JCTC Journal of Chemical Theory and Computation

Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions

Robert C. Rizzo,*[,†,§] Tiba Aynechi,[†,||] David A. Case,[‡] and Irwin D. Kuntz[†]

Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143-2240, and the Department of Molecular Biology TPC-15, The Scripps Research Institute, La Jolla, California 92037

Received April 12, 2005

Abstract: Absolute free energies of hydration (ΔG_{hyd}) for more than 500 neutral and charged compounds have been computed, using Poisson−Boltzmann (PB) and Generalized Born (GB) continuum methods plus a solvent-accessible surface area (SA) term, to evaluate the accuracy of eight simple point-charge models used in molecular modeling. The goal is to develop improved procedures and protocols for protein−ligand binding calculations and virtual screening (docking). The best overall PBSA and GBSA results, in comparison with experimental ΔG_{hyd} values for small molecules, were obtained using MSK, RESP, or ChelpG charges obtained from ab initio calculations using 6-31G* wave functions. Correlations using semiempirical (AM1BCC, AM1CM2, and PM3CM2) or empirical (Gasteiger-Marsili and MMFF94) methods yielded mixed results, particularly for charged compounds. For neutral compounds, the AM1BCC method yielded the best agreement with experimental results. In all cases, the PBSA and GBSA results are highly correlated (overall $r^2 = 0.94$), which highlights the fact that various partial charge models influence the final results much more than which continuum method is used to compute hydration free energies. Overall improved agreement with experimental results was demonstrated using atom-based constants in place of a single surface area term. Sets of optimized SA constants, suitable for use with a given charge model, were derived by fitting to the difference in experimental free energies and polar continuum results. The use of optimized atom-based SA constants for the computation of ΔG_{hyd} can fine-tune already reasonable agreement with experimental results, ameliorate gross deficiencies in any particular charge model, account for nonoptimal radii, or correct for systematic errors.

Introduction

The quantification of how a solute will partition into two different phases, A and B, is widely used in drug design.[1,2] Notable examples include using *n*-octanol-to-water partitioning ($\log P_{octanol/water}$) as a model for cell membrane permeability and gas-to-water partitioning ($\log P_{gas/water}$) to estimate desolvation penalties associated with protein−ligand binding. The two quantities are related, from the perspective of continuum models of solvation, in that they quantify partitioning between phases with low (gas ∼ 1 and octanol ∼ 17) and high (water ∼ 80) dielectric constants. Experimental $\log P_{gas/water}$ measurements, often expressed as free energies of hydration ($\Delta G_{hyd} = -2.3RT \log P_{gas/water}$), have been compiled by several research groups for both neutral and charged species (see Table S1 in the Supporting

* Corresponding author e-mail: rizzo@ams.sunysb.edu; phone: (415) 902-3549.
† University of California at San Francisco.
‡ The Scripps Research Institute.
§ Current address: Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794-3600.
|| Graduate Group in Biophysics.

10.1021/ct050097l CCC: $33.50
Published on Web 11/15/2005

Information).[3−9] These experimental data make the computation of ΔG_{hyd} an attractive thermodynamic property for validating continuum simulation methods and can be used to guide the choice of parameters employed in such calculations.

The ultimate goal of this study is to optimize computational methods for protein−ligand binding calculations and virtual screening (docking). The recently reported Molecular Mechanics Poisson−Boltzmann Surface Area (MM-PBSA) and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) methods[10−12] incorporate a ΔG_{hyd}-like term as a measure of the change in desolvation ($\Delta\Delta G_{hyd}$) for the receptor−ligand binding event.[13−20] In MM-PBSA and MM-GBSA analysis,[10−12] PBSA or GBSA continuum energy terms for a given species (complex, receptor, or ligand) are formally equivalent to an absolute ΔG_{hyd} if, as is commonly done, dielectric constants of 1 (gas phase) and 80 (water phase) are specified. Therefore, the accuracy of computed ΔG_{hyd} terms directly affect the final computed binding energies. Unfortunately, experimental free energies of hydration are not available for proteins, most drugs, or protein−drug complexes. A reasonable alternative is to verify that the calculation methods and parameters yield good results for small organic molecules, for which experimental absolute free energies of hydration are available,[3−9] prior to using MM-PBSA and MM-GBSA methods.

Historically, the most accurate ΔG_{hyd} calculations have employed free energy perturbation (FEP) or thermodynamic integration (TI) simulations incorporating explicit models of water.[21,22] This was first done in 1985 by Jorgensen and Ravimohan[23] who used FEP methods to compute the relative free energy of hydration ($\Delta\Delta G_{hyd}$) for ethane and methanol in excellent agreement with experimental results using Monte Carlo simulations. The FEP and TI methods yield $\Delta\Delta G_{hyd}$ (or ΔG_{hyd}) directly and without the need for partitioning the free energy into separate components, as in other more-approximate approaches. However, such simulations can be tedious to set up and too computationally expensive for high-throughput structure-based drug design.

Continuum theories which treat solvent as a bulk macroscopic quantity[24] represent a complementary approach to the computation of solute hydration. In particular, Poisson−Boltzmann (PB)[25] and Generalized Born (GB)[26] are two widely used methods used to estimate the polarization energy associated with bringing any species from the gas phase to the bulk solvent phase. PB and GB calculation results are typically augmented by a solvent-accessible surface area term (SA) to account for nonpolar contributions to the total free energy of hydration. A comprehensive study which compares the performance of various GB implementations to PB reference calculations has recently been reported by Feig et al.[27] In this paper, we instead focus on evaluating which commonly used partial charge models yield GBSA and PBSA absolute hydration free energies in agreement with experimental results.

Two early continuum studies that directly compare computed ΔG_{hyd} with experimental results include the original GBSA report by Still et al.[26] and the Sitkoff et al.[25] PARSE (parameters for solvation energy) study designed for use with PBSA methods. Excellent results were obtained in both cases; however, the number of molecules tested was relatively small (between 20 and 67 molecules).[25,26] Both prior studies employed charge models based on functional group assignment, which may be difficult to assign to compounds typically found in databases used for high-throughput virtual screening (docking). More recent efforts have focused on evaluating the accuracy of partial charge models that may be more easily assigned, in an automated fashion, to relatively large and diverse data sets.[4,28−31] For example, Jorgensen and co-workers have recently reported the implementation and validation of a generalized GBSA model in conjunction with the OPLS-AA force field employing charges obtained from AM1CM1 semiempirical calculations.[32] Excellent results were reported with a mean unsigned error of only 1 kcal/mol for 399 neutral compounds.[32] Levy and co-workers have also developed highly accurate GBSA models, termed SGBNP[33] and AGBNP,[34] which employ OPLS-AA charges and radii and incorporate optimized nonpolar contributions to minimize errors with experimental results. Cramer, Truhlar, and co-workers have developed numerous solvation models, validated using much of the same experimental data used here, which have been subsequently incorporated into the AMSOL program.[6,35] Although highly accurate, AMSOL models tend to be highly parametrized, which makes incorporating routines for the computation of ΔG_{hyd} into a general molecular mechanics force field somewhat cumbersome. Many continuum models have been optimized to yield good ΔG_{hyd} results for small molecules using multiple atom types, radii, charges, and various combinations of adjustable nonpolar parameters which can limit transferability. Less-empirical models use very high-level ab initio and PB calculations with partial charges computed from SCRF methods with polarization[36] but are not easily adapted for general high-throughput screening.

