JCTC Journal of Chemical Theory and Computation

Improving the Accuracy of the Linear Interaction Energy Method for Solvation Free Energies

Martin Almlöf,[†] Jens Carlsson,[†] and Johan Åqvist*

Department of Cell and Molecular Biology, Biomedical Center, Uppsala University, Box 596, SE-751 24 Uppsala, Sweden

Abstract: A linear response method for estimating the free energy of solvation is presented and validated using explicit solvent molecular dynamics, thermodynamic perturbation calculations, and experimental data. The electrostatic contribution to the solvation free energy is calculated using a linear response estimate, which is obtained by comparison to the free energy calculated using thermodynamic perturbation. Systematic deviations from the value of $1/2$ in the potential energy scaling factor are observed for some types of compounds, and these are taken into account by introducing specific coefficients for different chemical groups. The derived model reduces the rms error of the linear response estimate significantly from 1.6 to 0.3 kcal/mol on a training set of 221 molecules used to parametrize the model and from 3.7 to 1.3 kcal/mol on a test set of 355 molecules that were not used in the derivation of the model. The total solvation free energy is estimated by combining the derived model with an empirical size dependent term for predicting the nonpolar contribution. Using this model, the experimental hydration free energies for 192 molecules are reproduced with an rms error of 1.1 kcal/mol. The use of LIE in simplified binding free energy calculations to predict protein−ligand binding free energies is also discussed.

1. Introduction

Understanding the solvation properties of molecules in different environments is of tremendous importance for the understanding of various biological processes such as ligand binding, protein folding, and enzyme catalysis. Since a majority of these processes involve molecules in the aqueous phase, reliable estimates of hydration energies are crucial in order to accurately estimate the involved thermodynamic properties. Microscopic simulations can provide a detailed description of molecular interactions and are an efficient means of estimating the free energy changes of these processes. The most rigorous approaches to estimate solvation energies using molecular dynamics (MD) or Monte Carlo (MC) simulations are the free energy perturbation (FEP) and thermodynamic integration (TI) methods. The absolute free energies of solvation can be accurately estimated using these methods in combination with appropriate thermodynamic cycles. MD or MC simulations are carried out in both gas and aqueous phase and in each calculation the free energy is typically evaluated in two separate steps. First, the nonpolar part of the solvation energy is obtained by creating the van der Waals cavity formed by the solute. Then, the electrostatic contribution is calculated by gradually turning on the partial charges of the solute atoms. The transformation process is divided into several intermediate steps, and the total change in free energy is evaluated as the sum of these. FEP and TI calculations have shown that force fields are able to reproduce experimental solvation energies quite accurately, in particular those which have been specifically parametrized for this purpose.[1−10] However, application of these methods to more complex problems, such as estimating protein−ligand binding free energies,[11−13] has been shown to be difficult due to convergence and sampling problems mainly associated with creation/annihilation of particles and the large number of simulations that has to be carried out. For these reasons, a faster method for obtaining free energies of solvation would be extremely valuable.

---

* Corresponding author phone: +46 18 471 4109; fax: +46 18 53 69 71; e-mail: aqvist@xray.bmc.uu.se.
† These authors contributed equally to this work.

Prediction of Solvation Free Energies

J. Chem. Theory Comput., Vol. 3, No. 6, 2007 **2163**

The nonpolar contribution to the free energy of solvation is often estimated using linear relationships between solute size measures, e.g., molecular or solvent accessible surface area (MS or SASA), and free energies of solvation that have been observed for nonpolar compounds.[6,14-16] For the electrostatic contribution, there are a number of approaches that are based on a solvent linear response (LR) assumption. A classic example of this is the Born equation, which predicts solvation free energies of ions from a quadratic dependence on the ion charge and an inverse dependence on its radius.[17] Other continuum dielectric approaches employing LR to estimate solvation free energies are the Poisson−Boltzmann[18,19] and Generalized Born[20] methods. Another useful approach to estimate the electrostatic solvation free energy contribution from microscopic simulations, which also is derived from a LR assumption, is based on collecting the average electrostatic solute−solvent interaction energies from MD or MC simulations. The total free energy of turning on the electrostatic solute−solvent interactions can be approximated as

$$\Delta F_{el} = \beta[\langle U_{r-s}^{el}\rangle_{on} + \langle U_{r-s}^{el}\rangle_{off}] \quad (1)$$

where $\beta = 1/2$ and the $\langle U_{r-s}^{el}\rangle$ terms represent the average value of the solute−solvent electrostatic interaction energy evaluated by sampling with (on) and without (off) these interactions turned on.[21] The second term in eq 1 can further be neglected if the solute and solvent become randomly oriented with respect to each other in the absence of electrostatic interactions (see below). The above type of relationship has been used for estimating protein−ligand binding free energies with the linear interaction energy (LIE) method[22-24] and to calculate pK_a values of protein residues.[25,26] A similar approach has also been successfully used to predict experimental solvation energies of small organic compounds.[27-29]

Åqvist and Hansson performed a detailed investigation of the accuracy of eq 1 for estimating the free energy of turning on electrostatic solute−solvent interactions for various small solutes in different polar solvents. It was found that generally $\beta < 1/2$ for neutral molecules and that systematic deviations from the theoretical value could be identified for some chemical groups, e.g., for monoalcohols $\beta = 0.37$.[21] In this work we consider an alternate thermodynamic cycle, which includes both the electrostatic intra- and intermolecular interactions in the gas and aqueous phase, and investigate how the LR approximation can be used to predict the electrostatic component of solvation free energies from microscopic simulations. First, a general formula for the total electrostatic part of the solvation free energy, which includes both intra- and intermolecular contributions, is derived. Second, the scaling factor, β, is estimated for 211 small molecules that represent common neutral and ionic chemical groups. Different models for predicting β are then discussed, and, in particular, we again identify systematic deviations in the coefficient for specific chemical groups. The derived models are validated on a test set of 361 compounds. These molecules are more flexible and contain combinations of different chemical groups, for which the validity of LR was

not extensively studied by Åqvist and Hansson.[21] The total free energy of hydration is estimated by combining the calculated charging free energies with an empirical term for the nonpolar contribution. The use of the LR approximation in the LIE method to predict protein−ligand binding free energies is also discussed.

2. Theory

A Linear Response Approximation for Estimation of the Electrostatic Contribution to Solvation Free Energies. The interactions in a system will here be described using two classical potentials U_A and U_B, which represent states A and B of a solute. In state A all electrostatic interactions involving the solute are turned off, while B represents the state where these terms are turned on (compare eq 1). In addition, the two potentials have exactly the same force field parameters for bonded and van der Waals (Lennard-Jones) terms. Hence, the transformation from state A to state B will represent the turning on of solute electrostatic interactions and the difference between the two potentials can be expressed as

$$\Delta U = U_B - U_A = U_{r-s}^{el} + U_{r-r}^{el} \quad (2)$$

where U_{r-s}^{el} and U_{r-r}^{el} are the electrostatic solute−solvent (r−s) and solute−solute (r−r) interaction energies. One approach to derive eq 1 is to start with Zwanzig's expression for the free energy difference between two states[30]

$$\Delta F_{el} = -kT\ln\langle e^{-\Delta U/kT}\rangle_A \quad (3)$$

where T is the temperature, k is Boltzmann's constant, and $\langle...\rangle_A$ is an ensemble average on state A. Herein, we will make no distinction between the Helmholtz and Gibbs free energies since the difference between the two quantities has negligibile effects on the results. The cumulant expansion[31] of eq 3 yields

$$\Delta F_{el} = \langle\Delta U\rangle_A - \frac{1}{2kT}\langle(\Delta U - \langle\Delta U\rangle_A)^2\rangle_A + \frac{1}{6(kT)^2}\langle(\Delta U - \langle\Delta U\rangle_A)^3\rangle_A + ... \quad (4)$$

The corresponding expression utilizing a configurational average on state B is

$$\Delta F_{el} = \langle\Delta U\rangle_B + \frac{1}{2kT}\langle(\Delta U - \langle\Delta U\rangle_B)^2\rangle_B + \frac{1}{6(kT)^2}\langle(\Delta U - \langle\Delta U\rangle_B)^3\rangle_B + ... \quad (5)$$

Adding eqs 4 and 5 and discarding terms of order three and higher yields

$$\Delta F_{el} = \frac{1}{2}[\langle\Delta U\rangle_A + \langle\Delta U\rangle_B] - \frac{1}{4kT}\langle(\Delta U - \langle\Delta U\rangle_A)^2\rangle_A + \frac{1}{4kT}\langle(\Delta U - \langle\Delta U\rangle_B)^2\rangle_B \quad (6)$$

Furthermore, if equal fluctuations of the energy gaps are assumed (i.e., the parabolic free energy functions corresponding to states A and B have equal curvatures), the free energy can be evaluated from the averages of ΔU sampled

on states A and B, which gives us the linear response estimate of the free energy of charging

$$\Delta F_{el} = \frac{1}{2}[\langle\Delta U\rangle_A + \langle\Delta U\rangle_B] \qquad (7)$$

To estimate the electrostatic contribution to the free energy of hydration, the difference between the water and gas-phase free energy must be taken

$$\Delta F_{sol}^{el} = \Delta F_{el}^{w} - \Delta F_{el}^{g} =$$
$$\frac{1}{2}[\langle\Delta U\rangle_A^w + \langle\Delta U\rangle_B^w - \langle\Delta U\rangle_A^g - \langle\Delta U\rangle_B^g] \quad (8)$$

where superscripts w and g indicate that the averages are taken in aqueous and gas phase, respectively. Separation of eq 8 into intra- (r−r) and intermolecular (r−s) energies gives

$$\Delta F_{sol}^{el} = \frac{1}{2}[\langle U_{r-s}^{el}\rangle_A^w + \langle U_{r-s}^{el}\rangle_B^w + \langle U_{r-r}^{el}\rangle_A^w -$$
$$\langle U_{r-r}^{el}\rangle_A^g + \langle U_{r-r}^{el}\rangle_B^w - \langle U_{r-r}^{el}\rangle_B^g] \quad (9)$$

