Solvation Free Energy of Biomacromolecules: Parameters for a Modified Generalized Born Model Consistent with the AMBER Force Field

B. Jayaram,† D. Sprous, and D. L. Beveridge*

Department of Chemistry, Wesleyan University, Middletown, Connecticut 06459

Received: April 27, 1998

The generalized Born (GB) model provides rapid estimates of the electrostatic free energies of solvation for diverse molecules and molecular ions. This method is expected to be of considerable utility for studies of solvation in macromolecular and biological systems. Calculations on biological molecules are typically based on empirical energy functions, each of which have their own prescriptions for determining net atomic charges. For maximum compatibility, GB parameters tailored to specific force fields are required. The development of parameters compatible with the AMBER force field is described. The method is used to estimate free energies of A and B form structures of DNA obtained from molecular dynamics simulations. The results provide an account of the conformational preferences of right-handed DNA in solution.

I. Introduction

Generalized Born (GB) theory[1−4] is the basis of a computational method for estimating the electrostatic free energies of solvation of diverse molecules and molecular ions. In the GB model, a molecule in solution is represented as a set of point charges set in spherical cavities embedded in a polarizable dielectric continuum. GB calculations can be considered as a means of approximating finite difference Poisson−Boltzmann[5] free energies and related quantities. Previous studies have demonstrated that the GB method together with a simple treatment of nonelectrostatic effects can estimate solvation free energies that are generally within 5% of observed values,[1,4] and a recent modification, the MGB method,[6] has extended the purview of the GB theory to p*K* shifts of dicarboxylic acids as well as hydration energies.

The essential simplicity of the GB method results in rapid computation times in numerical calculations, and thus its extension to macromolecular and biological solvation problems is readily feasible. However, macromolecular energy calculations are typically based on empirical energy functions such as AMBER,[7] CHARMM,[8] or GROMOS,[9] each of which have their own prescription for specifying the net atomic charges on the individual atoms of the system. For a GB model to be successful on these systems, it is necessary that the effective GB radii for each atom type be parametrized in a manner fully consistent with the net atomic charges intrinsic to the assumed energy function. We describe herein the nature of the reparametrization process necessary to achieve this consonance and report the parameters required to obtain the effective Born radii compatible with the recently proposed force field of Cornell et al.,[10] incorporated into the AMBER suite of programs.[7] The reparametrized GB model is found to reproduce solvation free energies of 32 molecules, chosen as prototypes of protein and nucleic acid constituents, with a mean unsigned error of <1 kcal. Preliminary application of the method to treat the solvent dependent conformational preferences of a right-handed B-DNA double helix is described.

II. Background

The generalized Born model treats a molecule as a discrete set of overlapping charged spheres imbedded in a polarizable dielectric continuum. The defining equations of the Generalized Born theory are as follows:

$$G_{es} = 332\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{q_i q_j}{r_{ij}\epsilon} - 166\left(1-\frac{1}{\epsilon}\right)\sum_{i=1}^{n}\frac{q_i^2}{\alpha_i} \quad (1)$$

$$= 332\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{q_i q_j}{r_{ij}} + G_{pol} \quad (2)$$

$$G_{pol} = -166\left(1-\frac{1}{\epsilon}\right)\sum_{i=1}^{n-1}\sum_{j=1}^{n}\frac{q_i q_j}{f_{GB}} \quad (3)$$

$$f_{GB} = (r_{ij}^2 + \alpha_{ij}^2 e^{-D})^{0.5}; \quad \alpha_{ij} = (\alpha_i\alpha_j)^{0.5}; \quad D = r_{ij}^2/(4\alpha_{ij}^2) \quad (4)$$

$$G_{pol} = \underbrace{-166\left(1-\frac{1}{\epsilon}\right)\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{q_i q_j}{f_{GB}}}_{\text{shielding}} \underbrace{-166\left(1-\frac{1}{\epsilon}\right)\sum_{i=1}^{n}\frac{q_i^2}{\alpha_i}}_{\text{self}} \quad (5)$$

Equation 1 expresses the total electrostatic free energy G_{es} of a molecular system in kcal/mol as a sum of the Coulomb interaction energies between each pair of charges q_i and q_j separated by a distance r_{ij} in a solvent of dielectric constant ϵ (the first term) and the Born solvation (self) energy of each individual charge (the second term). The α_i in eq 1 are the Born radii, which are treated as disposable parameters. In eq 2, the free energy is rewritten as a sum of Coulomb interaction energies in a vacuum and polarization free energy G_{pol}. The GB polarization energy captures all the electrostatic effects due to solvent in one single term (eq 3), with a judicious choice of the effective distance parameter f_{GB} as provided by eq 4. The

* Author for correspondence. E-mail: dbeveridge@wesleyan.edu. Fax: (860) 685-2211. Tel.: (860) 685-2575.

