

**Chemical Theory and Computation Special Feature: Comparative study of generalized Born models: Protein dynamics**

Hao Fan, Alan E. Mark, Jiang Zhu, and Barry Honig

*PNAS* 2005;102;6760-6764; originally published online Apr 6, 2005;  
doi:10.1073/pnas.0408857102**This information is current as of March 2007.**

<b>Online Information &amp; Services</b>	High-resolution figures, a citation map, links to PubMed and Google Scholar, etc., can be found at: <a href="http://www.pnas.org/cgi/content/full/102/19/6760">www.pnas.org/cgi/content/full/102/19/6760</a>
	This article has been cited by other articles: <a href="http://www.pnas.org/cgi/content/full/102/19/6760#otherarticles">www.pnas.org/cgi/content/full/102/19/6760#otherarticles</a>
<b>E-mail Alerts</b>	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or <a href="#">click here</a> .
<b>Subspecialty Collections</b>	This article, along with others on similar topics, appears in the following collection(s): <b>Chemical Theory and Computation Special Feature</b> <a href="http://www.pnas.org/cgi/collection/chem_theory">www.pnas.org/cgi/collection/chem_theory</a>
<b>Rights &amp; Permissions</b>	To reproduce this article in part (figures, tables) or in entirety, see: <a href="http://www.pnas.org/misc/rightperm.shtml">www.pnas.org/misc/rightperm.shtml</a>
<b>Reprints</b>	To order reprints, see: <a href="http://www.pnas.org/misc/reprints.shtml">www.pnas.org/misc/reprints.shtml</a>

Notes:

# Comparative study of generalized Born models: Protein dynamics

Hao Fan<sup>\*†</sup>, Alan E. Mark<sup>\*\*‡</sup>, Jiang Zhu<sup>\*§</sup>, and Barry Honig<sup>\*§</sup>

<sup>\*</sup>Groningen Biomolecular Sciences and Biotechnology Institute, Department of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG, Groningen, The Netherlands; and <sup>§</sup>Department of Biochemistry and Molecular Biophysics, Columbia University and Howard Hughes Medical Institute, 630 West 168th Street, New York, NY 10032

Edited by Bruce J. Berne, Columbia University, New York, NY, and approved on March 11, 2005 (received for review December 16, 2004)

**In this work, we compare the results of molecular dynamics simulations involving the application of three generalized Born (GB) models to 10 different proteins. The three GB models, the Still, HCT, and modified analytical generalized Born models, were implemented in the computationally efficient GROMACS package. The performance of each model was assessed from the backbone rms deviation from the native structure, the number of native hydrogen bonds retained in the simulation, and the experimental and calculated radius of gyration. Analysis of variance (ANOVA) was used to analyze the results of the simulations. The rms deviation measure was found to be unable to distinguish the quality of the results obtained with the three different GB models, whereas the number of native hydrogen bonds and radius of gyration yielded a statistically meaningful discrimination among models. Our results suggest that, of the three, modified analytical generalized Born yields the best agreement between calculated and experimentally derived structures. More generally, our study highlights the need both to evaluate the effects of different variables on the results of simulations and to verify that the results of molecular dynamics simulations are statistically meaningful.**

analysis of variance | molecular dynamics

**M**olecular dynamics (MD) simulations are widely used in the study of protein structure and functions (1, 2). The results of a given simulation depend on a number of factors, such as the quality of the molecular force field, the treatment of solvent, the timescale of the simulation, and the sampling efficiency of the simulation protocol. There has been an enormous investment in the underlying technology in each of these areas, and the range of application of MD simulations has expanded greatly since the method was first introduced (3). However, by their very nature, MD simulations are not easy to compare to experiment. Simulations of protein systems involve a considerable amount of computer time, and the results in general cannot be directly compared with the experimental observables without additional processing. When such comparisons are made, the results are often very encouraging, but, given the multiplicity of parameters and computational methodologies that are used, it is hard to know whether the success of a particular protocol, e.g., using one of the available force fields, indicates that all other force fields will yield results of comparable quality. Fortunately, the growth in available computer power combined with the development of highly optimized computer code has made careful comparison of methods more feasible and, in addition, should make it increasingly possible to test whether a particular result is robust in terms of its sensitivity to the parameters of the calculation. Moreover, it would be highly desirable to know whether a particular methodology or combination of parameters is best suited to a particular application and to understand the reasons for performance differences, to the extent that they exist. This goal requires that different methods be compared so that their strengths and weaknesses can be critically evaluated.

