In Silico Prediction of Drug Solubility: 1. Free Energy of Hydration

Jan Westergren,[†] Lennart Lindfors,*[,†] Tobias Höglund,[†] Kai Lüder,[‡] Sture Nordholm,[‡] and Roland Kjellander[‡]

Pharmaceutical and Analytical R&D, AstraZeneca R&D, Mölndal, SE-431 83 Mölndal, Sweden, and Department of Chemistry, Göteborg University, SE-412 96 Göteborg, Sweden

Received: July 5, 2006; In Final Form: November 24, 2006

As a first step in the computational prediction of drug solubility the free energy of hydration, $\Delta G^{\bullet}_{vw}$, in TIP4P water has been computed for a data set of 48 drug molecules using the free energy of perturbation method and the optimized potential for liquid simulations all-atom force field. The simulations were performed in two steps, where first the Coulomb and then the Lennard-Jones interactions between the solute and the water molecules were scaled down from full to zero strength to provide physical understanding and simpler predictive models. The results have been interpreted using a theory assuming $\Delta G^{\bullet}_{vw} = A_{MS}\gamma + E_{LJ} + E_C/2$ where A_{MS} is the molecular surface area, γ is the water−vapor surface tension, and E_{LJ} and E_C are the solute−water Lennard-Jones and Coulomb interaction energies, respectively. It was found that by a proper definition of the molecular surface area our results as well as several results from the literature were found to be in quantitative agreement using the macroscopic surface tension of TIP4P water. This is in contrast to the surface tension for water around a spherical cavity that previously has been shown to be dependent on the size of the cavity up to a radius of ∼1 nm. The step of scaling down the electrostatic interaction can be represented by linear response theory.

1. Introduction

One important property of a drug is its crystalline or amorphous solubility in water. Since a poorly soluble drug molecule, which in many other aspects is promising, may require significant additional development resources or even be discarded it would be desirable to know the solubility at an early stage. Experimental determination of the solubility of drug molecules is unfortunately a very laborious activity. Much time and effort could be saved if solubility could be reliably predicted by computation. We have therefore decided to explore the possibility of predicting and understanding the drug solubility by a combination of numerical simulation and statistical mechanical analysis. In this paper we investigate the hydration of drugs as being a first step in the exploration of solubility.

The common practice today among those who wish to predict drug solubility is to use some kind of quantitative structure−property relationship (QSPR).[1,2] In such methods a correlation equation relates the solubility to a number of properties for a training set of molecules. To predict the solubility of a new molecule its corresponding properties are inserted into the correlation equation. The QSPR methods are often fast and may be accurate if the new molecule is similar to the ones in the selected training set. Otherwise the QSPR may fail badly. The reliability can be low because the input information and the statistical analysis are general and statistical rather than specific and physical.

Another way to predict the solubility is to use molecular simulations to calculate the chemical potential of the drug in crystalline or amorphous form (the pure solid phase) and as a function of the concentration of the solute in a solution, respectively. The solubility is the concentration of the drug at equilibrium, i.e., when equality in the chemical potential is obtained between the phases. By definition, the chemical potential is the change in Gibbs free energy when at constant temperature and pressure one extra drug molecule is added to the pure solid phase and the solution, respectively. Such processes can for the solution and the amorphous phase in principle be resolved by simulation employing, for instance, the free energy perturbation (FEP) method[3] as shown in this context by Åquist et al.[4] Unfortunately, FEP simulations for drug molecules in the molecular weight range from 400 to 600 Da, typical of modern drugs, may be so time-consuming that they cannot be used routinely. For the crystalline phase, the FEP method is not readily applicable. But there are methods that can calculate the chemical potential of one-component systems without inserting/removing any molecule, for instance, the reference system equilibration method.[5]

The free energy simulations are time-consuming, especially for the pure solid phases, and an alternative would be to establish a relation between the chemical potential on the one hand and more readily accessible quantities such as interaction energies and molecular volumes on the other hand. The latter quantities can be calculated in standard isobaric−isothermal (NPT) molecular simulations, which are considerably faster than free energy simulations. The model for the relation between chemical potential and energies and volumes differs fundamentally from QSPR relations as the former model is developed by physical reasoning rather than statistical correlation and regression. Not only can such a model improve computational efficiency, it can also provide understanding of the physical mechanisms.

Since the pioneering work of Hildebrand in the beginning of the 20th century[6] many scientists have been engaged in developing theories for understanding solubility. In the current

* Author to whom correspondence should be addressed. Phone: +46 31 7761065. Fax: +46 31 7763834. E-mail: Lennart.Lindfors@astrazeneca.com.

[†] AstraZeneca R&D, Mölndal.

[‡] Göteborg University.

10.1021/jp064220w CCC: $37.00
Published on Web 02/01/2007

discussion the focus is on cavity formation and molecular surface area and shape, solute−solvent interaction, and solvent reorganization.[7−11] Aqueous solubility is often discussed in the context of poorly soluble, or hydrophobic, molecules and is especially interesting for the pharmaceutical industry. What is the cause of hydrophobicity? One cause is that entropy of the surrounding water decreases as the water molecules close to the solute adopt more rigid structures to avoid wasting hydrogen bonds.[12,13] This is often called an iceberg effect[14] that decreases the solubility. There is also a bond energy effect when some water molecules find some of their water neighbors replaced by the solute. But it is not necessarily so that the bond energy between a poorly soluble solute and water is lower than that within water. The total interaction energy per volume is for some solutions lower than that in pure water, but for some solutes the situation is the opposite.

At phase equilibrium between the pure solid phase and the solution, there is no change in free energy when transferring one drug molecule from the pure phase to the solution. This transfer is conveniently divided into a first move of a molecule from the pure phase to an ideal gas and then from the ideal gas to the solution. When the solvent is water, the latter process is the hydration of the molecule.

Simulating the transfer of a molecule from the pure phase to an ideal gas is a complicated task from a computational point of view. If the pure solid phase is a crystal, then the thermodynamically stable polymorph first needs to be predicted. Although this is an active field of research it will take many years before this becomes routine.[15,16] Second, given the experimental polymorph obtained, e.g., from single-crystal X-ray studies, the computation of the free energy is not trivial. In our work we have therefore investigated an alternative computational route to the solubility of a crystal. This route includes the prediction of the aqueous solubility of the amorphous phase by simulation. Although these simulations are also complicated they are tractable. Approximate values of the solubility of a crystalline material can be obtained from the calculated values of amorphous solubility and via a thermodynamic cycle for the crystal → supercooled liquid process[17,18] using experimental melting data obtained, for example, from differential scanning calorimetry measurements. This will be the topic of future publications. Here we shall focus exclusively on the hydration step.

The solvation of a molecular species A in water, i.e., A(g) → A(aq), can conveniently be divided into three steps: (i) A cavity is created in the aqueous solution such that it would precisely fit the solute molecule A without any significant short-range repulsion; (ii) the solute A is inserted into the cavity, and electrostatic interactions due to the partial charges distributed in the solute and solvent molecules are turned on; (iii) the dispersion interactions between the solute and solvent are turned on. These three steps together represent the insertion of a solute molecule into the water solvent, and this procedure has also been used by others.[11,19,20]

Note that the corresponding free energy changes of the solution account for the hard repulsive, electrostatic, and dispersive interactions, respectively. They enter in order of strength of the response of the solvent. The hard repulsive interactions induce cavity formation, a strong nonlinear response. The electrostatic interactions can, as we shall verify below, to a good approximation be assumed to induce only a linear response on the solvent molecules when they reorient to minimize the corresponding free energy change. Finally, the dispersion attractions are expected to be sufficiently weak and/or uniform so that no significant structural response is induced in the solvent.

Despite the strong physical argument in the favor of the three insertion steps above we have, in fact, used a different separation of the insertion of the solute in two steps. The reason for this is technical and relates to the facilities available for the simulation study in the Biochemical and Organic Simulation System (BOSS) program.[21−23] First of all we have, in fact, removed the solute molecule from a water solution rather than inserting it. The removal is simpler than insertion since in the former no "hard core overlap" is encountered as would be expected in the process of insertion. There is, however, no significant disadvantage in doing removal rather than insertion. The changes in free energy in the steps simply change sign. Second, we have actually not separated the hard repulsive from the soft attractive interactions, which we both represent by Lennard-Jones interactions between the atoms of the solute and a solvent molecule. Thus our first and last steps above are, in fact, combined into one. We subdivide the removal of the solute molecule from the water solvent into two steps: (i) turning off the partial charges on the solute atoms and (ii) turning off the Lennard-Jones interactions between solute and solvent. The reason is that it is technically easier to manipulate the two separable potential terms in the simulation than to separate the Lennard-Jones interactions into a steeply repulsive and a soft attractive part. As we shall see below this inconsistency of the simulation practice and the physically sound procedure does lead to a minor problem in our analysis of our data, which we shall have to deal with in our theory.