† On leave from Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India.

10.1021/jp982007x CCC: $15.00 © 1998 American Chemical Society
Published on Web 11/03/1998

GB polarization energy is shown as a sum of solvent shielding and self-energy terms in eq 5. The agreement between the GB results and the experimental solvation energies for a wide range of molecules has been reported to be exceedingly good, due both to the choice of the functional form for f_{GB} and the calibration of the Born radii.[1−4] Thus the GB model provides an attractive and expeditious approach to calculate the total electrostatic energy and solvation energy of a molecule.

An inspection of the solvation and interaction energies in a prototype system of two charges embedded in a continuum solvent revealed earlier[6] that an accurate description of solvation energies did not guarantee a good description of intramolecular interactions. In particular, the performance of the GB model on pK_a shifts of dicarboxylic acids was noted to be below expectations which pointed to some inadequacies in the GB model per se. Self-terms when combined with shielding terms yield solvation energies and shielding terms when combined with vacuum Coulomb contributions give interaction energies.[6,11] Since vacuum Coulomb energies do not provide any leverage in calibrations once the geometry of the molecule and partial charge set is specified, both shielding and self-terms have to be of the right magnitude independently, for an accurate electrostatic treatment, not merely their algebraic sum. This led us to suggest the following modifications to the GB model.

$$f_{m1GB} = (r_{ij}^2 + \alpha_{ij}^2 e^{-D})^{0.5}; \quad D = r_{ij}^2/c\alpha_{ij}^2; \quad c = 2 \qquad (6)$$

$$f_{\mathrm{m2GB}} = f_{\mathrm{m1GB}}\{(\epsilon\gamma - \gamma)/(\epsilon\gamma - 1)\} \qquad (7)$$

where

$$\gamma = [1 - ((\epsilon - 4)/2)(\beta^2 + 2\beta + 2)\mathrm{e}^{-\beta}]; \quad \beta = (0.4r_{ij}+\alpha_{ij}) \qquad (8)$$

Equation 6 for effective distances does well on intramolecular interactions. Equation 7 shows a modest improvement[6] on both interactions and solvation over eq 4. The method incorporating the specifications of eq 7 is henceforth referred to as the "modified generalized Born" (MGB) method.

III. Calculations

The 32 molecules chosen as a parametrization set are listed in Figure 1. The electrostatic contributions to the solvation free energies are calculated based on eq 3. The input parameters are the charges q_i and q_j, the dielectric constant ϵ, the Cartesian coordinates of the atoms which specify the r_{ij}, and the effective Born radii α_i for each atom in the molecule. The α_i are computed following the procedure recommended by Hawkins, Cramer, and Truhlar,[4] which requires the van der Waals radii and a set of "screening parameters". The overall procedure adopted was as follows: (i) partial atomic charges of the 32 small molecules were derived using a procedure identical to that of Cornell et al.;[10] (ii) each atom in these molecules was identified with an appropriate parm94 AMBER atom type and the corresponding *r** values and van der Waals (vdW) radii assigned; (iii) the target values for electrostatic contribution to the solvation free energies were obtained via the finite difference Poisson−Boltzmann (FDPB) method;[5,12,13] (iv) the screening parameters required in the computation of effective Born radii were optimized. The details of each of the above steps are given below.

