JCTC Journal of Chemical Theory and Computation

Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model

Devleena Shivakumar,[†,§] Yuqing Deng,[†,‡,||] and Benoît Roux*[,†,‡]

Department of Biochemistry & Molecular Biology, University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, and Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439

Received October 20, 2008

Abstract: Accurate determination of absolute solvation free energy plays a critical role in numerous areas of biomolecular modeling and drug discovery. A quantitative representation of ligand and receptor desolvation, in particular, is an essential component of current docking and scoring methods. Furthermore, the partitioning of a drug between aqueous and nonpolar solvents is one of the important factors considered in pharmacokinetics. In this study, the absolute hydration free energy for a set of 239 neutral ligands spanning diverse chemical functional groups commonly found in drugs and drug-like candidates is calculated using the molecular dynamics free energy perturbation method (FEP/MD) with explicit water molecules, and compared to experimental data as well as its counterparts obtained using implicit solvent models. The hydration free energies are calculated from explicit solvent simulations using a staged FEP procedure permitting a separation of the total free energy into polar and nonpolar contributions. The nonpolar component is further decomposed into attractive (dispersive) and repulsive (cavity) components using the Weeks−Chandler−Anderson (WCA) separation scheme. To increase the computational efficiency, all of the FEP/MD simulations are generated using a mixed explicit/implicit solvent scheme with a relatively small number of explicit TIP3P water molecules, in which the influence of the remaining bulk is incorporated via the spherical solvent boundary potential (SSBP). The performances of two fixed-charge force fields designed for small organic molecules, the General Amber force field (GAFF), and the all-atom CHARMm-MSI, are compared. Because of the crucial role of electrostatics in solvation free energy, the results from various commonly used charge generation models based on the semiempirical (AM1-BCC) and QM calculations [charge fitting using ChelpG and RESP] are compared. In addition, the solvation free energies of the test set are also calculated using Poisson−Boltzmann (PB) and Generalized Born model of solvation (GB), which are two widely used continuum electrostatic implicit solvent models. The protocol for running the absolute solvation free energy calculations used throughout is automated as much as possible, with minimum user intervention, so that it can be used in large-scale analysis and force field optimization.

1. Introduction

An accurate determination of solvation free energy of a molecule plays a critical role in numerous areas of biomedical, chemical, and industrial research, and its evaluation is a long-standing and constantly evolving challenge in computational chemistry.[1,2] For example, estimating the desolvation

* Corresponding author phone: (773) 834-3557; fax: (773) 702-0439; e-mail: roux@uchicago.edu.

[†] University of Chicago.

[‡] Argonne National Laboratory.

[§] Current address: Schrodinger, Inc., 120 West 45th Street, 29th Floor, New York, NY 10036-4041.

[||] Current address: Zymeworks Inc., 540-1385 West Eighth Ave., Vancouver, BC V6H 3V9, Canada.

10.1021/ct800445x CCC: $40.75
Published on Web 03/24/2009

penalty for a ligand and its receptor upon formation of a bound complex is an important component in drug design.[2] Another example is in the area of pharmacokinetics, where partitioning between various solvents (the so-called "log *P*") and acid−base equilibrium is useful in determining absorption properties.[3–5] Solvation free energies are highly sensitive to the details of the molecular mechanical force field parameters used to describe the solvent and the solute molecules, such as atomic partial charges, Lennard-Jones radii, and the well depth. The calculation of the absolute solvation free energy is, thus, a critical test of the accuracy of a force field, and the underlying method to represent the interaction between the solute and the solvent. Because absolute solvation free energies have been experimentally determined for several small molecules, it provides a direct comparison between the experimental and calculated values.

Despite its importance, the calculation of "absolute" solvation free energies presents a challenge to computational chemists. Both the dynamic and the thermodynamic properties of a molecular solute are strongly influenced by the microscopic structure and organization of the solvent molecules in its surrounding. The current methods to treat solvation generally follow one of two approaches. The first one involves the explicit simulation of a large number of solvent molecules with periodic boundary conditions.[6] Alternatively, the effect of solvation can be incorporated implicitly, via continuum approximations.[7] While molecular dynamics free energy perturbation (FEP/MD) with explicit solvent molecules provides the most realistic, and arguably most accurate, treatment of solvation,[8] the approach is often limited in practice by the computational cost.[9] On the other hand, there is a vast range of available implicit solvent methods differing widely in sophistication, computational cost, and accuracy.[7] The family of methods constructed on the basis of continuum electrostatic solvation models are particularly appealing because of their ability to accurately reproduce solvation in polar liquids while remaining computationally inexpensive.[10–13]

Several groups have shown that the Poisson−Boltzmann (PB) models[14,15] are capable of reproducing polar explicit solvent forces between solute and solvent.[16,17] The performance of continuum electrostatic models is highly dependent on the input parameters, such as the effective atomic Born radii.[11,18–21] A good agreement with PB has been observed for the electrostatic or polar component of the solvation energy using an optimized set of radii.[18,21,22] In implicit solvation models based on continuum electrostatics, the free energy to insert the uncharged solute into the solvent, the so-called nonpolar contribution, must be treated separately. Such free energy contribution to solvation has been traditionally approximated from models based on the solvent-accessible surface area (SASA).[22–25] However, more sophisticated statistical mechanical treatments have shown that the cavity creation term is dependent on both the SASA and the solvent-accessible volume term, with a crossover at large solute sizes (>10 Å).[26,27] Furthermore, decomposition of the nonpolar solvation free energy in simulation with explicit water molecules indicates that the effects from the repulsive and the dispersive van der Waals interactions are both considerable and of opposite sign.[28] In fact, not accounting for the attractive van der Waals interactions between solvent and solute atoms in the SASA-based model introduces important inaccuracies,[29–32] and efforts have been made to design reasonably accurate and computationally tractable approximations to the nonpolar free energy contribution. Along those lines, Tan et al. have recently developed an implicit solvent model methodology treating explicitly the nonpolar solvation energy corresponding to the cavity creation and the attractive dispersion.[33] This implicit solvation model was implemented in the Amber simulation program (Amber9 and newer versions).

In the present study, we compute the absolute hydration free energy using FEP/MD with explicit water molecules for a set of 239 small molecules that are representative of the chemical functionalities found in drug design. The set includes saturated hydrocarbons, unsaturated hydrocarbons, strained systems, conjugated and aromatic systems, all major polar functional groups, as well as heterocyclic systems. The FEP/MD simulations with explicit solvent molecules are based on the TIP3P water model.[34] To increase the computational efficiency, the FEP/MD simulations are generated using the Spherical Solvent Boundary Potential (SSBP) approach,[35] which consists of keeping a small number of explicit solvent molecules in the vicinity of the solute, and representing the remaining bulk implicitly via an effective solvent boundary potential. Two widely used fixed charge general force fields are considered to model the small molecules. The first is the generalized Amber force field (GAFF)[36] established recently, and the second is the CHARMm-MSI force field,[37] which was developed almost a decade earlier. Both are biomolecular force fields constructed from simple analytical functions. Because they share a similar functional form, a comparison of the calculated absolute solvation free energy using these two force fields is meaningful and of general interest.