The brute force simulation of the true free energy changes associated with solute removal in water has obvious disadvantages in terms of computational inefficiency and lack of insight. We shall therefore complement brute force simulation with a simplified simulation method based on an approximate physical analysis with room for empirical input, if needed. Following the analysis first introduced by van der Waals and later extended into the generalized van der Waals (GvdW)[24] theory we shall estimate the free energy changes associated with solute insertion in water in terms of free energy of cavity formation related to the surface tension of water, an electrostatic free energy change obtained by linear response theory, and a dispersive binding energy obtained by mean field approximation. Besides using quantities obtained from simulations we will also study the effect of using experimentally determined surface tension for water in our model. An alternative method commonly used for estimating the cavity cost is the scaled particle theory.[25]

In the estimation of the free energy contribution due to electrostatic interaction the linear response assumption means that we need only estimate the energy change since the entropy loss will merely compensate for half of the binding energy. Åqvist and Hansson[26] and Carlson and Jorgensen[11] found that linear response is a good first approximation but deviations from linearity for neutral molecules were observed. Also Sandberg et al. observed non-negligible nonlinear effects for small molecules that they propose are due to saturation of solvent polarization.[20] However, nonlinear effects may arise even if the response does not reach saturation. The applicability of linear response is addressed in this paper.

Finally the mean field approximation completely ignores any structural response and related entropy changes again leaving us with an energy evaluation only. In this way our theory eliminates the need for free energy evaluations replacing them by evaluations of surface areas, surface tensions, and interaction energies, i.e., a much simpler and computationally efficient

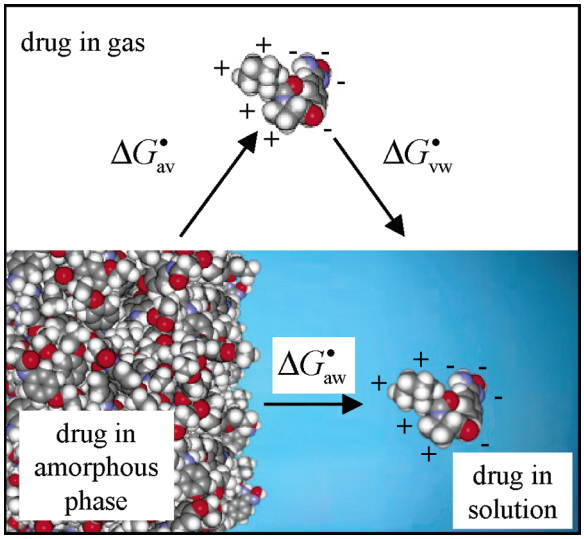

Figure 1. Equilibrium between amorphous drug, drug vapor, and drug aqueous solution.

method. If it is found to work in comparison with brute force simulation results for complex drug molecules it may be useful in a semiempirical theory of the solubility of molecules for which no satisfactory interaction potentials are known. We might even in very difficult cases be able to extract some understanding of the origin of the difficulty.

The paper is organized in the following way. In section 2 the fundamental thermodynamics of solubility and the description of our simulation method are given. The results are given in section 3 and followed by a discussion in section 4.

2. Theory and Simulations

2.1. Equilibrium between Amorphous Phase and Solution. Below the melting point of a one-component system the amorphous phase is in a metastable state and will eventually restructure to a crystal. The relaxation may, however, be slow, and on the lab scale it is possible to observe an apparent equilibrium between the amorphous phase and a solution. Hence the amorphous solubility is experimentally measurable in several cases. In any case we can imagine that the amorphous phase is in equilibrium with the solution although the system is really metastable.

The equilibrium concentration of the drug is conveniently expressed by the chemical potential μ, which is defined as the change in free energy of a system when one drug molecule is added. The aqueous solubility of a compound at $T = 298$ K and $P = 1$ atm can schematically be described by the processes shown in Figure 1 and explained below.

The system consists of three phases. One is the amorphous pure drug phase, and its chemical potential is denoted by μ_a. The second phase is the solution consisting of water and dissolved drug. For modern drugs the solubility is generally very low, and at the corresponding low concentration the chemical potential of the solute thus satisfies Henry's law

$$\mu_s(x_s) = \mu_{s,const} + RT \ln x_s \quad \text{when} \quad x_s \ll 1 \tag{1}$$

Here x_s is the mole fraction of solute in the solution, R is the gas constant, and $\mu_{s,const}$ is dependent on T and P but not on x_s. The third phase is the gas (vapor) phase, which is assumed to be ideal. The partial pressure of the drug equals the vapor pressure of the amorphous drug. This vapor pressure is generally low, but the presence of air in the gas will add to a total pressure of 1 atm. We do not account for any further effects of air in the system. The chemical potential of the drug vapor is

$$\mu_v(p_v) = \mu_v^{\ominus} + RT \ln p_v/p^{\ominus} \tag{2}$$

where p_v is the partial pressure of the drug and $\mu_v^{\ominus}$ is the chemical potential of the drug at the reference pressure $p^{\ominus} = 1$ atm and $T = 298$ K, i.e., at standard state conditions.

At equilibrium x_s and p_v equal the solubility (in terms of mole fraction) and the vapor pressure of the amorphous phase, respectively. They are determined by the equality of chemical potentials

$$\mu_a = \mu_s(x_s) = \mu_v(p_v) \tag{3}$$

To obtain the solubility from simulations, i.e., to solve the right-hand equation in eq 3, we need to find molecular processes that can be simulated that give information about the chemical potential. The change in free energy for the process of inserting a solute molecule with its center of mass fixed in the solution is given by[27]

$$\Delta G^{\bullet}_{\times w} = \mu_s^{\bullet} = \mu_s(x_s) - RT \ln N_A c_s \Lambda^3 \tag{4}$$

Here N_A is Avogadro's constant, $\times$ denotes pure water, w denotes the aqueous solution, and $\bullet$ denotes a fixed center of mass. The solute concentration c_s is for highly dilute solutions related to the mole fraction by $c_s = c_{H_2O} x_s$ where $c_{H_2O} = 55.3$ mol/dm^3 (at 298 K) is the concentration of water. The de Broglie thermal wavelength is defined as $\Lambda = \sqrt{h^2/2\pi m k_B T}$ where h is Planck's constant, m is the mass of the molecule, and k_B is Boltzmann's constant. But since Λ is equal for the drug molecule in all phases, it will cancel in all calculations. The advantage of using this fixed-center process instead of having a free center of mass is that in FEP simulations the center of mass of the inserted/annihilated molecule can be considered fixed. Hence FEP simulations yield $\Delta G^{\bullet}$ with the exception that kinetic energy is not included in the Monte Carlo simulations. But as complete annihilation is never simulated, i.e., the number of atoms in the system is fixed, the kinetic energy does not change in the processes that we consider.

The analogue of eq 4 holds also for the drug in the amorphous phase. Hence, transferring a molecule from the pure drug to the solution, having the center of mass fixed in both systems, yields

$$\begin{aligned}\Delta G^{\bullet}_{aw} &= \Delta G^{\bullet}_{a\times} + \Delta G^{\bullet}_{\times w} = -\mu_a^{\bullet} + \mu_s^{\bullet} \\ &= -(\mu_a - RT \ln c_a \Lambda^3) + (\mu_s(x_s) - RT \ln c_s \Lambda^3) \\ &= (\mu_s(x_s) - \mu_a) + RT \ln \frac{c_a}{c_s}\end{aligned} \tag{5}$$

At equilibrium $\mu_s(x_s) - \mu_a = 0$. Consequently, by obtained c_a and $\Delta G^{\bullet}_{aw}$ from simulations, the solubility c_s (in terms of concentration) is obtained. We have

$$c_s = c_a \, e^{-\Delta G^{\bullet}_{aw}/RT} = c_a \, e^{-(\mu_s^{\bullet} - \mu_a^{\bullet})/RT} \tag{6}$$

In the Henry's law regime where the solution is very dilute, the distance between solute molecules is on average long, and the only environment that each drug molecule experiences is pure water. Therefore only one solute molecule need be present

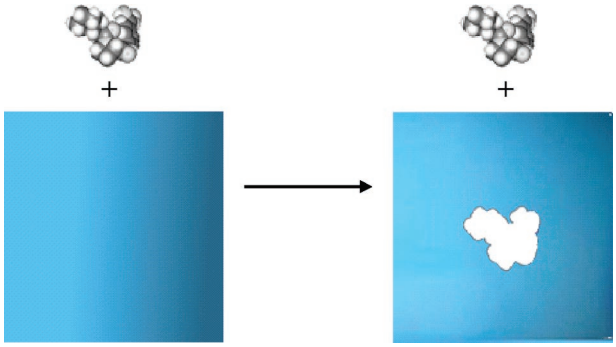

Figure 2. Step i: Creation of a cavity in pure water.