The most straightforward validation of MD simulations of proteins is based on comparisons between experimentally

determined structures and the conformations that are generated in a simulation that uses the experimental structure as an initial condition. However, there are a number of complications in even this kind of test. The most accurate experimental structures are obtained from x-ray crystallography, but most simulations do not mimic crystalline conditions. Comparisons with NMR structures obtained in solution would appear more appropriate, but NMR structures tend to be less accurate than x-ray structures and are more sensitive to the energy model (e.g., force field) used in generating coordinates. Much remains to be done in comparing theory and experiment and in comparing the one theoretical approach to another. An issue that naturally arises in such comparisons is whether a particular conclusion is statistically meaningful. Has a large enough data set been used, has the simulation been run for a long enough time, and have appropriate statistical tests been applied to the output of the simulations?

In this article, we address some of these issues by using a comparison of continuum solvent models as an example. Specifically, we evaluate the ability of three generalized Born (GB) models, used in conjunction with the GROMOS96 force field (4, 5) and stochastic dynamics (SD) (6) simulations, to reproduce experimentally determined protein structures. The three models are those of Still and coworkers (7), the HCT model from the Cramer/Truhlar group (8), and a modified AGB (mAGB) model of Levy and coworkers (9) developed in our laboratory (10). The three models have been implemented in the GROMACS 3.0 package (11–13). Five-nanosecond simulations using each of the three models were carried out on 10 different proteins (Table 1), and the structural properties averaged over the last 1 ns of each simulation were compared with the corresponding properties obtained from the experimentally determined structures. The extent to which differences in the results obtained from different GB modes were statistically meaningful was determined through analysis of variance (ANOVA). Our analysis reveals that some structural measures are better than others in their ability to discriminate among models, and, in addition, it emphasizes the importance of applying statistical analysis to the results of MD simulations. Proper analysis also requires the use of a sufficiently large data set, something that is no longer a problem given the computational efficiency of modern MD technologies and the availability of large-scale computational resources.

The results reported here and in a previous study (10) indicate that there are important differences between GB models. In addition, model performance is found to depend on simulation

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: DEV, relative deviation; GB, generalized Born; mAGB, modified analytical generalized Born; MD, molecular dynamics; RMSD, rms deviation; SA, surface area; SD, stochastic dynamics.

<sup>†</sup>H.F. and J.Z. contributed equally to this work.

<sup>‡</sup>To whom correspondence may be addressed. E-mail: a.e.mark@chem.rug.nl or bh6@columbia.edu.

© 2005 by The National Academy of Sciences of the USA

**Table 1. Proteins used in the simulations**

PDB	Description	Exp.	Resol.	$N_{res}$	$N_{\alpha}$	$N_{\beta}$	$N_{charge}$
1afi	Reduced form of MERP, mercuric transport protein	NMR	—	72	22	22	3
1ail	N-terminal fragment of Nsl, RNA-binding protein	X-ray	1.9	70	60	0	2
1ctf	C-terminal domain of the ribosomal protein L7/L12	X-ray	1.7	68	39	19	-2
1lea	Lex A repressor DNA binding domain	NMR	—	72	36	6	2
1pgb	B1 domain of protein G	X-ray	1.9	56	14	30	-4
1shg	Src homology 3 (SH3) domain	X-ray	1.8	57	3	28	1
1tuc	Src homology 3 (SH3) domain, circular permutant cut at S19-P20	X-ray	1.8	61	3	25	0
1ubi	Ubiquitin	X-ray	1.8	76	16	33	0
2bby	DNA-binding domain from human Rap30	NMR	—	69	32	12	3
2ci2	Chymotrypsin inhibitor 2	X-ray	2.0	63	13	22	-1

PDB, PDB ID code (two N-terminal residues of 2ci2 have been removed); Exp., experimental method for determining the protein structure; Resol., structural resolution for x-ray structures;  $N_{res}$ , number of residues in the protein;  $N_{\alpha}$ , number of helical residues as defined in the PDB file;  $N_{\beta}$ , number of strand residues as defined in the PDB file;  $N_{charge}$ , net charge of the protein at pH 7.0.

protocol; for example, the friction coefficient used in a simulation or initial conditions. Some of the implications of our results are discussed below.