Derivation of Charges. An earlier FDPB characterization[14] of the solvation electrostatics of AMBER force field wherein Cornell et al. charges were assigned as closely as possible to a set of small molecules comprising amino acid side chains (i.e., without a rederivation of charges) led to a mean unsigned error of 2.97 kcal/mol, suggesting that a rederivation of charges as appropriate for small molecules might be necessary. Partial atomic charges were thus derived consistent with AMBER 4.1 charge derivation protocols. We started by creating ideal geometries which matched the equilibrium bond lengths and angles for the atom type combinations in the Cornell et al. force field.[10] Dihedrals were set to match the approximate global minimum geometry for each molecule. We then determined the HF/6-31G* electrostatic potential using the MK option in GAUSSIAN-94.[15] The resulting potential was used as input by the RESP program[16−18] for determining the point (partial atomic) charges which recreated the electrostatic potential. This procedure was repeated for each of the 32 molecules investigated. The resulting charges are reported in Figure 1.

Assignment of the van der Waals Radii. All of the molecules in the set examined correspond to amino acid side chains or nucleic acid components. This facilitated matching the atoms with their corresponding atom types in the Cornell et al. force field for proteins and nucleic acids. The atom types H, HO, HP, and HS, cases where the van der Waals radii are either zero or less than 1 Å in AMBER, have been reassigned an *r** value of 1.2347 Å for determining the effective Born radii. The AMBER *r** values of all atoms in a molecule are then converted to van der Waals radii via $r_{\mathrm{vdw}} = r^*/(2)^{1/6}$. The resultant r_{vdw} are multiplied by the scale factors in Table 1 to obtain effective vdW radii for GB calculations. We found this additional step of premultiplication with the scale factors necessary to improve the quality of the results. The optimization of the scale parameters was done together with the calibration of the screening parameters as described below.

The FDPB Computations. The electrostatic contribution to the solvation free energy of each molecule was calculated by taking the difference between the total energy of the system obtained with $\epsilon_{\mathrm{int}} = 2$; $\epsilon_{\mathrm{ext}} = 80$ and that with $\epsilon_{\mathrm{int}} = 2$; $\epsilon_{\mathrm{ext}} = 1$. The solvent probe radius was set at 1.4 Å and the resolution employed was 4 grids/Å.

Calibration of the Screening Parameters. The pairwise dielectric descreening procedure of Hawkins et al.[4] is employed with an initial screening parameter of unity for each of the six elements. The effective Born radius α_i in this method is given by the expression

$$\alpha_i^{-1} = \rho_i^{-1} - \left(\frac{1}{2}\right)\sum_j\left[\frac{1}{L_{ij}} - \frac{1}{U_{ij}} + \frac{r_{ij}}{4}\left(\frac{1}{U_{ij}^2} - \frac{1}{L_{ij}^2}\right) + \frac{1}{2r_{ij}}\ln\frac{L_{ij}}{U_{ij}} + \frac{\rho_j^2}{4r_{ij}}\left(\frac{1}{L_{ij}^2} - \frac{1}{U_{ij}^2}\right)\right] \qquad (8)$$

where

$$L_{ij} = 1 \quad \text{if} \quad r_{ij} + r_j \le r_i$$

$$L_{ij} = r_i \quad \text{if} \quad r_{ij} - r_j \le r_i < r_{ij} + r_j$$

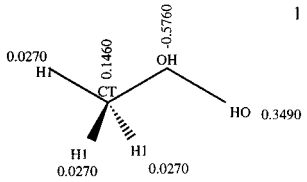
$$L_{ij} = r_{ij} - r_j \quad \text{if} \quad r_i \le r_{ij} - r_j$$

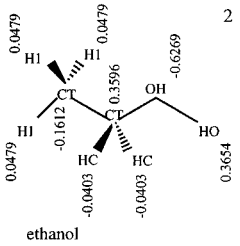
$$U_{ij} = 1 \quad \text{if} \quad r_{ij} + r_j \le r_i$$

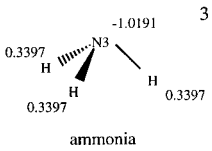
and

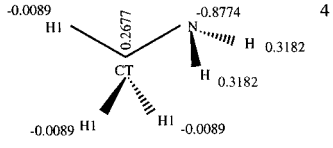
$$U_{ij} = r_{ij} + r_j \quad \text{if} \quad r_i < r_{ij} + r_j$$

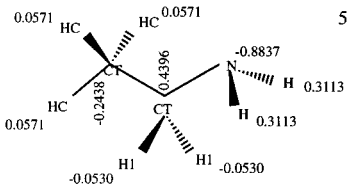
In the above equations, r_{ij} is the distance between atoms *i* and *j*, r_i is the van der Waals radius of atom i, and r_j is the product