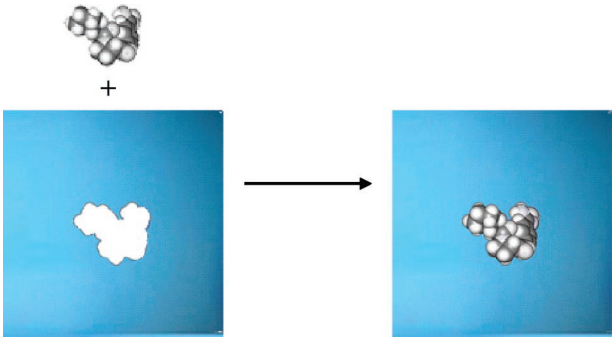

Figure 3. Step ii: Insertion of an alkane-like solute.

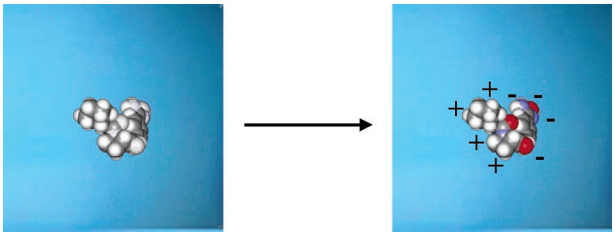

Figure 4. Step iii: Turning on atomic partial charges.

in the simulation of the solution. As a consequence, $\mu_s^\bullet$ is independent of the solute concentration.

In the w → × process (opposite to × → w) the solute is totally annihilated. In our simulations with fully flexible bonds in the solute, the very annihilation of the bonds implies a big change in energy and in free energy and therefore requires long FEP simulations to achieve high accuracy. Fortunately, the complete annihilation is not required in the simulations. It is better to end with all intramolecular degrees of freedom intact and just turn off the interaction with neighboring molecules. This equals a process of transforming the solute into an ideal gas molecule and the solvent into pure water. This process we denote as w → v (v for vapor). This incomplete annihilation of the molecules is perfectly fine as we do the same thing for both the solution and the amorphous drug phase. The equalities

$$\Delta G^\bullet_{a\times} + \Delta G^\bullet_{\times w} = \Delta G^\bullet_{aw} = \Delta G^\bullet_{av} + \Delta G^\bullet_{vw} \qquad (7)$$

therefore apply. The right-hand processes are illustrated in Figure 1. The processes of insertion or removal of a molecule from the amorphous drug phase we leave to future publications. The rest of this work is focused on the v → w process, i.e., the hydration of the solute.

2.2. Hydration Process. To understand our simulation results and explore the possibility of computational simplification we shall seek a relation between the change in potential energy, the molecular surface area (MSA), and the accompanying change in free energy for the hydration process, v → w. By dividing the hydration into three subprocesses we have found such relations that appear to work well. The three subprocesses in our model are as follows:

2.2.1. Step i: Creation of a Cavity in Pure Water. The drug molecule is in an ideal gas (vapor) throughout this step. The solvent is pure water at first, but a nonspherical cavity is created that is suitable in size and shape to contain the solute. The change in free energy for this process is denoted by ΔG_{cav}. Note that the solute is not involved in the process and thus • is omitted (Figure 2).

2.2.2. Step ii: Insertion of an Alkane-like Solute. The solute is brought into the cavity, and the Lennard-Jones interaction between the solute and water is turned on. However, there is no electrostatic interaction between the solute and water. In other words, the partial charges of the solute atoms are considered to be zero with respect to intermolecular interaction, i.e., $E_C = 0$ where E_C denotes the electrostatic solute−water interaction energy. Note that intramolecular electrostatic interaction is present. We call the solute alkane-like since at this stage the electrostatic interaction with water is zero. The behavior of alkanes can thus be used to understand the alkane-like molecules. The change in free energy for this process is denoted by $\Delta G^\bullet_{insert}$ (Figure 3).

2.2.3. Step iii: Turning on Atomic Partial Charges. In the last subprocess the electrostatic interaction between the solute and water is turned on, and the final state is the proper solution. The change in free energy for this process is denoted by $\Delta G^\bullet_{charge}$ (Figure 4).

In the models for the relation between change in potential energy and free energy we only want to include the intermolecular energy. There were, however, two technical constraints in our simulations that made intramolecular interactions contribute to the FEP free energy changes. First, we want to simulate fully flexible molecules; i.e., the bond lengths, bond angles, and dihedral angles are free to vary. This leads to the complication that the transfer of the molecule from gas to solution in general gives rise to a change in intramolecular energy. This change is expected to be small. We have chosen to ignore such changes in our analysis. Examination of our simulation results appears to justify our procedure.

The second simulation detail is that we could not avoid changing the intramolecular interaction due to restrictions in BOSS. In the process of decoupling the electrostatic interactions both the intramolecular and the intermolecular interactions were scaled equally and simultaneously. By analogy, both the intramolecular and the intermolecular Lennard-Jones energy was changed when decoupling the solute from the solvent. Furthermore, in the process of Lennard-Jones scaling the size of the solute was changed from 100% to 10% of its normal size. Hence, in contrast to our physical model, the intramolecular energy was not kept intact during the removal process. To compensate for the change in free energy due to intramolecular interactions, FEP simulations in the ideal vapor phase were performed. The free energy changes in these latter simulations were entirely due to intramolecular scaling and could thus be subtracted from the total free energy changes in FEP simulations in water to yield intermolecular free energy changes in water. Note that, strictly speaking, to separate the free energy into an "intermolecular" and an "intramolecular" part the intramolecular interaction must be independent of intermolecular interaction.

With the assumptions and compensations above, free energy changes due to intermolecular scaling were obtained, and by the dissection into subprocesses the total hydration free energy can be written as the sum

$$\Delta G^{\bullet}_{vw} = \Delta G_{cav} + \Delta G^{\bullet}_{insert} + \Delta G^{\bullet}_{charge} \quad (8)$$

However, after the compensation for intramolecular changes, the simulated processes still differ from processes i−iii in our physical model. First of all, in the simulation the solute was gradually removed instead of inserted. Second, the processes of cavity formation and insertion of alkane-like solute were simulated in one step.

2.3. Simulation Methodology. We have used the simulation program BOSS. To obtain the interaction energy and molecular volume by simulation (standard NPT Monte Carlo simulations) a box of approximately 500 TIP4P[28] water molecules and one drug molecule was prepared. The NPT simulations were run with periodic boundary conditions at fixed temperature and pressure.

The molecular surface area, A_{MS}, is calculated in BOSS by taking each atom in the solute to be a hard sphere. The A_{MS} is the area of the surface that the center of a hard sphere, representing water, describes when moving in contact with the solute. The radius of a solute atom is set to $r = 2^{1/6}\sigma/2$ where σ is one of the parameters in the Lennard-Jones (12−6) potential[29] (cf. eq 11). The reason for this definition of r is that the minimum-energy distance between two equal atoms according to the Lennard-Jones potential is $2r$. To move the interface close to the solute we have set the water radius to zero. Note that in our analysis we will use the molecular surface area of solutes without partial charges. However, in simulations we have found that the molecular surface area does not change when removing the partial charges.

A simulation consisted of an equilibration period and an equally long sampling period. In each period 60−480 million single molecule move attempts were performed, depending on number of solute atoms. Preferential sampling was used; i.e., water molecules far from the solute were moved more rarely. The solute molecule was flexible in contrast to the rigid water molecules. A solute displacement attempt consisted of (i) randomly selecting the molecule to be moved, (ii) translating the first three atoms in the selected molecule randomly in all three Cartesian directions, (iii) rotating the three first atoms in the molecule randomly about one randomly chosen Cartesian axis, and (iv) rebuilding the molecule from the applied changes and remaining atoms in the coordinate structure file. Changes in the internal variables were incorporated when the solute molecule was rebuilt from the coordinated structure file. A volume displacement attempt was periodically performed after 3125 attempts for the aqueous system. The acceptance ratio for displacements of molecules and volume changes was set to ∼40%. The displacements of the water molecule were done in the same manner except that the intramolecular structure was maintained.