For the same test set of small molecules and force field parameters, we compare the FEP/MD results to the free energies obtained from two commonly used implicit solvent models, PB/SA and GB/SA.[38] For the purpose of this comparison, the total solvation free energy is separated into polar (electrostatics) and nonpolar contributions to provide more insight, following a step-by-step reversible work to materialize the solute in solution.[7,39] The nonpolar component obtained in FEP/MD is further decomposed into purely repulsive and dispersive terms using the separation scheme for the Lennard-Jones 6-12 pair potential introduced in the WCA theory[40] and implemented in an earlier study of side chain analogues.[28] Correspondingly, the nonpolar component obtained with the implicit solvation models is separated into a repulsive and an attractive component using an optimized continuum decomposition.[33] Although such a decomposition scheme of the solvation free energy is path dependent by definition,[41] the analysis based on a specific step-by-step reversible work provides a useful framework for dissecting the free energy because each individual contribution can be associated with a microscopic process with a clear and well-defined physical meaning.[7] Here, the decomposition of the solvation free energy allows a comparison of the various

Table 1. Table Showing the Chemical Functional Group and Number of Small Molecules from Each Class Used in the Present Study

type	number	type	number
alkanes	8	esters	15
alkenes	10	ethers	11
alkynes	5	halogen, bromo	10
alcohols	17	halogen, chloroalkanes	11
aldehydes	6	halogen, chloroalkenes	5
aliphatic amines	16	halogen, chloroarenes	3
amides	5	halogen, flouro	6
arenes	14	halogen, iodo	8
aromatic amines	14	bifunamine	3
ketones	12	multiple halogens	15
bifunctional	5	nitriles	5
branched alkanes	7	cycloalkanes	5
carboxylic	5	sulfides	5
nitro	7		
disulfides	2		
total	239		

contributions terms from explicit as well as implicit solvent methods. Furthermore, decomposition of nonpolar solvation energy into repulsive and attractive contributions is advantageous to better understand the factors governing hydrophobic solvation and protein stability. In particular, it is shown that the nonpolar component of the free energy correlates strongly with log *P* and SASA values. All of the computational methodologies and force field parameters are described in the next section.

2. Method and Force Fields

a. Test Set. The absolute free energy of solvation for 239 neutral molecules (Table 1) was computed using three separate methods: molecular dynamics free energy perturbation (FEP/MD) simulation in explicit TIP3P solvent molecules,[34] finite-difference Poisson−Boltzmann (PB) continuum electrostatic,[10] and Generalized Born plus a surface area term (GB/SA).[11] This set of methods enables us to evaluate the accuracy of the commonly used point-charge models and force fields used in molecular modeling of small molecules. The test set consisting of 239 molecules was curated from earlier work in the literature.[42,43] The initial structures were obtained from the NIST Chemistry WebBook database. The molecules chosen for this comparative study have diverse chemical functional groups commonly encountered in drug design. These include saturated hydrocarbons, unsaturated hydrocarbons, conjugated and aromatic systems, all major polar functional groups, as well as heterocyclic and ionic systems. The diverse chemical functionalities include, alkanes, alkenes, alkynes, branched alkanes, alcohols, aldehydes, ketones, esters, carboxylic acid, cycloalkanes, arenas, aliphatic amines, aromatic amines, amides, bifunctional amines, ammonia, nitriles, nitro, thiols, sulfides, disulphides, ethers, fluoro, chloroalkanes, chloroalkenes, chloroarenes, bromo, iodo, multiple halogens, and bifunctional groups. A table summarizing the list of compounds under each category and their experimental solvation energy is shown in Supporting Information Table T1.

b. The Force Fields. The GAFF (Generalized Amber Force Field)[36] and the CHARMm-MSI force field[37] considered for the current study share a similar functional form:

$$V = \sum_{\text{bonds}} k_b(b - b_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\varphi(1 + \cos(n\varphi - \delta)) + \sum_{\text{nonbonded}} \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon r_{ij}} \quad (1)$$

Here, the first term accounts for the bond stretching energy when the atom moved $(b - b_0)$ from equilibrium. The second term in the potential energy function accounts for bond angle bending from equilibrium $(\theta - \theta_0)$. The third term corresponds to the twisting in the dihedral or the torsional angle with multiplicity n, and phase shift δ. The force constants for bond length, bond angle, and torsional angle are k_b, k_θ, and k_ϕ, respectively. The last term represents the standard 12-6 Lennard-Jones potential used to calculate the nonbonded interactions between pairs of atoms (i, j), where ε and $R_{\min}$ represent the well depth and radius parameters of the potential, respectively.

The ligand structures were read into the Antechamber program to generate the individual parameter and topology files based on the general Amber force field (GAFF). In addition, *tleap* program (Amber9 version) was used to generate topology and parameter file in Amber format for the implicit solvent simulations. The CHARMM format of the GAFF parameter file was used to generate the input topology and parameter file for small molecules compatible with the CHARMM program format. These files were used in all of the explicit solvent simulations using the CHARMM program. The GAFF potential function consists of 35 basic and 22 special atom types based on element type, hybridization, aromaticity, and chemical environment. The combination of basic and special atom types covers almost all of the organic chemical space that is made up of C, N, O, S, P, H, F, Cl, Br, and I atoms, making it a complete force field to study pharmaceutical ligand phase space. The van der Waals parameters of GAFF are the same as those used by the traditional Amber force field.[36] Also, GAFF uses a simple additive harmonic energy function consisting of bond, angle, dihedral, and Lennard-Jones 12-6 potential term as the traditional CHARMM and Amber force fields as described above, making it compatible in a broader perspective with these force fields. Both GAFF and CHARMm-MSI are complete force fields in the sense that they aim at covering most small organic molecules. GAFF can be assigned to a wide range of molecules in automatic fashion using the Antechamber package,[44] while the CHARMm-MSI force field can be automatically assigned to small molecules using the InsightII program (Accelrys). This makes both force fields practical to be applied to a large number of ligands in a database screening in docking studies.

The CHARMm-MSI force field parameters were assigned to the 239 small molecules in our test set using the InsightII program by specifying the force field file, CHARMM.cfrc (Insight II Version 2000 − Molecular Modeling System).

Perl scripts were used to automate the input file generation step with the InsightII program.

c. Charge Methods. The charge method plays a crucial role in any molecular mechanics calculations. While the set of atomic partial charges is closely associated with the rest of the force field, combining the nonelectrostatic part of a force field with sets of atomic partial charges derived from a variety of approaches is quite common when considering large databases of ligands. For the sake of clarity, we will be implicitly referring only to the nonelectrostatic part of the GAFF and CHARMm-MSI potential function when invoking those names hereon. Because the charges have a very large impact on the resulting solvation energies,[28,45,46] it is often worthwhile considering different strategies to generate them. In Amber's ff99 and ff94, the charges are derived by using the HF/6-31G* method. The idea behind using this as a preferred method is that it exaggerates the dipole moment of most of the residues by 10−20%, which in turn mimics the polarization in the condensed phase simulations. The CHARMM22 charges were obtained by fitting of solute−water dimer energies from SCF/6-31G*. In the present study, we compare three different charge models: two methods based on fitting partial charges to the electrostatic potential, ChelpG[47] and restrained electrostatic potential fit (RESP),[48] and a third method based on semiempirical AM1-BCC model.[49,50] The AM1-BCC method is a computationally inexpensive method used to generate high-quality atomic charges for condensed phase simulations. The AM1 atomic charges are "population" quantities based on occupation of the molecular orbital and do not reproduce the electrostatic potential (ESP) of a molecule. The simple additive bond charge corrections (BCC) are added to these AM1 atomic charges to generate the AM1-BCC charges, which better reproduce the ESP. The parametrizations of the BCC were carried out by Bayley and co-workers by fitting to the HF/6-31G* ESP of a training set of >2700 molecules.[50] Henceforth, the models considered will be GAFF/AM1-BCC, GAFF/ChelpG, GAFF/RESP, CHARMm-MSI/AM1-BCC, CHARMm-MSI/ChelpG, and CHARMm-MSI/RESP. The performance of three commonly used charge methods was compared in this work. They are ab initio methods such as RESP and ChelpG and a semiempirical method, AM1-BCC. For each molecule, these three methods are used to derive a set of atom-centered point charges.

AM1-BCC charges for our test set containing 239 molecules were obtained using the Antechamber package (ver 2.7). The Antechamber and mopac programs were used to calculate the AM1-Mulliken population charges for each small molecule. The am1bcc program was used to assign atom types and bond types according to AM1-BCC definitions, and then assign the BCC values for the atoms.

