

Draft Version

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*

Contribution from the 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, CA 92037

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 ($\log P_{\text{octanol/water}}$) as a model for cell membrane permeability, and gas to water partitioning ($\log P_{\text{gas/water}}$) 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, octanol~17) and high (water~80) dielectric constants. Experimental $\log P_{\text{gas/water}}$ measurements, often expressed as free energies of hydration ($\Delta G_{\text{hyd}} = -2.3 \log P_{\text{gas/water}}$), have been compiled by several research groups for both neutral and charged species.³⁻⁷ These experimental data make 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.

Historically, the most accurate ΔG_{hyd} calculations have employed free energy perturbation (FEP) or thermodynamic integration (TI) simulations incorporating explicit

models of water.^{8,9} This was first done in 1985 by Jorgensen and Ravimohan¹⁰ who used FEP methods to compute the relative free energy of hydration ($\Delta\Delta G_{\text{hyd}}$) for ethane and methanol in excellent agreement with experiment using Monte Carlo simulations. The FEP and TI methods yield $\Delta\Delta G_{\text{hyd}}$ (or ΔG_{hyd}) directly and without the need for partitioning the free energy into separate components, as in other more approximate approaches. However, the simulations can be tedious to setup, computationally prohibitive for high-throughput structure-based design, and absolute free energies of hydration can be difficult to obtain.

Continuum theories which treat solvent as a bulk macroscopic quantity¹¹ represent a complementary approach to computation of solute hydration. In particular, Poisson-Boltzmann (PB)¹² and Generalized Born (GB)¹³ 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 non-polar 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.¹⁴ In this report, we instead focus on evaluating which commonly used partial charge models yield GBSA and PBSA absolute hydration free energies in agreement with experiment.

For protein-ligand binding calculation, the recently reported Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) methods¹⁵⁻¹⁷ incorporate a ΔG_{hyd} -like term as a measure of the change in desolvation ($\Delta\Delta G_{\text{hyd}}$) for the receptor-ligand binding event.¹⁸⁻²⁵ In MM-

PBSA and MM-GBSA analysis,¹⁵⁻¹⁷ 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,^{3,7,26,27} prior to using such methods for estimating $\Delta\Delta G_{\text{hyd}}$ for protein-ligand binding. The ultimate goal is to obtain accurate free energies of hydration in the context of a general classical molecular mechanics force field for use in simulation of protein-ligand binding.

Two early continuum studies that directly compare computed ΔG_{hyd} with experiment include the original GBSA report by Still et al.¹³ and the Sitkoff et al.¹² 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 20-67 molecules).^{12,13} 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 coworkers have recently reported implementation and validation of generalized GBSA model in conjunction with the

OPLS-AA force field employing charges obtained from CM1A semiempirical calculations.³² Excellent results were reported with a mean unsigned error of only 1 kcal/mol for 399 neutral compounds.³² Numerous solvation models, developed and validated using much of the same experimental data used here, have also been reported by Cramer, Trular, and coworkers and subsequently incorporated into the powerful AMSOL program.^{6,33} Although highly accurate, the AMSOL models tend to be highly parameterized which makes incorporating routines for computation of ΔG_{hyd} into a general molecular mechanics force field somewhat cumbersome.

This study was prompted by the need for a generally useful classical force field, which includes terms to account for solvation effects, that can be easily employed for protein-ligand binding calculations and re-scoring of complexes obtained from high-throughput docking. We have focused on evaluating the accuracy of eight different point charge models, based on ab initio, semiempirical, and empirical calculations, through computation of hydration free energies. In particular we are interested in assessing the accuracy of different partial charge models, for both neutral and charged species, and comparison of PBSA and GBSA continuum methods across multiple data sets. Results for more than 500 compounds (460 neutral compounds, 42 polyatomic ions, and 11 monoatomic ions) are compared with experiment which to our knowledge represents the largest number of reference compounds employed for ΔG_{hyd} calculations. Other aims include the evaluation of non-polar contributions using an atom-based as opposed to molecule-based solvent accessible surface area term. 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.

Computational Methods

Free Energies of Hydration (ΔG_{hyd}). As in prior PBSA and GBSA continuum studies,^{12,13} the free energy of hydration is partitioned into two terms, polar and nonpolar, according to eq 1.