The water molecules are represented by the TIP4P model. In this model the water molecule consists of four points in a plane with fixed distances to each other. One point has Lennard-Jones interaction with neighbor molecules, and three points carry partial charges. The solute molecule is modeled by the optimized potential for liquid simulations all-atom (OPLS-AA) force field[28] for fully flexible molecules. Bond stretching and bond angle bending energies are represented by harmonic functions of deviation from the equilibrium distance and angle, respectively. The torsions of dihedral angles are given by a four-term sine series. Lennard-Jones (12−6) potential and electrostatic interaction apply between atoms within the molecule as well as with surrounding molecules. The cutoff distance r_c for interaction was set to 10 Å. To verify that the solute and a water molecule are within r_c all non-hydrogen atoms are selected on the solute. If any of the selected atoms are closer than r_c to the water molecule, the molecules are regarded as being inside the cutoff. Tail corrections are applied for the Lennard-Jones interactions where the radial distribution function is set to unity beyond r_c. The electrostatic interactions are in a quadratic manner smoothly brought to zero in the range from $r_c - 0.5$ Å to r_c. The TIP4P water is optimized without the use of Ewald summation,[31] and consequently Ewald summation was not employed in the simulations. The partial charges were calculated using Austin Model 1 (AM1)[32] and Charge Method 1A (CM1A)[33] of an optimized three-dimensional structure in vacuum taken from a BOSS library. To take into account the polarization of the molecule when brought into water, the AM1−CM1 charges were multiplied by 1.14 as has been recommended by Udier-Blagović et al.[34] The molecules had zero net charge. In version 4.6 of BOSS equivalent atoms, such as the three hydrogen atoms in a methyl group, are given the same charges.

A second kind of simulation was carried out to allow us to calculate free energy changes upon removing interactions between the solute and water. For this purpose the FEP[3] method with double-wide sampling[35] was employed. The final configuration of a standard NPT simulation was taken as the initial one in the FEP simulation. In the first round of the FEP simulation the electrostatic interaction was gradually turned off by reducing λ in the expression

$$E_C(\lambda) = \lambda \sum_{\text{all water molecules}} \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (9)$$

where the $1 \rightarrow 0$ scaling of λ was performed in 10 simulations. Here q_i and q_j are the partial charges of atom i in the solute and point j in the water molecule, and r_{ij} is the distance between them. The vacuum permittivity is ϵ_0. In the first simulation the reference state was $\lambda = 0.95$, and the free energy change to the perturbed state $\lambda = 1.00$ was calculated as

$$\Delta G^{\bullet}_{\lambda=0.95\rightarrow\lambda=1.00} = G^{\bullet}_{\lambda=1.00} - G^{\bullet}_{\lambda=0.95} = -RT \ln \langle \exp(-[E(\lambda = 1.00) - E(\lambda = 0.95)]/RT) \rangle_{\lambda=0.95} \quad (10)$$

Here $E(\lambda)$ is the total energy of the system at scaling factor λ. The sampled average $\langle \cdots \rangle_\lambda$ is taken over the configurations generated by the Boltzmann distribution corresponding to $E(\lambda)$ at its reference value. As the reference state is the same, the analogous value $\Delta G^{\bullet}_{\lambda=0.95\rightarrow\lambda=0.90}$ can be calculated in the same simulation, hence the name double-wide sampling. The final configuration of this simulation was taken as the initial state for the next simulation, $\lambda = 0.85$, etc. In each simulation free energy changes are calculated to perturbed states both more and less charged. But since the initial configuration in each simulation is taken from a simulation with higher λ we can say that there is an overall direction of removing the charges in the procedure. By summing over these stepwise changes in free energy we obtain the total free energy change $\Delta G^{\bullet}_{charge}$ corresponding to step iii above.

In the second round of the FEP simulation, the drug molecule was gradually shrunk by scaling all of the equilibrium bond lengths linearly with λ from 100% to 10%. The Lennard-Jones energy was not scaled linearly by λ since in that case the water molecules would, due to the infinite value of the Lennard-Jones

potential at the origin, be excluded from the centers of the drug atoms for all $\lambda > 0$. Instead the interaction was scaled down according to

$$E_{\mathrm{LJ}}(\lambda) = \lambda \sum_{\text{all water molecules}} \sum_i 4\epsilon_{i,\mathrm{w}} \left(\left(\frac{\sigma_{i,\mathrm{w}}(0.9\lambda + 0.1)}{r_{i,\mathrm{w}} + (1-\lambda) r_{\mathrm{shift}}} \right)^{12} - \left(\frac{\sigma_{i,\mathrm{w}}(0.9\lambda + 0.1)}{r_{i,\mathrm{w}} + (1-\lambda) r_{\mathrm{shift}}} \right)^{6} \right) \quad (11)$$

where we have used $r_{\mathrm{shift}} = 1.5$ Å. Here $r_{i,\mathrm{w}}$ is the distance between solute atom i and the TIP4P Lennard-Jones point, and the corresponding Lennard-Jones parameters are $\epsilon_{i,\mathrm{w}} = \sqrt{\epsilon_i \epsilon_{\mathrm{w}}}$ and $\sigma_{i,\mathrm{w}} = \sqrt{\sigma_i \sigma_{\mathrm{w}}}$. This scaling allows the water molecules to spatially overlap with the drug molecule and thus mimic the pure water system better. The $1 \rightarrow 0$ scaling of λ was performed in 10 simulations with double-wide sampling analogous to that with the electrostatic FEP simulations. By summing over these stepwise changes in free energy we obtain the total free energy change $\Delta G^{\bullet}_{\mathrm{cav+insert}}$ corresponding to steps i and ii above. The FEP simulations for decoupling the electrostatic interaction consisted of a total of 100 million single molecule move attempts for the equilibration and 200 million attempts for sampling. The FEP simulations for decoupling the electrostatic interaction and for decoupling the Lennard-Jones interaction were equally long.

The following 48 solutes were investigated by simulations: acetaminophen, acetylsalicylic acid, atropine, barbital, benzene, benzocaine, bifonazole, caffeine, chloramphenicol, chlorpromazine, cocaine, codeine, desipramine, diazepam, diethylstilbestrol, estradiol, ethyl-*p*-hydroxybenzoate, fenbufen, fenclofenac, fluconazole, flurbiprofen, griseofulvin, ibuprofen, imipramine, indomethacin, indoprofen, ketoprofen, lidocaine, lorazepam, morphine, naproxen, nitrofurantoin, oxazepam, perphenazine, phenacetin, phenobarbital, phenytoin, procaine, progesterone, promazine, prostaglandin E2, pseudoephedrine, salicylic acid, sulindac, testosterone, triclosan, trifluoperazine, and triflupromazine.

3. Results and Analysis

3.1. Cavity Creation and Introduction of Lennard-Jones Interaction. In this section the total process of cavity formation and insertion of an alkane-like solute in water is considered. As mentioned previously, these two steps were simulated in one process. Thus it is not possible for us to directly separate the contributions from each of the two steps. However, the two processes are modeled in the semiempirical theory below in a manner that is supported by the simulation results. For reasons that will be apparent soon we start by making a mean field estimate of $\Delta G^{\bullet}_{\mathrm{insert}}$ for insertion of a solute molecule into a ready-made cavity. Then we will proceed to the formation of this cavity in the solvent.

Let us start with the insertion step, i.e., the imaginary process of inserting an alkane-like solute in an existing suitable cavity. The process involves the coupling of Lennard-Jones solute−solvent interaction and the simultaneous restructuring of the solvent. In the mean field approximation, there is no significant restructuring of the surrounding solvent, which implies a negligible change in entropy. Since the pressure−volume work is also negligible when turning on the Lennard-Jones interaction, we use the ansatz

$$\Delta G^{\bullet}_{\mathrm{insert}} \cong E_{\mathrm{LJ}}(q = 0) \quad (12)$$

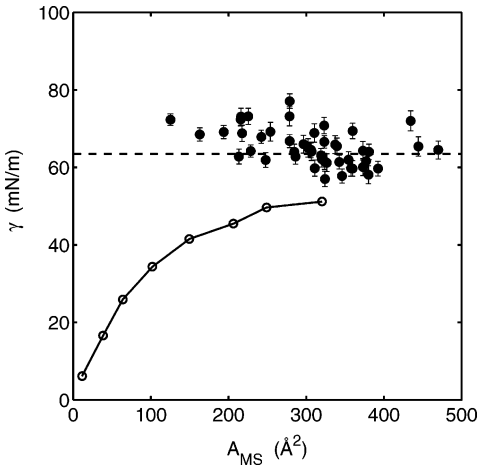

Figure 5. Simulation results for the 48 solutes in our data set. Filled circles represent the surface tension calculated by eq 13. The error bars show standard deviations in the simulations. Open circles represent the hard sphere surface tension as determined by Floris et al.[36] The dashed line shows the bulk surface tension of TIP4P water.[37]

Here $E_{\mathrm{LJ}}(q = 0)$ is the Lennard-Jones interaction energy between alkane-like solute and water. There is no need for the superscript • for interaction energies; they are the same with fixed or free solute center of mass.

