of freedom are taken into account implicitly, estimating energies of solvated structures is much more straightforward than with explicit water models.

The total (free) energy of a solvated molecule can be written as $E_{tot} = E_{vac} + \Delta G_{solv}$, where E_{vac} represents the molecule's energy in vacuum (gas-phase), and ΔG_{solv} is the free energy of transferring the molecule from vacuum into solvent (i.e., solvation free energy). In what follows, we assume that E_{vac} is represented by a classical potential function that breaks the interaction down into various physical components, such as bond and angle stretching, torsional twist, and Van der Waals (VDW) and Coulomb interactions between its atoms.⁸ To estimate the total solvation free energy of a molecule, ΔG_{solv} , one typically

Grant sponsor: National Institutes of Health; Grant number: M57513.

*Correspondence to: David A. Case, Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037. E-mail: case@scripps.edu

Received 14 May 2003; Accepted 5 October 2003

Published online 5 March 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20033

assumes that it can be decomposed into the “electrostatic” and “nonelectrostatic” parts:

$$\Delta G_{\text{solv}} = \Delta G_{\text{el}} + \Delta G_{\text{surf}}, \quad (1)$$

where ΔG_{surf} is the free energy of solvating a molecule from which all charges have been removed (i.e., partial charges of every atom are set to zero), and ΔG_{el} is the free energy of first removing all charges in the vacuum, and then adding them back in the presence of the solvent environment. The above decomposition is the basis of the widely used PB/SA scheme.⁹ In this approach, ΔG_{el} is computed by numerical Poisson–Boltzmann (PB) methodology, and ΔG_{surf} is taken to be proportional to the total solvent-accessible surface area (SA) of the molecule, with a proportionality constant derived from experimental solvation energies of small nonpolar molecules. In this work, we follow the same approach to compute ΔG_{surf} and concentrate on developing an algorithm for obtaining a reasonable, computationally efficient estimates of ΔG_{el} , to be used in MD simulations.

The analytic Generalized Born (GB) approximation is a way to calculate the electrostatic part of the solvation free energy, ΔG_{el} . The methodology has become popular,^{1,10–17} especially in MD applications,^{18–23} due to its relative simplicity and computational efficiency, compared to the more standard numerical solution of the PB equation. Within our GB model, we represent each atom in a molecule as a sphere of radius ρ_i with a charge q_i at its center; the interior of the atom is assumed to be filled uniformly with material of dielectric constant 1. The molecule is surrounded by a solvent of a high dielectric value ϵ_w (80 for water at 300 K). The GB model approximates ΔG_{el} by an analytical formula,^{10,24}

$$\Delta G_{\text{el}} \approx \Delta G_{\text{GB}} = -\frac{1}{2} \sum_{ij} \frac{q_i q_j}{f^{\text{GB}}(r_{ij}, R_i, R_j)} \left(1 - \frac{e^{-\kappa r_{ij}^{\text{GB}}}}{\epsilon_w} \right), \quad (2)$$

where r_{ij} is the distance between atoms i and j , the R_i are the so-called *effective Born radii* of atoms i and j , and f^{GB} is a certain smooth function of its arguments. A common choice¹⁰ of f^{GB} is

$$f^{\text{GB}} = [r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j)]^{1/2}, \quad (3)$$

although other other expressions have been tried.^{13,25} The effective Born radius of an atom reflects the degree of its burial inside the molecule: For an isolated ion R_i is equal to its VDW radius ρ_i , while for a deeply buried one $R_i \gg \rho_i$. The effective radii depend on the molecule’s conformation, and have to be recomputed every time the conformation changes: The procedure is described in detail below. Finally, the electrostatic screening effects of (monovalent) salt are incorporated²⁴ into formula Eq. (2) via the Debye–Huckel screening parameter $\kappa [\text{\AA}^{-1}] \approx 0.316 \sqrt{[\text{salt}]/[\text{mol/L}]}$.

This article has the following structure: We begin by discussing the main problems with some previous versions of the GB approximation, and outline the improvement strategy. Next, we introduce the modifications that are the core of the new model and describe approaches to testing

its performance. We test the model against the PB treatment on a variety of structures, and then evaluate its performance in native-state MD simulations of a number of proteins, as well in a simulation of Barnase/Barstar complex formation from an unbound state. Various underlying gas-phase force fields from the AMBER package are explored, with one version in particular giving consistently good results.

IMPROVING THE GENERALIZED BORN APPROXIMATION

The GB approximation was originally¹⁰ developed and optimized for small molecules, where it performed quite well.^{10,13,24,26,27} However, its performance on larger molecules has generally not been as good, with the trend being more pronounced for molecules having large interior regions.^{24,26} Since the GB model shares the same underlying physical approximation—continuum electrostatics—with the PB approach, it is natural, in optimizing the GB performance, to use the PB model as a reference. Agreement with PB calculations is not the only criterion of optimal GB performance; other tests include comparisons to explicit solvent simulations results, and to experiment. In particular, if the GB approach is to be used in MD simulations, we require that analytical expressions used to compute the effective Born radii that enter Eq. (2) be simple enough, and the resulting electrostatic energy component ΔG_{el} be “well-behaved,” so as not to cause any instabilities in numerical integration of Newton’s equations of motion. Since the continuum model is by itself an approximation, extreme efforts to achieve a perfect agreement between the GB and PB theories, especially at the expense of computational efficiency, may not pay off. Ultimately, one should seek best agreement with experiment, and as an intermediate goal, make sure that the model yields MD trajectories of protein native states that stay close to the corresponding X-ray or NMR structures. In light of these considerations, we have balanced our efforts equally between improving the agreement between the GB and PB approaches on a variety of test systems and satisfying the other criteria mentioned above.

A New Model for Effective Born Radii

In our previous work,²⁵ we demonstrated that a very good agreement between the GB and PB models can be achieved if the effective Born radii match those computed exactly using the PB approach—these are the so-called “perfect” radii. We therefore concentrate first on improving the way the radii are computed within the analytic GB approximation,^{15,16} which typically involves at least two levels of approximation. First, it is assumed that the energy density of the electrostatic field of the molecule in solvent can be approximated as the energy density of a Coulomb field; that is, the contribution of the reaction field is neglected. Second, approximations are introduced for the integration of this energy density over the molecular region, typically in order to obtain an expression involving only a sum over atom pairs.^{27,28} The result is summarized as

$$R_i^{-1} = \tilde{\rho}_i^{-1} - I, \quad (4)$$

where

$$I = \frac{1}{4\pi} \int_{\text{VDW}} \theta(|\vec{r}| - \tilde{\rho}_i) \frac{1}{r^4} d^3\vec{r}. \quad (5)$$

In a popular GB model,²⁷ which we refer to as GB^{HCT} (Hawkins, Cramer, Truhlar), the three-dimensional (3D) integral in the above equation is performed over the VDW spheres of solute atoms outside of the atom i whose effective radius is being calculated (the step-function θ ensures that the volume of atom i itself is excluded from the integration). This implies a definition of the solute volume in terms of a set of spheres, rather than the complex molecular surface²⁹ commonly used in the PB calculations. This allows one to obtain a relatively simple analytical form for R_i . A closed-form analytical expression for two overlapping spheres, from which I can be derived in the pairwise approximation, is given in Hawkins et al.²⁷ and Schaefer and Froemmel.²⁸ It was also shown¹⁰ that using slightly reduced values of $\tilde{\rho}_i = \rho_i - 0.09 \text{ \AA}$ instead of ρ_i in Eqs. (4) and (5) gives, for small molecules, a better agreement with the corresponding PB calculations based on ρ_i .