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

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, $\varepsilon(r)$ is the dielectric constant, and $\phi(r)$ the electrostatic potential. 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 gas-phase ($\varepsilon = 1$) to water ($\varepsilon = 80$).

$$\nabla \varepsilon(r) \nabla \phi(r) + 4\pi \rho(r) = 0 \quad (2)$$

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

For GB calculations, G_{polar} contributions were obtained using eq 4-5. Here, ε is the dielectric constant (80 for 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.^{34,35}

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

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

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.³⁶ In the present work 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 coworkers.^{16,17} 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,³⁷ and Scheraga and coworkers.³⁸

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

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

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 report will therefore be only a function of the two different continuum theories.

Computation Details. The atomic radii used in the calculations were assigned using mbondi (modified bondi) parameters^{39,40} from the AMBER7 program.⁴¹ In the mbondi scheme, hydrogen atoms connected to carbon, sulfur, nitrogen, or oxygen (types HC, HS, HN, or HO respectively) have unique radii (Table 1). Dielectric constants for all calculations (PB and GB) were set to 1 representing gas-phase and 80 representing water-phase. PB calculations were performed using the program Delphi4^{42,43} with the

following parameters: boundary conditions = 4, internal dielectric constant = 1.0, external dielectric constant = 80.0, 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.^{34,35} pairwise de-screening model with scaling parameters (Sx values) adopted from Tsui and Case (Table 1).³⁹ The DMS program was used for all the SA calculations.⁴⁴ 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 = ΣSA_i).

Table 1. Parameters for PB and GB continuum calculations.

Type	mbondi radii ^a	Sx value (GB only) ^b	No. atoms ^c
HC	1.30	0.85	4215
HN	1.30	0.85	98
HO	0.80	0.85	93
HS	0.80	0.85	13
C	1.70	0.72	2678
N	1.55	0.79	128
O	1.50	0.85	299
F, F ⁻	1.50	0.88	53
P	1.85	0.86	6
S	1.80	0.96	26
Cl, Cl ⁻	1.70	0.80	114
Br, Br ⁻	1.85	0.80	27
I, I ⁻	1.98	0.80	12
Li ⁺	1.82		1
Na ⁺	2.27		1
K ⁺	2.75		1
Mg ²⁺	1.73		1
Ca ²⁺	1.70		1
Fe ²⁺	1.50		1
Zn ²⁺	1.39		1

^aMbondi radii taken from Tsui and Case,^{39,40} the AMBER package,⁴¹ or Bondi's original compilation.⁴⁵ ^bScaling values from AMBER. ^cSee supporting information Table S1 for a listing of all compounds.

Molecular Structures and Experimental Data. Bordner et al.³¹ have generously made available 410 neutral molecular structures along with the corresponding experimental $\log P_{\text{gas/water}}$ partition coefficients from the tabulated work of Abraham and coworkers³ (converted to free energies at 25 °C using $\Delta G_{\text{hyd}} = -2.3 \log P_{\text{gas/water}}$). 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 charged (± 1) polyatomic compounds and 11 ionic monoatomic species. All additional compounds were constructed using the MOE program.⁴⁶

Partial Charge Models. Eight charge models were evaluated in this study: Gasteiger-Marsili (Gast),⁴⁷ MMFF94,⁴⁸ AM1BCC,^{29,30} AM1CM2,⁴⁹ PM3CM2,⁴⁹ Merz-Singh-Kollman (MSK),⁵⁰ Restrained Electrostatic Potential (RESP),^{51,52} and ChelpG.⁵³ While the preceding list is not exhaustive, it does include methods currently implemented in several molecular modeling packages and allow for the calculation of partial atomic charges for diverse organic molecules. For comparison, Udier et al. have recently evaluated the accuracy of partial charges based on CM1 and CM3 procedures.²⁸ In the present work several different software packages were used to assign the different charge models to the data sets. Gast and MMFF94 charges were assigned using the program MOE.⁴⁶ AM1BCC charges were determined using the ANTECHAMBER module in AMBER⁷⁴¹ from MOPAC⁵⁴ calculations. The AMSOL³³ program was used to compute AM1CM2 and PM3CM2 partial charges.⁴⁹ AMSOL calculation incorporated the SM5.42R⁶ 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 Gaussian98.⁵⁵ 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 variation in atomic charges due to differences in 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 513 compounds, more than half (53%) contain 2 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 experiment 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 ($\gamma = 0.00542$, $\beta = 0.92$). In every case the correlations between the experimental and theoretical free energies in Table 2 are due solely to the G_{polar} term; molecular SAs show no correlation with experiment as discussed below. 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.