The other imagined process in the FEP simulation is the process of only forming a cavity in water. The free energy must increase as the water molecules next to the cavity surface lose interaction energy. Furthermore, these molecules also lose entropy as their structure becomes more restricted. The standard approach for estimating the free energy change upon cavity creation is based on the process of making a planar macroscopic surface between water and vapor. The cost in free energy making a planar surface is $\Delta G = A\gamma_{\infty}$, where A is the surface area and γ_{∞} is the surface tension for the macroscopic water−vapor interface. To translate this expression to cavities of molecular size we need to consider both the molecular surface area and the surface tension of a small curved surface. Since the cavity is of molecular size, its area is highly dependent on where to put the interface between solute and solvent. Huang and Chandler noted that when water is the solvent and the solute is an alkane the interface is drawn close to the solute.[9] Note that this is not necessarily true for other solvents. As the free energy of cavity formation actually is to be calculated for alkane-like solutes, we used their observation and defined the molecular surface area, A_{MS}, as described in section 2.3.

If the mean field approximation and the estimation of A_{MS} are correct, then we would obtain the molecular surface tension by

$$\gamma = \frac{\Delta G_{\mathrm{cav}}}{A_{\mathrm{MS}}} = \frac{\Delta G^{\bullet}_{\mathrm{cav+insert}} - E_{\mathrm{LJ}}(q = 0)}{A_{\mathrm{MS}}} \quad (13)$$

where all quantities on the right-hand side are obtained in the simulations. The calculated γ values for the 48 molecules (with partial charges removed) are plotted in Figure 5. The open circles in Figure 5 show the FEP results of creating a "cavity" in TIP4P water performed by Floris and co-workers[36] by inserting a hard sphere. In their simulations they observed that the surface tension decreases with radius for small hard spheres. Surprisingly, this behavior of the cavities is not observed for the simulated molecules. The filled circles in Figure 5 show a more constant value of the surface tension. Furthermore, the average value of $\gamma = 64.8$ mN/m is in good agreement with the calculated

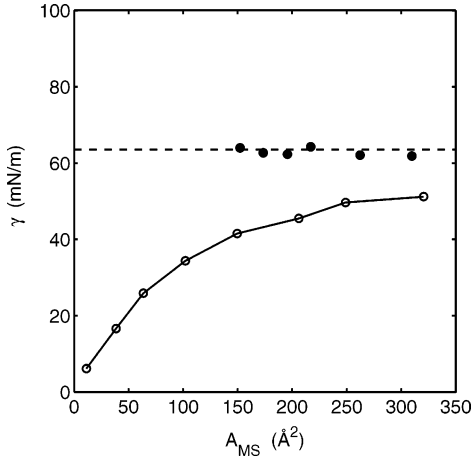

Figure 6. Simulation results for alkanes, with data partly from Gallicchio et al.[7] Filled circles represent the surface tension as calculated by eq 13. Open circles represent the hard sphere surface tension as determined by Floris et al.[36] The dashed line shows the bulk surface tension of TIP4P water.[37]

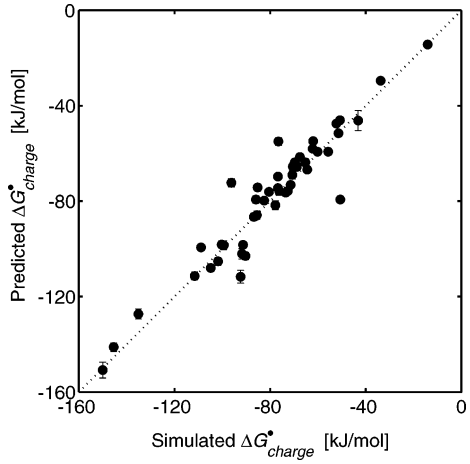

Figure 7. Simulation results for the 48 solutes in our data set. The predicted $\Delta G^{\bullet}_{\text{charge}}$ values are calculated from interaction energies and surface area using eq 16, and simulated $\Delta G^{\bullet}_{\text{charge}}$ values are from FEP simulations.

macroscopic TIP4P surface tension found to be 63.5 ± 5 mN/m,[37] which is approximately 15% lower than the surface tension at 25 °C of real water, which is 72 mN/m.[38]

Let us now see if alkanes give the same γ. Since real alkanes have only small partial charges, eq 13 should be applicable to the complete hydration of such molecules. By combining $\Delta G^{\bullet}_{\text{cav+insert}}$ from FEP simulations in TIP4P water by Gallicchio and co-workers[7] and values of A_{MS} and E_{LJ} calculated in our simulations in the same way as for the solutes in our data, we have calculated γ and found the same behavior as for our alkane-like molecules (Figure 6). The simulated hydration free energies for methane to hexane form a linear function of the number of carbon atoms. We can then by linear extrapolation obtain hydration free energies for larger alkanes up to dodecane. The extrapolation is also justified by experimental data giving linear behavior for the hydration free energy versus the number of carbon atoms.[39] Then γ values were calculated according to eq 13 and plotted in Figure 6. The result is an almost constant γ value with an average of 62.8 mN/m, which again is close to the macroscopic surface tension of the TIP4P water model.

The data in both Figures 5 and 6 seem to contradict the data of Floris et al. The rodlike shape of the alkane molecules may, however, imply that the average curvature of the cavity surface is nearly independent of rod length. The nonspherical shape of the molecules can thus plausibly explain why the surface tension maintains a macroscopic value for also quite small molecules. On the basis of these findings we propose the following model for the change in free energy upon inserting an alkane-like molecule

$$\Delta G^{\bullet}_{\text{cav+insert}} \cong A_{\text{MS}}\gamma_{\infty} + E_{\text{LJ}}(q = 0) \tag{14}$$

3.2. Introduction of Electrostatic Interaction. The last subprocess is to turn on the electrostatic interaction between the solute and water. This process has been studied in detail by, for example, Åqvist and Hansson.[26] A mean field approximation is not applicable as the onset of electrostatic interaction leads to a significant change in entropy of the solvent. In the linear response theory the solvent is assumed to respond linearly to the change in interaction energy. Generally, a linear response will lead to a simultaneous change in entropy that will compensate for half of the change in interaction energy. Hence, using the assumption found valid for a few selected solutes in our data set, namely, that the electrostatic interaction energy would normally vanish due to electroneutrality in the absence of solvent response, we have

$$\Delta G = \frac{E_{\text{C}}}{2} \tag{15}$$

Here E_{C} is the electrostatic interaction energy between water and the fully coupled solute. The change in water structure will, however, bring about a change in Lennard-Jones interaction energy as well. We believe that this change should be incorporated in the ΔG value associated with charging since the physically logical stepwise procedure of turning on interaction would start with hard interactions producing a nonlinear response, then add electrostatic interactions producing a linear response and with Lennard-Jones attractions assumed not to induce a structural response. Neglecting the effect of Lennard-Jones interactions on the electrostatic energy and evaluating the former as in the final structure we obtain the free energy change as

$$\Delta G^{\bullet}_{\text{charge}} \cong \Delta E_{\text{LJ}}(q = 0 \rightarrow 1) + \frac{E_{\text{C}}}{2} \tag{16}$$

where $\Delta E_{\text{LJ}}(q = 0 \rightarrow 1)$ is the change in Lennard-Jones solute−water interaction energy brought by the restructuring of the solvent in response to the electrostatic interaction. In Figure 7 we can see how the predicted $\Delta G^{\bullet}_{\text{charge}}$ values calculated using eq 16 versus $\Delta G^{\bullet}_{\text{charge}}$ values obtained directly from FEP simulations fall close to a straight line of proportionality with a slope of 1.0 and an average value of 0.51 for $(\Delta G^{\bullet}_{\text{charge}} - \Delta E_{\text{L}}(q = 0 \rightarrow 1))/E_{\text{C}}$. The coefficient of correlation is $R^2 = 0.86$. If we instead calculate the ratio $\Delta G^{\bullet}_{\text{charge}}/E_{\text{C}}$, then we obtain an average ratio of 0.4 with $R^2 = 0.83$.

Åqvist and Hansson found for the free energy of hydration that the ratio $\Delta G^{\bullet}_{\text{charge}}/E_{\text{C}}$ falls between 0.29 and 0.44 for a set of neutral organic molecules. However, they did not consider $\Delta E_{\text{LJ}}(q = 0 \rightarrow 1)$. If we follow their analysis for our data set, then we obtain $\Delta G^{\bullet}_{\text{charge}}/E_{C} = 0.45$ on average.