## Methods

**Protein Data Set.** Table 1 provides information about the 10 proteins used in the simulations. Seven of the structures were solved with x-ray crystallography, and seven were solved with NMR. For the latter, the first structure in each PDB file was used, except for 1lea, for which only one energy-minimized average structure is available. All 10 proteins are small (between 56 and 76 residues), have no disulfide bonds, and have different compositions of secondary structure elements.

**Model Implementation.** The HCT, modified analytical generalized Born (mAGB), and Still models were implemented in the GROMACS 3.0 package (11–13). The precision of the implementation was examined by evaluating both energies and forces. The energy calculation was tested through a comparison with values obtained from the same models within the GROMOS package. The force was tested by comparing the GB force calculated from analytic derivatives with that obtained from finite interpolation. In finite interpolation, the energy change upon a perturbation in a coordinate should approach the analytic derivative when the perturbation is sufficiently small. A coordinate increment of 0.01 Å was used, and the protein 1pgb was tested. The nonpolar solvation term, estimated from an analytical surface-area term calculation (10, 14, 15), was combined with each GB model (GB/surface area model, GB/SA model) to represent the complete solvent potential of mean force. It should be emphasized that the nonpolar term used in the original AGBNP model (9) has been replaced in the mAGB implementation (10) by a simpler surface-based term for convenience of comparison between methods.

**Simulation Methodology.** The GROMOS96 force field (4, 5) was used in all simulations. All ionizable amino acids were assumed to be in protonation states appropriate to a pH of 7 assuming standard  $pK_a$ . Each GB/SA model was combined with a SD algorithm (6) to represent the effect of the solvent. Bound ions and crystal waters were removed, and no counter ions were added to neutralize the protein/implicit solvent systems. All protein structures were first optimized through steepest descent energy minimization and then subjected to 10 ps of positionally restrained dynamics with a force constant of 2.4 kcal/molÅ<sup>2</sup>. Then, 5-ns SD simulation with each GB/SA model was performed. The equations of motion were integrated with a leapfrog SD scheme and a time step of 0.002 ps. The temperature was kept at 300 K,

and, except where noted (see below), the friction coefficient was set to 10 ps<sup>-1</sup>. Covalent bonds were constrained with the LINCS algorithm (16). Nonbonded interactions were calculated every step without cutoffs. The nonpolar surface-area term and its derivative were updated every 10 time steps. The surface coefficient  $\sigma$  was set to 0.005 kcal/molÅ<sup>2</sup> for all atom types except for sulfur (-0.005 kcal/molÅ<sup>2</sup>) and polar hydrogens (0.0 kcal/molÅ<sup>2</sup>) (10, 15).

Three additional 5-ns simulations were performed on 1pgb by using different initial atomic velocities, resulting in a total of four runs for this protein. Another four simulations by using otherwise identical conditions were performed on 1pgb by using a friction coefficient of 50 ps<sup>-1</sup>. All simulation trajectories were analyzed with tools in the GROMACS3.0 package. Different properties were extracted from the trajectories and averaged over the last period (1 ns) of the simulation. These average properties were subjected to the statistical analysis.

**ANOVA.** ANOVA (17) is a statistical technique that can evaluate the significance of differences between two or more groups. Two-way ANOVA is appropriate to cases where there are two independent variables. The “null hypotheses” to be tested are that the effect of the first (or second) variable is not significant and that the two variables are not correlated. In this work, the GB model and the protein correspond to the two independent variables, and the null hypothesis of interest is that three GB models are equivalent. ANOVA will calculate the probability ( $P$ ) of the null hypothesis being true. A sufficiently low  $P$  value indicates that a particular property can differentiate among the GB models. A high  $P$  value suggests that the property in question should not be used to infer differences in model quality. ANOVA results are valid only when the samples are normally distributed. This conclusion was confirmed through an analysis of residual. Detailed descriptions of ANOVA can be found online or in most statistics textbooks. We used a LINUX version of the statistical package, R ([www.r-project.org](http://www.r-project.org)) for ANOVA and validation tests.