Equation 9 can be further simplified by noting that $\langle U_{r-s}^{el}\rangle_A^w$ is close to zero.[21] This average is calculated from a simulation carried out on potential A, where the electrostatic solute interactions are switched off. The solvent will therefore in general be randomly oriented with respect to the solute, and the net electrostatic contribution to the solute−solvent interaction energy will be close to zero (note, however, that ionic solutes are an exception that will be further addressed below[21,32−34]). In addition, it could be expected that the solute conformations will be similar in both phases when sampling is carried out on state A, i.e. $\langle U_{r-r}^{el}\rangle_A^w - \langle U_{r-r}^{el}\rangle_A^g \approx 0$. Introducing these two approximation results in a LR approximation of the electrostatic contribution to the free energy of solvation

$$\Delta F_{sol}^{el} = \frac{1}{2}(\langle U_{r-s}^{el}\rangle_B^w + \Delta\langle U_{r-r}^{el}\rangle_B) = \beta(\langle U_{r-s}^{el}\rangle_B^w + \Delta\langle U_{r-r}^{el}\rangle_B) \quad (10)$$

where $\Delta\langle U_{r-r}^{el}\rangle_B$ is the difference between the intramolecular energies in aqueous and gas phase and $\beta = 1/2$. For rigid molecules, eq 10 can be further simplified. In this case $\Delta\langle U_{r-r}^{el}\rangle_B = 0$ and gives

$$\Delta F_{sol}^{el} = \beta\langle U_{r-s}^{el}\rangle_B^w \qquad (11)$$

While this expression requires the approximation $\Delta\langle U_{r-r}^{el}\rangle_B = 0$ using the thermodynamic cycle employed in the present work, this is not the case for the cycle used by Åqvist and Hansson.[21] With their approach, in which only solute−solvent interactions are perturbed, a different thermodynamic cycle can be used to calculate the free energy of solvation. A combined thermodynamic cycle, which illustrates the difference between the two approaches for calculating the total solvation free energy, is shown in Figure 1. The vertical legs in Figure 1 can be estimated using the linear response approximation and, depending on if the intramolecular energies are included in the perturbation, the two cycles differ slightly in how the nonpolar contribution would be calculated. In the upper cycle, the nonpolar contribution is the free energy of transferring the solute between the phases with
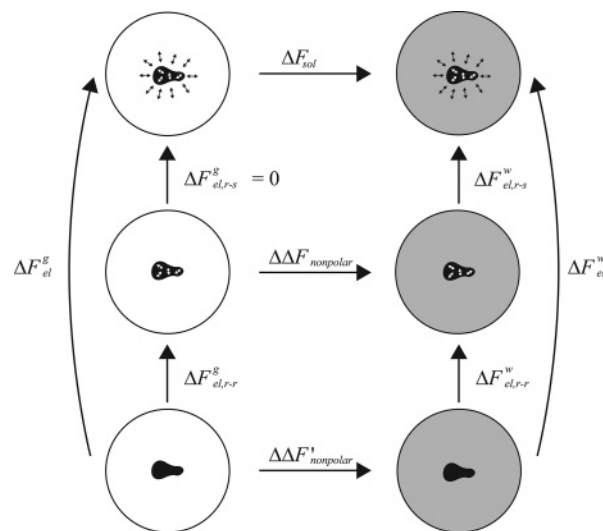

Figure 1. A thermodynamic cycle describing how the free energy of hydration can be obtained from the linear response approximation derived in this work and the approach devised by Åqvist and Hansson.[21] The upper cycle corresponds to the approach of Åqvist and Hansson, from which the total free energy of hydration can be estimated from $\Delta F_{sol} = \Delta F_{el,r-s}^w + \Delta\Delta F_{nonpolar}$, where $\Delta F_{el,r-s}^w$ is the free energy of turning on the intermolecular electrostatic interactions in water. In the derivation presented in this work the entire cycle is used, which yields the hydration energy $\Delta F_{sol} = \Delta F_{el,r-s}^w + \Delta F_{el,r-r}^w - \Delta F_{el,r-r}^g + \Delta\Delta F_{nonpolar}$. The top row represents states in which all interactions are turned on, while the middle row represents states in which the electrostatic solute−solvent interactions are turned off and all other interactions are on. The bottom row represents states where all electrostatic interactions involving the solute are turned off.

its intermolecular electrostatic interactions turned off, while in the complete cycle (used here) it is the free energy of transferring the solute with all electrostatic interactions involving the solute turned off.

Models for Including Systematic Deviations in β. By calculating the free energy of turning on the electrostatic solute interactions and also extracting the corresponding average electrostatic solute−solvent and solute−solute interaction energies in gas and aqueous phase, the value of the β coefficient can be evaluated from

$$\beta_{FEP} = \frac{\Delta F_{sol}^{el}}{\langle U_{r-s}^{el}\rangle_B^w + \Delta\langle U_{r-r}^{el}\rangle_B} \qquad (12)$$

where the subscript FEP has been added to indicate that the β coefficient is obtained from FEP calculations.

In order to investigate and account for systematic deviations in the β_{FEP} coefficient, several different models were investigated. The first model (A) uses the theoretically derived value of $\beta = 1/2$. The second model (B) was proposed by Hansson et al.[23] and is based on β values calculated for a small series of model compounds. In this model, $\beta = 1/2$ is only used for ionic compounds. For neutral molecules without any hydroxyl groups $\beta = 0.43$ is used, while 0.37 or 0.33 is used for molecules with one or several hydroxyl groups, respectively. Models C−E are new models presented

Prediction of Solvation Free Energies

J. Chem. Theory Comput., Vol. 3, No. 6, 2007 **2165**

in this work, and these are parametrized on a set of training compounds. Model C assumes that a single value of β, which is parametrized on the training set, can be used for all molecules. In models D and E, β can assume different values depending on the chemical nature of the compound. This is somewhat similar to model B, but models D and E take into account more chemical groups than model B (i.e., not just hydroxyls) and provide a method for determining β's for compounds containing a mixture of chemical groups.

In the derivation of a β value for solutes containing several chemical groups we will assume that the contributions from each group are additive, and the total electrostatic hydration free energy can then be written as

$$\Delta F_{\text{sol}}^{\text{el}} = \beta_1 [\langle U_{\text{r-s}}^{\text{el}} \rangle^{\text{w}} + \Delta \langle U_{\text{r-r}}^{\text{el}} \rangle]_1 + \beta_2 [\langle U_{\text{r-s}}^{\text{el}} \rangle^{\text{w}} + \Delta \langle U_{\text{r-r}}^{\text{el}} \rangle]_2 + \\ \beta_3 [\langle U_{\text{r-s}}^{\text{el}} \rangle^{\text{w}} + \Delta \langle U_{\text{r-r}}^{\text{el}} \rangle]_3 + ... \quad (13)$$

where $[\langle U_{\text{r-s}}^{\text{el}} \rangle^{\text{w}} + \Delta \langle U_{\text{r-r}}^{\text{el}} \rangle]_i$ is the change in electrostatic solute energy, and β_i is the FEP derived scaling factor for group i. If each scaling factor is rewritten as $\beta_i = \beta_0 + \Delta\beta_i$, eq 13 can be expressed as

$$\Delta F_{\text{sol}}^{\text{el}} = \left(\beta_0 + \frac{\sum_i [\langle U_{\text{r-s}}^{\text{el}} \rangle^{\text{w}} + \Delta \langle U_{\text{r-r}}^{\text{el}} \rangle]_i \Delta\beta_i}{\langle U_{\text{r-s}}^{\text{el}} \rangle^{\text{w}} - \Delta \langle U_{\text{r-r}}^{\text{el}} \rangle} \right) (\langle U_{\text{r-s}}^{\text{el}} \rangle^{\text{w}} - \\ \Delta \langle U_{\text{r-r}}^{\text{el}} \rangle) = \beta (\langle U_{\text{r-s}}^{\text{el}} \rangle^{\text{w}} - \Delta \langle U_{\text{r-r}}^{\text{el}} \rangle) \quad (14)$$

From this expression, a β for the total molecule can be estimated if the electrostatic energies for each group are known. In order for this expression to be useful, however, weighting factors predetermined from FEP calculations are introduced. The molecular β coefficient can now be identified as

$$\beta = \beta_0 + \frac{\sum_i w_i \Delta\beta_i}{\sum_i w_i} \quad (15)$$

where w_i, β_0, and $\Delta\beta_i$ are derived from explicit solvent FEP calculations of single chemical groups. To test this approach two different models are investigated. Model D uses weighting factors that are equal to the average solute potential energies of each chemical group in the training set. The second model (E) uses $w_i = 1.0$ for all neutral groups and one single weighting factor for anions and cations. The latter ($w_{\text{anion/cation}}$) is determined as the ratio between the average $[\langle U_{\text{r-s}}^{\text{el}} \rangle^{\text{w}} + \Delta \langle U_{\text{r-r}}^{\text{el}} \rangle]_i$ of the ionic and neutral groups in the training set and is found to be 11.0 ± 1.7.

Prediction of the Total Free Energy of Solvation. In order to make a comparison to experiment, the nonpolar contribution to the free energy of solvation has to be added to eq 10. In approximations of the nonpolar free energy contribution it is, in most cases, assumed that there is a linear relationship between size measures and the free energy. This is based on experimental observations that solvation free energies of hydrophobic compounds, for which there should be a negligible electrostatic contribution, generally depend linearly on size measures such as surface area.[14−16] Hence,
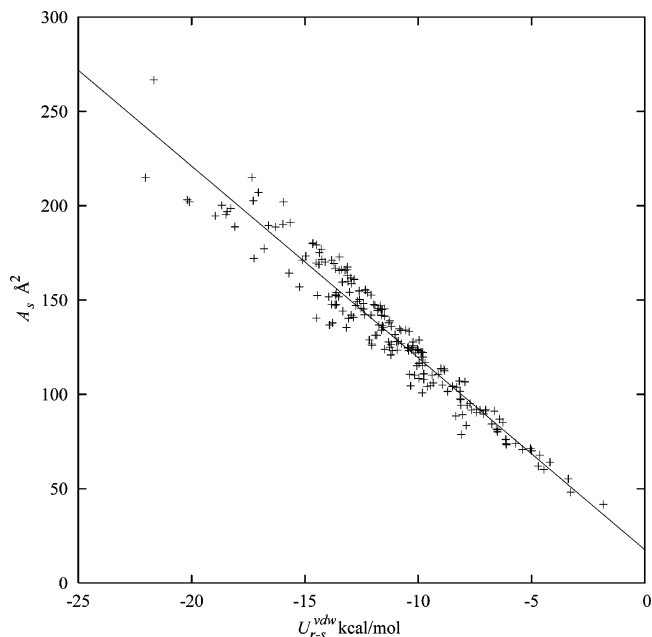

Figure 2. The average van der Waals solute−solvent interaction energy (kcal/mol) sampled on state A and the molecular surface (A_S, Å^2) for the molecules in the parametrization set. Linear regression yields the relation $A_S = -10.2 \langle U_{\text{r-s}}^{vdw} \rangle_A + 17.7$ (solid line).