ChelpG (charges from electrostatic potentials using a grid) charges were obtained using the Gaussian program.[51] In this method, the atomic charges were fitted to reproduce the molecular electrostatic potential (MEP) at a number of grid points around the molecule. After the MEP was evaluated at all valid grid points, atomic charges were derived that reproduce the MEP in the most optimum way. The only additional constraint in the fitting procedure was that the sum of all atomic charges equals that of the overall charge of the system.

RESP (restrained electrostatic potential fit) charges are based on a restrained least-squares fitting algorithm to best reproduce the quantum mechanical electrostatic potential of a molecule. The potential was evaluated at a large number of points defined by four shells of surfaces at 1.4, 1.6, 1.8, and 2.0 times the VDW radii. These distances have been shown to be appropriate for deriving charges that reproduce typical intermolecular interactions (energies and distances). Also, during the RESP charge fitting, the equivalent atoms were made to bear the same charges because equivalent atoms are indistinguishable during the molecular dynamical simulations. The value of the electrostatic potential at each grid point was calculated from the quantum mechanical wave function using the Gaussian 03 program.[51] The charges derived using this procedure are basis set dependent. The electrostatic potential for the small molecules in our test set was calculated at the HF/6-31G* level of theory. The 3-21G* basis set was used for the iodine atom. The Respgen utility of the Antechamber package was used to generate the input files for two-stage fitting. The partial atomic charges were fitted to the electrostatic potential using the RESP program and Antechamber package.

d. Generating Input Files. Because our ultimate goal is the practical application of these methods for screening databases containing multiple small molecules used in protein−ligand docking studies, emphasis was given to making the input file generation steps automated with the help of programs such as Antechamber and InsightII, and simple perl scripts. To efficiently handle hundreds of small molecules in any molecular mechanics calculations, one needs to automatically assign the parameters and generate the residue topology file rather than hand-editing individual molecule. The stand-alone version of Antechamber program v2.7 serves this purpose by automatically perceiving atom types and assigning bond type and bond order from the three-dimensional geometry of the molecule.[44] However, it should be kept in mind that when using Antechamber, the user has to provide the total charge of the molecule for accurate charge calculation; otherwise the program perceives the default total charge of 0. Because in our case only neutral molecules were used, we used the default total charge. The Antechamber program was used to generate the input topology and parameter files for both the Amber and the CHARMM programs when using GAFF. The InsightII program and perl scripts were used to automate the input file preparatory step when the CHARMm-MSI force field was used. All of the explicit solvent simulations were set up and performed with the CHARMM program in an automated manner using a suite of perl scripts, keeping in mind the broader application of this method in screening large databases of ligands. The only input that was supplied was a coordinate file for the small molecule, for example, mol2 or PDB format.

e. Solvation Free Energy Protocols. The absolute free energy of solvation for the small molecules was calculated using FEP/MD with explicit solvent molecules. The FEP/

MD simulations were carried out using the PERT module of the program CHARMM v31.[52] Three staging or coupling parameters, s, ξ, and λ, were used to separate the absolute solvation free energy in terms of its polar and nonpolar components.[28] The latter was further decomposed into purely repulsive and attractive terms using the WCA scheme. The free energy contribution from the core repulsion was calculated by setting the staging parameter s to 0.0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. For each value of s, two trajectories of 70 ps were run for both the initial and the final states. Only the last 60 ps of each trajectory was used in the free energy calculations. The free energy contributions from electrostatics and from the van der Waals dispersion interactions were both calculated using a standard linear coupling scheme. For the dispersive attraction, the coupling parameter ξ was set to 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. For the electrostatic free energy contribution, the coupling parameter λ was set to 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. In both cases, interaction energy samples were collected from the last 20 ps of two 30 ps trajectories, and the samples were processed by the weighted histogram analysis method (WHAM).[53] All free energies were calculated with 400 explicit TIP3P water molecules, with the influence of the remaining bulk being incorporated via the SSBP.[35] The system temperature was held fixed at 300 K with Langevin dynamics. A friction constant of 5 ps^{-1} was applied on the oxygen atoms of all of the water molecules. The integration time step of the dynamics was 2 fs. All bonds containing hydrogen atom were fixed with SHAKE constraints. The center of mass of the solute was constrained to the center of the solvent sphere with a harmonic potential with a force constant of 1.0 kcal $Å^{-2}$ mol^{-1}.

The solvation energies for the 239 small molecules were obtained using two implicit solvent methods, the GB/SA model in Nucleic Acid Builder (NAB)[54] and the PB. The NAB molecular modeling package was used to perform the MM-GB/SA calculations with the Amber force field. NAB is a high-level language that can perform the force field calculations, such as molecular dynamics and minimization similar to the Amber program.[54] The NAB and Amber program have the same GB implementation. The ΔG_{polar} was obtained using the GB method as implemented in the NAB program using a dielectric constant of 80 for the water phase. Each atom in a molecule was represented as a sphere of radius R_i with a charge q_i at its center; the interior of the atom was assumed to be filled uniformly with a material of dielectric constant 1. The Born radii were computed according to the method of Onuchic, Bash, and Case or the "OBC" model (implemented as igb = 5 in NAB, version 5).[55] The effective Born radius of an atom reflects the degree of its burial inside the molecule. In the "OBC" model, the effective radii are rescaled using parameters proportional to the degree of the atom's burial.

The $\Delta G_{nonpolar}$ contribution to the total solvation free energy was estimated using a term proportional to the total solvent accessible surface area (SASA) of the molecule, with proportionality constant derived from experimental solvation energies of small nonpolar molecules. A fast LCPO algorithm[56] was used to compute an analytical approximation to the SASA of the molecule.

The PB continuum solvent method with new and improved nonpolar solvation model in Amber9 was used in this study. This implicit solvent model decomposes the total solvation free energy into polar and nonpolar. The polar component of the total solvation free energy was obtained by using a grid-based finite difference solution to the PB equation with zero salt concentration and modified Bondi radii (corresponding to mbondi2 in Amber) for small molecules.[55] A grid spacing of 0.2 was used for generating the grids. The $\Delta G_{nonpolar}$ was obtained as a combination of two terms representing separate cavity or repulsive term and a dispersive or attractive term using an optimized 6-12 decomposition scheme using the methodology implemented as igb = 10 in Amber9.[33]

For the continuum solvent calculations, the starting structures of the small molecules were subjected to 500 steps each of SD and ABNR energy minimizations, followed by 1 ns of MD simulations using GB (igb = 5 in Amber). The PB was performed on the final snapshot of the equilibrated structure. As a comparison, we also calculated the PB electrostatic energy using PBEQ solver[18,21,57] in the CHARMM program using the same set of radii (mbondi2 in Amber) and PB parameters. The effect of conformational fluctuations on the magnitude of the solvation free energy was examined in the case of four small molecules: methyl hexanoate, methyl propyl ether, octanal, and 5-nonanone. Those molecules were chosen because they have the highest number of the rotatable bonds and are among the most flexible in the test set of 239 molecules studied here. The MD simulations for 10 and 20 ns duration for each of the molecule in vacuum were carried out to generate an ensemble of conformations using both the CHARMM and the NAB programs. The solvation free energy for the snapshots was then calculated perturbatively for the implicit solvent methods described above, PB/CHARMM and GBSA/NAB.