## Results

**rms Deviation (RMSD) of Backbone Atoms.** Table 2 lists the RMSDs of the backbone of secondary structures averaged over the last 1 ns of a given simulation. The RMSDs are in the range observed for previously reported explicit (15, 18) and implicit (15, 19–26) solvent simulations. Many of the values are <2 Å, and for only one protein (1pgb) is the RMSD close to 3 Å. Table 2 (multiple simulations, 10-ps<sup>-1</sup> friction coefficient) shows results for the four sets of simulations carried out on 1pgb with different initial velocities. The first of these simulations corresponds to the one

**Table 2. The backbone RMSD of secondary structure elements**

	Simulations with a 10-ps <sup>-1</sup> friction coefficient for selected PDB entries, Å										Multiple simulations of 1 pgb with a 10-ps <sup>-1</sup> friction coefficient, * Å					Multiple simulations of 1 pgb with a 50-ps <sup>-1</sup> friction coefficient, * Å				
	1afi	1ail	1ctf	1lea	1pgb	1shg	1tuc	1ubi	2bby	2ci2	1	2	3	4	Avg.	1	2	3	4	Avg.
HCT	1.5	2.5	1.7	2.0	3.3	1.6	<b>1.0</b>	1.4	2.3	<b>1.4</b>	3.3	2.1	1.4	2.8	<b>2.4</b>	—	—	—	—	—
mAGB	1.8	<b>1.6</b>	<b>1.3</b>	2.2	<b>2.7</b>	1.3	1.1	1.0	<b>1.2</b>	1.6	2.7	2.6	1.8	2.3	<b>2.4</b>	1.1	1.6	2.2	1.2	<b>1.5</b>
Still	1.6	2.0	1.8	<b>1.1</b>	2.9	<b>1.1</b>	1.8	<b>0.9</b>	2.0	2.7	2.9	1.9	2.3	2.8	<b>2.5</b>	—	—	—	—	—

The RMSD is calculated with respect to the experimental structure and averaged over the last 1-ns simulation. The value in each cell is the backbone RMSD of the secondary structures (as defined in the PDB file). For 1shg, we used the definition in the PDB file of 1aef because there is no definition in the 1shg PDB file, and 1shg and 1aef are the same protein. The lowest values for each protein are shown in bold. HCT, RMSD calculated from the SD simulation with the HCT GB/SA model; mAGB, RMSD calculated from the SD simulation with the mAGB GB/SA model; Still, RMSD calculated from the SD simulation with the Still GB/SA model.

\*Results from four simulations on 1pgb with each of the three GB models and a 10-ps<sup>-1</sup> friction coefficient and four simulations with the mAGB model and a 50-ps<sup>-1</sup> friction coefficient.

also listed in column 6. As can be seen, the RMSDs for the three additional simulations are lower than for the first. The average RMSD of the 10 proteins is 1.9, 1.6, and 1.8 Å for the HCT, mAGB, and Still models, respectively. The lowest RMSD for each protein (marked in bold) is scattered among the three GB models, and visual inspection does not identify a best performing GB model.

Two-way ANOVA confirms this conclusion; the probability of the null hypothesis being true (i.e., that the three GB models are equivalent) is 0.32, which indicates that differences among models detected by RMSD results are not statistically significant. This finding suggests that RMSD is not a useful discriminator of model quality in the current study. Although we have not analyzed the effect of the protein on the results in detail, it is of interest to consider whether there are differences between proteins that lead to different RMSDs. Possible factors include the location of salt bridges for which the subtle balance of relevant forces (e.g., Coulomb and desolvation) is difficult to calculate or, perhaps, the energetics of secondary structure formation.

**Number of Native Hydrogen Bonds.** We evaluated all hydrogen bonds involving residues that are part of secondary structure elements. Tables 3 lists the number of native hydrogen bonds observed in the experimentally determined structures, the av-

erage number of these hydrogen bonds in the last 1 ns (obtained by averaging the number seen in each snapshot), and the relative deviation of the two values (DEV). DEV is defined as  $Q_{sim}/Q_{exp} - 1$ , where  $Q_{sim}$  is the value for a particular quantity obtained from the calculation and  $Q_{exp}$  is the value of the same quantity obtained from the experimental structure. [The mAGB model is seen to have the lowest deviations for every protein except 1pgb, where, based on the results shown in Table 3 (multiple simulations with a 10-ps<sup>-1</sup> friction coefficient), the performance of all three models is essentially the same.] The average value of DEV for the 10 proteins is -0.368, -0.245, and -0.331 for the HCT, mAGB, and Still models, respectively. The HCT simulations, in particular, tend to lose a relatively large number of native hydrogen bonds. Two-way ANOVA indicates that the probability of the null hypothesis being correct is  $1.8 \times 10^{-4}$ , suggesting that differences among model performance, as measured by the number of native hydrogen bonds, are statistically significant.