these observations indicate that the nonpolar contribution to the solvation free energy can be estimated from

$$\Delta F_{\text{sol}}^{\text{np}} = a_{\text{w}} \sigma + b_{\text{w}} \quad (16)$$

where np denotes that it is the nonpolar contribution, σ is a size measure, and a_{w} and b_{w} are empirically derived parameters. By extrapolating eq 16 to zero solute size, b_{w} can be interpreted as the free energy of inserting a point particle into the solvent.[35] The first term in eq 16 corresponds to the free energy change of increasing the size of the point particle. A less empirical way of formulating eq 16 is to introduce the (macroscopic) surface tension, that determines the cost of cavity creation, together with the van der Waals or dispersion energy associated with introducing a nonpolar solute into the cavity[36,37]

$$\Delta F_{\text{sol}}^{\text{np}} = \gamma_S A_S + \langle U_{\text{r-s}}^{\text{vdW}} \rangle_A \quad (17)$$

where γ_S is the surface tension, and A_S the molecular surface area. Note, however, that eq 17 formally neglects the free energy of inserting a point particle into the solvent. The surface area is also very strongly correlated with the van der Waals interaction energy (cf. Figure 2), making the latter an equally useful size measure.[22] The regression equation obtained between molecular surface area and van der Waals energy in water (see below and Figure 2) is $A_S = -10.2 \langle U_{\text{r-s}}^{\text{vdW}} \rangle_A^{\text{w}} + 17.7$ in kcal/mol and Å^2, which together with the experimental value for the surface tension of water of 73 mN/m = 105 cal/(mol Å^2) predicts that $\Delta F_{\text{sol}}^{\text{np}} = -0.07 \langle U_{\text{r-s}}^{\text{vdW}} \rangle_A^{\text{w}} + 1.9$ kcal/mol. We will return to this prediction later on but can conclude that using the solute−solvent van der Waals energy as a size measure should give us the nonpolar contribution to the hydration energy as

$$\Delta F_{\text{sol}}^{\text{np}} = \alpha_{\text{w}}^{\text{vdW}} \langle U_{\text{r-s}}^{\text{vdW}} \rangle_B^{\text{w}} + \gamma_{\text{w}}^{\text{vdW}} \qquad (18)$$

where $\alpha_{\text{w}}^{\text{vdW}}$ and $\gamma_{\text{w}}^{\text{vdW}}$ are parameters that could either be derived empirically or be obtained as above. In accordance with the thermodynamic cycle of Figure 1, the nonpolar contribution to the hydration free energy (bottom row) should be estimated using $\langle U_{\text{r-s}}^{\text{vdW}} \rangle_A^{\text{w}}$ instead of $\langle U_{\text{r-s}}^{\text{vdW}} \rangle_B^{\text{w}}$ as a size measure. However, by using $\langle U_{\text{r-s}}^{\text{vdW}} \rangle_B^{\text{w}}$ in eq 18 the total hydration free energy can be estimated using a single simulation, and as will be shown in the Results and Discussion these two alternatives yield similar results.

A semiempirical expression for the total free energy of hydration can now be obtained by combining eqs 10 and 18

$$\Delta F_{\text{sol}} = \alpha_{\text{w}}^{\text{vdW}} \langle U_{\text{r-s}}^{\text{vdW}} \rangle_B^{\text{w}} + \beta(\langle U_{\text{r-s}}^{\text{el}} \rangle_B^{\text{w}} + \Delta \langle U_{\text{r-r}}^{\text{el}} \rangle_B) + \gamma_{\text{w}}^{\text{vdW}} \quad (19)$$

Here, the free energy of hydration can be estimated using two free parameters ($\alpha_{\text{w}}^{\text{vdW}}$ and $\gamma_{\text{w}}^{\text{vdW}}$), while β is determined according to one of the above derived models.

3. Method

Training and Test Set. The 211 compounds listed in Table 1 of the Supporting Information, hereafter referred to as the "training set", were used to parametrize the β_0's, $\Delta\beta_i$'s, and w_i's of models C−E. These compounds consist of hydrocarbon groups and at most one non-hydrocarbon moiety. For the set, 19 different groups were defined: alcohols, 1° amides, 2° amides, 3° amides, 1° amines, 2° amines, 3° amines, ketones/aldehydes, thiols, ethers, esters, carboxylic acids, nitriles, nitros, sulfides, anions, and cations. The parametrization of β_0's and $\Delta\beta_i$'s for the different models (C−E) was performed in a least-squares fashion by minimizing the squared error between the free energy of charging as calculated from the FEP simulations (in gas and water) and as predicted by eq 10. A second data set, referred to as the "test set", was created by combining fragments (Figure 3) consisting of the 19 types of moieties used in the training set (19·19 = 361 molecules in total). These were used to validate the models parametrized on the training set.

Molecular Dynamics and Free Energy Calculations. All molecular dynamics (MD) simulations were carried out with the program Q[38] using the OPLS all atom (OPLS-AA) force field.[39] The simulations were carried out at 300 K in an 18 Å sphere centered on the geometrical center of the solute, and the system was solvated with TIP3P[40] water molecules. Water molecules were subjected to radial and polarization restraints according to the SCAAS method.[38,41] A nonbonded cutoff of 10 Å was used, and long-range electrostatic interactions were treated with the local reaction field multipole expansion method,[42] except for the solute interactions that were calculated without any cutoff. The time step was set to 1 fs, and nonbonded pair lists were updated every 25 steps.

Electrostatic free energies of solvation were determined using the FEP method. In this method the free energy of charging the solute is determined by simulating several intermediate states of the charged and uncharged solute. The potentials governing the intermediate states are defined by $U_m = \lambda_m U_A + (1 - \lambda_m)U_B$ where A and B represent the uncharged and charged solute, respectively, and λ_m is a

mapping parameter which varies from $\lambda_1 = 0$ to $\lambda_n = 1$. The free energy difference between state A and B can then be calculated by summing up the free energy differences of the intermediate states as calculated using the Zwanzig expression (eq 3).

$$\Delta F_{A \to B} = -kT \sum_{m=1}^{n-1} \ln \langle e^{-(U_{m+1} - U_m)/kT} \rangle_m \qquad (20)$$

Each calculation comprised a 16 ps heating scheme followed by 50 ps of equilibration and then 41 data collection simulations at evenly spaced λ-values. The trajectories at the FEP endpoints ($\lambda=0$ and 1) were 200 ps in length, while at intermediate λ-values they were 20 ps in length. Energies were extracted every 25 steps, and the simulations were carried out for the charging and uncharging process in both gas and aqueous phases. Energies from the first 5 ps of each λ-simulation were discarded in the FEP calculations. Due to slower convergence for the zwitterionic compounds these simulations were run considerably longer (1 ns at each λ-value).

Standard errors of the FEP calculations were estimated as half the difference between the free energy of charging and uncharging the compound. The standard errors of the electrostatic potential energies were estimated as half the difference in average potential energy between the charging and uncharging simulations of the appropriate endpoint.

4. Results and Discussion

Parametrization of β. The electrostatic contribution to the free energy of solvation, electrostatic solute energies, and the calculated value of β_{FEP} for each molecule in the training set are presented in Table 1 of the Supporting Information. For the alkanes, the β_{FEP} values vary widely from -1.1 to 2.3, but this is simply due to the small electrostatic energies in these cases, which results in large errors when the quotient in eq 12 is evaluated. The β_{FEP} values for the other compounds are very similar to those obtained by Åqvist and Hansson.[21] In order to evaluate the LR approximation, several models were tested, and these are summarized in Table 1. The models differ from each other in that they have different rules to determine what value of β to use in eq 10. The model based on the theoretically derived $\beta = \frac{1}{2}$ (model A) consistently overpredicts the absolute charging energy for anions and neutral solutes (rms = 1.64 kcal/mol), which indicates that lower values of β are necessary for these compounds. For cations, on the other hand, the estimated free energies are more positive than those calculated using FEP, which indicates that higher values of β are required to reproduce the FEP calculations. It should be noted that the deviations from $\beta = \frac{1}{2}$ for ionic and neutral molecules have different origins. For the neutral molecules, the lower values of β were found to arise from nonquadratic free energy functions with unequal curvatures[21] (approximation in eqs 6 and 7). For the ionic molecules, on the other hand, the deviations from $\beta = \frac{1}{2}$ are primarily a result of neglecting the contribution from $\langle U_{\text{r-s}}^{\text{el}} \rangle_A^{\text{w}}$ [21,32,34] (one of the approximations made in going from eq 9 to 10). The results for model A are shown in Figure 4.

In the work of Hansson et al. it was suggested that $\beta = 0.43$ was appropriate for most neutral molecules except

Table 1. Models for Predicting β Studied in This Work

model	treatment of β and w_i	rms training[a]	rms test[a]
A	$\beta = 0.5$	1.64	3.72
B	β dependent on number of hydroxyls and net charge as in ref 23	1.21	3.29
C	one β parametrized for entire training set ($\beta = 0.48$)	1.52	3.68
D	$$\beta = \beta_0 + \frac{\sum_i w_i \Delta b_i}{\sum_i w_i} \quad w_i \propto \langle U_i^{el} \rangle_B$$	0.32	1.22
E	$$\beta = \beta_0 + \frac{\sum_i w_i \Delta b_i}{\sum_i w_i} \quad w \begin{cases} 1 \text{ for net charge} = 0 \\ 11 \text{ for net charge} = \pm 1 \end{cases}$$	0.32	1.26

[a] In kcal/mol.
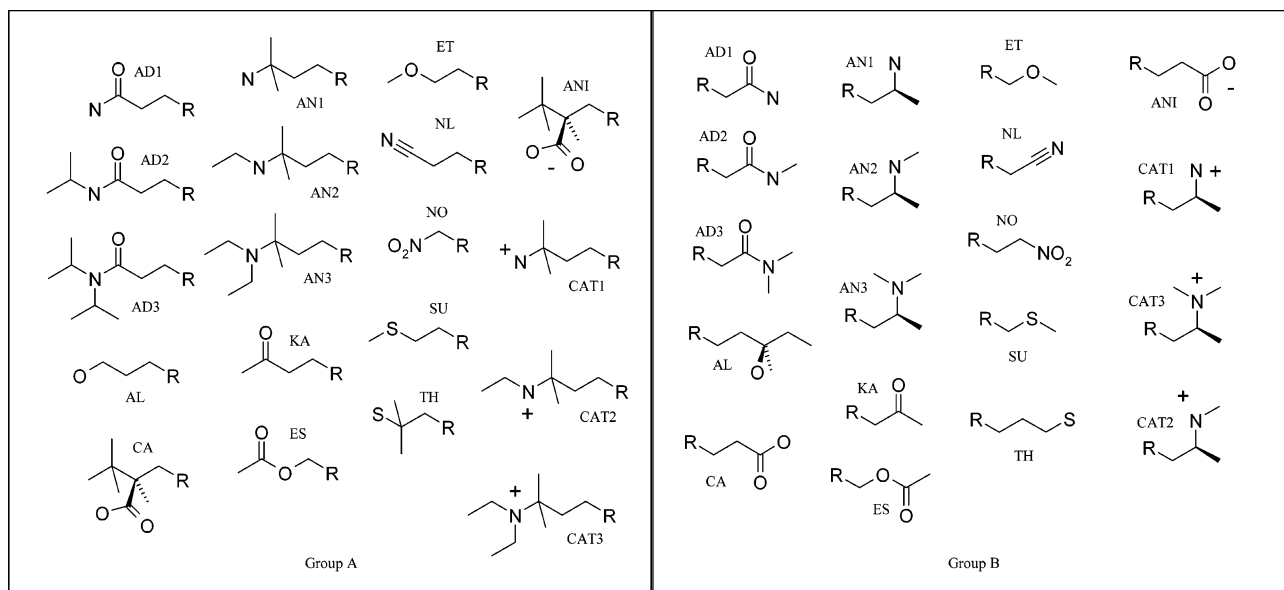
Figure 3. The fragments used to build the test set. Fragments from group A are combined with fragments from group B.

alcohols ($\beta = 0.33$ or 0.37), cations ($\beta = 0.5$), and anions ($\beta = 0.5$).[21,23] This scheme (model B) fares slightly better (rms = 1.21 kcal/mol) than using the theoretically derived $\beta = \frac{1}{2}$. It also performs slightly better compared to using a single β for the entire set (model C). For model C an optimized value of $\beta = 0.48$ was obtained, and the estimated charging free energies are again in reasonable agreement with the FEP calculations (rms = 1.52 kcal/mol).