Prompted by the need for a general continuum method with minimal optimized parameters, we have evaluated eight different point-charge models, based on ab initio, semiempirical, and empirical calculations in conjunction with a simple set of radii, through the computation of ΔG_{hyd} for small organic molecules. Computational results for more than 500 compounds (460 neutral compounds, 42 polyatomic ions, and 11 monatomic ions) are compared with experimental results, which, to our knowledge, represents the largest number of reference compounds employed for ΔG_{hyd} calculations. The evaluation of nonpolar contributions using an atom-based as opposed to a molecule-based solvent-accessible surface area term was also explored. Parameter set validation is critical since the use of different theoretical methods, atomic partial charge models, atomic radii, and nonpolar SA parameters will lead to different calculated ΔG_{hyd} results. The primary goal here was to assess the accuracy of the different charge models for neutral and charged compounds and to compare the results from both PBSA and GBSA continuum method calculations. The ultimate goal is to be able to easily incorporate accurate solvation effects using MM-GBSA methods for protein−ligand binding calculations and the rescoring of complexes obtained from high-throughput docking.

Computational Methods

Free Energies of Hydration (ΔG_{hyd}). As in prior PBSA and GBSA continuum studies,[25,26] the free energy of hydration is partitioned into two terms, polar and nonpolar, according to eq 1.

$$\Delta G_{hyd} = G_{polar} + G_{nonpolar} \quad (1)$$

Polar energies (G_{polar}) for PB calculations were obtained using a grid-based finite difference solution to the Poisson−Boltzmann equation with zero salt concentration (eq 2), where $\rho(r)$ is the charge distribution of the molecule, $\epsilon(r)$ is the dielectric constant, and $\phi(r)$ is the electrostatic potential. The solution of the PB equation for systems described by a classical force field yields the electrostatic potential at every grid point, and G_{polar} is then evaluated as a sum over all atoms (eq 3), where the partial atomic charge for atom i is multiplied by the difference in the computed grid-point potential ϕ_i for the transfer from the gas phase ($\epsilon = 1$) to water ($\epsilon = 80$).

$$\nabla[\epsilon(r)\,\nabla\phi(r)] = -4\pi\rho(r) \quad (2)$$

$$G_{polar} = \frac{1}{2}\sum_{i}^{N} q_i(\phi_i^{80} - \phi_i^{1}) \quad (3)$$

For GB calculations, G_{polar} contributions were obtained using eqs 4−5. Here, ϵ is the dielectric constant (80 for the water phase), q the partial atomic charges, r_{ij} the interatomic distance, and α_i are the Born radii, which are computed according to the pairwise descreening algorithm of Hawkins et al.[37,38]

$$G_{polar} = -166\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{q_i q_j}{f_{GB}} - 166\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{N}\frac{q_i^2}{\alpha_i} \quad (4)$$

$$f_{GB} = \{r_{ij}^2 + \alpha_{ij}^2 \exp[-r_{ij}^2/(4\alpha_{ij}^2)]\}^{0.5} \quad (5)$$

Nonpolar contributions ($G_{nonpolar}$) to ΔG_{hyd} are estimated using only a simple solvent-accessible surface area term. Alternative procedures, which treat $G_{nonpolar}$ as two terms representing separate cavitation and solute−solvent dispersion contributions, have been recently reported by Gallicchio and Levy.[34] In the present paper, $G_{nonpolar}$ is estimated using either the total molecular SA (eq 6) or atomic-based SA_i (eq 7). Prior MM-PBSA and MM-GBSA binding energy protocols typically employed a molecular SA (eq 6) with $\gamma = 0.00542$ and $\beta = 0.92$, as recommended by Kollman and co-workers.[11,12] An alternative method, which was pursued in the present work, is to compute atom-based SA_i and optimize each SA constant using multiple linear regression to improve agreement with experiment (eq 7). Using atom-based SA_i contributions to estimate free energies of solvation was first proposed by Eisenberg and McLachlan,[39] and Scheraga and co-workers.[40]

$$G_{nonpolar} = (\gamma SA) + \beta \quad (6)$$

$$\Delta G_{hyd}(exptl) - G_{polar} = G_{nonpolar} = \sum_i \gamma_i SA_i \quad (7)$$

Table 1. Parameters for PB and GB Continuum Calculations

type	mbondi radii	Sx value (GB only)[a]	number of atoms[b]
HC	1.30[c]	0.85	4215
HN	1.30[c]	0.85	98
HO	0.80[c]	0.85	93
HS	0.80[c]	0.85	13
HP	1.30[c]	0.85	10
C	1.70[a]	0.72	2678
N	1.55[a]	0.79	128
O	1.50[a]	0.85	299
F	1.50[a]	0.88	53
P	1.85[a]	0.86	6
S	1.80[a]	0.96	26
Cl	1.70[a]	0.80	114
Br	1.85[d]	0.80	27
I	1.98[d]	0.80	12

[a] From AMBER version 7.[43] [b] See Supporting Information Table S1 for a listing of all compounds. [c] From refs 41 and 42. [d] From Bondi's original work.[47]

For a given set of calculations, PBSA or GBSA, the same structures, partial charges, and atomic radii were employed. Any differences in the final calculation results in this paper will, therefore, be only a function of the two different continuum theories.

Computation Details. A simple set of atomic radii based on the mbondi (modified bondi) scheme,[41,42] from the AMBER7 program,[43] was used in the calculations for neutral and polyatomic charged species. In the mbondi scheme, hydrogen atoms connected to carbon, sulfur, nitrogen, phosphorus, or oxygen (types HC, HS, HN, HP, or HO, respectively) can have unique radii (Table 1). Dielectric constants for all calculations (PB and GB) were set to 1, representing the gas phase, and 80, representing the water phase. PB calculations were performed using the program Delphi4[44,45] with the following parameters: boundary conditions = 4, internal dielectric constant = 1.0, external dielectric constant = 80.0, and scale = 4 grids/Å. Other Delphi parameters were assigned automatically using default values. Generalized Born calculations were performed using an in-house version of the Hawkins et al.[37,38] pairwise descreening model with scaling parameters (Sx values) adopted from Tsui and Case (Table 1).[41] The DMS program was used for all of the SA calculations.[46] In addition to the total SA value for a compound, DMS can be used to estimate atom-based surface areas (SA_i). For a given compound, the total solvent-accessible surface area should be equivalent to the sum of each atom-based solvent-accessible surface area ($SA = \sum SA_i$).

Molecular Structures and Experimental Data. Bordner et al.[31] have generously made available 410 neutral molecular structures along with the corresponding experimental log $P_{gas/water}$ partition coefficients from the tabulated work of Abraham et al.[3] (converted to free energies at 25 °C using $\Delta G_{hyd} = -2.3RT \log P_{gas/water}$). However, the Bordner set did not contain compounds with polar hydrogens connected to sulfur (HS; Table 1) or include charged species. We augmented the neutral set with 50 additional neutral compounds (including compounds containing HS), as well as 42

Table 2. Correlation Coefficients (r^2) and Average Unsigned Errors (aue) for Experimental[a] vs Calculated[b] (PBSA or GBSA) Free Energies of Hydration (ΔG_{hyd}) Using Standard SA Constants[c]

	neutral molecules, $N = 460$; part I				charged (± 1) molecules, $N = 42$; part II			
model	r^2 PBSA	aue	r^2 GBSA	aue	r^2 PBSA	aue	r^2 GBSA	aue
Gast	0.53	3.20	0.49	3.36	0.68	7.52	0.67	8.15
MMFF94	0.29	3.26	0.26	3.41	0.73	7.44	0.72	8.27
AM1BCC	0.74	1.36	0.70	1.38	0.56	8.28	0.53	9.64
AM1CM2	0.71	3.09	0.67	2.81	0.39	11.67	0.34	13.63
PM3CM2	0.69	2.79	0.64	2.61	0.62	10.84	0.62	11.90
MSK	0.77	1.54	0.72	1.63	0.74	6.42	0.72	7.30
RESP	0.77	1.47	0.72	1.51	0.75	6.34	0.73	7.20
ChelpG	0.73	1.61	0.69	1.67	0.74	6.36	0.72	7.28

[a] See Supporting Information Table S1 for experimental references. [b] Calculated values obtained using eq 1. G_{polar} from either PB or GB calculations. [c] $G_{nonpolar} = (0.00542 \times SA_{total}) + 0.92$. Energies in kcal/mol.