For macromolecules, the approach based on Eqs. (4) and (5) tends to underestimate the effective radii for buried atoms,¹⁶ arguably because the standard integration procedure in Eq. (4) treats the small vacuum-filled crevices between the VDW spheres of protein atoms as being filled with water, even for structures with a large interior.²⁵ This error is expected to be greatest for deeply buried atoms characterized by large effective radii, while for the surface atoms, it is largely canceled by the opposing error arising from the Coulomb approximation, which tends^{11,15,28} to overestimate R_i . Since the GB^{HCT} model was parameterized and shown^{13,27} to perform very well on small molecules with few buried atoms (small effective radii), we focus on improving the procedure to calculate the effective radii for buried atoms in larger compounds, while preserving the good performance of GB^{HCT} for small R_i 's. In particular, we retain the use of $\tilde{\rho}_i = \rho_i - 0.09 \text{ \AA}$ instead of ρ_i .

In our previous work,¹⁶ we introduced a simple correction factor into Eq. (4) that increased the effective radii R_i of buried atoms and brought the GB model in closer agreement with the PB in calculating pK_a values of protein side-chains, which was the primary goal of that study. This simple correction did not, however, perform as well in MD simulations, where it led to numerical instabilities. The main difficulty was that small changes in conformation could sometimes lead to large alterations of effective radii, requiring very careful MD simulations with extremely short timesteps. The problem also exists, but is less pronounced, with the GB^{HCT} model, where the R_i 's were routinely underestimated, and as a consequence, were less sensitive to structural variations. Furthermore, the one-parameter correction we¹⁶ introduced was not optimal in keeping the GB^{HCT} model's established performance on

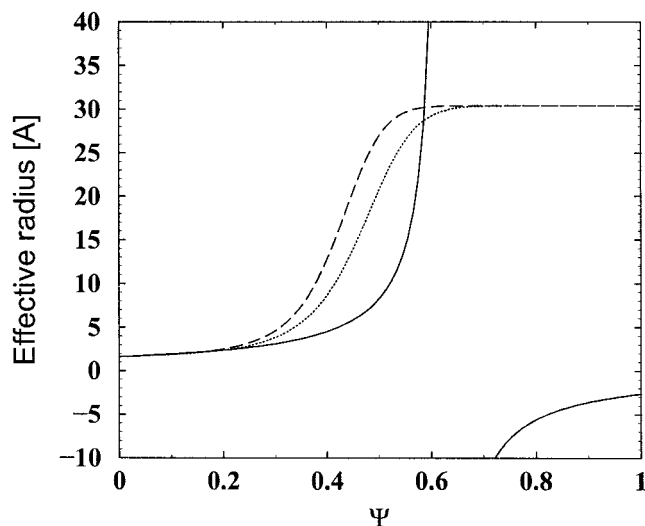


Fig. 1. Graphical representation of different mathematical expressions used to compute effective radius from the scaled volume integral Ψ . The broken lines correspond to Eq. (6) with parameters from Eq. (7) (dashed lines) and Eq. (8) (dotted lines) used, respectively, by GB^{OBC} (I) and (II) models. Original formula [Eq. (4)] used to compute effective radii in GB^{HCT} model is shown as a solid line. All curves correspond to $\rho_i = 1.7 \text{ \AA}$ (carbon atom).

small molecules. Here, we propose to rescale the R_i [Eq. (4)], with the rescaling parameters being proportional to the degree of the atom's burial, as quantified by the value of I in Eq. (5). The latter is large for the deeply buried atoms and small for exposed ones. Consequently, we seek a well-behaved rescaling function, such that $R_i \approx (\tilde{\rho}_i^{-1} - I)^{-1}$ for small I , and $R_i > (\tilde{\rho}_i^{-1} - I)^{-1}$ when I becomes large. We also want to have a "smooth" upper bound on R_i versus I to ensure numerical stability (see below). While there is certainly more than one way to ratify these constraints, we have chosen the following simple, infinitely differentiable rescaling function to replace the original expression for the effective radii [Eq. (4)]:

$$R_i^{-1} = \tilde{\rho}_i^{-1} - \rho_i^{-1} \tanh(\alpha\Psi - \beta\Psi^2 + \gamma\Psi^3), \quad (6)$$

where $\Psi = I\tilde{\rho}_i$, and α, β, γ are treated as adjustable dimensionless parameters to be optimized. In this article, we consider two sets of $\{\alpha, \beta, \gamma\}$ that are optimal according to criteria described further in the text:

$$GB^{\text{OBCI}}: \alpha = 0.8, \beta = 0, \gamma = 2.91 \quad (7)$$

$$GB^{\text{OBCII}}: \alpha = 1.0, \beta = 0.8, \gamma = 4.85 \quad (8)$$

We call this model—based on Eq. (6)— GB^{OBC} (Onufriev, Bashford, Case). Compared to GB^{HCT} , which is based upon the original expression in Eq. (4), the use of Eq. (6) results in a considerable increase of the effective radii for $\Psi > 0.3$, but has little effect in the region of small $\Psi < 0.2$, where GB^{HCT} already works well. This point is illustrated in the $\rho_i = 1.7 \text{ \AA}$ case in Figure 1, which compares Eq. (6) to Eq. (4). Figure 1 also illustrates the source of sensitivity of large effective radii to subtle conformational changes in the GB^{HCT} model based on Eq. (4): For Ψ approaching

~ 0.6 (i.e., for $I \rightarrow \tilde{\rho}_i^{-1}$), the formula diverges and can even change sign, which was observed in some large proteins.³⁰ Although the latter is a consequence of the approximations used to evaluate Ψ , and could not happen if the integral in Eq. (5) were computed exactly, the sensitivity of R_i to conformational changes would still remain, since Eq. (4) does diverge for $I \rightarrow \tilde{\rho}_i^{-1}$. The use of the hyperbolic tangent as a capping function in Eq. (6) limits the variation of the R_i and prevents occasional numerical instability in the MD integrator, which otherwise may result from large values of $\delta R_i / \delta r_j$. Since $|\tanh(y)| < 1$, the upper limit of an R_i computed from Eq. (6) is $(\tilde{\rho}_i^{-1} - \rho_i^{-1})^{-1}$, which is about 30 Å for a carbon atom ($\rho = 1.7$ Å), as in Figure 1. In our experience, perfect effective radii almost never exceed 15 Å even for the most deeply buried residues of typical proteins.²⁵ The above restriction on the upper values of R_i should therefore have very little effect on the accuracy of computed solvation energies, especially since the contribution to ΔG_{el} from deeply buried atoms is small, $\sim 1/R_i$.

The distribution of effective radii produced by the GB^{OBC} model (with either set I or II) is considerably closer to the distribution of the perfect (PB-based) radii than that produced by GB^{HCT} , as illustrated in Figure 2 (top). In particular, the systematic underestimation of the effective radii produced by GB^{HCT} is largely corrected in GB^{OBC} , especially for R_i 's between 2 and 5 Å [Fig. 2 (bottom)], although the model is still not an exact match of PB.