In an attempt to improve model B, the compounds in the training set were grouped depending on their chemical nature, i.e., alcohols, 1° amides, 2° amides, 3° amides, 1° amines, 2° amines, 3° amines, ketones/aldehydes, thiols, ethers, esters, carboxylic acids, nitriles, nitros, sulfides, anions, and cations. The average signed error of using $\beta = 0.43$ in eq 10 compared to the FEP calculations was then calculated for each group of compounds. The result of using $\beta = 0.43$ is shown in Figure 5, where each group of compounds is indicated by using different symbols. The groups deviating by more than 0.5 kcal/mol were alcohols,

1° amines, 2° amines, 1° amides, carboxylic acids, cations, and anions. The charging free energies for the alcohols, 1° amines, 2° amines, 1° amides, and carboxylic acids are all underestimated, which indicates that lower values of β are required for these groups. For the anions and cations the effect of charging the solutes is underestimated, suggesting $\beta > 0.43$ is necessary. Åqvist and Hansson showed that the lower values of β obtained for alcohols derives from their ability to form hydrogen bonds to the solvent.[21] Therefore it was not surprising that other hydrogen bond donating groups (amines, amides, and acids) displayed similar properties. While the values of β that were obtained for neutral molecules have been shown to reflect deviations from LR, the $\Delta\beta_i$'s for the anions and cations are due to the fact that $\langle U_{r-s}^{el} \rangle_A^w$ is not negligible.[21] However, compared to $\langle U_{r-s}^{el} \rangle_B^w$ the contribution from $\langle U_{r-s}^{el} \rangle_A^w$ is relatively small, and since the contribution is of a systematic nature,[21,32–34] it can be taken into account by adjusting $\Delta\beta_i$ for these groups. For
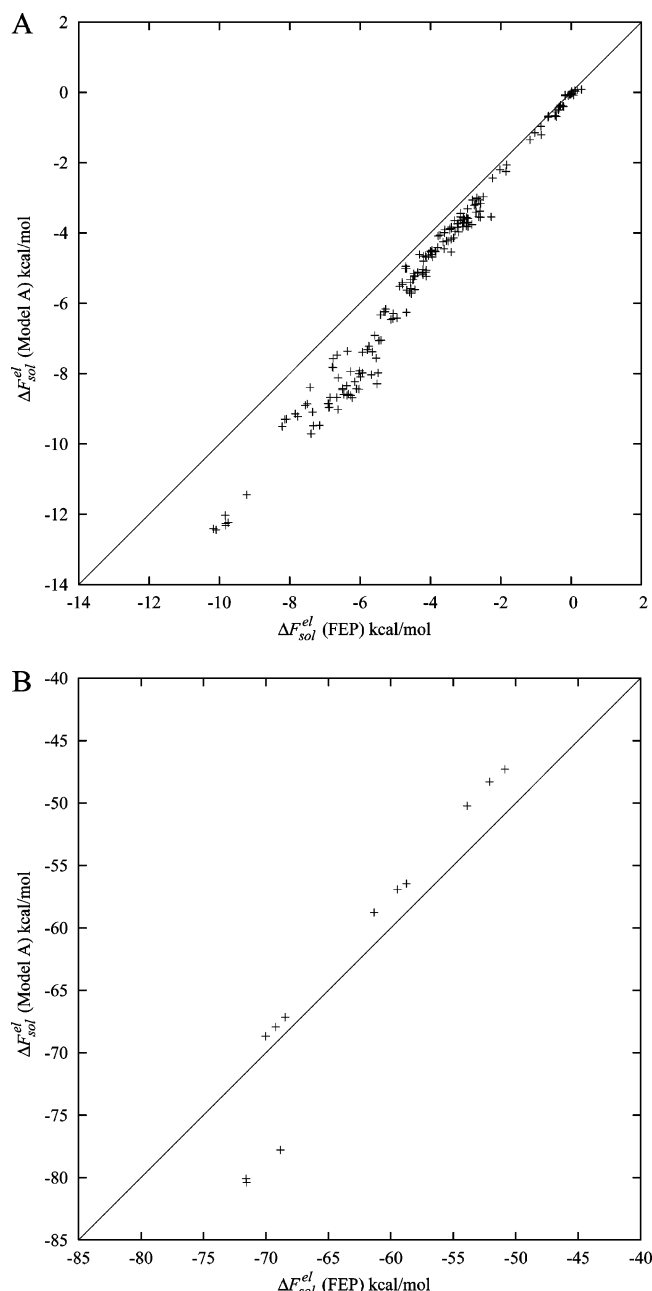
Figure 4. FEP calculated and estimated electrostatic components of the free energy of hydration (ΔF_{sol}^{el}) for model A for the neutral (A) and ionic (B) molecules in the training set. All values are in kcal/mol.

anions $\langle U_{r-s}^{el}\rangle_A^w = 8.4 \pm 0.2$ kcal/mol, and this results in a β value lower than 0.5, while the opposite effect is observed for cations ($\langle U_{r-s}^{el}\rangle_A^w = -8.7 \pm 0.2$ kcal/mol). Thus models D and E have $\beta_0 = 0.43$ and $\Delta\beta_i \neq 0$ for six chemical groups: alcohols ($i = 1$), 1° and 2° amines ($i = 2$), 1° amides ($i = 3$), carboxylic acids ($i = 4$), anions ($i = 5$), and cations ($i = 6$). Optimization of $\Delta\beta_i$ for these six groups yields $\Delta\beta_1 = -0.06$, $\Delta\beta_2 = -0.04$, $\Delta\beta_3 = -0.02$, $\Delta\beta_4 = -0.03$, $\Delta\beta_5 = 0.02$, and $\Delta\beta_6 = 0.09$ (Table 2). For models D and E (Figure 6) the rms for the full data set is in remarkably good agreement with the FEP calculations (rms = 0.32 kcal/mol) and is significantly better than models A−C. Note that weighting factors are not necessary for the parametrization set since each molecule only contains one of the defined
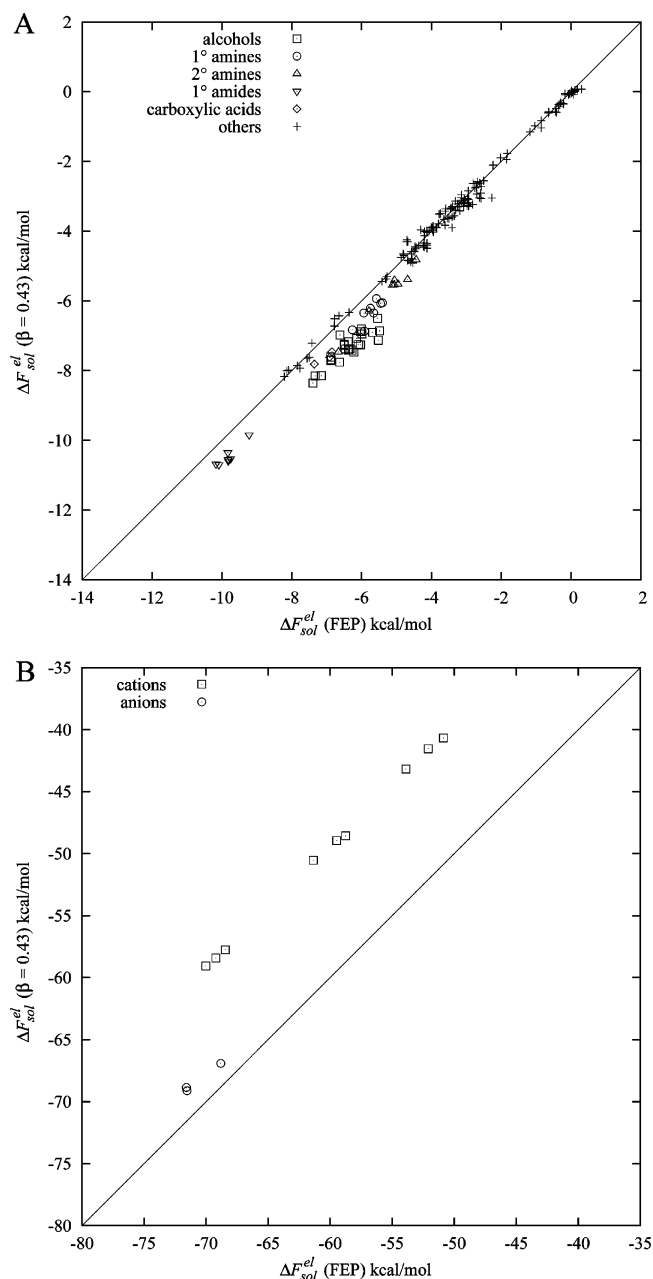
Figure 5. FEP calculated and estimated electrostatic components of the free energy of hydration (ΔF_{sol}^{el}) using $\beta = 0.43$ for the neutral (A) and ionic (B) molecules in the training set. All values are in kcal/mol.