**Radius of Gyration.** Table 4 lists the radius of gyration obtained from the experimental structures and from different simulations together with the corresponding DEV values. All GB models yield comparable values with a maximal deviation from the experiment of 0.5 Å. The average DEV value for the 10 proteins is -0.016, 0.001, and 0.012 for the HCT, mAGB, and Still models, respectively. The lowest DEV values, indicated in bold

**Table 3. No. of native hydrogen bonds involving secondary structure elements**

	Simulations with a 10-ps <sup>-1</sup> friction coefficient for selected PDB entries, Å										Multiple simulations of 1 pgb with a 10-ps <sup>-1</sup> friction coefficient, * Å					Multiple simulations of 1 pgb with a 50-ps <sup>-1</sup> friction coefficient, * Å				
	1afi	1ail	1ctf	1lea	1pgb	1shg	1tuc	1ubi	2bby	2ci2	1	2	3	4	Avg.	1	2	3	4	Avg.
EXP	29	49	41	25	31	19	16	22	21	17	—	—	—	—	—	—	—	—	—	—
HCT	18	31	28	18	14	12	11	16	11	11	14	19	22	16	<b>18</b>	—	—	—	—	—
	-0.379	-0.367	-0.317	-0.280	-0.548	-0.368	-0.312	-0.273	-0.476	-0.353	—	—	—	—	—	—	—	—	—	—
mAGB	19	40	34	20	14	15	14	19	16	12	14	18	18	22	<b>18</b>	25	23	24	24	<b>24</b>
	<b>-0.345</b>	<b>-0.184</b>	<b>-0.171</b>	<b>-0.200</b>	-0.548	<b>-0.210</b>	<b>-0.125</b>	<b>-0.136</b>	<b>-0.238</b>	<b>-0.294</b>	—	—	—	—	—	—	—	—	—	—
Still	18	37	28	20	15	13	10	18	12	11	15	21	21	17	<b>18</b>	—	—	—	—	—
	-0.379	-0.245	-0.317	-0.200	<b>-0.516</b>	-0.316	-0.375	-0.182	-0.429	-0.353	—	—	—	—	—	—	—	—	—	—

The number of native hydrogen bonds involved in the secondary structures retained in the simulation is calculated and averaged over the last 1-ns simulation. The native hydrogen bonds are defined as those in the experimental structure (given by EXP). In HCT, mAGB, and Still, the first line gives the number of nature hydrogen bonds, and the second line gives their relative deviation (DEV) from the experimental value. The lowest deviations among simulations are shown in bold. A hydrogen bond is considered to exist if the donor-hydrogen-acceptor angle is <60° and the distance between the hydrogen and the acceptor is <2.5 Å. EXP, value calculated from the experimental structure; HCT, value and DEV calculated from the SD simulation with the HCT GB/SA model; mAGB, value and DEV calculated from the SD simulation with the mAGB GB/SA model; Still, value and DEV calculated from the SD simulation with the Still GB/SA model.

\*Results from four simulations on 1pgb with each of the three GB models and a 10-ps<sup>-1</sup> friction coefficient and four simulations with the mAGB model and a 50-ps<sup>-1</sup> friction coefficient.