In many applications of the GB approximation, such as the computation of forces in MD simulations, the absolute value of ΔG_{el} of a molecule is not so important as its conformational dependence. Compared to PB calculations, the GB model was known¹⁶ to reproduce variations in ΔG_{el} associated with small conformational changes quite well, but the deviation became significant when large changes were involved, such as in protein folding.²⁵ To address this issue in optimizing parameters $\{\alpha, \beta, \gamma\}$ of GB^{OBC} , we use snapshots from MD simulations that describe global conformational changes, such as the complete unfolding of a protein. The fit to the PB model has been performed on an MD trajectory that describes acid-unfolding of apomyoglobin, and includes both native and completely unfolded states at 300 K.³¹ The latter has twice the size of the native state and about 20% residual helical structure, in agreement with the experiment.³² At this point, we seek best agreement between the GB and PB models in computing the change in ΔG_{el} in going from the native (N) to unfolded (U) ensembles of structures (see Table I).

The transferability of the optimized parameter sets I and II is verified using the unfolding trajectory of protein-A at neutral pH. Note a good agreement with the PB and GB^{OBC} for very different charge states: the apomyoglobin at low pH bears a total charge of +36, while the charge of protein-A is -3 at neutral conditions.

In contrast to these results, the GB^{HCT} model based on Eq. (5) shows a considerable energetic bias toward the native structures. This bias reflects the underestimation of effective radii for buried atoms resulting in too negative solvation energies of protein compact states, and is elimi-

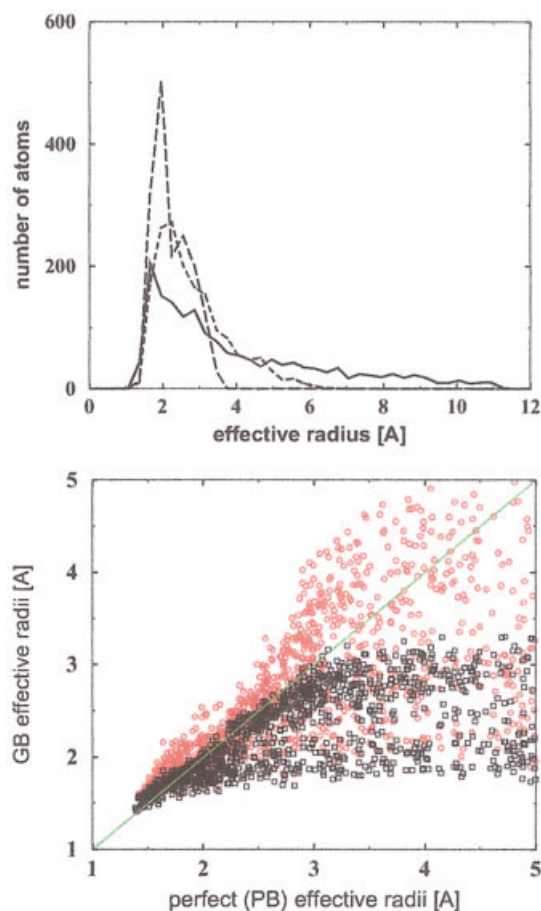


Fig. 2. **Top:** Distribution of effective radii in thioredoxin computed with the PB (solid line), GB^{HCT} (dashed line), and GB^{OBC} (I) (dotted line) models. **Bottom:** Effective radii in thioredoxin computed for each atom, with GB^{HCT} (black squares) and GB^{OBC} (red circles) models versus those calculated by the PB. An exact match between GB and PB models is indicated by the diagonal line.

nated almost completely by the GB^{OBC} model, as shown in Table I.

By construction, the GB^{OBC} model is intended to correct the deficiencies of the GB^{HCT} on compounds with pronounced interior regions, while preserving its performance on extended structures with little or no interior. To further test the model along these lines, we compare the performances of the GB^{HCT} and GB^{OBC} models as a function of a protein's degree of compactness, using structures of protein L obtained in an MD simulation, from the native to completely unfolded state at 450 K.¹⁷ As Figure 3 shows, for this protein, the overall effect of the rescaling procedure introduced in the GB^{OBC} model is to largely reduce the undue bias toward folded structures exhibited by GB^{HCT} . The improvement appears to be distributed fairly uniformly across the range of protein compactness.

The importance of conformational averaging in comparing ΔG_{el} between different theoretical models is emphasized by the thin dotted line in Figure 3, which is the difference between ΔG_{el} calculated by the PB model based on a coarse, 1 Å, and a finer, 0.25 Å, grid spacing: The

TABLE I. The Change in the Electrostatic Part of Solvation Free Energy, $\Delta G_{\text{el}}(N) - \Delta G_{\text{el}}(U)$ [kcal/mol], of Apo-Myoglobin and Protein-A in Going from the Unfolded (U) to the Native (N) States Calculated with PB and GB Models Based on the Bondi Radii Set

	Electrostatic model				
	PB (1 Å)	PB (0.5 Å)	GB^{OBC} set I	GB^{OBC} set II	GB^{HCT}
(Apo)myoglobin, pH = 2	-2082.0	-2088.8	-2089.9	-2093.8	-2161.1
Protein-A, pH = 7	+143.9	+143.0	+145.1	+142.9	+131.1

Since a perfect GB model would correspond to zero grid-size PB, and since we observe a decrease in $\Delta G_{\text{el}}(N) - \Delta G_{\text{el}}(U)$ upon reducing the grid size from 1 Å to 0.5 Å, we adjust the parameters of GB^{OBC} to produce values of $\Delta G_{\text{el}}(N) - \Delta G_{\text{el}}(U)$ slightly below the PB predictions based on 0.5 Å grid spacing.

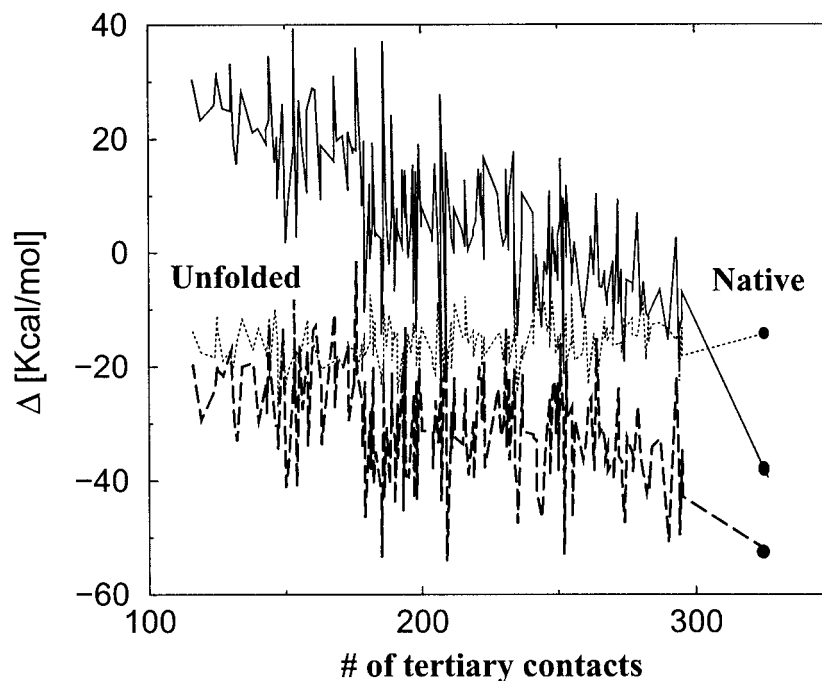


Fig. 3. Performance of the GB^{HCT} (solid line) and GB^{OBC} (dashed line) models relative to PB model as a function of a protein's degree of compactness represented by the number of residue-residue tertiary contacts in the protein. We use structures of protein L, obtained in an MD simulation,¹⁷ from the native to completely unfolded at 450 K. $\Delta G_{\text{el}}(GB)$ is calculated, within the GB model, for each of the 215 snapshots, and compared to the corresponding $\Delta G_{\text{el}}(PB)$ estimated using the PB model with grid spacing of 0.25 Å, $\Delta = \Delta G_{\text{el}}(GB) - \Delta G_{\text{el}}(PB)$. The dependence of $\Delta G_{\text{el}}(PB)$ on the grid spacing is illustrated by the thin dotted line, representing the difference between ΔG_{el} computed using coarse (1 Å) and fine (0.25 Å) grid PB models, $\Delta = \Delta G_{\text{el}}(PB, 1 \text{ Å}) - \Delta G_{\text{el}}(PB, 0.25 \text{ Å})$. Thick black dots correspond to the native structure.

variation of ΔG_{el} from snapshot to snapshot can be as high as 20 kcal/mol, but no systematic bias is present.