3. Results and Discussion

a. Comparison of the Charge Methods. For each set of Lennard-Jones parameters from GAFF and CHARMm-MSI force fields, we calculate the absolute solvation free energies using atomic partial charges obtained from three different methods. The first two methods are based on the electrostatic potential from the quantum mechanics (QM): ChelpG and RESP. The third method, AM1-BCC, is based on a semiempirical AM1 determination of charge followed by bond charge correction (BCC). The correlation plot between the experimental (ΔG_{exp}) and calculated (ΔG_{calc}) absolute solvation free energies is shown in Figure 1. When using GAFF/RESP and GAFF/ChelpG, the correlation coefficient R^2 between the absolute ΔG_{exp} and ΔG_{calc} corresponds to 0.84 and 0.80, respectively. Thus, the two QM methods show good overall performance in terms of reproducing the experimental results, the GAFF/RESP being slightly better than the GAFF/ChelpG. The major outliers using the GAFF/RESP model are (the error in absolute solvation free energy is shown in parentheses) 1,4-dimethylpiperazine (5.1), aze-

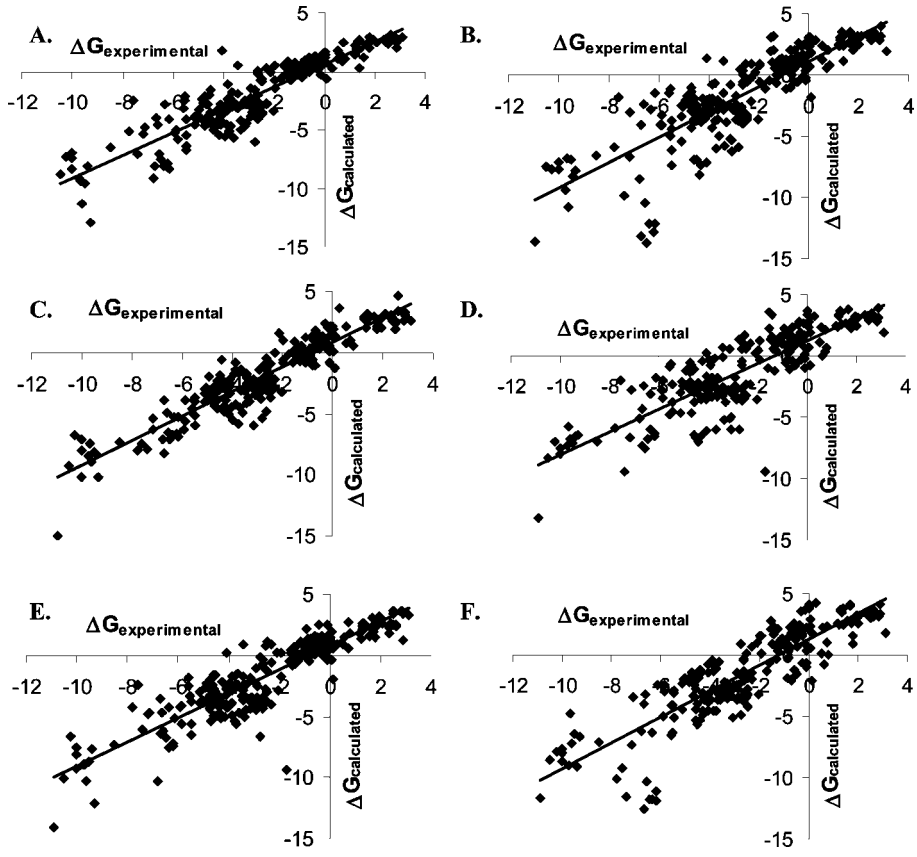

Figure 1. Plot showing correlation between experimental (*x*-axis; $\Delta G_{\text{experimental}}$, kcal/mol) and calculated (*y*-axis; $\Delta G_{\text{calculated}}$, kcal/mol) absolute solvation free energies obtained using explicit solvent simulations on 241 neutral molecules. Absolute solvation free energy calculated using the following models: GAFF/RESP ($R^2 = 0.84$) (A), CHARMm-MSI/RESP ($R^2 = 0.72$) (B), GAFF/ChelpG ($R^2 = 0.80$) (C), CHARMm-MSI/ChelpG ($R^2 = 0.72$) (D), GAFF/AM1-BCC ($R^2 = 0.87$) (E), and CHARMm-MSI/AM1-BCC ($R^2 = 0.76$) (F). The energies are in kcal/mol.

tidine (3.4), 1,2-ethanediol (−3.6), *N*-methylacetamide (3.0), 2-ethylpyrazine (3.1), propylaminoformamidine (−5.2), and methyl formate (−3.3). Upon discarding these top outliers, the R^2 between the absolute solvation free energy (ΔG_{exp} and ΔG_{calc}) improved from 0.82 to 0.88. On the other hand, the semiempirical charge AM1-BCC method is able to reproduce the absolute solvation free energies for our test set with a higher correlation coefficient ($R^2 = 0.87$). The major outlier with the GAFF/AM1-BCC method was 1-heptanol (3.8). When using the CHARMm-MSI Lennard-Jones parameters, we see that the overall correlation between the absolute ΔG_{exp} and ΔG_{calc} is worse for all three charge models, CHARMm-MSI/AM1-BCC, CHARMm-MSI/RESP, and CHARMm-MSI/CHelpG. However, a similar trend was observed between the performances of the three charge models. The R^2 between the absolute ΔG_{exp} and ΔG_{calc} from the CHARMm-MSI/RESP, CHARMm-MSI/ChelpG, and CHARMm-MSI/AM1-BCC corresponds to 0.72, 0.72, and 0.76, respectively (Figure 1B). The results show that for both force fields, the two QM methods, RESP and ChelpG, performed similarly in reproducing the experimental absolute solvation free energies, whereas the semiempirical AM1-BCC performed slightly better than the former two.

In an attempt to compare the performance of different charge models, the compounds were grouped depending on their chemical nature, and the average unsigned error (AUE) in the absolute solvation energy was calculated for each group (Figure 2). We observe that for small molecules containing hydrocarbons, such as linear and branched alkanes, alkenes, alkynes, cycloalkanes, polar groups such as aldehydes, ketones, carboxylic acids, esters, ethers, aliphatic amines, nitro, sulfides, and the majority of the halogenated molecules, the AUE in solvation energy is similar between the two QM charge methods (Figure 2). The major outliers are the small molecules containing hydrophobic ring structure (arenes), and a polar functional group like alcohols, amides, groups containing amines and another chemical group (bifunctional), and compounds with iodine. For the bulky arene hydrocarbons and compounds with an iodo functional group, the calculated solvation free energies with RESP charges yield better agreement with the experimental solvation free energies as compared to the ChelpG charges.

In the calculation of absolute solvation energies, the results obtained from the semiempirical AM1-BCC charges perform as well as those obtained with QM charges. In case of nitrogen-containing polar functional groups, such as amines, amides, nitro, and bifunctional groups with at least one amino function, the AM1-BCC performs better than the ChelpG and RESP, for a given force field. It is somewhat surprising that the AM1-BCC charges outperform the RESP for the above functional groups. This can be attributed to the five BCC parameters that were adjusted in the original AM1-BCC model to improve agreement with experimental data of amines, nitro, and unsaturated aromatic hydrocarbon