charged (± 1) polyatomic compounds and 11 ionic monatomic species (see Table S1, Supporting Information). All additional compounds were obtained from the NIST Chemistry WebBook database[48] or constructed using the MOE program.[49]

Partial Charge Models. Eight charge models were evaluated in this study: Gasteiger−Marsili (Gast),[50] MMFF94,[51] AM1BCC,[29,30] AM1CM2,[52] PM3CM2,[52] Merz−Singh−Kollman (MSK),[53] Restrained Electrostatic Potential (RESP),[54,55] and ChelpG.[56] While the preceding list is not exhaustive, it does include methods that are currently implemented in several molecular modeling packages and allow for the calculation of partial atomic charges for diverse organic molecules. For a comparison with the present work, Udier-Blagovic et al. have recently evaluated the accuracy of partial charges computed using CM1 and CM3 procedures.[28] The molecule database was maintained with the MOE program,[49] and several software packages were used to assign the different charge models. Gast and MMFF94 charges were assigned using MOE. AM1BCC charges were determined using the ANTECHAMBER module in AMBER7[43] from MOPAC[57] calculations. The AMSOL[35] program was used to compute AM1CM2 and PM3CM2 partial charges.[52] AMSOL calculations incorporated the SM5.42R[6] water solvent model, which allows the charges to be computed in a simulated condensed phase. The MSK, RESP, and ChelpG charges were computed at the HF/6-31G*//HF/6-31G* level of theory using the program Gaussian 98.[58] Molecules containing iodine used the 3-21G* basis set for iodine and 6-31G* for all other atoms. The ANTECHAMBER module in AMBER7 was used for two-stage RESP fittings. It should be mentioned that different software packages may yield slight variations in atomic charges because of differences in the implementation of a particular partial charge model. Only the above-named program implementations were evaluated in this report.

Molecule Geometries. For each compound, the partial charges obtained using the eight different methods were mapped back to one set of standard geometries. Using one set of conformations allows for a direct comparison of the accuracy of the partial charge models and removes the possibility that different geometries would affect the results. Here, the standard geometries were taken as those obtained from a gas-phase geometry optimization using the MMFF94 force field as implemented in the MOE program. In general, the optimizations yielded extended structures. Other geometries could have been used, although this was not explored. Given that the data set contains mostly rigid compounds, the effect of including multiple conformations on the computed free energies of hydration was not investigated. Of the 502 polyatomic compounds, more than half (53%) contain two or fewer rotatable bonds. Averaging over multiple conformations in the previous Bordner study changed the computed free energies by only a small amount.[31]

Results and Discussion

Charge Model Evaluation. Free energies of hydration were computed for comparison with experimental results for compounds employing one of eight partial charge models (Gast, MMFF94, AM1BCC, AM1CM2, PM3CM2, MSK, RESP, and ChelpG). Table 2 lists the correlation coefficients (r^2) and average unsigned errors (aue) between experiment and theory as obtained from PBSA and GBSA calculations. In Table 2, the $G_{nonpolar}$ term is computed from molecular SA (eq 6) using the standard MM-PBSA and MM-GBSA constants ($\gamma = 0.00542$ and $\beta = 0.92$). Results for charged and neutral compounds are always reported separately since artificially high r^2 squared values may result when correlations are computed using both species together. This is primarily due to the large difference in magnitude of the experimental data for charged versus neutral species.

The correlation coefficients for neutral compounds in Table 2 (part I) track with the eight different charge schemes in roughly the following order: ab initio (MSK, RESP, and ChelpG) > semiempirical (AM1BCC, AM1CM2, and PM3CM2) > empirical (Gast and MFF94). Ab initio charges yield PBSA and GBSA r^2 values from 0.69 to 0.77, semiempirical r^2 values from 0.64 to 0.74, and empirical r^2 values from 0.26 to 0.53. Average unsigned errors (aue) follow the r^2 trends; ab initio charges yield smaller errors (1.47−1.67 kcal/mol) than semiempirical (1.36−3.09 kcal/mol) or empirical (3.20−3.41 kcal/mol) charges. For comparison, results from various parametrizations of the AMSOL SM5.42R universal solvation models from Cramer, Truhlar, and co-workers yield small unsigned errors of 0.43−0.46 kcal/mol for 275 neutral solutes.[6] Gallicchio et al. have also reported errors of less than 0.5 kcal/mol using the optimized SGB/NP model.[33] The primary reason AMSOL and SGB/NP methods yield much smaller errors than the results

Table 3. Average Unsigned Errors (aue) for Experimental[a] vs GBSA Calculated[b] Free Energies of Hydration (ΔG_{hyd}) Using Standard SA Constants[c]

type	number	Gast	MMFF94	AM1BCC	AM1CM2	PM3CM2	MSK	RESP	ChelpG
alkanes	19	0.33	0.33	0.30	0.32	0.28	0.60	0.49	0.38
alkenes/dienes	11	0.83	0.83	0.23	0.22	0.14	1.23	1.21	0.70
alkynes	4	1.90	1.38	0.56	1.95	1.77	2.23	2.18	1.55
arenes	19	3.64	3.64	0.30	2.08	1.35	0.46	0.47	1.96
alcohols/phenols	26	4.66	4.66	0.88	1.07	0.85	1.96	1.77	1.59
ethers	15	2.20	2.20	2.07	1.24	1.72	1.87	1.97	2.03
ketones/aldehydes	20	2.87	2.87	2.73	6.13	5.59	1.48	1.37	1.33
carboxylic acids	3	4.32	4.32	3.14	10.61	10.46	7.01	6.93	7.29
esters	14	2.75	2.75	0.87	5.36	4.19	1.50	1.40	1.43
amines	21	5.86	5.86	2.88	2.97	3.47	2.89	2.97	3.02
amides	2	8.71	8.70	5.19	7.59	6.41	1.53	1.39	1.63
nitriles	4	5.95	5.95	1.14	8.18	8.92	4.39	3.76	4.26
nitrohydrocarbons	6	4.98	6.74	1.75	5.44	7.63	4.92	4.38	4.58
nitrogen heterocyclic	10	4.98	4.98	0.96	3.41	3.35	0.90	1.02	1.87
thiols	3	2.73	2.73	0.34	0.80	1.03	0.82	0.51	0.21
sulfides	3	3.43	3.43	1.32	0.76	0.85	0.65	1.04	1.46
all 180 molecules	180	3.43	3.48	1.37	2.85	2.75	1.76	1.69	1.85

[a] See Supporting Information Table S1 for experimental references. [b] Calculated values obtained using eq 1. G_{polar} from GB calculations. [c] $G_{nonpolar} = (0.00542 \times SA_{total}) + 0.92$. Energies in kcal/mol.

presented here in Table 2 is due to the fact that many parameters have been optimized to minimize errors with respect to experimental results. Similar approaches could be adopted by the present GBSA method through the incorporation of schemes which allow separate radii and nonpolar contributions to be optimized on the basis of unique atom types (e.g., aromatic vs aliphatic carbons). A very simple optimization, based on the elemental atom types listed in Table 1, is presented in a later section of this paper. In the present study, the best agreement with experiment, for all 460 neutral compounds, was obtained using AM1BCC charges, which yielded aue's of 1.36 and 1.38 kcal/mol from PBSA and GBSA calculations, respectively (Table 2). For comparison, in the recent Jorgensen et al. study,[32] a mean unsigned error of 1.01 kcal/mol was reported for GBSA results using OPLS-AA radii with scaled AM1CM1A charges for 399 neutral compounds. A smaller subset of 75 molecules in that report yielded a larger mean unsigned error of 1.51 kcal/mol, which dropped to 1.16 kcal/mol if nitro compounds and DMSO were excluded.[32]