Macromolecular complex formation is another area where large conformational changes are expected, and the GB^{OBC} model may be advantageous, at least in providing a reasonable approximation to the PB. As a recent study³³ suggests, the GB^{OBC} model shows considerable improvement (compared to the GB^{HCT}) in estimating the electrostatic component of free energies of unbound proteins Ras and Raf, and their complex, relative to the PB treatment.

Finally, we note that for molecules with little interior region, ΔG_{el} calculated by the GB^{OBC} model are not expected to differ much from those computed by the original GB^{HCT} model, which should give a reasonable estimate in this case. For example, when applied to an ensemble of 120 structures representing³⁰ various confor-

mational states of the small, rather loosely packed protein villin head-piece, the GB^{OBC} shows only marginal improvement over GB^{HCT} (Fig. 4). Specifically, ΔG_{el} computed by GB^{OBC} differs from that computed by PB (0.5 Å grid) by 5.41 kcal/mol [root-mean-square deviation (RMSD)], whereas for GB^{HCT} , this number is 7.16 kcal/mol. These numbers represent, respectively, 0.8% and 1.1% of the average electrostatic solvation energy of the protein ($\Delta G_{\text{el}} = -653$ kcal/mol), and are similar³⁰ to those computed by the new GB(MV) model¹⁷ implemented in CHARMM. We feel that a word of caution would be appropriate at this point: The quality of an approximation should be judged in the context of the particular application of the method and characteristic energy scales involved. While a 1% or smaller error in the total solvation energy relative to a standard (PB in our case) may be negligible in many cases, that

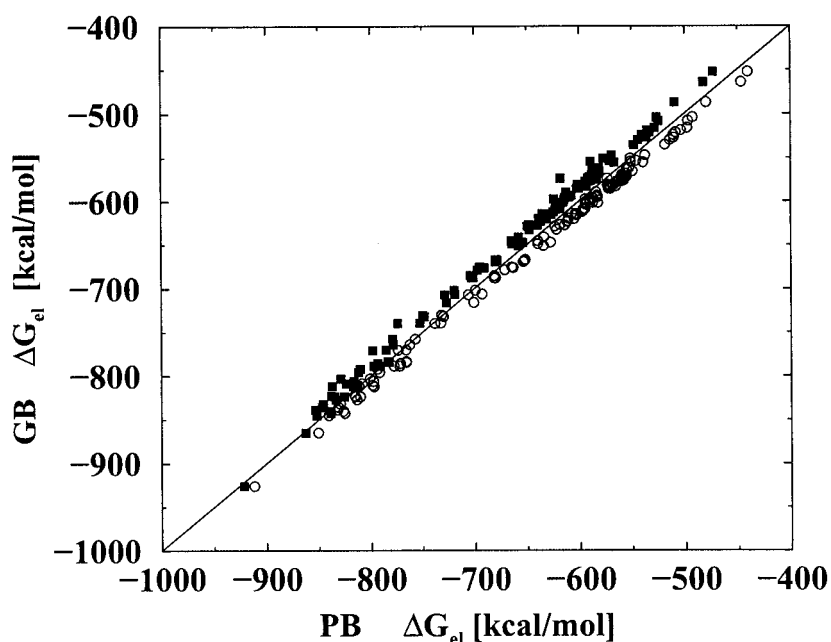


Fig. 4. Performance of the GB^{HCT} (solid squares) and GB^{OBC} (II) (open circles) models relative to the PB model in calculating the electrostatic part of solvation free energy for 120 conformers of villin head-piece. An exact match between the GB and PB models is given by the diagonal line. Both GB models approximate the PB results quite well for this small protein lacking significant interior region.

same error may become appreciable in others. For example, the 2–5 kcal/mol difference between GB and PB results in Table I translates into 20–50% of the protein’s experimental folding free energy of about 10 kcal/mol—clearly a non-negligible error if one is considering theoretical estimates of the folding free energy.

MD SIMULATIONS WITH THE IMPROVED GB MODEL

Thioredoxin

One of the most attractive features of a GB model is the possibility to use it in MD simulations where it can effectively mimic solvation effects and therefore eliminate the need for a more costly explicit water representation. The GB^{OBC} model is particularly attractive in this respect due to its algorithmic simplicity and reasonably good performance relative to PB treatments. The latter cannot, however, serve as a sole measure of the quality of a GB model, so as the next step we test it directly in MD simulations.

We begin by performing 6 MD simulations, ~ 7 -ns each, of thioredoxin, a 108-residue protein that had been used previously to test GB models.^{25,34} The simulations, summarized in Table II, use the GB^{OBC} model based on either of the parameter sets [Eqs. (7) and (8)], with either the Bondi or a modified Bondi radii set. We have also explored three different underlying gas-phase force fields from AMBER.

While both of the GB^{OBC} parameterizations show similar performance with respect to deviations from the crystal structure, the *parm99* force field performed appreciably better than *parm94* at the 6-ns timescale (Table II). In the *parm94* case, a large and abrupt jump in RMSD from ~ 1.7

TABLE II. Summary of the Six MD Trajectories of Thioredoxin

MD trajectory	I	II	III	IV	V	VI
Final RMSD (Å)	3.00	2.94	2.14	2.16	2.02	1.48

The simulation protocol is the same in each case (see Methods), except for three parameters that define the potential function: gas-phase force field, GB model (I or II), and radii set used to estimate the solvation energy (which includes both the electrostatic and the surface energy parts). These parameters listed in the above order for each trajectory are: I, *parm94*, GB^{OBC} (I), Bondi; II, same as I, but different random seed; III, *parm99*, GB^{OBC} (I), Bondi; IV, *parm99*, GB^{OBC} (II), Bondi; V, *parm99*, GB^{OBC} (I), mBondi2; VI, *parm99MOD2*, GB^{OBC} (I), mBondi2. The mBondi2 set is prepared from the Bondi set by increasing radius of every hydrogen atom bound to a nitrogen from 1.2 to 1.3 Å. The final RMSD is reported as an average over 1-ns time interval between $t = 5$ and 6 ns.

to ~ 2.7 Å was observed, with the maximum RMSD reaching 3.5 Å over the course of the simulation. (Trajectories I and II were generated using the same MD protocol but a different random number generator seed to prepare distribution of initial velocities; the rapid increase in the RMSD occurred at $t \approx 2$ ns in trajectory I and at $t \approx 3.6$ ns in trajectory II.) The structural changes responsible for this jump are localized in residues 82–92. These residues were initially in a β strand conformation but underwent a fast transition into a helix. Removing residues 82–92 from the calculation of the RMSD removes the corresponding “jumps” and lowers the final RMSD for both trajectories by about 1 Å. This kind of behavior does not appear to be specific to the GB model we use—similar observations were made in earlier MD simulations of thioredoxin using