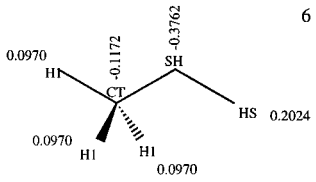

1
methanol

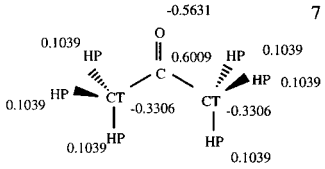

2
ethanol

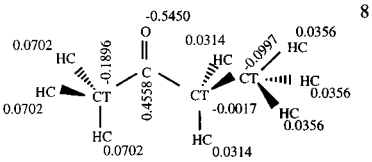

3
ammonia

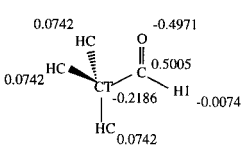

4
methylamine

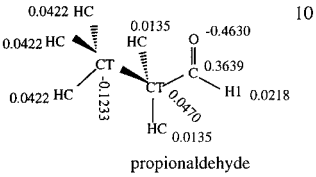

5
ethylamine

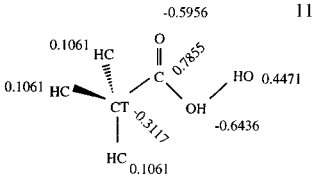

6
methylthiol

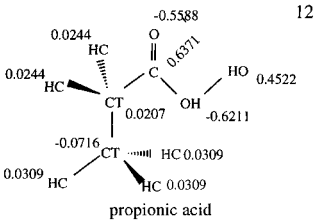

7
acetone

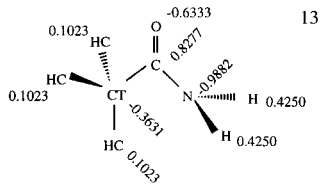

8
butanone

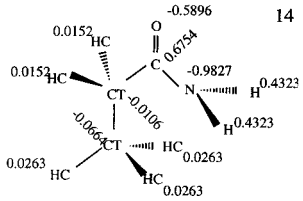

9
acetaldehyde

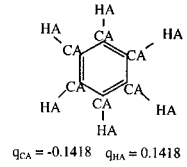

10
propionaldehyde

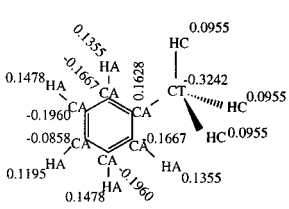

11
acetic acid

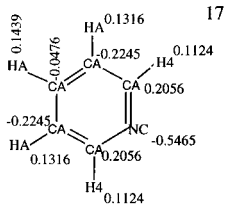

12
propionic acid