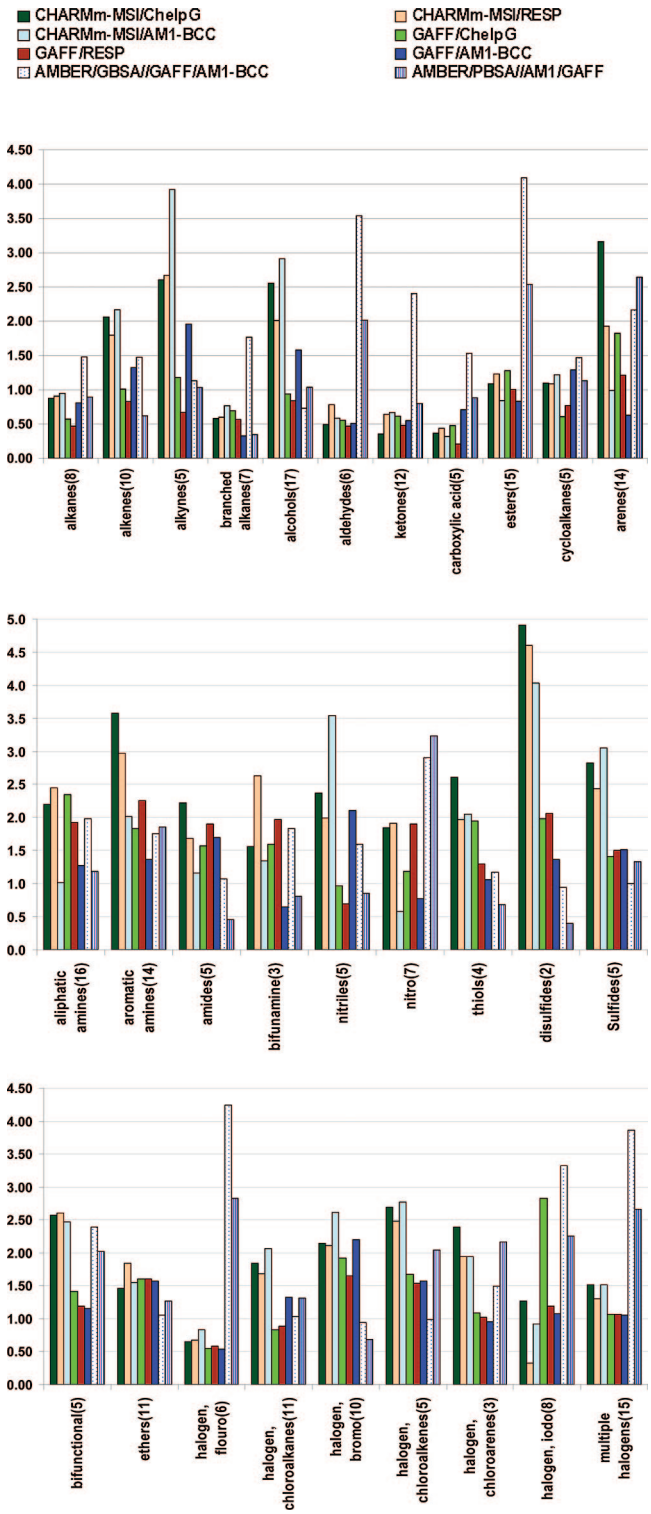

Figure 2. Average unsigned error [AUE] in the absolute solvation free energies. The AUE is shown in the *y*-axis, and the chemical functionalities in the small molecules are plotted in the *x*-axis. The solid bars represent the solvation free energies calculated using explicit solvent/FEP method in CHARMM. The bars with dotted line and stripes represent the solvation free energy calculated using GB and PB model in Amber9.

groups.[50] The AM1-BCC model also outperforms the RESP and ChelpG in case of bulky arene molecules, such as 1,3-dimethyl naphthalene, benzene, biphenyl, etc., for the above-mentioned reason.

For the small molecules containing alkyne and nitrile functional groups, AM1-BCC underperforms as compared to the two QM charge models. A possible explanation can be the AM1 population charges are insufficient to model the electron delocalization in the highly unsaturated compounds. The nitriles were also among the major outliers in a previous study using implicit solvent GB/SA and SGB/NP model[58] and indicate the challenges associated with obtaining accurate charge distributions for nitrogen-containing species. Overall, the AM1-BCC charge model appears to perform well with both the GAFF or CHARMm-MSI force field and the explicit solvent FEP/MD in reproducing the experimental absolute solvation free energies. Considering the fact that the AM1-BCC model was not parametrized to work with the GB/SA (or PB) model, the low AUE in solvation free energies calculated with implicit solvent models is extremely encouraging. Because the AM1-BCC charge generation method is very fast as compared to the RESP and ChelpG methods, the present results suggest that it can be a preferred practical method for screening database containing thousands of diverse small molecules.

b. Comparison of the Lennard-Jones Parameters/Force Fields. We compare the absolute solvation free energies for small molecules using two popular fixed charge force fields that are commonly used as a general force field for small molecules, GAFF and CHARMm-MSI. The former is a general version of the traditional biomolecular Amber force field that was recently developed, and the latter is a part of the CHARMm force field distributed commercially by MSI. Both of them use a similar functional form of the harmonic potential energy function to describe their additive force field equations based on bond, angle, dihedral, and nonbonded Lennard-Jones parameters.

Comparing the overall reproduction of the experimental absolute solvation free energies with any given charge method, we find that GAFF performs better than CHARMm-MSI. For example, using the GAFF/AM1-BCC method ($R^2 = 0.87$), we obtain a better correlation between the experimental and calculated solvation free energy when using the CHARMm-MSI/AM1-BCC ($R^2 = 0.76$, Figure 1). The AUE in absolute solvation free energies obtained from GAFF and CHARMm-MSI for different chemical functions is shown in Figure 2. For a given charge method, GAFF performs better than CHARMm-MSI for several functional groups, such as alkenes, alkynes, arenes, aromatic amines, nitriles, bifunctional groups, and small molecules with chloride functional group. Interestingly, in this study, the disulfides and sulfides are one of the major outliers in the CHARMm-MSI force field, which point in the direction of need for improvement in the sulfur atomic parameters in MSI force field. GAFF performs well even for those small molecules for which CHARMm-MSI has trouble reproducing the absolute solvation free energies. On the other hand, GAFF has high AUE [AUE > 1.5 kcal/mol] in reproducing the absolute solvation free energy for chemical functional groups such as amides, sulfides, ethers, bromo, and chloro-alkenes.

c. Comparison of the Explicit and Implicit Solvent Models. The accuracy of commonly used continuum solvent models, such as PB and GB, depends sensitively upon the

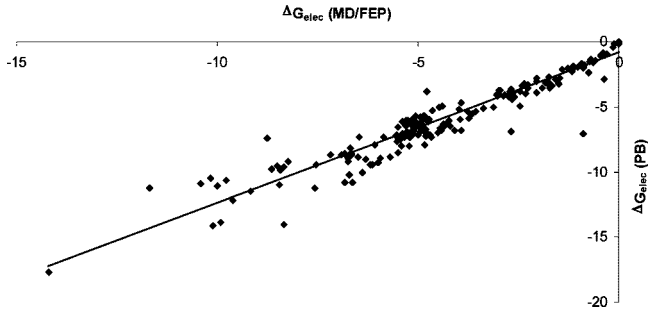

Figure 3. Comparison of the electrostatic component of the absolute solvation free energy between explicit solvent and implicit solvent model. The implicit solvent is modeled using PB in the Amber9 program. The explicit solvent results are from the FEP/MD and CHARMM v31 program. The correlation coefficient is 0.92. The calculations were done with the GAFF/AM1-BCC model. The energies are in kcal/mol.

parameters like atomic charges, the solute and solvent dielectric coefficients, and the atomic Born radii, which are used to define the solute−solvent dielectric boundary. Here, we compare the electrostatic component of the absolute solvation free energies obtained using the GB and PB continuum solvation models and compare it with its counterpart from the explicit solvent simulation, FEP/MD. We find that the electrostatic components obtained by solving the PB equation are in better agreement ($R^2 = 0.91$) with the explicit solvent simulation as compared to the GB model ($R^2 = 0.81$) [Figure 3]. Such a good performance of continuum PB in approximating the electrostatic free energy from FEP/MD has been observed in several previous studies.[18,22,29]

The decomposition scheme implemented in the implicit solvent module (igb = 10) in the Amber9 program was used to separate the nonpolar solvation free energy into repulsive ($\Delta G_{\text{rep-PB}}$) and attractive or dispersive ($\Delta G_{\text{att-PB}}$) components.[33] In this optimized 6-12 decomposition scheme, the SASA has been used to correlate the repulsive (cavity) term only, and a surface-integration approach is used to compute the attractive (dispersion) term. We benchmark the corresponding values for the repulsive and attractive terms with the similar counterparts obtained using WCA decomposition from the explicit solvent simulation using CHARMM ($\Delta G_{\text{rep-explicit}}$ and $\Delta G_{\text{att-explicit}}$). The correlation plot comparing the two components of the nonpolar solvation free energy is shown in Figure 4A and B. The $\Delta G_{\text{att-PB}}$ shows a correlation coefficient of 0.94 with $\Delta G_{\text{att-explicit}}$, whereas the $\Delta G_{\text{rep-PB}}$ has a correlation coefficient of 0.96 with $\Delta G_{\text{rep-explicit}}$.