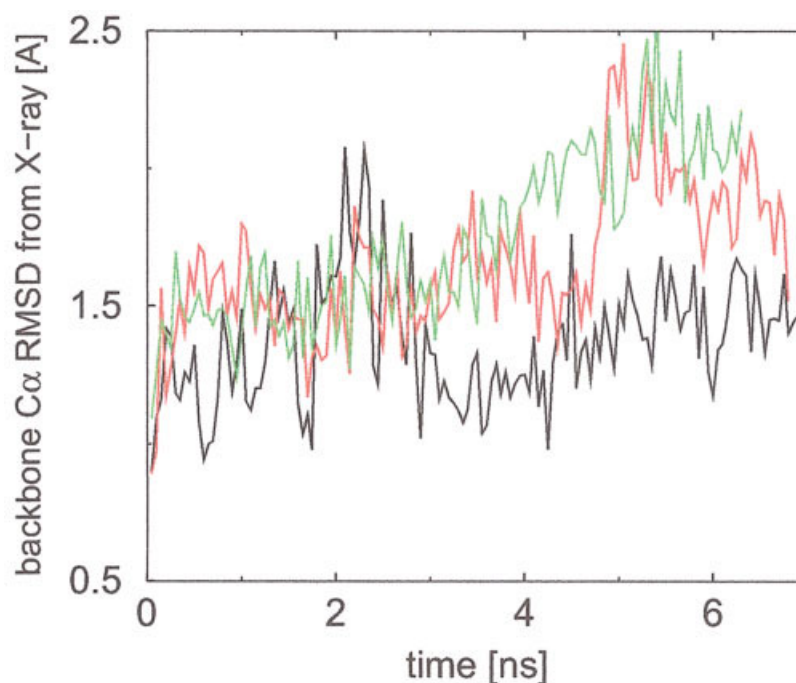


Fig. 5. Backbone RMSD (residues 4–107) from X-ray structure during a MD simulation of thioredoxin using GB^{OBC} (l) model based on different radii sets and underlying force fields. Green trace: the original Bondi set and AMBER *parm99*. Red trace: a modified Bondi radii set (radius of hydrogens bound to a nitrogen is increased to 1.3 Å versus 1.2 Å in the original Bondi set) and AMBER *parm99*. Black trace: modified Bondi set and *parm99MOD2* underlying force field.

parm94 and the GB^{HCT} model.²² Poor performance of *parm94* on structures containing β -strands was reported before.^{35,36} The use of *parm99* with either of the GB^{OBC} parameterizations considered here yields smaller structural deviations, backbone RMSD being less than 2.2 Å by the end of the simulation, with no abrupt changes, and never exceeding 2.5 Å (Fig. 5). In earlier simulations³⁴ of thioredoxin using the GB^{HCT} model, the RMSD was also about 2.4 Å after 5 ns. Note that the GB^{OBC} model appears to have much less energetic bias toward the compact (native) states, at least as judged from its comparison with the PB treatment discussed above.

On the other hand, an earlier explicit water simulation of thioredoxin based on *parm94* resulted³⁴ in RMSD from X-ray of only about 1.1 Å on a 6-ns timescale. This may indicate problems with the GB model. Or it may be that the structural changes seen in the GB-based MD simulations may not have occurred in the corresponding explicit water simulation on the same timescale of a few nanoseconds—that much longer simulation times are required in explicit water simulations than in GB-based ones to observe similar change. Indeed, conformational transitions in explicit water are expected to happen considerably slower than if the essentially zero viscosity GB model is used. For example, it was shown earlier²² that the $A \rightarrow B$ conformational transition in DNA happens about 20 times faster in MD based on GB compared to an equivalent explicit water simulation. These observations suggest that one can use a reasonable GB model, even with imperfections, in testing of underlying molecular mechanics force

fields: For example, here we have been able to identify a possible problem with *parm94* parameterization of β -structures in proteins; the use of *parm99* instead apparently has partially alleviated the problem.

Although the use of a newer gas-phase force field (*parm99*) did result in a more stable MD trajectory of thioredoxin, the structural drift is still not negligible, with C_{α} RMSD reaching 2.2 Å in 6.5 ns (Fig. 5, green trace). There are at least two distinct levels at which one may seek further improvements to the GB model's performance in MD simulations: at the level of the underlying electrostatic model and/or at the level of the gas-phase force field. In GB^{HCT} -based MD simulations of nucleic acids, it was shown³⁴ that using a larger value for the hydrogen radius improves the model's performance, and here we explore a possibility along those lines. Namely, we have performed an MD simulation based on a slightly modified radii set used to calculate the solvation energy: it is prepared from the Bondi set by increasing radius of every hydrogen atom bound to a nitrogen, H(N), from 1.2 Å to 1.3 Å. This modification is intended to strengthen hydrogen-bonded interactions, which may be underestimated in the mean-field continuum models relative to explicit solvent. The result for thioredoxin is shown in Figure 5 (red trace). Compared to an equivalent simulation based on the original Bondi radii (green trace), the use of the modified radii set yields only a slightly more stable trajectory: C_{α} RMSD averaged over the entire trajectory has shown only a marginal decrease from 1.7 Å to 1.6 Å. Longer simulations are required to make a decisive choice in favor of one or the

TABLE III. Effect of Radius Set Modification on $\Delta G_{\text{el}}(N) - \Delta G_{\text{el}}(U)$ [kcal/mol]

	Electrostatic model				
	PB (1 Å)	PB (0.5 Å)	GB^{OBC} set I	GB^{OBC} set II	GB^{HCT}
(Apo)myoglobin, pH = 2	-2090.4	-2094.6	-2095.2	-2093.4	-2149.8
Protein-A, pH = 7	+142.3	+141.9	+142.8	+141.2	+127.9

Radii of all hydrogens bound to a nitrogen are increased to 1.3 Å from 1.2 Å in the Bondi set. All other radii are Bondi.

other radii set for thioredoxin. However, it is important to stress that no undue bias toward the folded form is introduced by the radii adjustment (Table III), and since it was shown before to improve GB’s performance on nucleic acids, here we continue to explore the modified set on other systems.

Note that from a general standpoint, the task of balancing the intramolecular H-bonds versus H-bonds to implicit solvent is nontrivial. While in explicit solvent models their relative strength can be estimated from quantum mechanical calculations on small water/solute clusters, the strategy would not be directly applicable to the continuum solvent case.

The performance of the GB^{OBC} model in MD simulations can be significantly improved by modifications to the Φ/Ψ torsional potentials of the underlying *parm99* gas-phase force field, according to a proposal³⁷ aimed at a better energetic balance between α -helical and β -sheet regions of the Ramachandran plot. This force field, *parm99MOD2*, is described in the Methods section. Its use with the GB^{OBC} model results in additional stabilization of thioredoxin trajectory, C_{α} RMSD is 1.5 Å at 6 ns (Fig. 5, black trace), with the trajectory averaged value of 1.3 Å. Overall, for thioredoxin, we find the last combination (i.e., modified Bondi set and *parm99MOD2*) to perform best in MD simulation based on the GB^{OBC} model.

Ubiquitin and Protein-A

To test the performance of the GB^{OBC} model further, MD simulations on the native states of ubiquitin and protein-A have been carried out, using the same protocol as for thioredoxin. In both simulations, we use *parm99MOD2* force field and the modified Bondi radii set. Results are presented in Figure 6. In both cases, backbone C_{α} RMSDs from corresponding X-ray structures are stable over the last 7 ns of the trajectory, being around 1.4 Å for ubiquitin and 1.7 Å for protein-A. A short “spike” in the protein-A RMSD observed in the very beginning of the trajectory quickly disappears. Protein-A, being the smallest (46 residues) of the proteins we have used, is likely to be sensitive to residual strain left in the structure after equilibration; the effect is more pronounced in the implicit water environment, where there is nothing to “hold” the proteins during the first steps of unconstrained dynamics.