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}). Nonpolar contributions obtained using molecule-based solvent accessible surface areas with standard constants.^c Energies in kcal/mol.

Charge Model	Neutral molecules, N = 460				Charged (± 1) molecules, N = 42			
	Part I				Part II			
	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

^aSee supporting information Table S1 for experimental references. ^bCalculated values obtained using eq 1. G_{polar} from either PB or GB calculations. ^c $G_{\text{nonpolar}} = (\text{SA}_{\text{total}} * 0.00542) + 0.92$.

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,

ChelpG) > semiempirical (AM1BCC, AM1CM2, PM3CM2) > empirical (Gast, MFF94). Ab initio charges yield PBSA and GBSA r^2 values from 0.69 to 0.77, semiempirical r^2 from 0.64 to 0.74, and empirical r^2 from 0.26 to 0.53. Average unsigned errors (AUE) follow the r^2 trends; ab initio charges yield smaller errors (1.47 to 1.67 kcal/mol), than semiempirical (1.36 to 3.09 kcal/mol), or empirical (3.20 to 3.41 kcal/mol). For comparison, results from various parameterizations of the AMSOL SM5.42R universal solvation models from Cramer, Truhlar, and coworkers model yield small unsigned errors of 0.43-0.46 kcal/mol for 275 neutral solutes in 91 solvents.⁶ The recent Jorgensen study³² reported a mean unsigned error of 1.01 kcal/mol for 399 neutral compounds from the same experimental data set from Abraham and coworkers³ studied here.

Surprisingly the three semiempirical methods tested here yield poorer agreement with experiment 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, semiempirical r^2 from 0.34 to 0.62, and empirical r^2 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 scaled CM1A charges³² the poor results obtained here using AM1CM2 and PM3CM2 is unexpected. AM1CM2 and PM3CM2 methods should yield qualitatively similar partial charges to those obtained from the CM1A procedure. Larger errors for charged compounds reported here may arise from other differences such as the number and type of compounds tested, the functional form of the GB equation used to compute G_{polar} , or the atomic radii used in the calculations. In the present work we only considered element-based radii, with the exception of hydrogen atoms connected to C, N, O, and S (Table 1). As was the case for neutral compounds, the

AUE errors also track with the correlation coefficients. Again ab initio partial charges yield the lowest errors (6.34 to 7.30 kcal/mol) but for the charged species, semiempirical yield the largest errors (8.28 to 13.63 kcal/mol). Empirical AUEs are in the middle (7.44 to 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 these three ab initio methods cluster around 0.75 for both neutral and charged species (Figure 1). For the expanded dataset tested here, and in contrast with recently published results,³² the use of semiempirical charges with GBSA and PBSA methods yields good agreement with experiment only for neutral compounds.

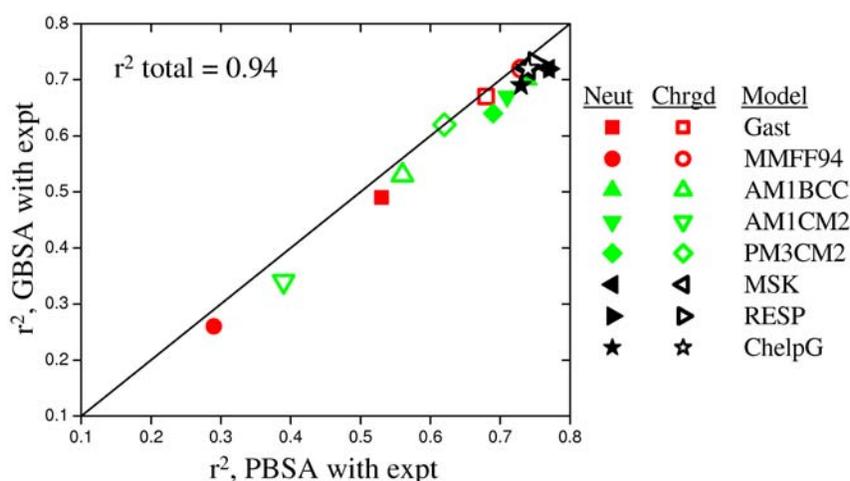


Figure 1. Comparison of correlation coefficients (r^2 values) for calculated versus experimental free energies of hydration from PBSA and GBSA calculations. For each partial charge model two r^2 values are plotted representing results for 460 neutral compounds (filled symbols) and 42 charged compounds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