We observe a very good correlation between the repulsive and attractive component of the nonpolar solvation free energy ($\Delta G_{\text{rep-explicit}}$) calculated using the WCA decomposition via FEP/MD simulation with explicit solvent and the SASA ($R^2 = 0.95$) approximation (Figure 5). However, the total nonpolar contribution obtained from both PB and explicit solvent simulations shows poor correlation with the SASA approximation (Figure 5). This suggests that implicit solvent models that rely exclusively on the SASA approximation to model the total nonpolar contribution might lead to considerable errors and should be used with caution.

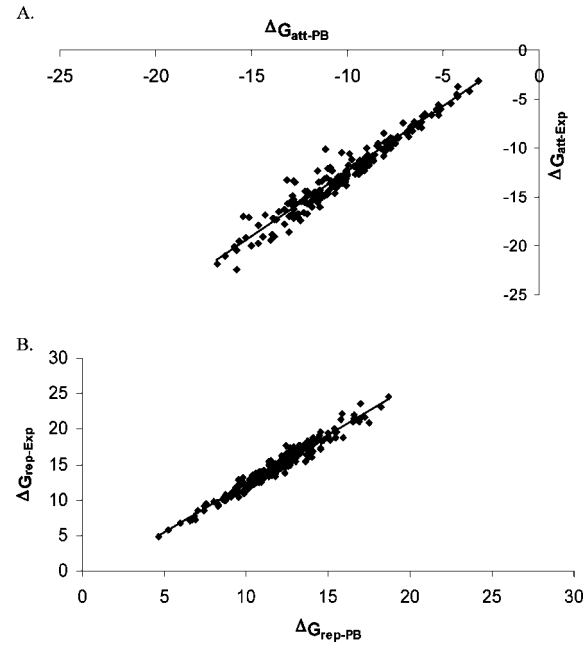

Figure 4. Comparison of the nonpolar components of the absolute solvation free energy between explicit solvent and implicit solvent model. The correlation between the attractive component of the nonpolar solvation energy obtained from FEP/MD with explicit solvent ($\Delta G_{\text{att-exp}}$) and PB implicit continuum solvent ($\Delta G_{\text{att-PB}}$) is 0.95. In the continuum PB model, the nonpolar part of the solvation energy is obtained from the Amber9 program. The FEP/MD results are from the CHARMM v31 program. The calculations were done with the GAFF/AM1-BCC model. The energies are in kcal/mol.

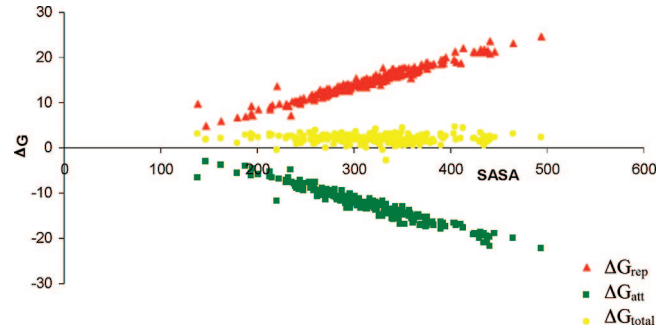

Figure 5. Plot showing correlation between total nonpolar solvation free energy (kcal/mol), and its repulsive and attractive components calculated using explicit solvent FEP/MD with solvent accessible surface area (SASA, Å^2).

In an attempt to better understand the performance of the PB solver implemented in the two popular molecular mechanics program, CHARMM and Amber, we compared the electrostatic solvation free energies between the two. The modified Bondi radii (corresponding to mbondi2 in Amber) were extracted from Amber and read in as a stream file in the PBEQ solver in CHARMM. With a grid spacing of 0.2, reentrant water surface, and probe density of 1.4, we find an excellent correlation between the electrostatic component of the solvation energy using the PB solvers from CHARMM and Amber ($R^2 > 0.99$) (Supporting Information, Figure S1).

In the present study, we used a single energy-minimized conformation to calculate the solvation free energies with the implicit solvation model. This is actually a common

Table 2. Table Showing the Average Solvation Free Energy and Single Snapshot Using Implicit Solvent Models, PB and GB[a]

molecule	$\Delta G_{experiment}$	PB (10 ns)	PB (20 ns)	PB (1 snapshot)	GB (10 ns)	GB (20 ns)	GB (1 snapshot)
methyl hexanoate	−2.49	−4.27	−4.25	−4.59	−6.07	−5.93	−6.07
methyl propyl ether	−1.59	−1.21	−1.22	−0.97	−2.18	−2.12	−2.14
octanal	−2.29	−3.71	−3.77	−4.14	−5.66	−5.58	−5.7
5-nonanone	−2.67	−2.78	−2.77	−2.97	−4.78	−4.68	−4.6

[a] All energies are reported in kcal/mol.

practice in docking and scoring with implicit solvation models. However, the neglect of conformational flexibility can cause some problems, as shown in a previous study.[59] To ascertain the impact of conformational flexibility on the solvation free energy, we compared the solvation free energy obtained using the single solute conformation with that from a proper ensemble average. The solvation free energy was calculated pertubatively from an ensemble of conformations generated by an extensive MD simulation of the molecule in a vacuum using the following expression:

$$e^{-\Delta G/k_B T} = \frac{1}{N}\sum_i e^{-\Delta W(\mathbf{X}_i)/k_B T} \quad (2)$$

where $\Delta W(\mathbf{X}_i)$ is the potential of mean force incorporating the effect of solvent implicitly for the conformation $\mathbf{X}_i$, k_B is the Boltzmann constant, and T is absolute temperature. We obtained the average solvation free energy using both the PB and the GB implicit solvent simulations, for four of the most flexible molecules in our set (number of rotatable bonds larger than 5). These molecules are methyl hexanoate, 5-nonanone, octanal, and methyl propyl ether. The results are shown in Table 2. Comparing the average solvation free energy obtained from the above method with that of the single solute conformation reported in this Article, we find that the solvation free energy estimated from a single energy-minimized snapshot is within 0.5 kcal/mol for all four flexible molecules. These estimates are comparable to the results from a previous study where for the energy-minimized conformation in vacuum (called "BestVac" in Mobley et al.[59]) the rms error was on the order of 0.34 kcal/mol. They showed that an approach based on energy-minimum conformations in vacuum gives the lowest errors.

d. log *P*. The molecular hydrophobicity or log P of a small molecule is an important descriptor to measure differential solubility of a solute in two immiscible solvent media, hydrophobic and hydrophilic solvent.[4] For example, the partition coefficient of a solute between octanol and water is represented by

$$\log P_{oct/wat} = \log([\text{solute}]_{oct}/[\text{solute}]_{wat}) \quad (3)$$

It is commonly used in the QSAR studies and drug design to study the drug absorption, bioavailability, metabolism, and toxicity. However, the force field-based simulations are not commonly used to calculate this quantity as the simulations in octanol solvent pose significant challenges due to complex structure formations, such as micelles.

Here, we compare the correlation between the experimental log $P_{oct/wat}$ with the repulsive component of the nonpolar solvation free energy ($\Delta G_{rep\text{-}explicit}$) obtained using explicit solvent FEP/MD. The experimental data for log $P_{oct/wat}$ have been taken almost entirely from the recommended values in the compilation by Hansch and co-workers.[60] The data cover a range of 7.0 log units from ca. −2 to +5. It shows high correlation with the cavity or repulsive term, $\Delta G_{rep\text{-}explicit}$. Such a high correlation is intriguing. A possible explanation is that log P is a measure of hydrophobicity, and therefore is correlated with $\Delta G_{rep\text{-}explicit}$. Also, the SASA is correlated with the $\Delta G_{rep\text{-}explicit}$ as well as log P with a correlation coefficient of 0.95.