Barnase/Barstar Complex Formation

So far, we have been testing the modified GB model in MD simulations of native proteins. The usefulness of the GB^{OBC} model in simulating large conformational changes

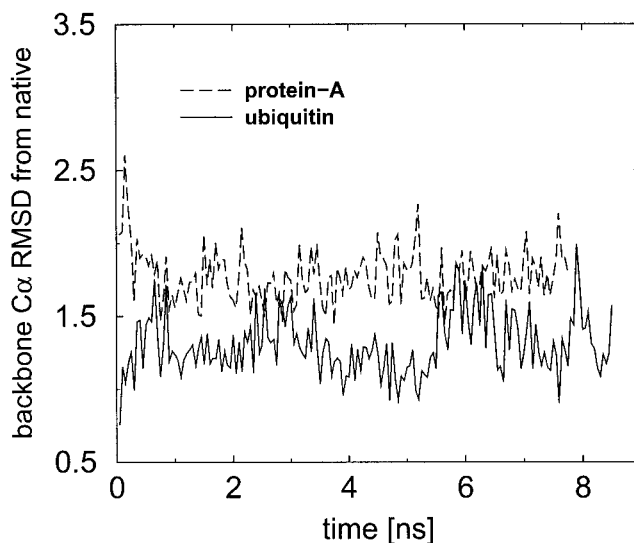


Fig. 6. Backbone RMSD from X-ray structure during the MD simulation of ubiquitin (solid line) and protein-A (dashed line) using GB^{OBC} (I) model based on the modified Bondi radii set and AMBER *parm99MOD2* force field.

and protein–protein interactions can be explored by simulating the formation of Barnase/Barstar complex from the unbound state. Unlike native-state simulations, considerable change in the amount of buried surface is expected upon docking, providing a good test for the model.

The results of this kind of simulation based on the GB^{OBC} model are presented in Figure 7. Initially, the two proteins are separated by 4.5 Å relative to their position in the X-ray structure of the complex (see Methods section). During the next 0.2 ns of the simulation, Barstar docks back into Barnase, and stays docked over the rest of the simulation. The decrease in the potential energy correlates well with the formation of the complex (Fig. 7). The structure corresponding to the lowest potential energy at $t = 0.409$ ps has a C_{α} RMSD from the X-ray (complex) of 1.88 Å. This observation suggests that one can further explore the possibility to use the AMBER-7 molecular mechanics potential energy, which, in the case of an implicit solvation model, includes solvation free energy as a reasonable scoring function in predicting the correct conformation of the complex. One test along these lines that we have carried out is as follows: We have performed 22 different MD simulations in which barnase and barstar are initially separated by 7 Å relative to their position in the X-ray structure of the complex (see Methods section).

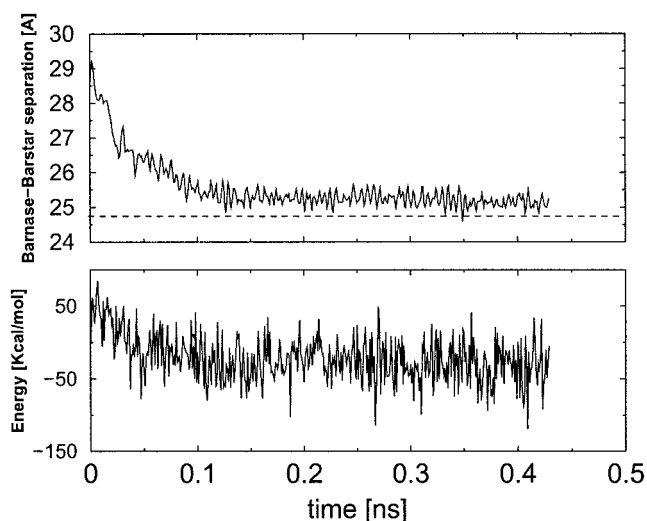


Fig. 7. Separation between Barnase and Barstar, measured as the distance between their centers of mass (**top**) and system's potential energy (**bottom**) during the MD simulation of the complex formation. GB^{OBC} (II) model is based on the original Bondi radii set and AMBER *parm99MOD2* force field is used. Thin dashed lines in the upper panel indicates the distance between Barnase and Barstar centers of mass in the X-ray structure of the complex.

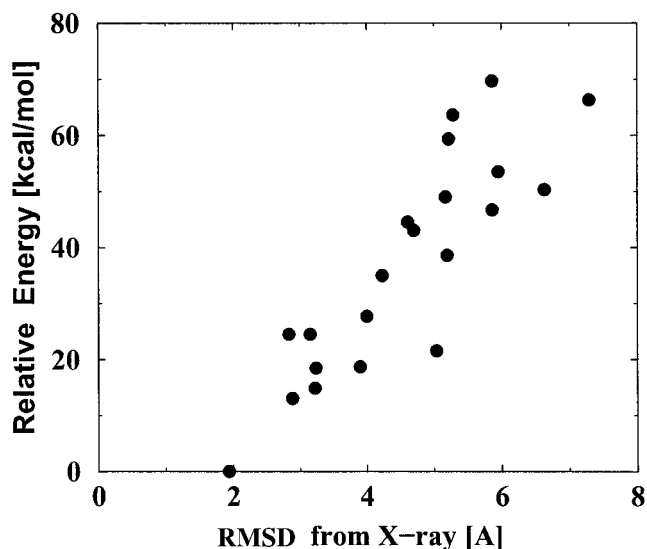


Fig. 8. Correlation between the energy of the predicted docked conformation of the Barnase/Barstar complex (AMBER-7 molecular mechanics potential energy + solvation free energy) versus the RMSD from the known X-ray structure of the complex. Each point represents an average over 10 lowest energy snapshots from the corresponding simulation. The 22 MD simulations differ by the initial position of Barstar relative to Barnase. The lowest energy/lowest RMSD point corresponds to Figure 7. GB^{OBC} (II) model based on the original Bondi radii set and AMBER *parm99MOD2* force field is used.

Results are presented in Figure 8, where the deviation (RMSD from the X-ray) from the correct structure is plotted as a function of the average energy of the 10 lowest energy snapshots for each trajectory. Clearly, there is a correlation between the two quantities: Structures with the lower energy tend to deviate less from the experimentally determined conformation of the complex.

Recently, Wang and Wade²³ reported a “run away” problem in GB-based simulations of barnase/barstar docking: After formation of a few intermolecular contacts, the molecules quickly moved away from each other. While a number of factors may account for the observed differences between the simulations based on GB and other models used in that study, as well as our own results presented above, we would like to point out that the effects of the initial conditions and the equilibration protocol may be more pronounced in the implicit solvent simulations relative to what is expected from one’s experience with explicit solvent; we have already seen a possible manifestation of these effects in the previous section (protein-A). In the case of Barnase/Barstar docking, the bound conformation corresponding to the lowest potential (Fig. 7) may not be reached if the relative kinetic energy of the unbound components is large enough. In fact, we observed this behavior with a protocol (different from what we have used in the present work) in which ballistic motion of the parts was not explicitly removed.