Table 2. Obtained Parameters for Models D/E

β_0	0.43
$\Delta\beta_1$(alcohols)	−0.06
$\Delta\beta_2$(1°, 2°-amines)	−0.04
$\Delta\beta_3$(1°-amides)	−0.02
$\Delta\beta_4$(COOH)	−0.03
$\Delta\beta_5$(anions)	0.02
$\Delta\beta_6$(cations)	0.09
$\Delta\beta_7$(other)	0

chemical groups, i.e., models D and E are equivalent for the training set. The a priori assumption of assigning $\beta_0 = 0.43$[23] was tested by optimizing all parameters ($\Delta\beta_1$, $\Delta\beta_2$, $\Delta\beta_3$, $\Delta\beta_4$, $\Delta\beta_5$, $\Delta\beta_6$, and β_0) and yielded an optimal β_0 of 0.43,

Prediction of Solvation Free Energies

J. Chem. Theory Comput., Vol. 3, No. 6, 2007 **2169**

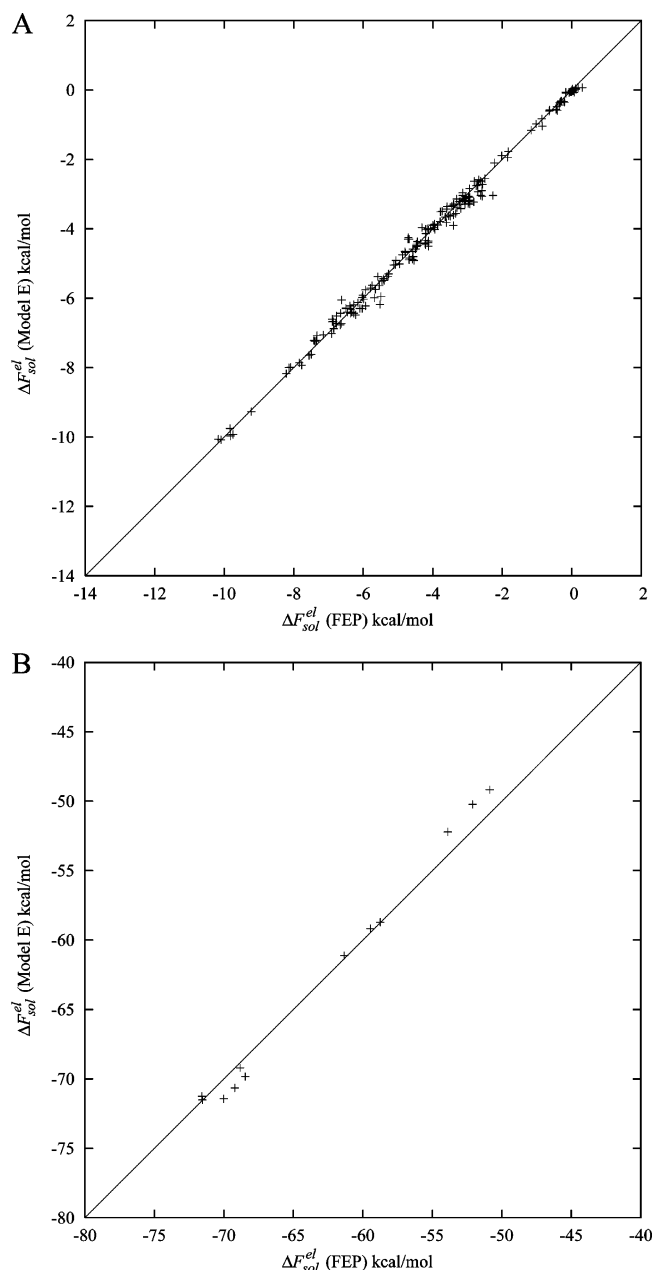**Figure 6.** FEP calculated and estimated electrostatic components for the free energy of hydration (ΔF_{sol}^{el}) for model E for the neutral (A) and ionic (B) molecules in the training set. All values are in kcal/mol.

confirming the results of Åqvist and Hansson.[21] Using $\beta_0 = 0.43$ instead of $\beta_0 = 0.50$ also has its practical reasons since it allows for a smaller number of $\Delta\beta_i$'s.

Validation of the Parametrization. In order to test the proposed schemes to estimate β_{FEP} for molecules containing mixed chemical groups we calculated charging free energies for a test set of compounds, which consisted of all possible pairwise combinations of group A and B in Figure 3. The results from these calculations are shown in Table 2 of the Supporting Information. In total there were 361 test compounds comprising combinations of all chemical groups in the training set. To elucidate which model is most useful and predictive, models A–E (as parametrized on the training set) were tested. The results for the different models are

summarized in Table 1. In this case models D and E use the values of β_0 and $\Delta\beta_i$ parametrized on the training set, but the models have different weighting factors (w_i). In model D the weighting factors were taken as the average solute electrostatic energies for each group. This is compared to using $w_i = 1.0$ for all neutral groups and $w_i = 11.0$ for the anions and cations in model E. Note that even though models D and E only have nonzero $\Delta\beta_i$'s for six types of groups, the other chemical groups will still influence the estimation of β through their weighting. For example, a compound containing only one alcohol moiety will receive an estimated β of 0.37 (0.37 = 0.43 + (−0.06)/1, using $w_i = 1$) in model E, while a compound containing an alcohol and keto moiety will receive an estimated β of 0.40 (0.40 = 0.43 + (−0.06 + 0.00)/2). A problematic case is the six zwitterions in the test set. The additivity assumed in eq 13 might not hold for these combinations due to the field canceling effect of the opposite charges and/or strong electrostatic intramolecular interactions. Therefore the results for these molecules will be presented separately at the end of this section.

The rms errors for models A–E on the test set (355 combinations in total, zwitterions excluded) are 3.72, 3.29, 3.68, 1.22, and 1.26 kcal/mol, respectively. The differences in rms between the models tested on the test set should not be seen as a statistical effect of increasing the number of free parameters since the parametrizations were performed on a different set of compounds. Thus it is clear that models D and E outperform the other models and hence 1°, 2° amine, 1° amide, alcohol, carboxylic acid, anion, and cation moieties need different β coefficients than other compounds. The introduced complexity of using specific weighting factors for each group of the neutral moieties does not seem justified considering that model E yields very similar accuracy. The best model is hence considered to be model E, and the result of using this approach is shown in Figure 7.

Westergren et al. recently suggested that the deviations in LR observed by Åqvist and Hansson[21] were due to neglect of the change in solute−solvent van der Waals interactions upon charging of the solute.[36] Instead they proposed that ΔF_{sol}^{el} should be approximated by the expression

$$\Delta F_{sol}^{el} = \frac{1}{2} \langle U_{r-s}^{el} \rangle_B^w + \Delta \langle U_{r-s}^{vdW} \rangle^w \qquad (21)$$

where $\Delta \langle U_{r-s}^{vdW} \rangle^w$ is the difference between the intermolecular van der Waals energies in states B and A. However, as can be seen from Zwanzig's expression (eq 3) and from the derivation of the LR approximation, all explicit contributions from the change in van der Waals interactions cancel when the difference between U_A and U_B is taken (while they, of course, implicitly affect the Boltzmann factor in all ensemble averages). Thus, there appears to be no theoretical support for additionally including van der Waals energy differences upon charging as was done by Westergren et al. (eq 21). Nevertheless, this ad hoc approximation of the electrostatic contribution is somewhat better than the strict LR result, yielding an rms of 3.53 kcal/mol compared to 3.72 kcal/mol using model A.

For the zwitterions in the test set, which were excluded from the above analysis, $\beta = 0.48$ is obtained using model
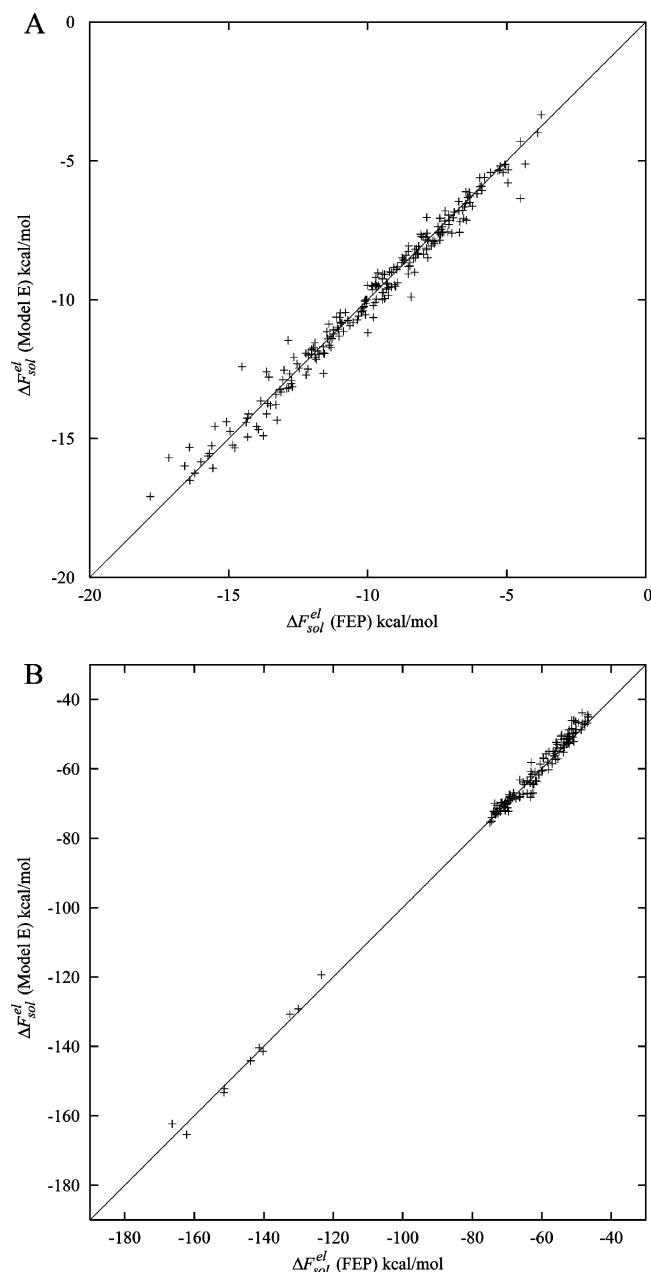
A

B

Figure 7. FEP calculated and estimated electrostatic components of the free energy of hydration (ΔF_{sol}^{el}) for model E for the neutral (A) and ionic (B) molecules in the test set. All values are in kcal/mol.