Within a particular chemical series, the log P is linearly correlated with $\Delta G_{rep\text{-}explicit}$ with a high correlation coefficient (Figure 6). For example, for alcohols, such as methanol, ethanol, propanol, butanol, pentanol, hexanol, heptanol, and octanol, the correlation coefficient between log P and $\Delta G_{rep\_Explicit}$ is 0.996. For amines, such as methylamine, ethylamine, 1-propanamine, 1-butylamine, 1-pentanamine, *N*-ethylethamine, *N*,*N*-dimethylamine, *N*,*N*-diethylethanamine, *N*-propylpropan-1-amine, trimethylamine, piperidine, and pyrrolidine, the correlation coefficient is 0.93. For nitro-containing small molecules, such as nitroethane, 1-nitropropane, 2-nitropropane, 1-nitrobutane, nitrobenzene, 3-nitrophenol, and 1-methyl-2-nitro-benzene, the correlation coefficient is 0.99.

The physical significance for such a correlation may be that a larger SASA facilitates solvation in a more hydrophobic solvent, such as octanol. It reflects the importance of van der Waals interaction in the organic solvent, which in turn is related to the cavitation of the solute. The nature of the chemical function governs the slope and the intercept of this log P versus $\Delta G_{rep\text{-}explicit}$ linear curve. We believe that the FEP/MD simulations, similar to the ones used in this study, could provide important information about the molecular descriptor properties such as hydration and log P. Further studies in making a physics-based accurate log P predicting tool are currently underway and will be reported elsewhere.

4. Conclusion

In this study, we computed the absolute solvation free energy for a diverse set of 239 neutral molecules using FEP/MD with explicit solvent and compared the results with experimental data as well as with two widely used implicit solvent models (GB/SA and PB). We evaluated the performance of two popular general force fields, GAFF and CHARMm-MSI, used to model small molecules and ligands in drug design. Also, we examined the sensitivity of the free energy to the atomic partial charges generated via ab initio QM (ChelpG and RESP) and semiempirical (AM1-BCC) methods.

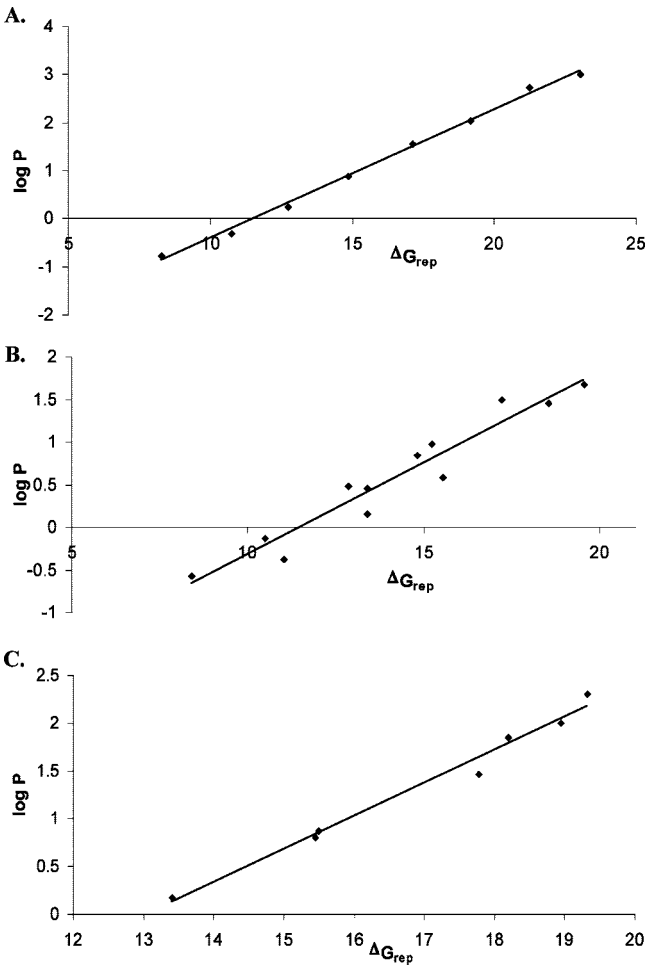

Figure 6. Correlation of log *P* with nonpolar solvation free energy ($\Delta G_{rep\_explicit}$). The correlation coefficient for alcohols, such as methanol, ethanol, propanol, butanol, pentanol, hexanol, heptanol, octanol, is 0.996 (A). For amines, such as methylamine, ethylamine, 1-propanamine, 1-butylamine, 1-pentanamine, *N*-ethylethamine, *N,N*-dimethylamine, *N,N*-diethylethanamine, *N*-propylpropan-1-amine, trimethylamine, piperidine, and pyrrolidine, the correlation coefficient is 0.93 (B). For nitro-containing small molecules, such as nitroethane, 1-nitropropane, 2-nitropropane, 1-nitrobutane, nitrobenzene, 3-nitrophenol, and 1-methyl-2-nitro-benzene, the correlation coefficient is 0.99 (C).

Our analysis indicates that the GAFF force field with the AM1-BCC atomic charges generally outperforms the others for a majority of chemical functionalities. However, the AM1-BCC charges need to be revisited for some unsaturated chemical functional groups, such as nitriles and alkynes, to provide better agreement with experiment. The origin of the problem is clearly the partial charges because those same functional groups show good correlation with experimental values when RESP charges are used. Also, our studies show that the sulfur-containing groups are particularly troublesome in case of the CHARMm-MSI force field.

Concerning the performance of implicit solvent models, it is observed that the electrostatic component obtained by solving the finite-difference PB equation is in better agreement with the explicit solvent simulations as compared to the GB model. The underperformance of GB as compared to PB noted here is somewhat surprising because the accuracy of GB is generally believed to deteriorate for large complex macromolecules, but to be quite adequate in the case of small molecules such as those considered here. On the other hand, the repulsive and the attractive components of the nonpolar solvation free energy show excellent correlation between the implicit and the explicit solvent simulations. The repulsive components of nonpolar solvation free energy obtained using the implicit and explicit solvent models are also highly correlated with the SASA. However, the total nonpolar solvation free energy is poorly correlated with SASA. This suggests that implicit solvent models that use only SASA to approximate the total nonpolar contribution are less accurate and should be used with caution.

Interestingly, log *P* appears to be highly correlated with the cavity or repulsive component of the nonpolar solvation free energy, $\Delta G_{rep\text{-}explicit}$, as well as the SASA. Within particular chemical series, log *P* displays an almost perfect linear correlation with the $\Delta G_{rep\text{-}explicit}$. Although the reasons for such a high correlation are not entirely clear, it is possible that it reflects the importance of van der Waals interaction in the organic solvent such as octanol, which is related to the size of the cavity needed to insert the solute.

Continued efforts are needed to assess the accuracy of atomic partial charges and other aspects of molecular mechanical force fields. Such assessment is critically important for the progress of computational structural biology. The present study shows how force field development and validation could benefit from extensive free energy computations performed on a large-scale basis.

Acknowledgment. This work was supported by the National Science Foundation through Grant MCB-0630140, and by Grant 2006-264-R2 from the Laboratory-Directed Research and Development (LDRD) program at Argonne National Laboratory, and by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, under Contract no. DE-AC02-06CH11357. D.S. thanks Jeff Klauda for supplying the AMBER99 force field conversion for use in the CHARMM program.

Supporting Information Available: Please contact authors for the decomposition of the nonpolar estimates, molecular structures of the molecules, and assigned atomic partial charges. The absolute solvation free energy for the test set is provided, along with the correlation plot between the electrostatic component of the solvation energy using the PB solvers from CHARMM and Amber. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Karplus, M. Molecular Dynamics Simulations of Biomolecules. *Acc. Chem. Res.* **2002**, *35*, 321–323.

(2) Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Ligand Solvation in Molecular Docking. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 4–16.

(3) Vedani, A.; Dobler, M.; Lill, M. A. The Challenge of Predicting Drug Toxicity in silico. *Basic Clin. Pharm. Toxicol.* **2006**, *99*, 195–208.

(4) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3–26.

(5) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813–1818.