Table 3 shows GBSA results obtained here, broken down by molecular class for 180 out of 199 neutral compounds in common with the Gallicchio et al. study, which employed the SGB/NP model.[33] As previously noted, the SGB/NP model is a fitted method and, therefore, yields aue errors much lower than those reported here (aue = 0.32 kcal/mol fitted, 0.50 kcal/mol jackknife).[33] However, Table 3 highlights the fact that, for many molecule classes, low errors can, in fact, be obtained using a very simplistic GBSA model not fit a priori to reproduce experimental results. In particular, excellent results are obtained with the AM1BCC model, which yields a low aue of only 1.37 kcal/mol. The three ab initio methods yield errors between 1.69 and 1.85 kcal/mol. Other models yield larger aue's between 2.75 and 3.48 kcal/mol. Notably, the AM1BCC charges yield errors of less than 1 kcal/mol for more than half the compounds tested in Table 3 and include alkanes, alkenes/dienes, alkynes, arenes, alcohols/phenols, esters, nitrogen heterocyclics, and thiols. The major outliers for AM1BCC are amides ($N = 2$), carboxylic acids ($N = 3$), amines ($N = 21$), and ketones/aldehydes ($N = 20$). The largest errors using ab initio charges in Table 3 are for carboxylic acids, amines, nitriles ($N = 4$), and nitrohydrocarbons ($N = 6$). The largest errors reported in the prior SGB/NP[33] study also included nitriles (jackknifed aue = 1.15 kcal/mol) and nitro compounds, (jackknifed aue = 2.57 kcal/mol), in addition to nitrogen heterocyclics (jackknifed aue = 1.22 kcal/mol), and indicate the challenges associated with obtaining accurate charge distributions for nitrogen-containing species.

It should be noted that, during parametrization of the AM1BCC method, adjustments were made to the way partial charges are computed so that calculated relative free energies of solvation for amines, nitros, and unsaturated hydrocarbons were in closer agreement to experimental results.[29,30] However, the AM1BCC method was not optimized for use with the GBSA model or mbondi radii utilized here. Given this fact, the results in Tables 2 and 3 are extremely encouraging given that low errors can be obtained for most molecules and that AM1BCC charges are extremely fast to generate for databases containing even hundreds of thousands of diverse molecules.[30]

Surprisingly, the three semiempirical methods tested here yield poorer correlations with experimental results than do the two empirical methods for charged (± 1) molecules (Table 2, part II). Ab initio charges yield the strongest correlations, with r^2 values from 0.72 to 0.75, compared to semiempirical r^2 values from 0.34 to 0.62 and empirical r^2 values from 0.72 to 0.73. Given that Jorgensen et al. obtained good results for 17 polyatomic charged compounds using the OPLS-AA GBSA models augmented with unscaled AM1CM1 charges,[32] the poor results obtained here using AM1CM2 and PM3CM2 are unexpected. AM1CM2 and PM3CM2 methods should yield partial charges qualitatively similar to those obtained from the AM1CM1 procedure. Results for 22 charged

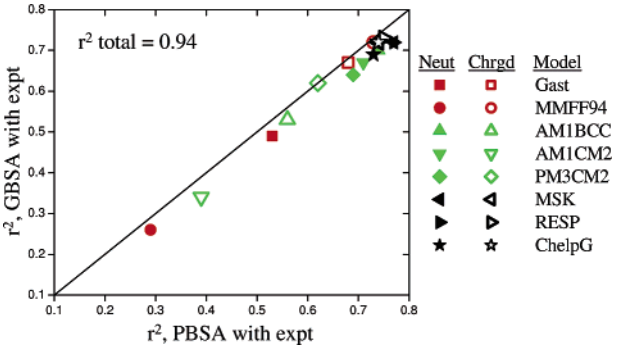

Figure 1. Comparison of correlation coefficients (r^2 values) for calculated versus experimental free energies of hydration from PBSA and GBSA calculations. For each partial charge model, two r^2 values are plotted representing results for 460 neutral compounds (filled symbols) and 42 charged compounds (open symbols)(see Table 2). The overall correlation between the total PBSA and GBSA results is $r^2 = 0.94$.

compounds using the more sophisticated SGB/NP model with OPLS-AA charges yielded jackknifed aue's of 4.12 and 2.23 kcal/mol for 3 carboxylate anions and 19 ammonium cations, respectively. The larger errors reported here for charged species compared with those of other studies could arise from differences such as the number and type of compounds tested, the functional form of the GB equation used to compute G_{polar}, the atomic radii used in the calculations, or a lack of fitted nonpolar contributions. As emphasized previously, only element-based radii, with the exception of hydrogen atoms connected to C, N, O, P, and S (Table 1), were used. Li et al. estimate that the typical uncertainty in the experimental data for ions is larger, at about 5 kcal/mol, in comparison with neutral compounds, which is typically 0.2 kcal/mol.[6]

As was the case for neutral compounds, the aue errors reported here (Table 2) track with the correlation coefficients. Again, ab initio partial charges yield the lowest errors (6.34−7.30 kcal/mol), but for the charged species, semiempirical charges yield the largest errors (8.28−13.63 kcal/mol). Empirical aue's are in the middle (7.44−8.27 kcal/mol). Thus, using MSK, RESP, and ChelpG, partial charges for neutral and charged species consistently yield the strongest correlations and lowest average unsigned errors with experimental free energies of hydration regardless of which continuum method was employed for the computation (Table 2). The r^2 values for three ab initio methods cluster around 0.75 for both neutral and charged species (Figure 1).

PBSA versus GBSA. The PBSA and GBSA results are highly correlated and independent of the charge model used for the calculations (Table 2; Figure 1). The strong agreement between PBSA and GBSA r^2 values (obtained from computed versus experimental results) suggests that a given partial charge model will influence the final free energies much more than which continuum method (PBSA or GBSA) is used for the calculations. Correlation coefficients between PB and GB polar energies are always very strong, $r^2 > 0.94$, and independent of which partial charge model or data set (neutral or charged compounds) was employed in the calculations. These trends continue to provide strong support for using GBSA methods as a reasonable alternative to the more computationally demanding PBSA calculations for free energy calculations.

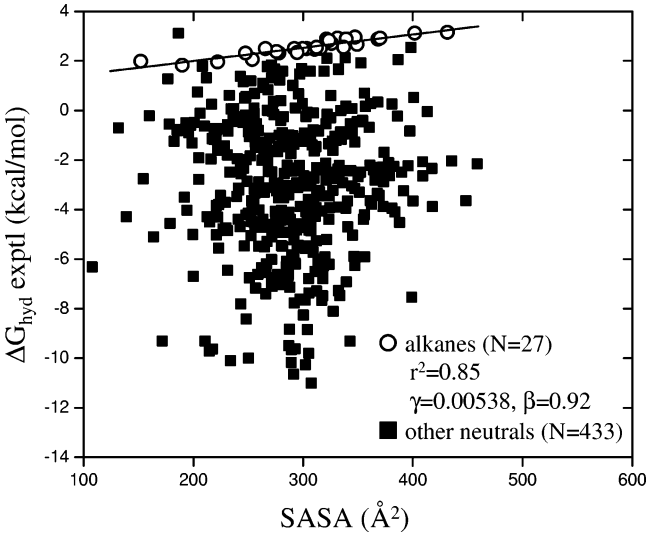

Figure 2. Experimental free energies of hydration vs total molecular solvent-accessible surface area (SASA). The best fit line to the 27 linear and branched alkanes (○) yields a correlation coefficient $r^2 = 0.85$, $\gamma = 0.005\,38$, and $\beta = 0.92$. Other compounds are represented as filled squares (■).