Our tentative conclusion is that including $\Delta E_{\text{LJ}}(q = 0 \rightarrow 1)$ yields a more accurate model and that indeed the solutions feature a linear response to the onset of electrostatic interaction and a mean field response to Lennard-Jones interactions after the restructuring caused by the electrostatic interaction. Our account of the effect of changing Lennard-Jones interactions

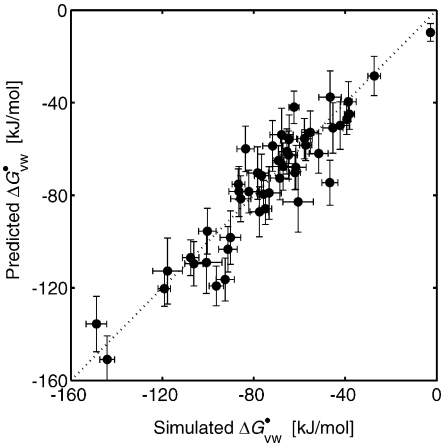

Figure 8. Simulation results for the 48 solutes in our data set. The predicted $\Delta G^{\bullet}_{vw}$ values are calculated from interaction energies, surface area, and TIP4P macroscopic surface tension using eq 17. The simulated $\Delta G^{\bullet}_{vw}$ values are from FEP simulations.

upon charging is, however, still ad hoc. A more rigorous treatment of the coupling between solute and solvent and between energy end entropy changes is possible and desirable but will have to wait until a more fundamental investigation has been completed in which the turning on or off of the interactions in the simulations could be done in the physically most logical order.

3.3. Total Hydration Process. The total hydration process (subprocesses i−iii) can now be modeled by the relation

$$\Delta G^{\bullet}_{vw} \cong A_{MS}\gamma_{\infty} + E_{LJ}(q=1) + \frac{E_C}{2} \qquad (17)$$

where $E_{LJ}(q = 1) = E_{LJ}(q = 0) + \Delta E_{LJ}(q = 0 \rightarrow 1)$ is the solute−water Lennard-Jones interaction for the fully charged solute. A plot of predicted $\Delta G^{\bullet}_{vw}$ values calculated using eq 17 versus $\Delta G^{\bullet}_{vw}$ values obtained directly from FEP simulations is shown in Figure 8. The results show that the free energy of hydration for most of our solutes can be accurately predicted by using interaction energies and our definition of molecular surface area. Furthermore, the final model for the hydration energy in eq 17 has been tested against simulation data for small organic molecules generated by Carlson and Jorgensen.[11] The values of $\Delta G^{\bullet}_{vw}$, $E_{LJ}(q = 1)$, and E_C are taken from their work, but A_{MS} is calculated using our definition of molecular surface area. By rearranging eq 17 the surface tension can be calculated by

$$\gamma = \frac{\Delta G^{\bullet}_{vw} - E_{LJ}(q=1) - \frac{E_C}{2}}{A_{MS}} \qquad (18)$$

As illustrated in Figure 9, the surface tension is again close to the macroscopic TIP4P surface tension. Possibly, a trend of a small decrease in surface tension can be seen for the smallest molecules, which are CH_3OH and CH_3NH_2. Using eq 17 on the same data using the surface tension of TIP4P for γ_{∞}, the relation between free energy changes determined by FEP simulations and those obtained using our model is as shown in Figure 10.

3.4. Comparison with Experimental Data. It seems that we have reached our goal to make a semiempirical model, which with good accuracy can let us substitute the time-consuming FEP simulations by quicker interaction energy simulations. But

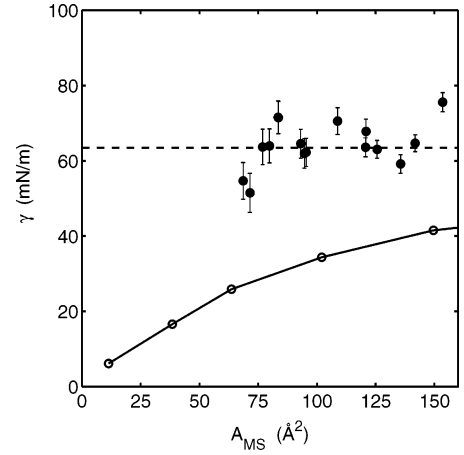

Figure 9. Simulation results for 16 small organic molecules, for which data is partially from work by Carlson and Jorgensen.[11] Filled circles represent the surface tension calculated by eq 18. The error bars show standard deviations in the simulations. Open circles represent hard sphere surface tension by Floris et al.[36] The dashed line shows the bulk surface tension of TIP4P water.[37]

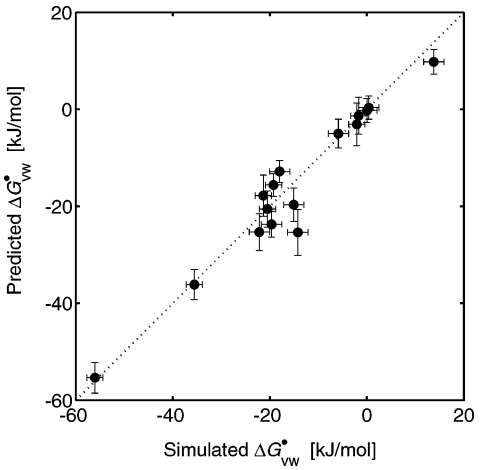

Figure 10. Simulation results for 16 small organic molecules, for which data is partially from work by Carlson and Jorgensen.[11] The predicted $\Delta G^{\bullet}_{vw}$ values are calculated from interaction energies, surface area, and TIP4P macroscopic surface tension using eq 17. The simulated $\Delta G^{\bullet}_{vw}$ values are from FEP simulations.

we can actually reach even further. For seven of the drugs in our database (acetylsalicylic acid, ethyl-*p*-hydroxybenzoate, flurbiprofen, ibuprofen, ketoprofen, naproxen, and salicylic acid) and benzene we have found literature data on the hydration free energy derived from vapor pressure[40−41] and crystalline solubility.[42]

An appropriate hydration free energy can be calculated using eq 17 and simulated interaction energy, molecular surface area, and surface tension. This value is compared to the experimental value in Figure 11. Since the literature value of the macroscopic surface tension of TIP4P water is not accurately determined, the error bars of the simulated values are large (filled circles in Figure 11). However, since the experimental surface tension is available we have used it to improve the prediction from eq 17. The open circles in Figure 11 show $\Delta G^{\bullet}_{vw}$ when the experimental surface tension of 72 mN/m has been used in eq 17. These results actually show a significantly improved agreement with experimental data. The possibility of inserting experimental data in the semiempirical model is a significant advantage of the model compared to the use of FEP simulations. It must, however, be noted that there is an inconsistency in the insertion of the experimental surface tension. The interaction

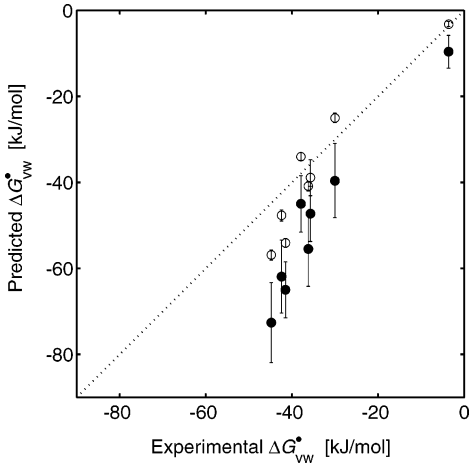

Figure 11. Simulation results for 8 of our 48 solutes. The predicted $\Delta G^{\bullet}_{vw}$ values are calculated from interaction energies, surface area, and TIP4P macroscopic surface (tension filled circles) and experimental surface tension (open circles). The experimental $\Delta G^{\bullet}_{vw}$ values are calculated from experimental vapor pressure measurements and crystal solubilities.[40−42]

energy and molecular surface area terms in eq 17 are still taken from simulations with TIP4P simulations. Thus properties of TIP4P and real water are both used in the same equation. It is noteworthy that a relatively minor difference in surface tension between the TIP4P model and real water plays an important role in determining the free energy of hydration especially for large solutes such as modern drug molecules.

4. Discussion

Our purpose has been to find an in silico prediction strategy that is so fast that it can be used routinely but is still robust; i.e., the prediction must not be gravely in error for any drug compound. Due to our demand for robustness QSPR methods are not useful as they are not reliable for compounds not similar to those in the selected training set. Nor is there a reliable definition of "similarity" in this context. An alternative is to calculate the chemical potential in an aqueous solution and in pure drug using free energy simulation. Although possibly robust such methods would require long simulation times. This is true for the hydration simulations presented in this publication and even more so for simulations of pure drug systems that we will present in future publications. Therefore we have sought an approximative but still reliable model allowing us to replace the FEP simulations by much more efficient standard NPT simulations. According to our present results it appears to be the case that the intermolecular energies and molecular surface areas obtained in NPT simulations can be used to estimate the free energy of hydration.