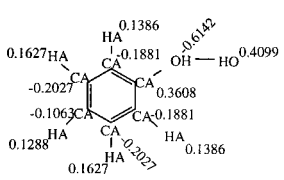

13
acetamide

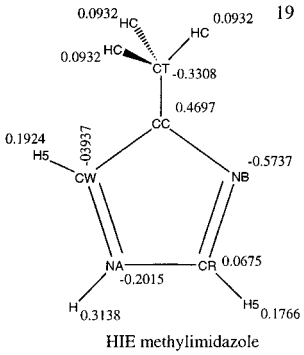

14
propionamide

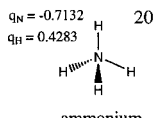

15
Benzene

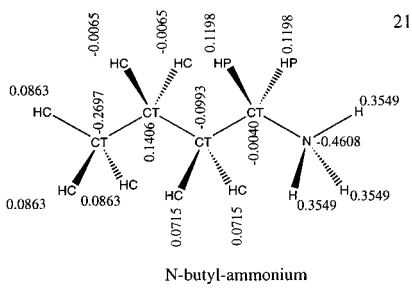

16
toluene


17
pyridine


18
phenol


19
HIE methylimidazole


20
ammonium


21
N-butyl-ammonium

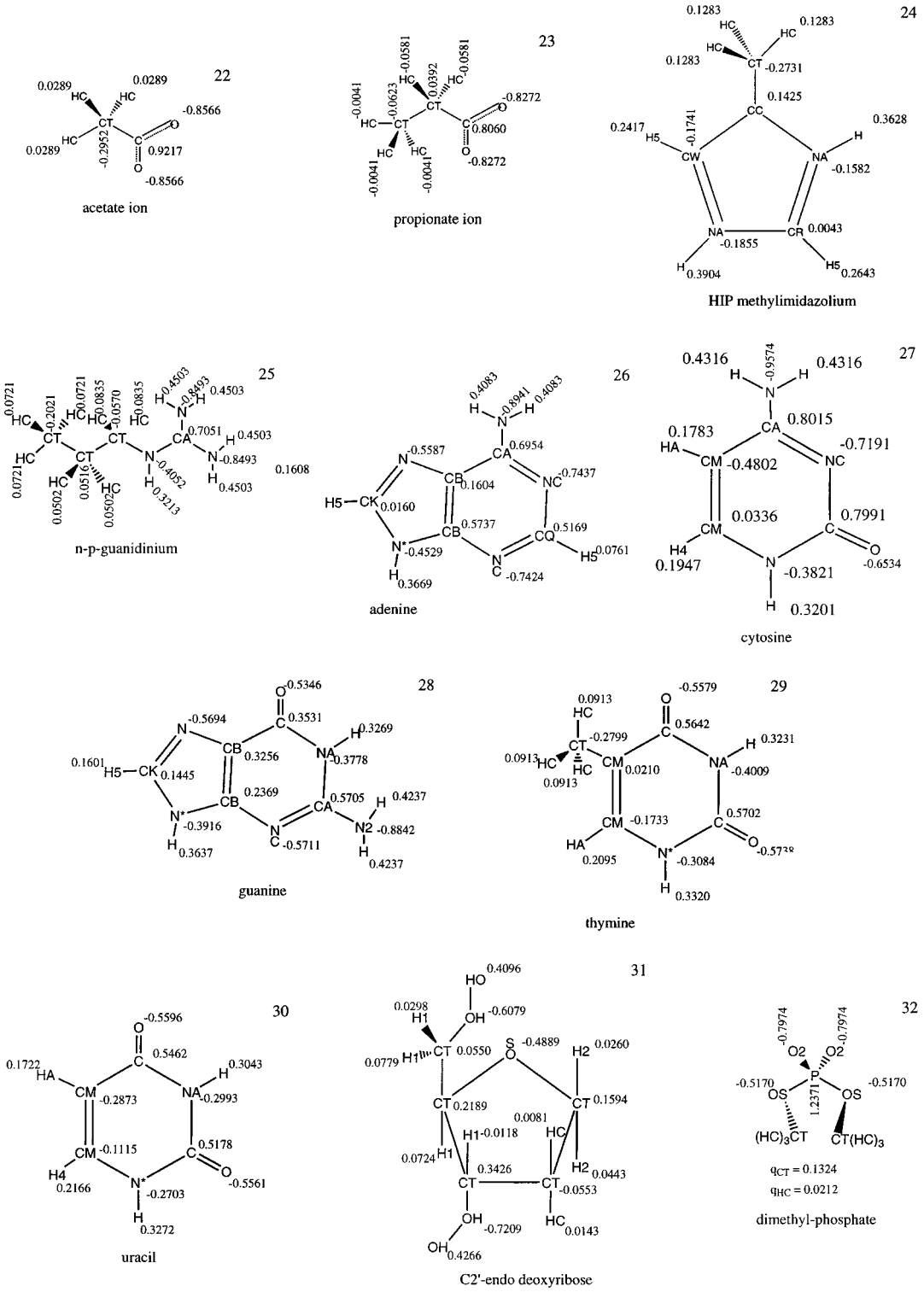


Figure 1. Computed partial atomic charges for the set of small molecules (constituents of proteins and nucleic acids) investigated.

of the van der Waals radius of atom *j* and the screening parameter of atom *i*. The screening parameters were introduced by Hawkins et al.[3,4] to correct for systematic errors introduced by the pairwise screening approximation.