E, and this yields an rms error of 13.1 kcal/mol. The predicted absolute electrostatic component of the solvation energy is overestimated in each case and shows that lower β values are appropriate for these molecules.[21] A possible explanation is that $\langle U_{r-r}^{el}\rangle_A^w - \langle U_{r-r}^{el}\rangle_A^g$ is not negligible, which was assumed in the derivation of eq 10. However, including this term in eq 10 does not improve the results significantly (rms = 12.5 kcal/mol). Optimizing the coefficient gives $\beta = 0.39$, which is similar to that obtained for neutral compounds, and reduces the rms error significantly (rms = 3.8 kcal/mol). Further analysis, however, indicates that the zwitterions do not behave like the other neutral solutes. A plausible explanation is the inability of LR to properly describe the free energy differences associated with the large conformational changes upon charging flexible
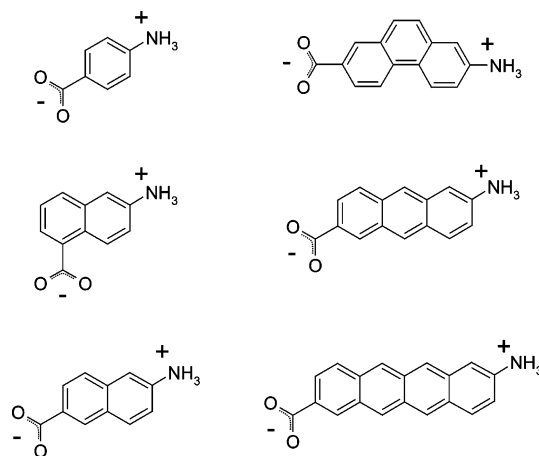

Figure 8. The compounds used to investigate the linear response approximation for rigid zwitterions.

Table 3. Electrostatic Solute Energies ($\langle U_{r-s}^{el}\rangle_B^w$ and $\Delta\langle U_{r-r}^{el}\rangle_B$) and Electrostatic Contribution to the Free Energies of Hydration from FEP Calculations (ΔF_{sol}^{el}(FEP)) for the Rigid Zwitterions (Figure 6)

solute	ΔF_{sol}^{el}(FEP)[a,b]	$\langle U_{r-s}^{el}\rangle_B^w$ [a,c]	$\Delta\langle U_{r-r}^{el}\rangle_B$ [a,d]	β_{FEP}
ZW1	−101.8	−220.2	1.7	0.47
ZW2	−116.3	−249.2	1.2	0.47
ZW3	−108.7	−234.0	1.6	0.47
ZW4	−125.3	−266.1	1.0	0.47
ZW5	−125.6	−267.3	0.9	0.47
ZW6	−131.9	−280.0	0.8	0.47

[a] kcal/mol. [b] Average uncertainties are 0.0 and 0.4 kcal/mol for the gas and water phase, respectively. [c] Average uncertainties are 0.7 kcal/mol. [d] Average uncertainties are 0.1 and 0.0 kcal/mol for the gas and water phase, respectively.

zwitterions (to some extent this probably also applies to FEP calculations in general, where charging of multiply ionized flexible compounds can be associated with severe convergence problems). In order to test this hypothesis, the electrostatic contribution to the free energies of solvation was calculated for a series of rigid zwitterions (Figure 8), for which the intramolecular energy contribution will be negligible. In excellent agreement with model E, the calculated β_{FEP} values for the rigid zwitterions are all 0.47 (Table 3). The observed differences in β_{FEP} values for flexible and rigid zwitterions suggests an alternative formulation of the LR approximation for solvation free energies, with intra- and intermolecular contributions scaled separately using the expression

$$\Delta F_{sol}^{el} = \beta^{inter}\langle U_{r-s}^{el}\rangle_B^w + \beta^{intra}\Delta\langle U_{r-r}^{el}\rangle_B \qquad (22)$$

where β^{inter} and β^{intra} are scaling factors for inter- (r−s) and the change in intramolecular (r−r) energies, respectively. For the flexible zwitterions, this yields $\beta^{inter} = 0.48$, in excellent agreement with model E, and a $\beta^{intra} = 0.66$, with an rms of 1.9 kcal/mol. This clearly shows that the anomalous β_{FEP} values obtained for the flexible zwitterions originate from inaccuracies in the LR assumption for solute−solute energies and not from solute−solvent energies. Adding another scaling factor to eq 10 would, of course, add more complexity to the model, and, as shown above, model E reproduces the

Prediction of Solvation Free Energies

J. Chem. Theory Comput., Vol. 3, No. 6, 2007 **2171**

FEP calculated values rather well and suggests that such a model is not necessary for nonzwitterionic compounds. In fact, using $\beta^{inter} = \beta_{model\ E}$ and parametrizing β^{intra} on the test set (excluding zwitterions) yields $\beta^{intra} = 0.48$, but this approach does neither improve nor diminish the agreement with the electrostatic component of the solvation free energies calculated using FEP. Hence, eq 10 is preferable to eq 22 for all molecules except flexible zwitterions, for which separate scaling factors appear necessary in order to obtain accurate results.

All the models investigated in this work are parametrized using eq 10 and the complete thermodynamic cycle in Figure 1, which differs from the approach used by Åqvist and Hansson[21] who used eq 11 along with the upper thermodynamic cycle of Figure 1. As described above, the change in intramolecular energies in going from gas to water for the training set are close to zero for all compounds, and thus the derived β values can be expected to be identical using these two approaches. The test set, however, contains compounds which may change significantly in intramolecular energies when going from gas to water phase. In order to test the benefit, if any, of using the thermodynamic cycle in which all electrostatic solute interactions are turned on as compared to the cycle of only turning on the solute−solvent electrostatic interactions, we have performed FEP calculations using both these approaches for all the nonionic compounds in the test set. Using eq 10 along with the full thermodynamic cycle on the nonionic compounds in the test set yields an rms error of 0.20 kcal/mol, while Åqvist and Hansson's approach[21] using eq 11 yields an rms error of 0.80 kcal/mol (data not shown). In particular, large errors are observed using Åqvist and Hansson's approach[21] for molecules which can form internal hydrogen bonds. Hence, for flexible molecules, the full thermodynamic cycle employed in this work seems to perform better.

Semiempirical Prediction of the Total Free Energy of Solvation. The total solvation free energy is estimated here by combining the predicted electrostatic part of the solvation energy with an empirical term for the nonpolar contribution. Experimental hydration free energies[43] were available for 194 of the molecules in the training set, and net neutral and ionic molecules are presented separately.

Using eq 19 combined with β values from model E to calculate hydration free energies of the neutral compounds results in a parametrization of $\alpha_w^{vdW} = 0.01$ and $\gamma_w^{vdW} = 1.18$ kcal/mol with an rms of 1.1 kcal/mol (Figure 9A) (using $\langle U_{r-s}^{vdW} \rangle_A^w$ as a size measure in eq 19 yields similar results (rms = 1.1 kcal/mol)). The largest individual errors are obtained for secondary and tertiary amides, for which OPLS-AA and many other force fields are known to have problems reproducing experimental hydration free energies.[44,45] Excluding secondary and tertiary amides yields an rms of 0.94 kcal/mol and $r^2 = 0.86$ for eq 19. The low value of the α_w^{vdW} coefficient suggests that the $\alpha_w^{vdW} \langle U_{r-s}^{vdW} \rangle_B^w$ term is almost insignificant in improving the prediction of hydration free energies, and exclusion of the $\alpha_w^{vdW} \langle U_{r-s}^{vdW} \rangle_B^w$ term does, in fact, yield similar results ($\Delta F_{sol} = \beta \Delta \langle U^{el} \rangle + \gamma$, rms = 1.1 kcal/mol). Optimizing all three coefficients in eq 19 yields $\beta = 0.42$, $\alpha_w^{vdW} = 0.05$, and $\gamma_w^{vdW} = 1.65$ and gives a
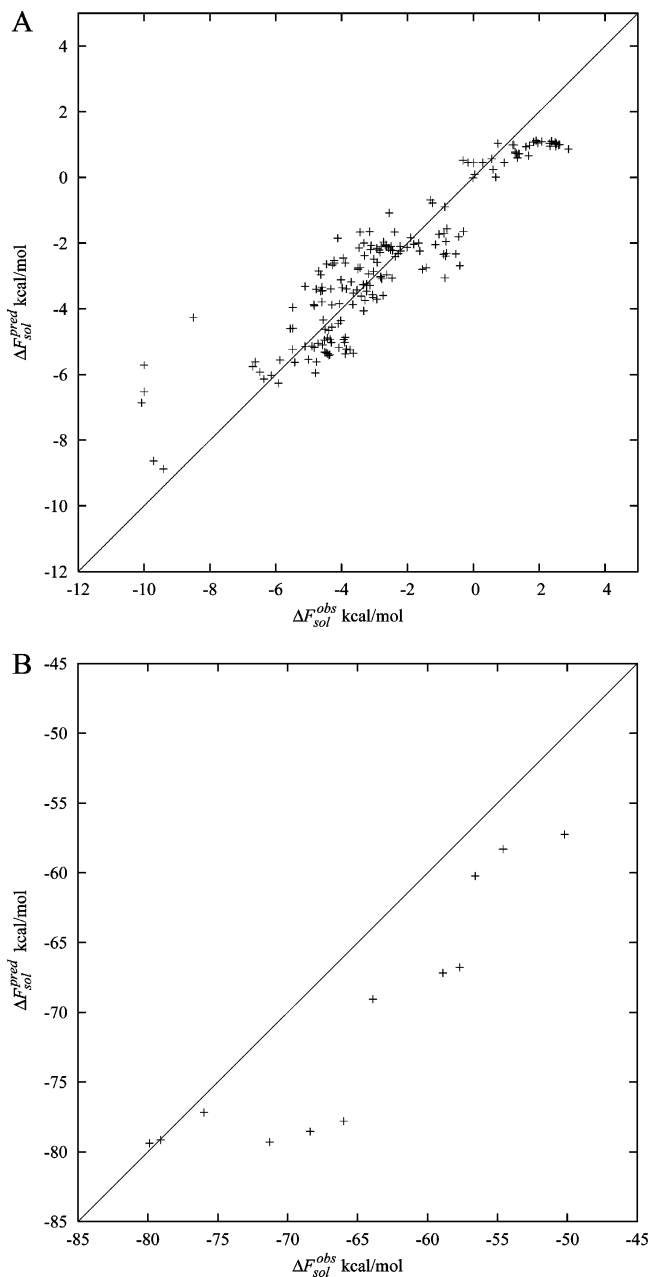
Figure 9. Experimental[43] and predicted absolute hydration free energies (kcal/mol) for the neutral (A) and ionic (B) molecules using eq 19 with $\alpha = 0.01$, $\beta = \beta_{model\ E}$ and $\gamma = 1.18$ kcal/mol.