Note that we have not yet addressed the issue of accuracy of the force field. The OPLS-AA force field does not account for polarization, and since the electrostatic interaction is significant the determination of partial charges that is done for the initial structure is crucial. As mentioned above, the original partial charges were calculated in vacuum, and the polarization effect in water has been accounted for by multiplying these partial charges by 1.14. When we now have found that linear response is applicable, the polarization factor may be further tuned by fitting the free energy of hydration to experimental data for a training set. At this stage, however, the free energy of hydration is experimentally determined for only eight compounds in our data set, which comprises a too small training set for tuning the polarization factor.

We have also seen that the underestimation of surface tension of water inherent in the TIP4P model can be partly compensated for by replacing the TIP4P value by the experimental one in eq 17. The agreement between experimental and simulated free energy of hydration thereby improved substantially. Note that this semiempirical modification is not completely consistent since, for example, the underlying error in the simulation is not limited to the surface tension term but would affect the interaction energy term as well.

In the work by Carlson and Jorgensen[11] the authors used the same subprocesses as in this work. However, they did not generally split up the FEP simulations in separate subprocesses but used a fitting procedure to obtain the correlation between hydration free energy on the one hand and interaction energies and cavity area on the other. They found that the response to electrostatic interaction was approximately linear, but the response to the van der Waals interaction did not support a mean field approximation. However, since, for instance, the cavity area and the van der Waals interaction energy correlate,[11] a good correlation might result even with unphysical coefficients in the correlation. We think the agreement that we have found for our data set does support a usage of the mean field approximation, although further investigation of this point is needed.

As of matter of fact, since the interaction energies and molecular size roughly correlate with each other, dissection of the simulations is important to find accurate models for the changes in free energy. Therefore, a few decades ago, models were proposed that correlate solubility with molecular volume or area only.[13] Such models would in our view be too simple for our purposes. Moreover, we can see that the three subprocesses in our approach are approximately equally important for the total $\Delta G^{\bullet}_{vw}$. The average of the free energy change upon electrostatic coupling for all of the molecules in our data set is $\Delta G^{\bullet}_{charge} \cong -80$ kJ/mol (Figure 7). Assuming mean field treatment of the coupling of Lennard-Jones interaction we obtain on average $\Delta G^{\bullet}_{insert} \cong -115$ kJ/mol (not shown in figure) and $\Delta G_{cav} \cong +120$ kJ/mol (derived from Figure 8). Hence, none of the subprocesses can be ignored nor clearly dominate over the other.

The free energy of cavity formation has drawn a lot of attention in the literature. Floris et al.[36] actually performed simulations of inserting hard spheres into TIP4P water as illustrated in Figures 5, 6, and 9. The data are also shown in Figure 12 as open circles. In the GvdW theory[25] the surface tension for creating a spherical cavity in a solvent can be estimated as

$$\gamma_{GvdW} = \frac{\Delta G_{GvdW}}{A_{MS}} = \gamma_{\infty}\frac{A_{MS}\left(1 - \frac{8}{27}\left(\frac{\sigma_w}{\sigma_s}\right)^2\left(2 + 3\ln\frac{\sigma_s}{\sigma_w}\right) - \frac{1}{81}\left(\frac{\sigma_w}{\sigma_s}\right)^8\right)}{A_{MS}} \quad (19)$$

where σ_s and σ_w are the diameters of the cavity and the solvent, respectively. The diameter of the water molecules can be calculated from an equation of state for pure water. The corresponding GvdW equation of state is

$$\frac{P}{RTc_w} = \frac{1}{1-\phi} + \frac{E_{coh,LJ} + E_{coh,C}/2}{RT}, \quad (20)$$

where the volume fraction is $\phi = \pi N_A c_w \sigma_w^3/6$ and N_A is Avogadro's constant. The cohesive energy of pure water is

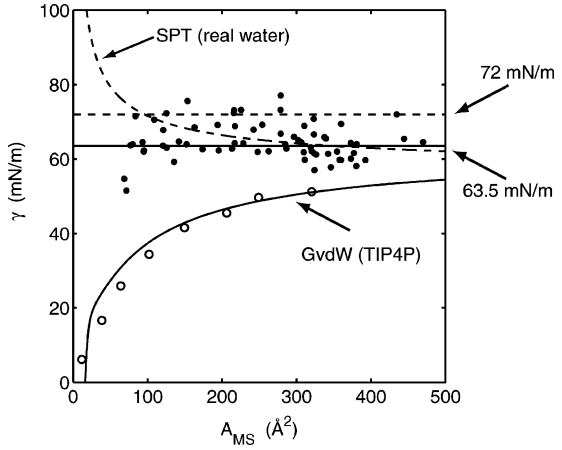

Figure 12. Surface tension vs molecular surface area. The upper dashed curve shows the SPT model for real water, and the lower dashed curve is the constant $\gamma = 72$ mN/m. The solid curves correspond to the TIP4P water model where the upper solid curve is the constant value $\gamma = 63.5$ mN/m and the lower solid curve shows the GvdW model with $\sigma_w = 3.66$ Å and $\gamma_\infty = 63.5$ mN/m. The open circles are the hard sphere surface tension data of Floris et al.[36] The filled circles are the simulation data in Figures 5, 6, and 9.

divided into a Lennard-Jones part, $E_{coh,LJ}$, and an electrostatic part, $E_{coh,C}$. From simulations with BOSS we obtain $E_{coh,LJ} = +7.7$ kJ/mol and $E_{coh,C} = -49.8$ kJ/mol, and the equation of state yields $\sigma_w = 3.66$ Å at 298 K and 1 atm. By inserting this value in eq 19 the bottom solid curves in Figure 12 result for the TIP4P value $\gamma_\infty = 63.5$ mN/m.

Another model for cavity formation is the scaled particle theory (SPT).[24] As in the GvdW theory the water molecules are not hard spheres but rather interact with each other attractively as well as repulsively. At pressure P the SPT yields the surface tension

$$\gamma_{SPT} = \frac{\Delta G_{SPT}}{A_{MS}} = \frac{RT}{N_A A_{MS}} \times \left(-\ln(1-\phi) + \frac{3\phi}{1-\phi}\frac{\sigma_s}{\sigma_w} + \left(\frac{3\phi}{1-\phi} + \frac{9}{2}\left(\frac{\phi}{1-\phi}\right)^2\right)\left(\frac{\sigma_s}{\sigma_w}\right)^2 + \frac{\phi P}{RTc_w}\left(\frac{\sigma_s}{\sigma_w}\right)^3\right) \quad (21)$$

The diameter of water may be set to 2.75 Å, which is close to the position of the first peak in the oxygen−oxygen radial distribution function in pure water.[41] The surface tension will then vary with A_{MS} as in the top dashed curve in Figure 12. However, this curve seems to be far too low as it should approach $\gamma = 72$ mN/m for large cavities. It appears that the GvdW theory best reproduces the data of Floris et al. both qualitatively and quantitatively. Note that the SPT model predicts a decrease in $\Delta G_{cav}/A_{MS}$ for increasing cavity size. For very small cavities this is of course correct since $\Delta G_{cav} > 0$ even for a pointlike cavity[24] with $A_{MS} = 0$, but the data of Floris et al. show clearly that this should occur for smaller cavities than those in which we are interested. Also Huang and Chandler[9] propose that the surface tension increases linearly with radius for small cavities and is constant for larger cavities.

The surface tensions obtained for the molecules in our data set (Figure 5), from Gallicchio et al.[7] (Figure 6) and Carlson and Jorgensen[11] (Figure 9), are also plotted in Figure 12. Remarkably, there is a considerable discrepancy between the surface tension of the hard spheres of Floris and the cavities accommodating the molecular solutes. The magnitude of the surface tension is of course a matter of how the molecular surface area is defined, as Southall[13] asks "What is the right measure of contact?" In simulation results of Huang and Chandler[9] the water−vapor interface appears to be very sharp using the SPC/E water model, which indicates a well-defined cavity−water interface. This is, however, something that other solvents might lack. Huang and Chandler[9] also showed that for interaction forces typical of water−alkanes the interface is in contact with the solute surface. Since our cavities correspond to the alkane-like solutes we were inspired to define a molecular surface area placed close to the solute atoms. With this definition of A_{MS} the surface tension of the cavities corresponding to the large molecules falls close to the macroscopic value of TIP4P. According to Figure 12 it appears that the macroscopic surface tension is also applicable for the smallest molecules. We believe the reason for this is that the molecules and thus the cavities are nonspherical. With, for instance, the oblate form of the cavity a higher fraction of the surface has less curvature than that for a sphere.