***$G_{nonpolar}$* from Molecular SA versus Atomic SA$_i$.** The constants ($\gamma = 0.00542$ and $\beta = 0.92$) typically used[13−20] in MM-PBSA and MM-GBSA calculations to convert SA (Å^2) to $G_{nonpolar}$ (kcal/mol) are based on fitting molecular SA results to experimental ΔG_{hyd} values for small straight-chain alkanes.[25] The rationale for this procedure exploits the fact that alkanes have low dipole moments and nonpolar contributions will, therefore, dominate ΔG_{hyd}. Figure 2 shows the molecular SA for the 460 neutral molecules studied here versus experimental ΔG_{hyd} values along with the best fit regression line using only the 27 linear and branched alkanes.

The constants obtained from this linear regression fit (Figure 2, open circles, $r^2 = 0.85$, $m = 0.00538$, $b = 0.92$) are essentially identical to the standard constants ($\gamma = 0.00542$ and $\beta = 0.92$).[13−20] However, as a group, molecular SAs have no correlation with experimental ΔG_{hyd} values (Figure 2, filled squares). Rankin et al. have reported similar results based on an analysis of 210 neutral compounds.[59] In most cases, G_{polar} contributions are the dominant factor for the final correlations with experimental results for the neutral molecules reported in Table 2. This is illustrated in Figure 3 for 460 neutral compounds in which the polar energies (G_{polar} $r^2 = 0.77$, filled squares) computed from PB calculations with RESP charges strongly correlate with experimental ΔG_{hyd} values. However, nonpolar contributions computed using standard SA constants (eq 6; $\gamma = 0.00542$ and $\beta = 0.92$) yield no correlation ($G_{nonpolar}$ $r^2 = 0.00$, open circles) and, for this data set, do not contribute to any improvement or diminishment in the final correlation coefficient with experimental results (ΔG_{hyd} $r^2 = 0.77$).

To improve the agreement with experimental results, we explored a procedure first proposed by Eisenberg and McLachlan[39] and Scheraga and co-workers,[40] which optimizes nonpolar contributions using atom-based coefficients for a given SA type through multiple linear regression

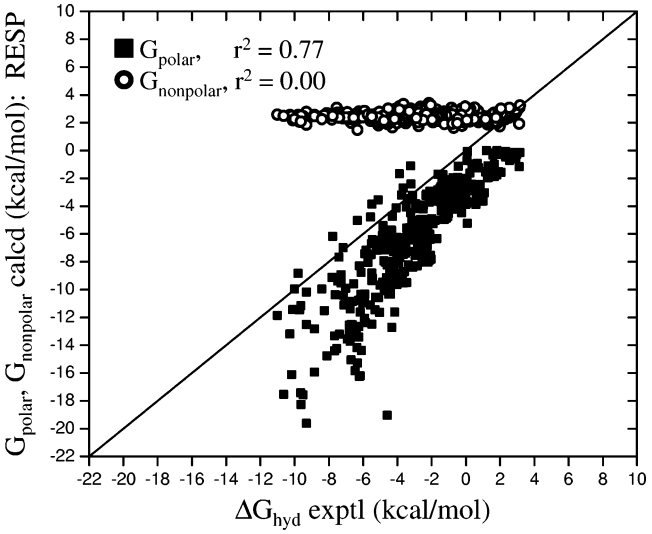

Figure 3. Correlation of individual components with experimental free energies of hydration for neutral compounds ($N = 460$) using RESP-derived partial charges. Polar (■) energies G_{polar} from PB calculations. Nonpolar (○) energies from molecular solvent-accessible surface area calculations $G_{nonpolar} = (0.00542 \times SA_{total}) + 0.92$.

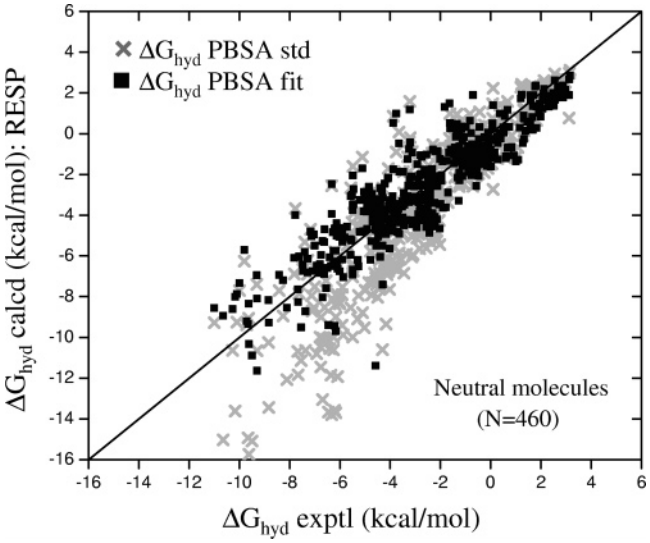

Figure 4. Predicted free energies of hydration (ΔG_{hyd} calcd) vs experimental free energies of hydration (ΔG_{hyd} exptl) from PBSA calculations with RESP charges for neutral compounds ($N = 460$). Nonpolar energies from molecule-based SAs using standard constants (×) or atom-based SAs using fitted constants (■).

fitting.[39,40] In the present work, SA γ_i coefficients were optimized for each mbondi type (HC, HN, HO, HS, HP, C, N, O, F, P, S, Cl, Br, and I) using multiple linear regression (eq 7) by fitting to the residuals in exptl $\Delta G_{hyd} - \Delta G_{polar}$ computed using the eight different charge models from either GB or PB calculations. After the fittings, new $G_{nonpolar}$ contributions were recomputed using the atom-based constants (γ_i's) so that optimized ΔG_{hyd} values could then be compared with experimental results. Levy and co-workers have reported an alternative functional form for $G_{nonpolar}$ in which atom-based coefficients are used to compute separate cavitation and solute−solvent dispersion interactions, as implemented in the SGB/NP[33] and AGBNP[34] models.

In most cases, utilizing the new SA_i constants to estimate nonpolar energies improves the overall agreement with experimental ΔG_{hyd} values (Table 4 versus Table 2). In particular, overall aue's are substantially reduced and r^2 values improve in general. However, using fitted constants actually reduces correlations (r^2) for neutral compounds for the three semiempirical models (AM1BCC, AM1CM2, and PM3CM2) despite the fact that the aue's show dramatic improvement (Table 4 versus Table 2). Here, degradations in r^2, for neutral molecules with semiempirical charges, arise because the fitting procedure attempts to minimize the overall error (neutral and charged) with experimental results, and these models originally performed poorly for charged species. The most robust and best overall improvement is obtained for the ab initio methods that consistently yielded the best agreement with experimental results for both neutral and charged species.

Figures 4 and 5 highlight favorable cases where the use of atom-based constants yield improved results even if a particular charge model already leads to good agreement with experimental results. Ab initio charges (MSK, RESP, and ChelpG) yield G_{polar} energies in strong correlation with those of the experiment for neutral and charged compounds in all cases. However, using molecule-based constants (gray crosses) to compute $G_{nonpolar}$ values can lead to a systematic overestimate (absolute error) of the hydration free energies for species with ab initio charges in the experimental range from −11 to −2 kcal/mol for neutrals and −90 to −60 kcal/mol for charged species. As an example, for neutrals, Figure

Table 4. Correlation Coefficients (r^2) and Average Unsigned Errors (aue) for Experimental[a] vs Calculated[b] (PBSA or GBSA fit) Free Energies of Hydration (ΔG_{hyd}) Using Fitted SA Constants[c]

	neutral molecules, $N = 460$; part I				charged (±1) molecules, $N = 42$; part II			
model	fitted r^2 PBSA	aue	fitted r^2 GBSA	aue	fitted r^2 PBSA	aue	fitted r^2 GBSA	aue
Gast	0.67	1.43	0.56	1.62	0.69	8.60	0.69	8.99
MMFF94	0.36	1.91	0.28	2.07	0.70	8.24	0.68	8.60
AM1BCC	0.68	1.26	0.58	1.49	0.61	6.71	0.60	6.83
AM1CM2	0.62	1.71	0.54	1.83	0.55	7.35	0.58	7.55
PM3CM2	0.61	1.66	0.52	1.83	0.68	7.24	0.71	7.47
MSK	0.81	0.99	0.69	1.32	0.79	4.46	0.77	4.68
RESP	0.80	1.02	0.69	1.33	0.80	4.45	0.78	4.69
ChelpG	0.81	0.99	0.70	1.30	0.79	4.46	0.77	4.67

[a] See Supporting Information for experimental references. [b] Calculated values obtained using eq 1. G_{polar} from either PB or GB calculations. [c] $G_{nonpolar} = \sum \gamma_i SA_i$. Energies in kcal/mol.