Table 4. Radius of gyration

	Simulations with a 10-ps <sup>-1</sup> friction coefficient for selected PDB entries, Å										Multiple simulations of 1pgb with a 10-ps <sup>-1</sup> friction coefficient,* Å					Multiple simulations of 1pgb with a 50-ps <sup>-1</sup> friction coefficient,* Å				
	1afi	1ail	1ctf	1lea	1pgb	1shg	1tuc	1ubi	2bby	2ci2	1	2	3	4	Avg.	1	2	3	4	Avg.
EXP	10.9	12.8	11.2	11.4	10.6	10.3	10.9	11.7	12.0	11.2	—	—	—	—	—	—	—	—	—	—
HCT	11.1	12.5	10.9	11.1	10.5	10.4	10.5	11.7	11.6	10.8	10.5	10.4	10.5	10.5	<b>10.5</b>	—	—	—	—	—
	<b>0.018</b>	-0.023	-0.027	-0.026	<b>-0.009</b>	<b>0.010</b>	-0.037	<b>0.0</b>	-0.033	-0.036										
mAGB	11.4	12.8	11.2	11.5	10.5	10.6	10.8	11.7	11.7	10.9	10.5	10.6	10.6	10.5	<b>10.6</b>	10.6	10.6	10.5	10.6	<b>10.6</b>
	0.046	<b>0.0</b>	<b>0.0</b>	<b>0.009</b>	<b>-0.009</b>	0.029	<b>-0.009</b>	<b>0.0</b>	-0.025	-0.027										
Still	11.3	13.0	11.3	11.5	10.5	10.6	11.2	11.8	11.8	11.3	10.5	10.7	10.6	10.5	<b>10.6</b>	—	—	—	—	—
	0.037	0.016	0.009	<b>0.009</b>	<b>-0.009</b>	0.029	0.028	0.008	<b>-0.017</b>	<b>0.009</b>										

The radius of gyration is calculated and averaged over the last 1-ns simulation. The first HCT, mAGB, and Still line gives the average radius of gyration, and the second line gives the relative deviation from the experimental value (given by EXP). The lowest deviations among simulations are shown in bold. EXP, value calculated from the experimental structure; HCT, value and DEV calculated from the SD simulation with the HCT GB/SA model; mAGB, value and DEV calculated from the SD simulation with the mAGB GB/SA model; Still, value and DEV calculated from the SD simulation with the Still GB/SA model.

\*Results from four simulations on 1pgb with each of the three GB models and a 10-ps<sup>-1</sup> friction coefficient four simulations with the mAGB model and a 50-ps<sup>-1</sup> friction coefficient.

in the table, are evenly distributed and do not indicate a preference for a particular model. Nevertheless, two-way ANOVA reveals that the probability of the null hypothesis being correct is  $9.5 \times 10^{-5}$ , suggesting that the radius of gyration is a good discriminator of GB models. Indeed, the data in Table 4 suggest that the HCT model tends to make proteins slightly too compact and that the Still model tends to make them expand slightly, whereas mAGB exhibits no particular preference and has the smallest average deviations from experimental values.

**Friction Coefficient.** Four simulations using the mAGB model and a friction coefficient of 50 ps<sup>-1</sup> were performed on 1pgb with different initial atomic velocities. (Results are presented in the last five columns of Tables 2–4.) The average RMSD is reduced from 2.4 to 1.5 Å when the larger friction coefficient is used (Table 2). More native hydrogen bonds are retained (24 instead of 18; Table 3), but the radius of gyrations is not affected (Table 4). The results suggest that the structural deviations (RMSDs) observed in the simulations, in particular the large deviation of 1pgb, may be strongly influenced by the simulation protocol (e.g., the friction coefficient) and not just changes in the interaction function, in this case the GB model. The improved agreement with experiment of course does not imply that the GB model performs better with a higher friction coefficient. The friction coefficient affects only dynamic properties of a system, not thermodynamic properties such as the equilibrium structure. In this case, increasing the friction coefficient reduces the structural deviation by restricting the mobility of the protein, thus lowering the sampling efficiency.

## Discussion

**Criteria for Model Evaluation.** Despite the wide use of RMSD as a measure of the success of a particular set of parameters in applications to proteins, two-way ANOVA indicates that RMSD is in fact a poor discriminator of different GB models. Strictly speaking, this result is valid only for the comparison of the GB models used in this study, but its reduced sensitivity relative to factors such as the number of native hydrogen bonds or radius of gyration is noteworthy. Although this result does not prove that RMSD will always fail to discriminate among models, the calculated RMSD clearly results from a combination of factors that may reduce the amount of useful information this measure provides.