slightly worse agreement with experiment (rms = 1.2 kcal/mol), which emphasizes the importance of group specific β-values. Excluding the change in intramolecular energies, i.e., using eq 11 instead of eq 10 in eq 19, also gives similar results (rms = 1.1 kcal/mol), which shows that there are only small differences in solute conformations in gas and aqueous phase for the training set. Although several studies[46,47] have shown that a combination of $\langle U_{r-s}^{vdW} \rangle_B$ and surface area gives the best description of the nonpolar contribution to hydration, it is of questionable statistical merit to introduce free scaling factors for both these terms because they are strongly correlated (Figure 2).[27] An alternative approach is to replace the nonpolar contribution in eq 19 ($\alpha_w^{vdW} \langle U_{r-s}^{vdW} \rangle_B^w + \gamma_w^{vdW}$)
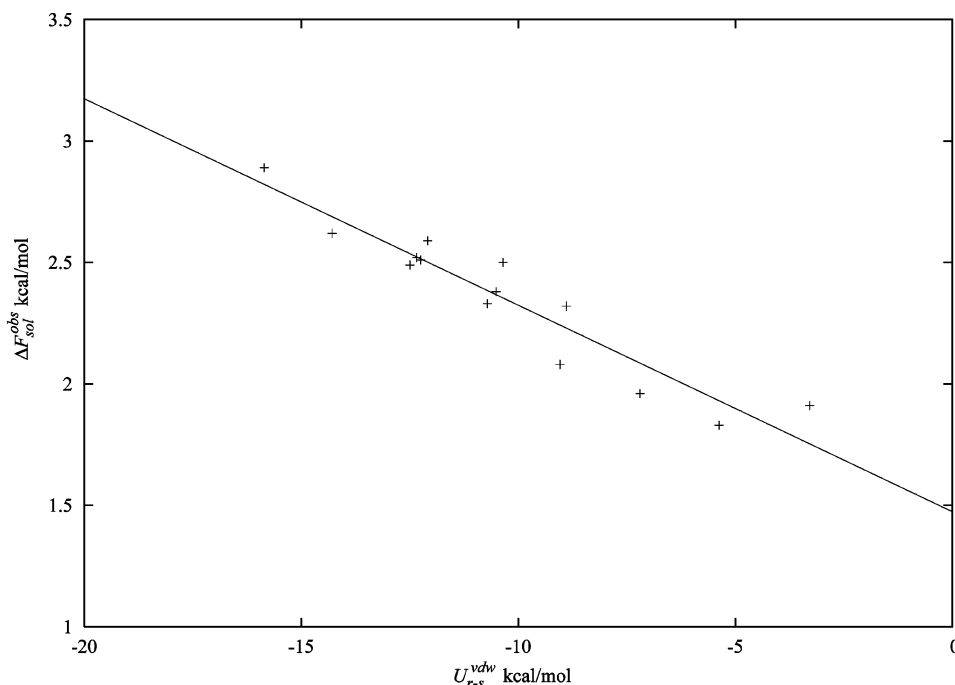
Figure 10. Correlation between hydration free energies,[43] $\Delta F_{\text{sol}}^{\text{obs}}$, (kcal/mol) for linear and branched alkanes and the solute−solvent van der Waals energy, $\langle U_{\text{r−s}}^{\text{vdW}} \rangle_A^{\text{w}}$ (kcal/mol).

with $\gamma_S \cdot A_S + \langle U_{\text{r−s}}^{\text{vdW}} \rangle_B + \gamma$.[37] Parametrization of γ_s and γ yields an rms error of 1.4 kcal/mol and $\gamma_s = 0.10$ kcal mol^{-1} Å^{-2} and $\gamma = -2.1$ kcal mol^{-1} which is slightly worse than the model using eq 19. Overall, the best model found in the course of this work, which, however, requires an additional simulation of the solute in the uncharged state, is

$$\Delta F_{\text{sol}} = \gamma_S \cdot A_S + \langle U_{\text{r−s}}^{\text{vdW}} \rangle_A^{\text{w}} + \beta(\langle U_{\text{r−s}}^{\text{el}} \rangle_B^{\text{w}} + \Delta \langle U_{\text{r−r}}^{\text{el}} \rangle_B) \quad (23)$$

Equation 23 incorporates the nonpolar approximation of eq 17, treating γ_S as a free parameter, along with the polar approximation of model E. Parametrizing eq 23 on experimental hydration free energies for the neutral compounds yields an rms error of 0.82 kcal/mol with $\gamma_S = 0.09$ kcal mol^{-1} Å^{-2}. Adding a constant to eq 23 does not improve the results significantly (rms = 0.82 kcal/mol). Interestingly the parametrized value of γ_S agrees very well with the experimental surface tension of 0.105 kcal mol^{-1} Å^{-2}.

For the ionic molecules, the $\alpha_{\text{w}}^{\text{vdW}}$ and $\gamma_{\text{w}}^{\text{vdW}}$ coefficients as parametrized above were used in eq 19. For each molecule the free energy contribution arising from interactions with water outside the simulation sphere boundary were calculated using the Born equation[17] which yields −9.1 kcal/mol. While reasonable agreement was obtained for anions (rms = 0.75 kcal/mol for eq 19), large errors compared to experiment were observed for the cations (rms = 9.66 kcal/mol for eq 19) using model E in eq 19 (Figure 9B). Since the calculated FEP energies were very well reproduced by model E (Figure 7), it must be the actual nonbonded force field parameters that do not reproduce experimental absolute solvation energies in this case. This has also been observed previously for ammonium.[48] Hence, in order to reproduce experimental solvation energies, reparameterization of the OPLS-AA force field for charged amines seems necessary.

The weak correlation between the nonpolar contribution to the hydration free energy of these compounds and the size descriptors was surprising considering how well this relationship has been documented experimentally.[16,49] The same observation was also made for SASA in a recent study using PB and GB electrostatic hydration energies.[50] One possible explanation is that although there is a clear correlation between nonpolar hydration energies and size, the difference in nonpolar hydration energy between the molecules in our set is relatively small. This combined with inaccuracies of force field parameters and experimental values makes it difficult to accurately predict this quantity. For example, in the case of linear and branched alkanes, where the electrostatic contribution is negligible, there is a strong correlation between $\langle U_{\text{r−s}}^{\text{vdW}} \rangle_A^{\text{w}}$ and the experimental hydration free energy. The slope is, however, relatively small, $\Delta F_{\text{sol}} = -0.08 \langle U_{\text{r−s}}^{\text{vdW}} \rangle_A^{\text{w}} + 1.5$ kcal/mol, and the difference in experimental hydration energy between the largest (octane) and the smallest (methane) molecule is only 1.1 kcal/mol (Figure 10).[43] For larger solutes or other solvents, e.g., n-hexane, where the corresponding relation between solute−solvent van der Waals interactions and nonpolar solvation energies is associated with considerably steeper slopes, a constant term would clearly not be accurate.[16] It is also noteworthy here, that our estimate in section 2 ($\Delta F_{\text{sol}}^{\text{np}} = -0.07 \langle U_{\text{r−s}}^{\text{vdW}} \rangle_A^{\text{w}} + 1.9$ kcal/mol) of the above relationship, from the experimental surface tension of water together with the correlation between molecular surface and van der Waals energy for nonpolar solutes, is surprisingly accurate.

Jorgensen and co-workers have proposed another LR variant to estimate free energies of hydration from microscopic simulations.[27−29] In their approach, eq 11 is combined

Prediction of Solvation Free Energies

J. Chem. Theory Comput., Vol. 3, No. 6, 2007 **2173**

with both SASA and the solute−solvent van der Waals interaction energy as in eq 24

$$\Delta F^{solv} = \alpha' \langle U_{r-s}^{vdW} \rangle_B + \beta' \langle U_{r-s}^{el} \rangle_B + \gamma' \langle SASA \rangle_B \quad (24)$$

but with α', β', and γ' treated as free scaling factors (the use of SASA instead of MS is not very important, although the latter quantity is more compatible with the experimental surface tension). Using eq 24 for the neutral compounds in our set yields $\alpha' = 0.38$, $\beta' = 0.49$, and $\gamma' = 0.02$ kcal mol^{-1} Å^{-2} with rms = 1.1 kcal/mol, which is the same quality as obtained using eq 19, and the obtained coefficients agree nicely with the coefficients obtained by Jorgensen and co-workers[28] using OPLS-AA charges (they obtained $\alpha' = 0.42$, $\beta' = 0.49$, and $\gamma' = 0.02$ kcal mol^{-1} Å^{-2}). The α', β', and γ' coefficients derived by Jorgensen and co-workers also appear to vary depending on the chosen scheme to assign partial charges for the molecules used in the parametrization when these are carried out on experimental hydration free energies.[27−29] It is important to note that this does not imply that there is a force field dependence of the β coefficient. The varying β parameters obtained by Jorgensen and co-workers rather reflect force field deficiencies, i.e., the chosen charge scheme does not reproduce experimental hydration free energies. For example, when EPS charges are used,[51] which are more polarized than OPLS-AA charges, smaller values of the β coefficient will be required to reproduce experiment. Since different values of the α' and γ' coefficients are obtained for the different charge schemes, they not only reflect nonpolar contributions to the solvation free energy but also provide compensation for possible force field errors.[27−29,51] Our models were instead optimized to reproduce FEP results, and thus the accuracy of the models when compared to experiment will be limited by the accuracy of the force field.

Implications for Calculations of Protein−Ligand Binding Free Energies. The scheme derived by Hansson et al. for predicting β_{FEP} (model B)[23] has successfully been used in the LIE method for predicting binding free energies of ligands binding to their receptors.[52−57] In the LIE method, the binding free energy is estimated in analogy with solvation energies as the free energy of transfer between water and protein environments. Simulations are carried out for the ligand in water and the solvated protein system, and the Gibbs free energy of binding is calculated from the ligand-surrounding (l-s) (the ligand's interactions with both protein and solvent atoms) electrostatic (el) and van der Waals (vdW) interaction energies

$$\Delta G_{bind}^{LIE} = \alpha \Delta \langle U_{l-s}^{vdW} \rangle + \beta \Delta \langle U_{l-s}^{el} \rangle + \gamma \quad (25)$$

where the Δ's refer to differences in protein and water simulations.[22] The standard parametrization of the model was derived with β values according to model B using a set of 18 protein−ligand complexes, and the optimal value of α was found to be 0.18.[23] For this data set the constant γ was found to be 0.0, which is not always the case.[58] Note that the derived values of α and γ in eq 25 cannot be directly compared to α_w^{vdW} and γ_w^{vdW} in eq 19, as attempted by Almlöf et al.[58] That is, if eq 18 is used to express a relation

between nonpolar solvation free energies in protein and solvent environments, we obtain

$$\Delta\Delta G_{sol}^{np} = \Delta G_{sol}^{np,p} - \Delta G_{sol}^{np,w} =$$
$$\alpha_p \langle U_{l-s}^{vdW} \rangle - \alpha_w \langle U_{l-s}^{vdW} \rangle + \gamma_p - \gamma_w \quad (26)$$

which cannot be rewritten in terms of $\Delta \langle U_{l-s}^{vdW} \rangle$ to identify α and γ in eq 25. Rather, α and γ can be derived from relations relating size to the change in ligand-surrounding van der Waals interactions ($\Delta \langle U_{l-s}^{vdW} \rangle$) and the nonpolar free energy of solvation ($\Delta\Delta G_{sol}^{np}$) between protein and water environments[59]

$$\Delta\Delta G_{sol}^{np} = a\sigma + b$$

$$\Delta \langle U_{l-s}^{vdW} \rangle = c\sigma + d$$

$$\Rightarrow \Delta\Delta G_{sol}^{np} = \frac{a}{c}(\Delta \langle U_{l-s}^{vdW} \rangle - d) + b = \alpha \Delta \langle U_{l-s}^{vdW} \rangle + \gamma \quad (27)$$

where σ is a size measure, such as MS, SASA, or the number of heavy atoms in the ligand, and a, b, c, and d are empirically derived parameters. From eq 27, the contributions from nonpolar solvation to α and γ in eq 25 can be identified as a/c and $b-ad/c$, respectively. Since the parametrization of LIE was performed using experimental binding free energies, $\alpha = 0.18$ takes into account van der Waals interactions and all other size dependent contributions to binding, e.g., the hydrophobic effect and relative translational and rotational entropies.[59] As noted above, the nonpolar contribution to the hydration free energy was small and could be well represented by a constant term for our data set. In contrast to hydration free energies however, the nonpolar term in eq 25 often makes a significant contribution to the binding free energy and cannot be represented by a constant term.