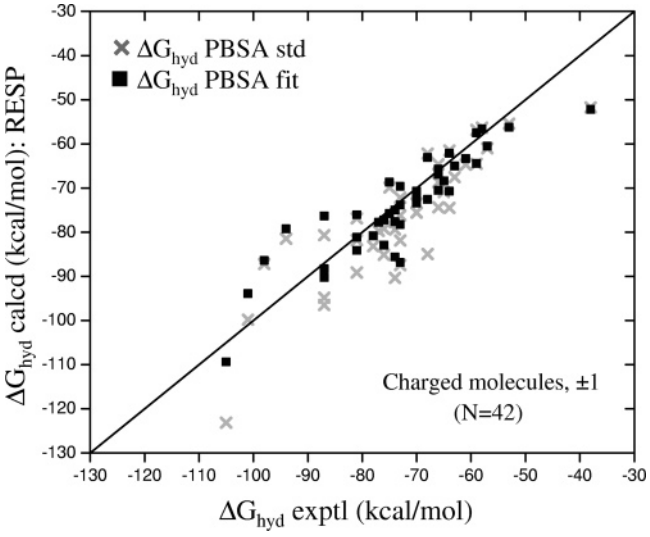

Figure 5. Predicted free energies of hydration (ΔG_{hyd} calcd) vs experimental free energies of hydration (ΔG_{hyd} exptl) from PBSA calculations with RESP charges for charged compounds ($N = 42$). Nonpolar energies from molecule-based SAs using standard constants (×) or atom-based SAs using fitted constants (■).

4 shows that using γ_i constants optimized from PBSA-RESP fits leads to an improvement of r^2 from 0.77 (ΔG_{hyd} std, gray crosses) to 0.80 (ΔG_{hyd} fit, black squares), and the aue error with the experiment drops from 1.47 to 1.02 kcal/mol (Figure 4). More dramatic results are observed for charged species; PBSA correlations increase from 0.75 (ΔG_{hyd} std, gray crosses) to 0.80 (ΔG_{hyd} fit, black squares), and the aue error with the experiment drops dramatically from 6.34 to 4.45 kcal/mol (Figure 5).

The primary motivation for using atom-based SA_i instead of molecule-based SA procedures is to reduce errors with respect to experimental results in three ways: (1) remedy gross deficiencies a particular charge model may have (r^2 and aue), (2) fine-tune an already reasonable agreement with experimental results (primarily aue), or (3) account for nonoptimal radii. On a case-by-case basis, simple atom-based constants (Figures 4−5, black squares) can correct for systematic errors. However, additional improvement, beyond what is presented in Tables 2, 3, and 4, would probably require changes to the model to include more atom-typing, adjustable radii, and nonpolar parameter optimization as in other methods.[6,33,34]

Optimized SA Coefficients. Tables 5 and 6 list "sets" of optimized SA_i constants (γ_i) obtained from multiple linear regressions using PB and GB G_{polar} results for all eight charge models employed in the calculations. For a new calculation that employs a particular charge model, atom-based γ_i values can be used to estimate $G_{nonpolar}$ energies that should lead to improved ΔG_{hyd} calculations. Despite the fact that G_{polar} results from both continuum methods show strong correlation (Table 2; Figure 1), for completeness, separate fits were performed for PB- (Table 5) or GB-derived (Table 6) G_{polar} energies.

As averaged over the entire data set of 502 molecules, the magnitude and sign for each γ_i value can give some indication as to the error with the experiment (and direction) associated with a particular charge model for a given atom (mbondi) type. However, caution should be exercised when trying to ascribe too much physical significance to any given SA coefficient. For some atom types listed in Table 1, HS ($N = 13$), P ($N = 6$), and I ($N = 12$), a lack of experimental data could potentially lead to SA optimizations that are underdetermined. Nevertheless, given the fact that related charge methods such as AM1CM2/PM3CM2 or MSK/RESP yield similar fitted SA constants (Tables 5 and 6), the multiple linear regression results appear robust. As an example, phosphorus (mbondi type P) coefficients from GB fits for AM1CM2 and PM3CM2 charged compounds are relatively large in magnitude compared with other types (Tables 5 and 6). Here, the negative coefficients are always in the range −1.8 to −2.8. The large negative sign indicates that, on average, G_{polar} terms computed using AM1CM2 and PM3CM2 charges underestimate the experimental ΔG_{hyd} values. Nonpolar contributions computed using atom-based SA_i will yield a favorable free energy to correct for this underestimation given that SA is always a positive value and, in this case, the γ_i for P atoms are negative. On the other hand, the GB γ_i coefficients for atom types P for ab initio-based methods (MSK, RESP, and ChelpG) are positive and

Table 5. Optimized Atomic SA Coefficients (γ_i Values)[a] Obtained Using Poisson−Boltzmann (PB) Derived G_{polar} Energies

type	Gast	MMFF94	AM1BCC	AM1CM2	PM3CM2	MSK	RESP	ChelpG
Hc	0.00093	−0.00002	0.00355	0.00962	0.00827	0.00679	0.00687	0.00649
Ho	−0.00434	−0.11172	0.25999	0.12379	0.11210	0.37414	0.36422	0.36037
Hs	0.28952	0.23307	0.33731	−0.53475	−0.45896	0.05493	0.06772	0.09424
Hn	−0.04103	−0.02779	−0.01058	−0.01094	−0.01857	−0.00574	−0.00436	−0.00813
Hp	−0.12342	0.00990	0.02589	0.47729	0.38605	−0.02415	−0.00164	−0.01025
C	−0.01634	−0.01610	0.02001	0.04395	0.03708	0.01765	0.01468	−0.00278
N	−0.00798	−0.01032	0.07251	0.05061	0.08398	0.04518	0.04440	0.05156
O	0.00759	0.04621	0.02409	0.09277	0.08863	0.03592	0.03292	0.04072
F	0.02036	0.02024	0.02256	0.02661	0.01954	0.01755	0.01643	0.01873
P	2.12323	0.36337	0.98863	−2.44577	−2.59507	0.92016	0.61608	0.79176
S	0.01477	0.02908	0.05082	0.15426	0.13041	0.04414	0.04145	0.03315
Cl	0.00336	0.00302	0.00384	0.00330	0.00662	0.00560	0.00527	0.00657
Br	−0.00532	−0.00455	0.00139	−0.00410	0.00415	0.00681	0.00550	0.00479
I	−0.00635	−0.00609	0.01495	−0.01134	−0.00775	0.00656	0.00562	−0.00116

[a] $G_{nonpolar} = \sum \gamma_i SA_i$ optimized using neutral ($N = 460$) and charged ($N = 42$) compounds.