In addition to RMSD, MD studies of protein dynamics often report structural features such as radius of gyration, hydrophilic and hydrophobic solvent-accessible surface area, hydrogen

bond, and the number of residue–residue contacts. Our results suggest that some of these properties are more sensitive to changes in the energy function than RMSD. Indeed, ANOVA indicates that hydrophilic and hydrophobic surface areas also provide significant tests of GB models. In contrast, the number of residue–residue contacts was not sensitive to the model differences. It may well prove to be the case that different tests are appropriate to different questions about the energy function. For example, hydrogen bonding is clearly sensitive to solvent model and, thus, is likely to be affected by changes in the GB model. In contrast, it may well be the case that the number of residue–residue contacts will provide a useful means of evaluating van der Waals potentials.

**Evaluation of GB Models.** In a previous study, we compared the Still, HCT, mAGB, and GBMV2 models in terms of their ability to reproduce Born radii obtained from the Poisson equation (DELPHI, refs. 27 and 28). We also compared the effectiveness of each model, used in conjunction with the GROMOS96 force field (4, 5), in the *ab initio* folding of two peptides. The GBMV2 model (29, 30) (which is computationally more demanding than the other three models) was found to be most effective in reproducing the DELPHI results, but the Still and mAGB models were most effective in the peptide folding, yielding some trajectories that produced native-like structures without the use of enhanced sampling techniques. This finding was attributed to the fact that the latter two models allowed for more efficient sampling even though they are somewhat less accurate in their ability to reproduce the results of a rigorous continuum model. This study also finds that mAGB is the most successful of the three fast GB models. In particular, mAGB maintains the largest number of native hydrogen bonds and yields radii of gyration in best agreement with experiment. In a previous study we found that, of the three models considered here, mAGB was the most effective in reproducing DELPHI-derived Born radii. In our implementation, mAGB is between 10% and 15% slower than the Still and is faster than HCT by about the same factor.

**General Implications.** Continuing improvements in access to computer resources and algorithms (programs) are enabling MD simulations of biomolecular systems to approach experimentally relevant time and length scales. With these factors, it is becoming increasingly possible to rigorously evaluate the available computational methodology against experimental data. In this paper, we have described the results of SD simulations on 10 proteins with the specific goal of evaluating the performance of

three fast GB models. A second goal was to demonstrate the importance of using statistical techniques in the analysis of the results of MD simulations. The more general purpose of both aspects of this study is to help establish standards that enable the comparison and validation of parameters used in the simulations. A number of our findings clearly indicate the need to go beyond the consideration of just individual molecules and/or well chosen test cases. Instead, we must strive to employ data sets that are sufficiently large and diverse to enable rigorous statistical analysis. For example, the B1 domain of protein G (1pgb) has been used as a model system in a variety of simulations. However, in our study, the results obtained for this protein appear atypical, and, indeed, agreement with calculated structural parameters appears better for other proteins in our test set.

A second finding of our study is the value of using statistical analysis in assessing the significance of conclusions that are derived from simulations. For example, we find that, despite its wide use in evaluating the relative quality of different simulations, RMSD does not appear to be an optimal measure for discriminating among different GB models. In contrast, the number of native hydrogen bonds and the radius of gyration appear to be better discriminators of GB solvent models. We note in passing that the number of native hydrogen bonds that are maintained in our simulations is significantly lower than the number observed experimentally. The situation appears to improve when a larger friction coefficient is used, simply highlighting the need to carefully evaluate the performance of a particular simulation protocol before specific conclusions are drawn. Still, the extent to which hydrogen bonds are broken raises concerns

as to conclusions that are often made in MD simulations as to the biological importance of individual hydrogen bonds.

Finally, it should be pointed out that the major goal of this work was not to comment directly on the accuracy of a particular set of parameters but rather to introduce criteria as to how comparisons should be carried out. In particular, we have shown that it is possible to use appropriate statistical tools such as ANOVA in the analysis of computational studies of proteins. As mentioned above, truly detailed experimental evaluation of simulation accuracy requires, on the one hand, that the crystal environment be properly accounted for and, on the other hand, that the accuracy of NMR structures continues to improve. The problem is particularly complex for side chains, whose conformations can be strongly influenced by crystal packing and can be difficult to determine accurately with NMR methods. However, in this work we have used backbone hydrogen bonds as a means of evaluation so that this issue should not be significant. Indeed, the overall level of agreement with experiments did not appear to be affected by the method used for structure determination.

We hope that our results indicate that the true merits of different interaction functions and methodologies may be assessed in an objective fashion so that we may progress beyond the current situation, where results are often presented without careful validation. We feel that the available programs, algorithms, and computer power now make it possible to carry out such tests that were often difficult to implement in the past.