(6) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. Proteins a Theoretical Perspective of Dynamics, Structure and Thermodynamics. In *Advances in Chemical Physics*; Prigogine, I., Rice, S. A., Eds.; John Wiley and Sons: New York, 1988.

(7) Roux, B.; Simonson, T. Implicit Solvent Models. *Biophys. Chem.* **1999**, *78*, 1–20.

(8) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(9) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins. *J. Chem. Phys.* **2003**, *119*, 5740–5761.

(10) Warwicker, J.; Watson, H. C. Calculation of the Electric Potential in the Active Site Cleft due to α-helix Dipoles. *J. Mol. Biol.* **1982**, *157*, 671–679.

(11) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(12) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. Solvation Free Energies of Peptides: Comparison of Approximate Continuum Solvation Models with Accurate Solution of the Poisson-Boltzmann Equation. *J. Phys. Chem. B* **1997**, *101*, 1190–1197.

(13) Schaefer, M.; Karplus, M. A Comprehensive Analytical Treatment of Continuum Electrostatics. *J. Phys. Chem. B* **1996**, *100*, 1578–1599.

(14) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. Focusing of Electric Fields in the Active site of Cu-Zn Superoxide Dismutase: Effects of Ionic Strength and Amino-acid Modification. *Proteins* **1986**, *1*, 47–59.

(15) Sharp, K. A.; Honig, B. Electrostatic Interactions in Macromolecules: Theory and Applications. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301–332.

(16) Wang, J.; Tan, C.; Tan, Y. H.; Lu, Q.; Luo, R. Poisson-Boltzmann Solvents inMolecular Dynamics Simulations. *Commun. Comput. Phys.* **2008**, *3*, 1010–1031.

(17) Jean-Charles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T.; Still, C. Electrostatic Contributions to Solvation Energies: Comparison of Free Energy Perturbation and Continuum Calculations. *J. Am. Chem. Soc.* **1991**, *113*, 1454–1455.

(18) Nina, M.; Beglov, D.; Roux, B. Atomic Radii for Continuum Electrostatics Calculations Based on Molecular Dynamics Free Energy Simulations. *J. Phys. Chem. B* **1997**, *101*, 5239–5248.

(19) Jayaram, B.; Sprous, D.; Beveridge, D. L. Solvation Free Energy of Biomacromolecules: Parameters for a Modified Generalized Born Model Consistent with the AMBER Force Field. *J. Phys. Chem. B* **1998**, *102*, 9571–9576.

(20) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.

(21) Banavali, N. K.; Roux, B. Atomic Radii for Continuum Electrostatics Calculations on Nucleic Acids. *J. Phys. Chem. B* **2002**, *106*, 11026–11035.

(22) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys. J.* **2004**, *86*, 67–74.

(23) Eisenberg, D.; McLachlan, A. D. Solvation Energy in Protein Folding and Binding. *Nature* **1986**, *319*, 199–203.

(24) Chothia, C. Hydrophobic Bonding and Accessible Surface Area in Proteins. *Nature* **1974**, *248*, 338–339.

(25) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. Reconciling the Magnitude of the Microscopic and Macroscopic Hydrophobic Effects. *Science* **1991**, *252*, 106–109.

(26) Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at Small and Large Length Scales. *J. Phys. Chem. B* **1999**, *103*, 4570–4577.

(27) Rajamani, S.; Truskett, T. M.; Garde, S. Hydrophobic Hydration from Small to Large Lengthscales: Understanding and Manipulating the Crossover. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9475–9480.

(28) Deng, Y.; Roux, B. Hydration of Amino Acid Side Chains: Nonpolar and Electrostatic Contributions Calculated from Staged Molecular Dynamics Free Energy Simulations with Explicit Water Molecules. *J. Phys. Chem. B* **2004**, *108*, 16567–16576.

(29) Wagoner, J. A.; Baker, N. A. Assessing Implicit Models for Nonpolar Mean Solvation Forces: The Importance of Dispersion and Volume Terms. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8331–8336.

(30) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. The SGB/NP Hydration Free Energy Model Based on the Surface Generalized Born Solvent Reaction Field and Novel Nonpolar Hydration Free Energy Estimators. *J. Comput. Chem.* **2002**, *23*, 517–529.

(31) Pitera, J. W.; van Gunsteren, W. F. The Importance of Solute-Solvent van der Waals Interactions with Interior Atoms of Biopolymers. *J. Am. Chem. Soc.* **2001**, *123*, 3163–3164.

(32) Floris, F.; Tomasi, J. Evaluation of the Dispersion Contribution to the Solvation Energy. A Simple Computational Model in the Continuum Approximation. *J. Comput. Chem.* **1989**, *10*, 616–627.

(33) Tan, C.; Tan, Y.-H.; Luo, R. Implicit Nonpolar Solvent Models. *J. Phys. Chem. B* **2007**, *111*, 12263–12274.

(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(35) Beglov, D.; Roux, B. Finite Representation of an Infinite Bulk System: Solvent Boundary Potential for Computer Simulations. *J. Chem. Phys.* **1994**, *100*, 9050–9063.

(36) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(37) Momany, F. A.; Rone, R. Validation of the General-Purpose Quanta3. 2/CHARMM force-field. *J. Comput. Chem.* **1992**, *13*, 888.

(38) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, J. K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(39) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. On the Nonpolar Hydration Free Energy of Proteins: Surface Area and Continuum Solvent Models for the Solute-Solvent Interaction Energy. *J. Am. Chem. Soc.* **2003**, *125*, 9523–9530.

(40) Weeks, J. D. C.; Andersen, D.; Hans, C. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *J. Chem. Phys.* **1971**, *54*, 5237–5247.

(41) Boresch, S.; Archontis, G.; Karplus, M. Free Energy Simulations: the Meaning of the Individual Contributions from a Component Analysis. *Proteins* **1994**, *20*, 25–33.

(42) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Model for Aqueous Solvation Based on Class {IV} Atomic Charges and First Solvation Shell Effects. *J. Phys. Chem.* **1996**, *100*, 16385–16398.

(43) Maple, J. R.; Cao, Y.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A. A Polarizable Force Field and Continuum Solvation Methodology for Modeling of Protein-Ligand Interactions. *J. Chem. Theory Comput.* **2005**, *1*, 694–715.

(44) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

(45) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B* **2007**, *111*, 2242–2254.

(46) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting Small-Molecule Solvation Free Energies: An Informal Blind Test for Computational Chemistry. *J. Med. Chem.* **2008**, *51*, 769–779.

(47) Breneman, C. M.; Wiberg, K. B. Determining Atom-centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.

(48) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well Behaved Electrostatic Potential based Method using Charge Restraints for Deriving Atomic Charges−The RESP Model. *J. Phys. Chem.* **1993**, *40*, 10269–10280.

(49) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

(50) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(52) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(53) Souaille, M. a. R. B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135*, 40–57.

(54) Macke, T.; Case, D. A. Modeling unusual nucleic acid structures. In *Molecular Modeling of Nucleic Acids*; Leontes, N. B., SantaLucia, J. J., Eds.; American Chemical Society: Washington, DC, 1998; pp 379−393.

(55) Onufriev, A.; Bashford, D.; Case, D. A. Exporling Protein Native States and Large-scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004**, *55*, 383–394.

(56) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217–230.

(57) Roux, B. Influence of the Membrane Potential on the Free Energy of an Intrinsic Protein. *Biophys. J.* **1997**, *73*, 2980–2989.

(58) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2006**, *2*, 128–139.

(59) Mobley, D. L.; Dill, K. A.; Chodera, J. D. Treating Entropy and Conformational Changes in Implicit Solvent Simulations of Small Molecules. *J. Phys. Chem. B* **2008**, *112*, 938–946.

(60) Leo, A.; Hoekman, D. H.; Hansch, C. *Explor. QSAR, Hydrophob., Electron., Steric Const.*; American Chemical Society: Washington, DC, 1995.

CT800445X