Table 6. Optimized Atomic SA Coefficients (γ_i Values)[a] Obtained Using Generalized Born (GB) Derived G_{polar} Energies

type	Gast	MMFF94	AM1BCC	AM1CM2	PM3CM2	MSK	RESP	ChelpG
Hc	−0.00049	−0.00146	0.00129	0.00719	0.00588	0.00489	0.00484	0.00436
Ho	0.02765	−0.07826	0.25937	0.13669	0.12418	0.36583	0.35732	0.35663
Hs	0.29006	0.23719	0.30314	−0.48392	−0.40299	0.07870	0.09415	0.13314
Hn	−0.04816	−0.02950	−0.02299	−0.02113	−0.02386	−0.01374	−0.01218	−0.01520
Hp	−0.07575	0.06191	0.07913	0.54968	0.46106	0.01841	0.03414	0.02993
C	−0.01537	−0.01529	0.01715	0.03967	0.03379	0.02164	0.01859	0.00328
N	0.01065	0.00709	0.10707	0.07294	0.10361	0.06938	0.06810	0.07333
O	0.01100	0.04920	0.02952	0.09624	0.09423	0.04269	0.03965	0.04760
F	0.02353	0.02559	0.02941	0.02826	0.02085	0.02082	0.01948	0.02374
P	1.50762	−0.30940	0.71401	−1.75879	−2.78904	0.47251	0.25635	0.40528
S	0.01889	0.03237	0.05437	0.15530	0.13185	0.04452	0.04131	0.03165
Cl	0.00536	0.00489	0.00662	0.00515	0.00913	0.00878	0.00784	0.00787
Br	−0.00329	−0.00275	0.00466	−0.00130	0.00710	0.01492	0.01301	0.00786
I	−0.00419	−0.00384	0.02054	−0.00733	−0.00400	0.01865	0.01703	0.00294

[a] $G_{nonpolar} = \sum\gamma_i SA_i$ optimized using neutral ($N = 460$) and charged ($N = 42$) compounds.

Table 7. GBSA Results for Monoatomic Ions

ion	ΔG_{hyd} exptl[a]	OPLS-AA radii[b]	ABS error[c]	adjusted radii	ABS error[c]
F^-	−107	1.540	4.54		
Cl^-	−78	2.090	2.21		
Br^-	−72	2.255	1.89		
I^-	−63	2.700	2.26		
Li^+	−122	1.350	6.66	1.370[d]	4.62
Na^+	−98	1.680	3.53		
K^+	−81	2.020	2.22		
Mg^{2+}	−456	1.455	22.89	1.515[d]	2.64
Ca^{2+}	−381	1.735	16.01	1.785[d]	4.23
Fe^{2+}	−456			1.515[e]	2.64
Zn^{2+}	−485			1.435[e]	1.04

[a] See Supporting Information Table S1 for experimental references. [b] From Jorgensen et al.[32] [c] Absolute error for ΔG_{hyd} exptl − ΔG_{hyd} calcd; calculated values obtained using eq 1 with $G_{nonpolar} = (0.00542 \times SA_{total}) + 0.92$. [d] Adjusted from reference 32. [e] This work. Energies in kcal/mol.

much smaller at about 0.26−0.47. The variation in the optimized coefficients in Tables 5 and 6 is a direct result of the differences that are obtained from the different partial charge methods used for the computation of G_{polar}. Because of this fact, optimized constants can be viewed as a SA-based correction factor to account for errors in any particular charge model in an average sense. Moreover, γ_i constants should only be used in conjunction with the partial charge model with which they were derived.

Monatomic Ions. We have also pursued free energy of hydration calculations for 11 monatomic ions using the same GBSA protocols for comparison with experimental results. Monatomic ions are a unique case given that only a single atom is present, and therefore, they are not charge-model-dependent; only the formal ion charge and radius needs to be specified. Radii for most monatomic ions were taken from Jorgensen et al.[32] and used as reported or adjusted slightly for use with the present model. In general, nonpolar contributions to the total ΔG_{hyd} for monatomic ions would be negligible given the large polarization energy (−63 to −485 kcal/mol; Table 7) compared to the small solvent-accessible surface area contribution. The solvent-accessible surface area for a monatomic species is simply $SA = 4\pi(r + 1.4)^2$, where 1.4 Å represents the standard probe radius for water and r is the radius.

In general, the absolute difference between GBSA computed and experimental values (ΔG_{hyd} exptl − ΔG_{hyd} calcd) is lower than the estimated uncertainty for ions[6] (5 kcal/mol) using OPLS-AA radii with standard MM-GBSA constants ($G_{nonpolar}$ $\gamma = 0.00542$ and $\beta = 0.92$), as shown in Table 7. However, for Li^+, Mg^{2+}, and Ca^{2+}, larger errors of 6.66, 22.89, and 16.01 kcal/mol, respectively, are observed. It should be noted that results from Jorgensen et al. for the monatomics agree exactly with experimental results.[32] The results presented here in Table 7 employed different SA constants than those used by Jorgensen et al.; thus, slight differences are not unexpected. For consistency and to optimize parameters for the present GBSA model, the radii for three ions, Li^+, Mg^{2+}, and Ca^{2+}, were adjusted slightly so that all errors for monatomics would be less than 5 kcal/mol. Since SA contributions to hydration free energies for monatomic species are assumed to be small, the dominant change from the adjustment of radii will be to the G_{polar} term. For ions in particular, continuum results are very sensitive to the choice of atomic radii. For example, GB results for the ±1 monatomic ions shown in Table 7 change by more than 7 kcal/mol with only a 0.1 Å change in radius (1.5−1.6 Å). The same change for ±2 monatomics changes GB results dramatically by more than 30 kcal/mol. Which radii to employ for PBSA and GBSA continuum calculations is the subject of considerable research.[25,60−62]

Conclusion

The primary goal of this study was to evaluate procedures for the computation of free energies of hydration, in the context of a general classical molecular mechanics force field, for use in the simulation of protein−ligand binding and virtual screening (docking). Improved computational procedures continue to advance the utility of structure-based drug design. Here, absolute free energies of hydration have been computed using continuum PBSA and GBSA methods for comparison with experimental results for a diverse set of 460 neutral compounds, 42 polyatomic ions, and 11 monatomic ions. A systematic evaluation of eight different models has revealed that continuum results for small organic

molecules with partial charges based on one of three ab initio methods consistently lead to the best overall correlation with experimental results for both neutral and charged species (Table 2; Figure 1). Correlation coefficients with the experiment using MSK, RESP, and ChelpG charges with GBSA yield r^2 values between 0.69 and 0.73 and with PBSA yield r^2 values between 0.72 and 0.77. The semiempirical AM1BCC model yields good results for neutral compounds with an r^2 value of 0.70−0.74 and the lowest aue's of all the models tested (aue = 1.36−1.38). However, the use of semiempirical (AM1BCC, AM1CM2, and PM3CM2) and empirical (Gast and MFF94) charge schemes yielded mixed results dependent on whether the compounds were charged or neutral (Table 2).

The computational results presented here clearly show that correlations with experimental ΔG_{hyd} values are independent of which implicit solvation model (PBSA or GBSA) is employed in the calculations. In all cases, the Hawkins pairwise GB results are strongly correlated (overall $r^2 = 0.94$) with the much more expensive PB calculations, provided that identical coordinates, radii, and atomic charges are used (Figure 1).

An examination of polar and nonpolar energy components shows that $G_{nonpolar}$ energies derived from molecule-based SAs and standard conversion constants have no correlation with experimental results for neutral compounds (Figure 3). The lack of a universal SA constant stems from the erroneous assumption that all exposed atoms contribute equally to nonpolar energies. In the present work, improved correlations with experimental results were obtained through simple optimizations of atom-based SA constants using multiple linear regression fits to the difference in experimental free energies and polar energy terms obtained from continuum calculations (Tables 4−6). On a case-by-case basis, using atom-based SA_i instead of molecule-based SA constants significantly reduces both relative (r^2) and absolute unsigned errors (aue) with respect to the experiment by eliminating any gross deficiencies a particular charge model may have (Tables 4 versus Table 2). In particular, aue's are substantially reduced (Table 4), and systematic errors can be corrected (Figures 4−5, black squares). As was the case using standard SA constants, the best agreement with experimental results using atom-based SA_i constants was obtained with ab initio partial charge models; improved r^2 values are between 0.69 and 0.78 and 0.79−0.81 from GBSA and PBSA calculations, respectively (Table 4).