We acknowledge Dr. Erik Lindahl for his help in the implementation of GB/SA models in GROMACS and Dr. Tsjerk Wassenaar for his help with the ANOVA. Support from National Institutes of Health Grant GM-30518 (to B.H.) is also acknowledged.

- Karplus, M. (2003) *Biopolymers* **68**, 350–358.
- Karplus, M. & McCammon, J. A. (2002) *Nat. Struct. Biol.* **9**, 788.
- McCammon, J. A., Gelin, B. R. & Karplus, M. (1977) *Nature* **267**, 585–590.
- van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hunenberger, P. H., Kruger, P., Mark, A. E., Scott, W. R. P., Tironi, I. G. (1996) *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (University of Groningen, Groningen, The Netherlands and ETH, Zurich, Switzerland).
- van Gunsteren, W. F., Daura, X. & Mark, A. E. (1998) in *Encyclopedia of Computational Chemistry* (Wiley, New York) Vol. 2, pp. 1211–1216.
- Shi, Y. Y., Wang, L. & van Gunsteren, W. F. (1988) *Molecular Simulation* **2**, 217–221.
- Qiu, D., Shenkin, P. S., Hollinger, F. P. & Still, W. C. (1997) *J. Phys. Chem.* **101**, 3005–3014.
- Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. (1996) *J. Phys. Chem.* **100**, 19824–19839.
- Gallicchio, E. & Levy, R. M. (2004) *J. Comput. Chem.* **25**, 479–499.
- Zhu, J., Alexov, E. & Honig, B. (2005) *J. Phys. Chem. B* **109**, 3008–3022.
- Berendsen, H. J. C., Vanderspoel, D. & Vandrunen, R. (1995) *Comput. Phys. Commun.* **91**, 43–56.
- Lindahl, E., Hess, B. & van der Spoel, D. (2001) *J. Mol. Model (Online)* **7**, 306–317.
- van der Spoel, D., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Hess, B., Feenstra, K. A., Lindahl, E., van Drunen, R. & Berendsen, H. J. C. (2001) *GROMACS 3.0* (University of Groningen, Groningen, The Netherlands).
- Richmond, T. J. (1984) *J. Mol. Biol.* **178**, 63–89.
- Zhu, J. A., Shi, Y. Y. & Liu, H. Y. (2002) *J. Phys. Chem. B* **106**, 4844–4853.
- Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. (1997) *J. Comput. Chem.* **18**, 1463–1472.
- Wonnacott, T. H. & Wonnacott, R. J. (1990) *Introductory Statistics* (Wiley, New York).
- Fan, H. & Mark, A. E. (2003) *Proteins Struct. Funct. Genet.* **53**, 111–120.
- Calimet, N., Schaefer, M. & Simonson, T. (2001) *Proteins Struct. Funct. Genet.* **45**, 144–158.
- Cornell, W., Abseher, R., Nilges, M. & Case, D. A. (2001) *J. Mol. Graphics* **19**, 136–145.
- Dominy, B. N. & Brooks, C. L. (1999) *J. Phys. Chem. B* **103**, 3765–3773.
- Ferrara, P., Apostolakis, J. & Caflisch, A. (2002) *Proteins Struct. Funct. Genet.* **46**, 24–33.
- Fraternali, F. & van Gunsteren, W. F. (1996) *J. Mol. Biol.* **256**, 939–948.
- Krol, M. (2003) *J. Comput. Chem.* **24**, 531–546.
- Onufriev, A., Case, D. A. & Bashford, D. (2002) *J. Comput. Chem.* **23**, 1297–1304.
- Tsui, V. & Case, D. A. (2000) *Biopolymers* **56**, 275–291.
- Rocchia, W., Alexov, E. & Honig, B. (2001) *J. Phys. Chem. B* **105**, 6507–6514.
- Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A. & Honig, B. (2002) *J. Comput. Chem.* **23**, 128–137.
- Lee, M. S., Salsbury, F. R. & Brooks, C. L. (2002) *J. Chem. Phys.* **116**, 10606–10614.
- Lee, M. S., Feig, M., Salsbury, F. R. & Brooks, C. L. (2003) *J. Comput. Chem.* **24**, 1348–1356.