The introduction of electrostatic interaction is of quantitative importance for drug molecules in contrast to the unpolar alkanes. It is customary to use linear response as a first approximation for the response of water. To be able to calculate the Lennard-Jones attraction by mean field theory before introducing partial charges in the solute molecule we would have to assume that the structural change imposed by charging is decoupled from the Lennard-Jones interaction in accord with the linear response treatment of charging. Clearly, that is not the case since the Lennard-Jones energy also changes significantly. This clearly suggests that we should have turned on the charges before the Lennard-Jones attractions. Still, the modeling $\Delta G^{\bullet}_{charge} \cong \Delta E_{LJ}(q = 0 \rightarrow 1) + E_C/2$ appears to be accurate. We therefore propose the following interpretation of the process: The coupling of electrostatic interaction causes a reorganization of the water molecules that gives a linear response. However, the reorganization causes a change in the Lennard-Jones interaction, but with the mean field approximation this change does not cause a further reorganization of the solvent. The Lennard-Jones interaction is thus only superimposed on the electrostatic response. When not taking into account the change in Lennard-Jones energy we obtain $\Delta G^{\bullet}_{charge}/E_C < 0.5$ in accordance with the findings of Åqvist and Hansson.[26] When, instead, using the full model in eq 16 $(\Delta G^{\bullet}_{charge} - \Delta E_{LJ}(q = 0 \rightarrow 1))/E_C$ is close to one-half. We therefore believe that the response in the work of Åqvist and Hansson also would appear more linear if they had used eq 16. Although good agreement is found in Figure 7 there are a few molecules with negative or positive errors up to 20 kJ/mol. Sandberg et al.[20] have also studied the electrostatic effects on free energy of hydration but with quantum mechanical treatment of the electrostatic interaction between small neutral organic molecules and continuum-modeled water. They proposed that nonlinear effects are important, at least for small molecular solutes, as they contribute as much as 10 kJ/mol to the free energy of hydration. However, they did not dissect the total hydration into subprocesses, and it is difficult for us to judge the accuracy of their calculations of cavity formation and Lennard-Jones interaction. It is therefore left for further study whether inclusion of nonlinear effects in our model would improve the agreement between predicted $\Delta G^{\bullet}_{charge}$ and FEP results.

In the future papers of our series on prediction of drug solubility we will investigate the process of bringing a drug solute from the vapor phase to the pure amorphous phase. This

process is analogous to hydration but is considerably more difficult to simulate. At 25 °C the amorphous phase of most solutes in our data set is a glass, and when removing a molecule from the glass the neighbor molecules cannot easily fill up the remaining hole. Therefore we will simulate the removal of a molecule at 400 °C, where all of the systems are molten, and then cool the systems to room temperature.

Acknowledgment. This work was supported by the Swedish Research Council.

Supporting Information Available: A summary of all important simulation results for 48 molecules. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Jain, N; Yalkowsky S. *J. Pharm. Sci.* **2001**, *90*, 234.

(2) Jorgensen, W. L.; Duffy, E. M. *Adv. Drug Delivery Rev.* **2002**, *54*, 355.

(3) Pearlman, D. A.; Rao, B. G. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley: New York, 1998; Vol. 2, p 1036.

(4) Åquist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng.* **1994**, *7*, 385.

(5) Westergren, J.; Nordholm, S. *Chem. Phys.* **2003**, *290*, 189.

(6) Hildebrand, J. *J. Am. Chem. Soc.* **1916**, *38*, 1452.

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

(8) Graziano, G. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3567.

(9) Huang, D.; Chandler, D. *J. Phys. Chem. B* **2002**, *106*, 2047.

(10) Guillot, B.; Guissani, Y. *J. Chem. Phys.* **1993**, *99*, 8075.

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

(12) Tanford, C. *The Hydrophobic Effect*; Wiley: New York, 1973.

(13) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. *J. Phys. Chem. B* **2002**, *106*, 521.

(14) Frank, F. S.; Evans, M. W. *J. Chem. Phys.* **1945**, *13*, 507.

(15) Gavezotti, A. *CrystEngComm* **2002**, *4*, 343.

(16) Day, G. M.; Motherwell, W. D. S.; Amon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Vinuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweitzer, B.; van Eijck, B. P.; Erk, P.; Vacelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwe, P. *Acta Crystallogr., Sect. B: Struct. Sci.* **2005**, *61*, 511.

(17) Neau, S. H.; Bhandarkar, S. V.; Hellmuth, E. H. *Pharm. Res.* **1997**, *14*, 601.

(18) Lindfors, L.; Forssén, S.; Skantze, P.; Skantze, U.; Zackrisson, A.; Olsson, U. *Langmuir* **2006**, *22*, 911.

(19) McDonanld, N. A.; Carlson, H. A.; Jorgensen, W. L. *J. Phys. Org. Chem.* **1997**, *10*, 563.

(20) Sandberg, L.; Casemyr, R.; Edholm, O. *J. Phys. Chem. B* **2002**, *106*, 7889.

(21) Jorgensen, W. L. *Biochemical and Organic Simulation System*, version 4.6; Yale University: New Haven, CT, 2004.

(22) Jorgensen, W. L. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley: New York, 1998; Vol. 5, p 3281.

(23) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689.

(24) Hooper, M.; Nordholm, S. *J. Chem. Phys.* **1984**, *81*, 2432.

(25) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717.

(26) Åquist, J.; Hansson, T. *J. Phys. Chem.* **1996**, *100*, 9512.

(27) (a) Ben-Naim, A. *Solvation Thermodynamics*, 1st ed.; Plenum Press: New York, 1987**.** (b) Ben-Naim, A; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.

(28) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M.; Tirado-Rives, J. *J. Chem. Phys.* **1983**, *79*, 926.

(29) Connolly, M. L. *J. Mol. Graphics* **1993**, *11*, 139.

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

(31) Jamshed, A.; Heyes, D. M. *J. Chem. Phys.* **2005**, *122*, 224117.

(32) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(33) Storer, J. W.; Giesen, D. J.; Cramer C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.

(34) Udier-Blagović, M.; De Tirado, P. M.; Pearlman, S. A.; Jorgensen, W. L. *J. Comp. Chem.* **2004**, *25*, 1322.

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

(36) Floris, F.; Selmi, M.; Tani, A.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 6353.

(37) Pohorille, A.; Wilson, M. *J. Mol. Struct. (THEOCHEM)* **1993**, *284*, 271.

(38) Abraham, M.; Nasehzadeh, A. *J. Chem. Soc., Faraday Trans. 1* **1981**, *77*, 321.

(39) McAuliffe, C. *J. Phys. Chem.* **1966**, *70*, 1267.

(40) (a) Perlovich, G. L.; Bauer-Brandl, A. *Curr. Drug Delivery* **2004**, *1*, 213. (b) Perlovich, G. L.; Rodionov, S. V.; Bauer-Brandl, A. *Eur. J. Pharm. Sci.* **2005**, *24*, 25. (c) Perlovich, G. L.; Bauer-Brandl, A. *Eur. J. Pharm. Sci.* **2003**, *19*, 423. (d) Perlovich, G. L.; Kurkov, S. V.; Hansen, L. K.; Bauer-Brandl, A. *J. Pharm. Sci.* **2004**, *93*, 654. (e) Perlovich, G. L.; Kurkov, S. V.; Kinchin, A. N.; Bauer-Brandl, A. *J. Pharm. Sci.* **2003**, *92*, 2502. (f) Perlovich, G. L.; Kurkov, S. V.; Kinchin, A. N.; Bauer-Brandl, A. *Eur. J. Pharm. Biopharm.* **2004**, *57*, 411. (g) Stephenson, R. M.; Malanowski, S. *Handbook of the Thermodynamics of Organic Compounds*; Elsevier: New York, 1987.

(41) Graziano, G. *Biophys. Chem.* **1999**, *82*, 69.

(42) (a) Bergström, C. A. S; Norinder, U.; Luthman, K.; Artursson, P. *Pharm. Res.* **2002**, *19*, 182. (b) Chen, X.-Q.; Cho, J. S.; Li, Y.; Venkatesh, S. *J. Pharm. Sci.* **2002**, *91*, 1938. (c) Jorgensen, W. L.; Duffy, E. M. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1155. (d) James, W. McFarland, J. W.; Avdeef, A; Berger, C. M.; Raevsky, O. A. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1355.