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

G_{nonpolar} from Molecular SA. The constants ($\gamma = 0.00542$, $\beta = 0.92$) typically used¹⁸⁻²⁵ in MM-PBSA and MM-GBSA calculations to convert SA (\AA^2) to G_{nonpolar} (kcal/mol) are based on fitting molecular SA results to experimental ΔG_{hyd} for small straight-chain alkanes.¹² 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} 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 ($\gamma = 0.00542$, $\beta = 0.92$).¹⁸⁻²⁵ However, as a group, molecular SAs have no correlation with experimental ΔG_{hyd} (Figure 2, filled squares). Rankin et al. have reported similar results based on analysis of 210 neutral compounds.⁵⁶

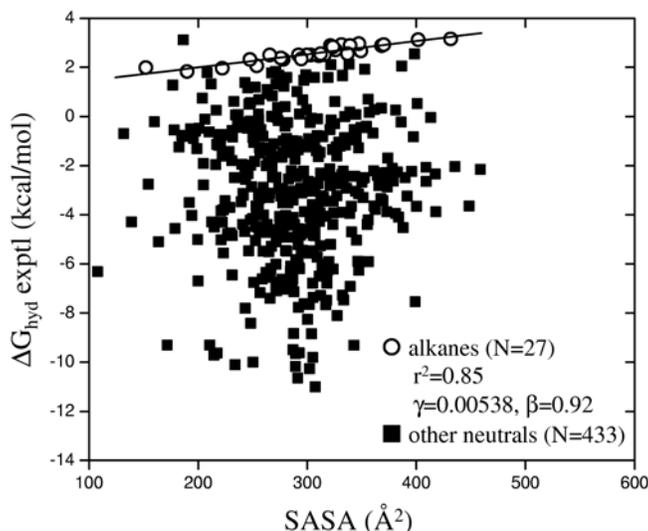


Figure 2. Experimental free energies of hydration versus total molecular solvent accessible surface area (SASA). The best fit line to the 27 linear and branched alkanes (○) yields a correlation coefficient $r^2 = 0.85$, slope $\gamma = 0.00538$, and intercept $\beta = 0.92$. Other compounds are represented as filled squares (■).

Although the results are clearly charge model dependent, in general, G_{polar} contributions should be linearly correlated with ΔG_{hyd} . This is illustrated in Figure 3 for RESP charged neutral compounds in which the polar energies (G_{polar} $r^2 = 0.77$, filled squares) were computed from PB calculations. Given that ΔG_{hyd} is estimated from the linear sum of two terms (eq 1), the sum of G_{polar} and G_{nonpolar} and both terms individually should be linear with experiment if there is to be agreement. However, using standard constants (eq 6, $\gamma = 0.00542$, $\beta = 0.92$) to compute nonpolar contributions yield no correlation (G_{nonpolar} $r^2 = 0.00$, open circles) and will therefore not contribute to any improvement or diminishment in the total correlation coefficient with experiment (ΔG_{hyd} $r^2 = 0.77$). In short, converting molecular-based SA to G_{nonpolar} energies (eq 6) for use in computing absolute free energies of hydration yields no advantage over using G_{polar} energies alone; correlations listed in Table 2 are due solely to the G_{polar} term.