Finally, studies that continue to assess the accuracy of atomic partial charges and other force field parameters are critically important for the field of computational structural biology. The results here show both the strengths and the weaknesses of various "Amber-like" approaches to force fields for docking and screening. Hence, the results will be of interest to those considering such calculations. The fact that hydration energies calculated by other models (e.g., AMSOL or SGB/NP) show better agreement with experimental results than those reported here is useful information that may help researchers in their choice of computational strategies. The primary motivation here was to determine generally useful force-field parameters for estimating changes in the free energy of hydration associated with molecular recognition for use with MM-PBSA and MM-GBSA and for docking calculations. On the basis of a comparison with experimental ΔG_{hyd} values for more than 500 diverse organic molecules, MSK, RESP, or ChelpG partial atomic charges obtained from ab initio calculations using 6-31G* wave functions would be recommended for both charged and neutral species if computational resources allow it. For molecule libraries requiring partial charge assignments for hundreds of thousands of compounds, the semiempirical AM1BCC method would be recommended over the more approximate alternatives tested.

Acknowledgment. Gratitude is expressed to Walter Mangel for helpful discussions and to the Department of Defense (Award Number DAMD17-00-1-0192, Modification P00001) and National Institutes of Health (GM-56531, P. Ortiz deMontellano, P. I.) for support of this research. T.A. was supported by a U. C. president's dissertation year fellowship.

Supporting Information Available: Experimental free energies of hydration (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chem. Rev.* **1971**, *71*, 525−616.

(2) 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.

(3) Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. Thermodynamics of Solute Transfer from Water to Hexadecane. *J. Chem. Soc. Perkin Trans. 2* **1990**, 291−300.

(4) 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.

(5) Gerber, P. R. Charge distribution from a simple molecular orbital type calculation and nonbonding interaction terms in the force field MAB. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 37−51.

(6) Li, J. B.; Zhu, T. H.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Extension of the platform of applicability of the SM5.42R universal solvation model. *Theor. Chem. Acc.* **1999**, *103*, 9−63.

(7) Åvist, J. Ion−Water Interaction Potentials Derived form Free Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94*, 8021−8024.

(8) Babu, C. S.; Lim, C. Theory of ionic hydration: Insights from molecular dynamics simulations and experiment. *J. Phys. Chem. B* **1999**, *103*, 7958−7968.

(9) Marcus, Y. A Simple Empirical Model Describing the Thermodynamics of Hydration of Ions of Widely Varying Charges, Sizes, and Shapes. *Biophys. Chem.* **1994**, *51*, 111−127.

(10) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability

of DNA, RNA, and phosphoramidate−DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.

(11) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889−897.

(12) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.

(13) Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43*, 3786−3791.

(14) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221−5230.

(15) Masukawa, K. M.; Kollman, P. A.; Kuntz, I. D. Investigation of Neuraminidase-Substrate Recognition Using Molecular Dynamics and Free Energy Calculations. *J. Med. Chem.* **2003**, *46*, 5628−5637.

(16) Huo, S.; Wang, J.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. *J. Med. Chem.* **2002**, *45*, 1412−1419.

(17) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Luo, R.; Bayly, C. I.; Kollman, P. A. An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis. *J. Am. Chem. Soc.* **2001**, *123*, 3986−3994.

(18) Suenaga, A.; Hatakeyama, M.; Ichikawa, M.; Yu, X.; Futatsugi, N.; Narumi, T.; Fukui, K.; Terada, T.; Taiji, M.; Shirouzu, M.; Yokoyama, S.; Konagaya, A. Molecular dynamics, free energy, and SPR analyses of the interactions between the SH2 domain of Grb2 and ErbB phosphotyrosyl peptides. *Biochemistry* **2003**, *42*, 5195−5200.

(19) Donini, O. A. T.; Kollman, P. A. Calculation and prediction of binding free energies for the matrix metalloproteinases. *J. Med. Chem.* **2000**, *43*, 4180−4188.

(20) Rizzo, R. C.; Toba, S.; Kuntz, I. D. A Molecular Basis for the Selectivity of Thiadiazole Urea Inhibitors with Stromelysin-1 and Gelatinase-A from Generalized Born Molecular Dynamics Simulations. *J. Med. Chem.* **2004**, *47*, 3065−3074.

(21) Jorgensen, W. L. Free Energy Calculations: A Breakthrough For Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184−189.

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

(23) Jorgensen, W. L.; Ravimohan, C. Monte Carlo Simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050−3054.

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

(25) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(26) 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.

(27) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004**, *25*, 265−284.

(28) Udier-Blagovic, M.; Morales De Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. Accuracy of free energies of hydration using CM1 and CM3 atomic charges. *J. Comput. Chem.* **2004**, *25*, 1322−1332.

(29) 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.

(30) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parametrization and validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(31) Bordner, A. J.; Cavasotto, C. N.; Abagyan, R. A. Accurate Transferable Model for Water, n-Octanol, and n-Hexadecane Solvation Free Energies. *J. Phys. Chem. B* **2002**, *106*, 11009−11015.

(32) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. Free Energies of Hydration from a Generalized Born Model and an All-Atom Force Field. *J. Phys. Chem. B* **2004**, *108*, 16264−16270.

(33) 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.

(34) Gallicchio, E.; Levy, R. M. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* **2004**, *25*, 479−499.

(35) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Winget, P.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *AMSOL*, version 6.6; University of Minnesota: Minneapolis, Minnesota. Based in part on *AMPAC*, version 2.1, by Liotard, D. A.; Healy, E. F.; Ruiz, J. M.; Dewar, M. J. S. and on *EF* by Jensen, F.

(36) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. New model for calculation of solvation free energies: Correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. *J. Phys. Chem.* **1996**, *100*, 11775−11788.

(37) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246*, 122−129.

(38) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824−19839.

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

(40) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. Accessible Surface-Areas as a Measure of the Thermodynamic Parameters of Hydration of Peptides. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086−3090.

(41) Tsui, V.; Case, D. A. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.

(42) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275−291.

(43) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(44) Rocchia, W.; Alexov, E.; Honig, B. Extending the applicability of the nonlinear Poisson−Boltzmann equation: Multiple dielectric constants and multivalent ions. *J. Phys. Chem. B* **2001**, *105*, 6507−6514.

(45) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23*, 128−137.

(46) *DMS*; University of California at San Francisco Computer Graphics Laboratory: San Francisco, CA.

(47) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441−451.

(48) NCI Public Database. http://dtp.nci.nih.gov.

(49) *MOE*, version 2002.03; Chemical Computing Group: Montreal, Canada.

(50) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity − a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.

(51) Halgren, T. A. Merck molecular force field. 1. Basis, form, scope, parametrization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(52) Li, J. B.; Zhu, T. H.; Cramer, C. J.; Truhlar, D. G. New class IV charge model for extracting accurate partial charges from wave functions. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.

(53) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(54) 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**, *97*, 10269−10280.

(55) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of Resp Charges to Calculate Conformational Energies, Hydrogen-Bond Energies, and Free-Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.

(56) 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.

(57) Stewart, J. J. P.; Rossi, I.; Hu, W.-P.; Lynch, G. C.; Liu, Y.-P.; Chuang, Y.-Y.; Li, J.; Cramer, C. J.; Fast, P. L.; Truhlar, D. G. *MOPAC*, version 5.09mn; University of Minnesota: Minneapolis, Minnesota.

(58) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.9; Gaussian Inc.: Pittsburgh, PA.

(59) Rankin, K. N.; Sulea, T.; Purisima, E. O. On the transferability of hydration-parametrized continuum electrostatics models to solvated binding calculations. *J. Comput. Chem.* **2003**, *24*, 954−962.

(60) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* **2002**, *23*, 1297−1304.

(61) 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.

(62) Banavali, N. K.; Roux, B. Atomic radii for continuum electrostatics calculations on nucleic acids. *J. Phys. Chem. B* **2002**, *106*, 11026−11035.

CT050097L