In our development of the LIE method, the ligand-surrounding electrostatic energies in both the protein and water simulations are scaled by the same factor. The original idea of LIE was that the β coefficient would not be used as a free parameter and, even though the β coefficient for the protein environment is to some extent uncertain and deserves further investigation, it is somewhat questionable to optimize β in eq 25 freely. In several published attempts at reproducing binding free energies using LIE, the electrostatic scaling factor in the LIE method is sometimes found to be very small and in a few cases even negative.[60−63] In the present work, it is clear that for ligands in aqueous phase a value of $\beta = 0.37−0.52$ is appropriate, and, therefore, in the above-mentioned problematic cases, it would hence make more sense to scale the electrostatic ligand-surrounding energies of the water and protein simulations separately and only treat the scaling of the electrostatics in the protein simulation (β_{prot}) as a free parameter

$$\Delta G_{bind}^{LIE} = \alpha \Delta \langle U_{l-s}^{vdW} \rangle + \beta_{prot} \langle U_{l-s}^{el} \rangle_p - \beta_{wat} \langle U_{l-s}^{el} \rangle_w + \gamma \quad (28)$$

This approach was actually suggested initially,[23] but to this date there has been no reason to introduce the increased amount of complexity into the model.

The results presented here also show how intramolecular energies can be explicitly included in the LIE method. In

analogy to solvation free energies, this would lead to the introduction of an intramolecular term in eq 25, yielding

$$\Delta G_{\text{bind}}^{\text{LIE}} = \alpha\Delta\langle U_{\text{l-s}}^{\text{vdW}}\rangle + \beta(\Delta\langle U_{\text{l-s}}^{\text{el}}\rangle + \Delta\langle U_{\text{l-l}}^{\text{el}}\rangle) + \gamma \quad (29)$$

$\Delta\langle U_{\text{l-l}}^{\text{el}}\rangle = \langle U_{\text{l-l}}^{\text{el}}\rangle_{\text{p}} - \langle U_{\text{l-l}}^{\text{el}}\rangle_{\text{w}}$, where $\langle U_{\text{l-l}}^{\text{el}}\rangle_{\text{p}}$ and $\langle U_{\text{l-l}}^{\text{el}}\rangle_{\text{w}}$ are the intramolecular ligand−ligand energies in the bound and free state, respectively. It should be noted, in analogy with the two different cycles used for solvation free energies (Figure 1), that eqs 25 and 29 are both rigorously derived. The accuracy of either equation will ultimately depend on the appropriateness of the approximations used in eqs 25 and 29 to predict the relevant legs of each thermodynamic cycle.

5. Conclusions

In this work, a LR approach to estimate the electrostatic component of the free energy of solvation has been presented. The main result is that derivation of scaling factors for specific chemical groups yields remarkable agreement with the exact results calculated using the FEP method. For molecules containing several chemical groups, a scheme for deriving specific values of β for each compound was proposed, and this was shown to yield impressive results on a large data set not included in the parametrization. For estimates of the total hydration free energy, the electrostatic component was combined with an empirical size dependent treatment of the nonpolar contribution to the free energy, and the results are in good agreement with experiment. The results reported herein should be useful for predicting free energies of solvation and also to improve the accuracy of simplified binding free energy calculations.

Abbreviations. MD, molecular dynamics; FEP, free energy perturbation; rms, root mean square; LIE, linear interaction energy; MC, Monte Carlo; TI, thermodynamic integration; LR, linear response.

Supporting Information Available: Electrostatic and Lennard-Jones interaction energies, free energies of charging, predicted free energies of charging using models D/E, calculated β_{FEP} values, and experimental hydration free energies for the training set and electrostatic interaction energies, free energies of charging, and predicted free energies of charging for the test set. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050−3054.

(2) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564−568.

(3) Straatsma, T. P.; Berendsen, H. J. C. *J. Chem. Phys.* **1988**, *89*, 5876−5886.

(4) Åqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021−8024.

(5) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161−185.

(6) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem. B* **2000**, *104*, 6271−6285.

(7) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740−5761.

(8) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656−1676.

(9) Oostenbrink, C.; Juchli, D.; van Gunsteren, W. F. *ChemPhysChem* **2005**, *6*, 1800−1804.

(10) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616−17626.

(11) Mezei, M.; Beveridge, D. L. *Ann. N. Y. Acad. Sci.* **1986**, *482*, 1−23.

(12) Åqvist, J. *J. Comput. Chem.* **1996**, *17*, 1587−1597.

(13) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707−2714.

(14) Abraham, M. H. *J. Am. Chem. Soc.* **1979**, *101*, 5477−5484.

(15) Abraham, M. H. *J. Am. Chem. Soc.* **1982**, *104*, 2085−2094.

(16) Blokzijl, W.; Engberts, J. B. F. N. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1545−1579.

(17) Born, M. *Z. Phys.* **1920**, *1*, 45−48.

(18) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671−679.

(19) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.

(20) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(21) Åqvist, J.; Hansson, T. *J. Phys. Chem.* **1996**, *100*, 9512−9521.

(22) Åqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng.* **1994**, *7*, 385−91.

(23) Hansson, T.; Marelius, J.; Åqvist, J. *J. Comput-Aided. Mol. Des.* **1998**, *12*, 27−35.

(24) Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. *Adv. Prot. Chem.* **2003**, *66*, 123−158.

(25) Del Buono, G. S.; Figueirido, F. E.; Levy, R. M. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 85−97.

(26) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458−4472.

(27) Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 2878−2888.

(28) Carlson, H. A.; Jorgensen, W. L. *J. Phys. Chem.* **1995**, *99*, 10667−10673.

(29) McDonald, N. A.; Carlson, H. A.; Jorgensen, W. L. *J. Phys. Org. Chem.* **1997**, *10*, 563−576.

(30) Zwanzig, R. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

(31) Kubo, R. *J. Phys. Soc. Jpn.* **1962**, *17*, 1100−1120.

(32) Åqvist, J.; Hansson, T. *J. Phys. Chem. B* **1998**, *102*, 3837−3840.

(33) Vorobjev, Y. N.; Hermans, J. *J. Phys. Chem. B* **1999**, *103*, 10234−10242.

(34) Kastenholz, M. A.; Hünenberger, P. H. *J. Chem. Phys.* **2006**, *124*, 124106.

(35) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717−726.

Prediction of Solvation Free Energies

J. Chem. Theory Comput., Vol. 3, No. 6, 2007 **2175**

(36) Westergren, J.; Lindfors, L.; Hoglund, T.; Luder, K.; Nordholm, S.; Kjellander, R. *J. Phys. Chem. B* **2007**, *111*, 1872−1882.

(37) Su, Y.; Gallicchio, E.; Das, K.; Arnold, E.; Levy, R. M. *J. Chem. Theory Comput.* **2007**, *3*, 256−277.

(38) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *J. Mol. Graphics Modell.* **1998**, *16*, 213-+.

(39) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(41) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647−3661.

(42) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100−3107.

(43) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563−595.

(44) Morgantini, P. Y.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 6057−6063.

(45) Udier-Blagovic, M.; De Tirado, P. M.; Pearlman, S. A.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 1322−1332.

(46) Zacharias, M. *J. Phys. Chem. A* **2003**, *107*, 3000−3004.

(47) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. *J. Am. Chem. Soc.* **2003**, *125*, 9523−9530.

(48) Luzhkov, V. B.; Almlöf, M.; Nervall, M.; Åqvist, J. *Biochemistry* **2006**, *45*, 10807−10814.

(49) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016−2027.

(50) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. *J. Chem. Theory Comput.* **2006**, *2*, 128−139.

(51) Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. *J. Comput. Chem.* **1993**, *14*, 1240−1249.

(52) Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Åqvist, J. *J. Comput-Aided. Mol. Des.* **1998**, *12*, 119−131.

(53) Graffner-Nordberg, M.; Kolmodin, K.; Åqvist, J.; Queener, S. F.; Hallberg, A. *J. Med. Chem.* **2001**, *44*, 2391−2402.

(54) Luzhkov, V. B.; Åqvist, J. *FEBS Lett.* **2001**, *495*, 191−196.

(55) Ersmark, K.; Feierberg, I.; Bjelic, S.; Hultén, J.; Samuelsson, B.; Åqvist, J.; Hallberg, A. *Bioorg. Med. Chem.* **2003**, *11*, 3723−3733.

(56) Leiros, H. K. S.; Brandsdal, B. O.; Andersen, O. A.; Os, V.; Leiros, I.; Helland, R.; Otlewski, J.; Willassen, N. P.; Smalås, A. O. *Protein Sci.* **2004**, *13*, 1056−1070.

(57) Ersmark, K.; Feierberg, I.; Bjelic, S.; Hamelink, E.; Hackett, F.; Blackman, M. J.; Hultén, J.; Samuelsson, B.; Åqvist, J.; Hallberg, A. *J. Med. Chem.* **2004**, *47*, 110−122.

(58) Almlöf, M.; Brandsdal, B. O.; Åqvist, J. *J. Comput. Chem.* **2004**, *25*, 1242−1254.

(59) Carlsson, J.; Åqvist, J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5385−5395.

(60) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.

(61) Tounge, B. A.; Reynolds, C. H. *J. Med. Chem.* **2003**, *46*, 2074−2082.

(62) Stjernschantz, E.; Marelius, J.; Medina, C.; Jacobsson, M.; Vermeulen, N. P. E.; Oostenbrink, C. *J. Chem. Inf. Model.* **2006**, *46*, 1972−1983.

(63) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517−529.