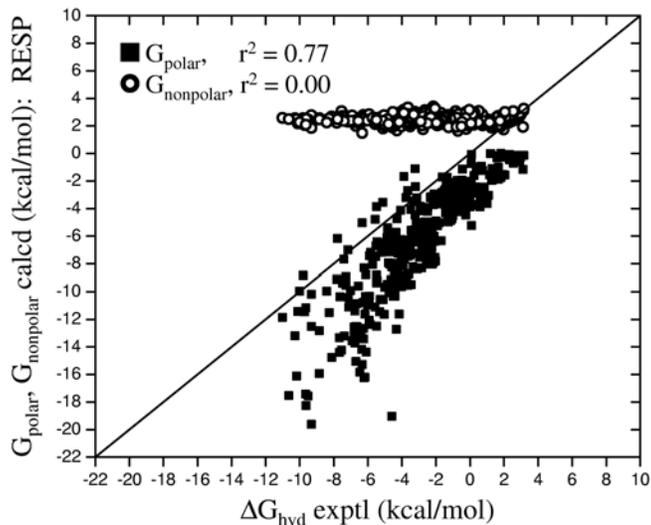


Figure 3. Correlation of individual components with experimental free energies of hydration for neutral compounds (N=460) using RESP derived partial charges. Polar (■) energies G_{polar} from PB calculations. Nonpolar (○) energies from molecular solvent accessible surface area

G_{nonpolar} from Atomic SA_i . The lack of correlation between G_{nonpolar} with ΔG_{hyd} is troublesome, especially given that molecule-based SA G_{nonpolar} energies are routinely added to G_{polar} to estimate ΔG_{hyd} (or alternatively $\Delta\Delta G_{\text{hyd}}$ for protein-ligand binding).¹⁸⁻²⁵ To improve agreement with experiment we explored a procedure to recompute G_{nonpolar} which includes calculation of atom-based SA_i and makes use of multiple linear regression fitting to determine an optimal coefficient for each SA type.^{37,38} Galliccio and Levy have reported an alternative procedure which uses atom-based coefficients to account for different solute-solvent dispersion interactions.³⁶ In the present work, optimized C_i coefficients for each mbondi type (HC, HN, HO, HS, C, N, O, F, P, S, Cl, Br) were obtained using continuum results (G_{polar} from GB or PB) for each of the eight charge models via multiple linear regression fittings of eq 7. After the fittings, new G_{nonpolar}

contributions were recomputed using the atom-based constants (C_i 's) so that optimized ΔG_{hyd} could be compared with experiment.

Initially, fits were pursued using only the neutral compounds. The resultant SA_i constants however lead to compute ΔG_{hyd} for the charged (± 1) species in poor agreement with experiment. These tests revealed that including both neutral and charged compounds together in the fitting procedure would yield the best overall agreement with experiment (Table 3), and this protocol was adopted for subsequent parameter optimizations.

Table 3. Correlation coefficients (r^2) and average unsigned errors (AUE) for experimental^a vs. calculated^b (PBSA fit or GBSA fit) free energies of hydration (ΔG_{hyd}). Nonpolar contributions obtained using atom-based solvent accessible surface areas with optimized constants.^c Energies in kcal/mol.

Charge Model	Neutral molecules, N = 460				Charged (± 1) molecules, N = 42			
	Part I		Part I		Part II		Part II	
	r^2 PBSA	AUE	r^2 GBSA	AUE	r^2 PBSA	AUE	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

^aSee supporting information for experimental references. ^bCalculated values obtained using eq 1. G_{polar} from either PB or GB calculations. ^c $G_{\text{nonpolar}} = \sum SA_i * C_i$.

In most cases, utilizing optimized SA_i constants to estimate nonpolar terms improves agreement with ΔG_{hyd} experiment (Table 3 versus 2). However, substantial improvement in AUE and correlations for charged species are coupled to diminishment in r^2 for neutral compounds that have utilized semiempirical partial charges (AM1BCC,

AM1CM2, PM3CM2). This diminishment is not surprising given that the semiempirical models originally performed quite poorly for charged compounds (Table 2, Part II, $r^2 = 0.34\text{--}0.62$). Optimization of SA_i constants using all data (neutral and charged) corrects for differences between experiment and theory in an average way. The primary motivation for using atom-based SA_i instead of molecule-based SA procedures is to reduce errors with respect to experiment in three ways: (1) remedy gross deficiencies a particular charge model may have (r^2 and AUE), (2) fine tune already reasonable agreement with experiment (primarily AUE), or (3) account for non-optimal radii.

Conclusion

Absolute free energies of hydration have been computed using continuum PBSA and GBSA methods for comparison with experiment for 460 neutral compounds, 42 polyatomic ions, and 11 monoatomic 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 better agreement with experiment for both neutral and charged species (Table 2, Figure 2). Correlations coefficients with experiment using MSK, RESP, and ChelpG charges with GBSA yield r^2 values between 0.69–0.73 and PBSA r^2 between 0.72–0.77. The use of semiempirical (AM1BCC, AM1CM2, PM3CM2) and empirical (Gast, MFF94) charge schemes yielded mixed results dependant on whether the compounds were charged or neutral (Table 2).