The solvation free energies of all the 32 molecules are computed using eq 3 with the effective Born radii given by eq 9 and compared with the FDPB values. This procedure is carried out iteratively optimizing the screening and scale parameters via simulated annealing till the root-mean-square deviation between the GB results and the FDPB results is at a

TABLE 1: Scale Parameters for AMBER van der Waals Radii

	GB	MGB
H	0.909	0.9021
C	0.828	0.8742
O	0.864	0.9486
N	1.035	0.9486
S	0.900	0.930
P	0.900	0.930
Na^+ (IP)	0.899	0.899

TABLE 2: Screening Parameters

	GB	MGB
H	0.8461	0.8846
C	0.9615	0.9186
O	1.0088	0.8836
N	0.9343	0.8733
S	1.1733	0.9323
P	1.0700	0.9604
Na^+ (IP)	1.0000	1.0000

minimum. The resultant scaling and screening parameters are given in Tables 1 and 2.

Nonelectrostatic Contribution to the Solvation Free Energy. The surface area of each molecule is computed with a probe radius of 1.4 Å using ACCESS.[19] The nonelectrostatic contribution to the solvation free energy of each molecule is then estimated with a coefficient[1] of 7.2 cal/Å^2. Results on the net solvation free energies of the data set are presented in Table 3.

Results and Discussion

The results of the calibration are collected in Table 3. The GB methodology incurs a mean unsigned error less than 1 kcal/mol, and the error on the total solvation energies of all the systems considered is almost negligible. Note, however, that the error may be larger on an individual basis, which suggests that some further compensatory effects arise in the GB theory. The resulting parameters are suitable for use in calculations of MGB solvation energies, which may then be combined with corresponding estimates of the intramolecular free energy contributions in studies of solvent effects on structure and conformation. A particular application we[20] and others[21] are pursuing is the use of MGB in lieu of computationally intensive free energy simulations to estimate the free energies of macromolecules with structures obtained from MD simulations.

As an example, we have employed the GB methodology (MGB) for solvation free energies together with entropies estimated in the quasiharmonic approximation[22] and the intramolecular energetics provided by the Cornell et al. force field in a post facto analysis of four molecular dynamics trajectories on A and B forms of DNA.[20,23−24] These MD simulations were carried out in water and in 85% ethanol solutions with explicit counterions using AMBER 4.1 protocols. Relative to the B form of DNA in water, the computed free energy differences were +359 kcal for the A form in water, +135 kcal for the B form in 85% ethanol and +122 kcal for the A form in 85% ethanol. The results clearly indicate that the B form is more stable in water than in 85% ethanol and that the A form is more stable in 85% ethanol than in water consistent with experiment.[25] Here the calculations reproduce the experimental trends in a complex system and serve to identify the major contributing components to an analysis of the molecular thermodynamics of DNA conformational preferences.

In a recent paper, Reddy et al.[26] applied generalized Born parameters derived for the OPLS−SA[27] united atom force with other charge sets from a variety of common force fields and for charge sets derived from common and respected quantum mechanical calculation procedures. Though all methods tested showed reasonable correlation with experimental solvation

TABLE 3: Calculated Solvation Free Energies[a] (in kcal/mol)

system	experiment[b]	FDPB(AMBER)	GB(AMBER)	MGB(AMBER)
1. methanol	−5.08	−3.96	−4.79	−3.90
2. ethanol	−4.90	−3.12	−3.46	−3.05
3. ammonia	−4.31	−6.12	−6.14	−7.94
4. methylamine	−4.57	−3.24	−2.57	−3.23
5. ethylamine	−4.50	−2.16	−1.17	−2.16
6. methylthiol	−1.24	−2.03	−2.04	−2.02
7. acetone	−3.85	−5.74	−6.59	−6.29
8. 2-butanone	−3.64	−5.07	−5.79	−5.30
9. acetaldehyde	−3.50	−4.21	−5.19	−4.53
10. propionaldehyde	−3.44	−3.23	−3.86	−3.32
11. acetic acid	−6.70	−7.33	−9.32	−7.92
12. propionic acid	−6.47	−6.16	−7.99	−6.51
13. acetamide	−9.72	−9.09	−9.23	−10.80
14. propionamide	−9.42	−7.97	−8.17	−9.44
15. benzene	−0.87	−1.17	−0.84	−0.32
16. toluene	−0.76	+0.07	+0.16	+0.35
17. pyridine	−4.70	−3.44	−2.98	−2.82
18. phenol	−6.62	−4.13	−4.82	−3.92
19. methyl imidazole	−10.25	−6.68	−6.68	−7.06
20. ammonium	−81.53	−95.54	−94.77	−95.56
21. *N*-butylammonium	−69.24	−66.52	−69.05	−66.97
22. acetate ion	−80.65	−81.11	−80.75	−80.75
23. propionate ion	−79.12	−77.48	−77.20	−77.47
24. methyl imidazolium	−64.13	−63.42	−61.23	−58.78
Total	−469.21	−469.51	−475.81	−470.84
%Error on Total	0.0	0.08	1.12	0.02
25. *N*-*p*-guanidinium		−59.48	−59.48	−59.37
26. adenine		−12.03	−11.77	−11.21
27. cytosine		−18.50	−18.89	−21.53
28. guanine		−21.34	−21.68	−21.33
29. thymine		−13.11	−12.60	−13.03
30. uracil		−14.87	−15.01	−14.90
31. C2′ endo D-ribose		−8.19	−6.27	−6.22
32. dimethyl phosphate anion		−74.81	−74.52	−74.90
Mean unsigned error		0.0	0.68	0.66