METHODS

Structures

The set of apo-myoglobin structures used in optimization of the GB model parameters was prepared from the holo-Mb coordinate set [Protein Data Bank (PDB) ID: 2Mb5] by heme removal and simulated acid unfolding in explicit solvent, as described elsewhere.³¹ The native state is represented by 50 consecutive snapshots (2 ps apart from each other) with near-native radius of gyration, ~ 16 Å, taken from the beginning of the acid-unfolding simulation. The unfolded state is represented by 50 consecutive snapshots from the end of that simulation, at which point the radius of gyration has approached ~ 30 Å—as is experimentally observed³² in the unfolded state. Protein-A structures were prepared from the NMR average coordinate set (PDB ID: 1BDD, residues 10–55). This structure was used as a starting point for all MD simulations reported here. The native-state ensemble is represented by 50 consecutive snapshots (2 ps apart from each other) from the implicit solvent simulation protocol described below, and deviations from the native coordinates are less than 2 Å for C_{α} atoms. The unfolded state was prepared by heating the protein to 450 K for 1 ns in an implicit solvent environment (Onufriev, unpublished data) and 50 consecutive snapshots with average RMSD from the native structure of no less than 15 Å were chosen to represent this state. Protein L (PDB ID: 1PTL) is represented by 215 snapshots from an explicit solvent high-temperature unfolding simulation,¹⁷ spanning various degrees of compactness, from close to native to largely unfolded. The villin head-piece is represented by 120 structures that model compact near-native, compact misfolded, and extended conformations, produced³⁰ from the native coordinate set (PDB ID: 1VII). The ubiquitin native structure is PDB ID: UBQ, and thioredoxin is PDB ID: 2TRX. Barnase/Barstar complex is given by X-ray coordinate set PDB ID: 1AY7, which is also used to prepare the unbound state by translating Barstar relative to Barnase.

MD Simulation Protocols/Force Fields

We used gas-phase potentials E_{vac} (force fields) provided by the AMBER(7) suite of programs, *parm94* and *parm99*, as well as a previously proposed modification of *parm99*, which we refer to as *parm99MOD2* (see below). The GB/SA methodology is used to represent solvation effects, with ΔG_{el} computed by a user-specified GB model. In AMBER-7, setting *igb* = 1 or 2 invokes, respectively, GB^{HCT} or GB^{OBC} (I) models used in this study. The GB^{OBC} (II) model also used here is currently not implemented in AMBER-7 but can be easily accessed within *igb* = 2 by changing appropriate values of the rescaling parameters α , β , γ (in *mdread.f*) to *galpha* = 1.0, *gbbeta* = 0.8, *gbgamma* = 4.85. The nonpolar contribution is computed via $\Delta G_{surf} = 0.005[\text{kcal/mol}] * A[\text{\AA}^2]$, where A is the solvent-accessible surface of the molecule. The SHAKE method is used to restrain hydrogen-heavy atom bond distances. The integration time-step is 2 fs, with no cutoff for long-range interactions (with the exception of the Protein-A and the Barnase/Barstar simulations, in which 20 Å and 24 Å cutoffs, respectively, are used). The average temperature of the system is maintained at 300 K by weak coupling (via the Berendsen algorithm) to a heat bath with a coupling constant of 2 ps. Each simulation begins with 100 steps of steepest descent minimization, followed by 20 ps of equilibration, during which the temperature is gradually raised to 300 K. The protein atom coordinates remain fixed by harmonic restraints (force constant 1 kcal/mol⁻¹ Å⁻² during minimization and 0.1 kcal/mol⁻¹ Å⁻² during equilibration) at their original (crystallographic) positions. After the equilibration is completed, the constraints are removed (except in the case of Barnase/Barstar, see below), and the simulation continues at 300 K. Overall translational (C.O.M.) and rotational motion is removed every 500 steps. The computations are performed on 16 CPUs of an R12000 SGI Origin 2800 machine. For thioredoxin (108 residues), 1 ns of MD simulation takes about 24 h. We find that computational speed, per CPU, of the GB^{OBC} model for this protein is only a few percentage points less than that of the GB^{HCT} model, although the difference may be platform specific. For protein-A (46 residues), the computational time is reduced to 8 h per nanosecond. A 0.5-ns Barnase/Barstar dynamics (185 residues) takes 24 h. The initial state for the first of the 22 simulations of Barnase/Barstar complex formation (Fig. 7) is prepared from the crystallographic coordinates of the complex by translating Barstar's coordinates (residues 97–185) by a vector (0, 3.0, -3.0)[Å], which is roughly perpendicular to their mutual interface. The other 21 simulations (650 ps each) are prepared by translating Barstar's coordinates (residues 97–185) by a vector [0, 7.0√2 cos(φ), 7.0√2 sin(φ)] [Å], φ = 120°, 125°...170°. For φ = 155°, 7 simulations were performed, with different seeds (AMBER default and *ig* = 1000, 4000, 5000, 6000, 7000, 8000) used to initialize the random distribution of initial velocities in the production run. For φ = 170°, 4 simulations were performed, with different seeds (AM-

BER default and *ig* = 1000, 2000, 3000). Positions of Barnase atoms remain fixed by weak harmonic restraints (0.01 kcal/mol Å⁻²), which, in combination with the removal of the overall C.O.M. translational velocity of the system, automatically eliminates possible ballistic motion of Barstar. The *parm99MOD2* modified force field is obtained from *parm99* by changing 4 torsional potentials describing Φ/Ψ dihedral motions. The modified lines in *parm99MOD2.dat* file are as follows:

```
N -CT-C -N 1 0.700 180.000 -1. parm99MOD2
N -CT-C -N 1 1.100 180.000 2. parm99MOD2
C -N -CT-C 1 0.000 180.000 -2. parm99MOD2
C -N -CT-C 1 1.000 0.000 1. parm99MOD2
```

The above changes were suggested in Simmerling et al.³⁷ and shown to yield good results for the folding of the “trp-cage” miniprotein.

Poisson-Boltzmann Calculations

All PB solvation energies were computed using DELPHI-II^{2,38} with a cubic box. The grid spacing used in each case is given in the main text. The dielectric constant for protein interior is 1, and the ionic strength is zero.

CONCLUSIONS

We have continued to improve the analytical pairwise GB approximation, with an eye toward using it in MD simulation of large-scale conformational changes. This has required a nonuniform rescaling of the effective radii computed by the original model. The modifications to the algorithm ensured its numerical stability in MD simulations, without significantly increasing the complexity of the algorithm. The model has performed quite well in native-state simulations of various proteins, as well in a simulation of Barnase/Barstar complex formation from an unbound state. Its computational efficiency in MD simulations is as good as that of the earlier GB model implemented in AMBER-6, in spite of a minor increase in the algorithm's complexity.

The adjustable parameters of the new model have been optimized to give best agreement with solvation energies computed via the numerical PB method on an ensemble of apo-myoglobin structures representing its simulated unfolding trajectory at low pH, from the native to the completely unfolded state. The optimized parameters have been tested on a set of structures representing the native and unfolded states of protein-A at neutral pH. Despite the large difference in the overall charge of these systems, the modified GB model agrees with the PB model to within 2–5 kcal/mol, for both the apo-myoglobin and protein-A, in computing the changes in electrostatic part of solvation free energy upon unfolding. Overall, the use of the improved GB model has substantially reduced, at a very modest computational expense, the bias (relative to PB model) toward compact structures found in the earlier GB model. Putting extra effort into improving the match with the PB may come at a price of increased complexity (and reduced computational efficiency) of the model. For the small villin head-piece structures, no appreciable difference between the new and earlier GB models was ob-

served, both predicting the values of the electrostatic part of solvation free energies within 1% of the corresponding PB values.