Examination of polar and non-polar energy components show that G_{nonpolar} energies derived from molecule-based SA's and standard conversion constants have no correlation with experimental results (Figure 3). The lack of a universal SA constant stems from the erroneous assumption that all exposed atoms contribute equally to the

non-polar energy. In the present work, improved correlations with experiment were obtained through optimization of atom-based SA's constants using multiple linear regression fits to the difference in experimental free energies and polar energy terms obtained from continuum calculations (Table 3). Overall, use of atom-based SA_i instead of molecule-based SA constants significantly reduce both relative (r^2) and absolute unsigned errors (AUE) with respect to experiment by eliminating any gross deficiencies a particular charge model may have (Tables 2 vs. 3). In particular, AUE errors for charged species were substantially reduced (Table 3) and systematic errors can be corrected. The best agreement with experiment for atom-based results is once again obtained using ab initio partial charge models; improved r^2 values are between 0.69–0.78 and 0.79–0.81 from GBSA and PBSA calculations respectively (Table 3).

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

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 primary motivation here was to determine generally useful force-field parameters for estimating changes in free energy of hydration associated with molecular recognition for use in MM-PBSA, MM-GBSA, and docking calculations. Based on comparison with experimental ΔG_{hyd} for more than 500 diverse organic molecules, MSK,

RESP, or ChelpG partial atomic charges obtained from ab initio calculations using 6-31G* wave functions can be recommended for both charged and neutral species over the more approximate alternatives tested.

Acknowledgments. 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. Tiba Aynechi was supported by a UC president's dissertation year fellowship.

References

- (1) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chemical Reviews* **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. Deliv. 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 non-bonding 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) 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.
- (8) Jorgensen, W. L. Free Energy Calculations: A Breakthrough For Modeling Organic Chemistry in Solution. *Accounts Chem. Res.* **1989**, *22*, 184-189.
- (9) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395-2417.
- (10) Jorgensen, W. L.; Ravimohan, C. Monte Carlo Simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050-3054.
- (11) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161-2200.
- (12) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978-1988.
- (13) Still, W. C.; Tempezyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation For Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.

- (14) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., 3rd
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.
- (15) 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.
- (16) 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.
- (17) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum
solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug
Discov. Design* **2000**, *18*, 113-135.
- (18) 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.
- (19) 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.

- (20) 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.
- (21) 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.
- (22) 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.
- (23) 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.
- (24) 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.
- (25) 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.
- (26) Abraham, M. H.; Andonianhaftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. Hydrogen Bonding. Part 34. The Factors That Influence the Solubility of Gases

- and Vapours in Water at 298 K, and a New Method for Its Determination. *J. Chem. Soc.-Perkin Trans. 2* **1994**, 1777-1791.
- (27) Marcus, Y. Thermodynamics of Solvation of Ions, Part 5. Gibbs Free Energy of Hydration at 298.15 K. *J. Chem. Soc.-Faraday Trans.* **1991**, 87, 2995-2999.
- (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. Parameterization 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) 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.; 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. *AMSOL version 6.6*; University of Minnesota: Minneapolis, Minnesota.
- (34) 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.
- (35) 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.
- (36) 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.
- (37) Eisenberg, D.; McLachlan, A. D. Solvation Energy in Protein Folding and Binding. *Nature* **1986**, *319*, 199-203.
- (38) 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. USA* **1987**, *84*, 3086-3090.
- (39) 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.
- (40) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275-291.
- (41) *AMBER Version 7*; University of California at San Francisco: San Francisco, CA.
- (42) 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.

- (43) 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.
- (44) *DMS*; University of California at San Francisco Computer Graphics Laboratory: San Francisco, CA.
- (45) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441-451.
- (46) *MOE Version 2002.03*; Chemical Computing Group: Montreal, Canada.
- (47) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219-3228.
- (48) Halgren, T. A. Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490-519.
- (49) 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.
- (50) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431-439.
- (51) 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.

- (52) 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.
- (53) 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.
- (54) 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.
- (55) 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.

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