[a] Includes nonelectrostatic contributions. [b] From Sitkoff et al. (ref 12).

energy measures, they noted that best results with the use of the OPLSA force field. This result indicates that some transferability exist for GB parameters, at least for united atom force fields, but that it is best to stay with the charge set for which the parameters were developed. Our approach in this paper takes the effective GB radii for each atom parametrized in a manner fully consistent with the net atomic charges intrinsic to the assumed energy function, AMBER parm94.

Conclusions

In this article, we determine GB parameters to be used along with the AMBER force field for estimating solvation free energies. The methodology and the parameters described offer a powerful tool to carry out a free energy component analysis on matters related to stability and feasibility in large biochemical systems at little extra computational expense. A recent application to the conformational preferences of A- and B-DNA in water and 85% ethanol solutions demonstrates that the force field, the parameters, and the methodology support the trends observed in experimental data, indicating that the applications of the method immediately under consideration are viable.

Acknowledgment. This project is funded by Grant GM 37909 from the National Institutes of Health. D.S. thanks NIH for NRSA fellowship Grant GM-18117. Supercomputer resources provided by FBSC are gratefully acknowledged.

References and Notes

(1) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc*. **1990**, *112*, 6127.

(2) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213.

(3) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122.

(4) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem*. **1996**, *100*, 19824.

(5) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144 and references therein.

(6) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys*. **1998**. In press.

(7) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; Fergusen, D. M.; Siebel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 4.1*; University of San Francisco: San Francisco, CA, 1995.

(8) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(9) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) System*; University Groningen: The Netherlands, 1987.

(10) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(11) Nakamura, H. *Quart. Rev. Biophys*. **1996**, *29*, 1.

(12) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem*. **1994**, *98*, 1978.

(13) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins*: *Struct. Funct. Genet*. **1988**, *4*, 7.

(14) Dixit, S. B.; Bhasin, R.; Rajasekaran, E.; Jayaram, B. *J. Chem. Soc., Faraday Trans*. **1997**, *93*, 1105−1113.

(15) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pope, J. A. GAUSSIAN-94; Gaussian, Inc.: Pittsburgh, PA, 1995.

(16) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem*. **1993**, *97*, 10269.

(17) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem*. **1995**, *16*, 1357.

(18) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.

(19) Lee, B.; Richards, F. M. *J. Mol. Biol*. **1971**, *55*, 379.

(20) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Am. Chem. Soc.* **1998**. In press.

(21) Srinivasan, J.; Case, D. **1997**. Private communication.

(22) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325.

(23) Young, M. A.; Ravishanker, G.; Beveridge, D. L. *Biophys. J*. **1997**, *73*, 2313.

(24) Sprous, D.; Young, M. A.; Beveridge, D. L.; *J. Phys. Chem*. **1998**. In press.

(25) Ivanov, V. I.; Krylov, D. Y. *Methods Enzymol.* **1992**, *211*, 111 and references therein.

(26) Reddy, M. R.; Erion, M. D.; Agarwal, A.; Viswanadhan, V. N.; McDonald, D. Q.; Still, W. C. *J. Comput. Chem.* **1998**, *19*, 769.

(27) Jorgenson, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc*. **1988**, *110*, 1657.