We have also explored the performance of the modified GB model in 8-ns-long MD simulations of three different proteins. A number of parameter options have been explored, and we found that a modification of the Φ/Ψ torsional parameters in AMBER *parm99* force field, aimed at improving the energetic balance between α -helical and β -sheet regions, results in smallest deviations (~ 1.5 Å backbone RMSD) from the X-ray structures, approaching the values expected from explicit water simulations. This is in contrast to the *parm94* force field, which showed poorer performance on these timescales when used for structures containing β -strands. These results demonstrate the usefulness of the modified GB model in fairly long-scale MD simulations of proteins. Another possible application of the model is the testing of underlying gas-phase force fields. There, it provides a relatively inexpensive way of exploring macromolecular dynamics on large, effective timescales, which may yet be inaccessible to explicit solvent simulations. The speed-up here is not purely computational: The absence of viscosity in implicit solvation models greatly facilitates conformational transitions compared to MD simulations based upon conventional explicit water models. This is not to say that the implicit solvent methodology is ready to eliminate the need for explicit solvent, but that a reasonable implicit solvation model, such as the proposed modified GB, may help identify subtle problems with force fields that only reveal themselves on long timescales that are much harder to access when using an explicit model.

The usefulness of the improved GB model in MD simulations of large-scale conformational changes is demonstrated in a simulation of Barnase/Barstar complex formation. The structure of the Barnase/Barstar complex is regenerated, starting from an unbound state, to within 1.9 Å relative to the crystal structure of the complex. Significantly, the decrease in molecular mechanics potential energy along the trajectory correlates well with the formation of the complex.

ACKNOWLEDGMENTS

We thank Michael Feig and Michael Lee for valuable discussions and for providing structural ensembles of villin and protein-L. We also thank Andreas Svrcsek-Seiler for reading the manuscript and making helpful comments.

REFERENCES

- Cramer CJ, Truhlar DG. Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem Rev* 1999;9:2161–2200.
- Honig B, Nicholls A. Classical electrostatics in biology and chemistry. *Science* 1995;268:1144–1149.
- Beroza P, Case DA. Calculation of proton binding thermodynamics in proteins. *Methods Enzymol* 1998;295:170–189.
- Madura JD, Davis ME, Gilson MK, Wade RC, Luty BA, McCammon JA. Biological applications of electrostatic calculations and Brownian dynamics. *Rev Comp Chem* 1994;5:229–267.
- Gilson MK. Theory of electrostatic interactions in macromolecules. *Curr Opin Struct Biol* 1995;5:216–223.
- Scarsi M, Apostolakis J, Cafilisch A. Continuum electrostatic energies of macromolecules in aqueous solutions. *J Phys Chem A* 1997;101:8098–8106.
- Luo R, David L, Gilson MK. Accelerated Poisson–Boltzmann calculations for static and dynamic systems. *J Comp Chem* 2002;23:1244–1253.
- Gelin BR. Testing and comparison of empirical force fields: techniques and problems. In: van Gunsteren WF, Weiner PK, Wilkinson AJ, editors. *Computer simulation of biomolecular systems*. Vol. 2. Leiden: ESCOM; 1993. p 127–146.
- Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Donini WWO, Cieplak P, Srinivasan J, Case DA, Cheatham TE. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 2000;33:889–897.
- Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 1990;112:6127–6129.
- Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. *J Phys Chem* 1996;100:1578–1599.
- Edinger S, Cortis C, Shenkin P, Friesner R. Solvation free energies of peptides: comparison of approximate continuum solvation models with accurate solution of Poisson–Boltzmann equation. *J Phys Chem B* 1997;101:1190–1197.
- Jayaram B, Liu Y, Beveridge DJ. A modification of the generalized Born theory for improved estimates of solvation energies and pK shifts. *J Chem Phys* 1998;109:1465–1470.
- Aghosh A, Rapp CS, Friesner RA. A generalized Born model based on a surface integral formulation. *J Phys Chem* 1998;102:10983–10990.
- Bashford D, Case D. Generalized Born models of macromolecular solvation effects. *Annu Rev Phys Chem* 2000;51:129–152.
- Onufriev A, Bashford D, Case D. Modification of the Generalized Born model suitable for macromolecules. *J Phys Chem B* 2000;104:3712–3720.
- Lee MS, Salsbury FR Jr, Brooks CL III. Novel generalized Born methods. *J Chem Phys* 2002;116:10606–10614.
- Dominy BN, Brooks CL. Development of a Generalized Born model parametrization for proteins and nucleic acids. *J Phys Chem B* 1999;103:3765–3773.
- David L, Luo R, Gilson MK. Comparison of generalized Born and Poisson models: energetics and dynamics of HIV protease. *J Comp Chem* 2000;21:295–309.
- Spassov VZ, Yan L, Szalma S. Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models. *J Phys Chem B* 2002;106:8726–8738.
- Calimet N, Schaefer M, Simonson T. Protein molecular dynamics with the generalized Born/ACE solvent model. *Proteins* 2001;45:144–158.
- Tsui V, Case D. Molecular dynamics simulations of nucleic acids using a generalized Born solvation model. *J Am Chem Soc* 2000;122:2489–2498.
- Wang T, Wade R. Implicit solvent models for flexible protein–protein docking by molecular dynamics simulation. *Proteins* 2003;50:158–169.
- Srinivasan J, Trevathan M, Beroza P, Case D. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor Chem Acc* 1999;101:426–434.
- Onufriev A, Case D, Bashford D. Effective Born radii in the Generalized Born approximation: the importance of being perfect. *J Comp Chem* 2002;23:1297–1304.
- Qiu D, Shenkin P, Hollinger F, Still WC. The GB/SA continuum model for solvation: a fast analytical method for the calculation of approximate Born radii. *J Phys Chem A* 1997;101:3005–3014.
- Hawkins GD, Cramer CJ, Truhlar DG. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J Phys Chem* 1996;100:19824–19836.
- Schaefer M, Froemmel C. A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution. *J Mol Biol* 1990;216:1045–1066.
- Richards FM. Areas, volumes, packing and protein structure. *Annu Rev Biochem Bioeng* 1977;6:151–176.
- Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL. Performance comparison of Generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J Comp Chem* 2004;25:265–284.
- Onufriev A, Case DA, Bashford D. Structural details, pathways, and energetics of unfolding apomyoglobin. *J Mol Biol* 2003;325:555–567.

32. Eliezer D, Yao J, Dyson HJ, Wright PE. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat Struct Biol* 1998;5:148–155.
33. Gohlke H, Case DA. Converging free energy estimates: MM-PB(GB)SA studies on the protein–protein complex Ras-Raf. *J Comp Chem* 2004;25:238–250.
34. Tsui V, Case D. Theory and applications of the Generalized Born solvation model in macromolecular simulations. *Biopolymers* 2001; 56:275–291.
35. Beachy MD, Chasman D, Murphy RB, Halgren TA, Friesner RA. Accurate ab initio quantum chemical determination of the relative energies of peptide conformations and assessment of empirical force fields. *J Am Chem Soc* 1997;119:5908–5920.
36. Garcia AE, Sanbonmatsu KY. Exploring the energy landscape of a β -hairpin in explicit solvent. *Proteins* 2001;42:345–354.
37. Simmerling C, Strockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. *J Am Chem Soc* 2002;124:11258–11259.
38. Nicholls A, Honig B. A rapid finite difference algorithm, utilizing successive over relaxation to solve the Poisson–Boltzmann equation. *J Comp Chem* 1991;12:435–445.