ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/6538873

In Silico Prediction of Drug Solubility: 2. Free Energy of Solvation in Pure Melts

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · MARCH 2007

Impact Factor: 3.3 · DOI: 10.1021/jp0642239 · Source: PubMed

CITATIONS
15

READS
44

5 AUTHORS, INCLUDING:

Jan Westergren
AstraZeneca
16 PUBLICATIONS **375** CITATIONS
SEE PROFILE

Sture Nordholm
University of Gothenburg
215 PUBLICATIONS **3,512** CITATIONS
SEE PROFILE

Available from: Jan Westergren
Retrieved on: 26 January 2016

In Silico Prediction of Drug Solubility: 2. Free Energy of Solvation in Pure Melts

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

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

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

The solubility of drugs in water is investigated in a series of papers and in the current work. The free energy of solvation, $\Delta G_{vl}^{\bullet}$, of a drug molecule in its pure drug melt at 673.15 K (400 °C) has been obtained for 46 drug molecules using the free energy perturbation method. The simulations were performed in two steps where first the Coulomb and then the Lennard-Jones interactions were scaled down from full to no interaction. The results have been interpreted using a theory assuming that $\Delta G_{vl}^{\bullet} = \Delta G_{cav} + E_{LJ} + E_C/2$ where the free energy of cavity formation, ΔG_{cav}, in these pure drug systems was obtained using hard body theories, and E_{LJ} and E_C are the Lennard-Jones and Coulomb interaction energies, respectively, of one molecule with the other ones. Since the main parameter in hard body theories is the volume fraction, an equation of state approach was used to estimate the molecular volume. Promising results were obtained using a theory for hard oblates, in which the oblate axial ratio was calculated from the molecular surface area and volume obtained from simulations. The Coulomb term, $E_C/2$, is half of the Coulomb energy in accord with linear response, which showed good agreement with our simulation results. In comparison with our previous results on free energy of hydration, the Coulomb interactions in pure drug systems are weaker, and the van der Waals interactions play a more important role.

1. Introduction

The requirements on new drugs with respect to activity and selectivity are constantly increasing, which calls for greater efforts in the process of finding new drug candidates with the possibility to become a final drug product. One important feature of a new drug candidate today is high selectivity in terms of binding preferentially to a certain desired receptor to obtain the desired bioeffect. Consequently, the trend among current drug candidates is that they increase in size and become more hydrophobic than previously.[1] As the active molecules become increasingly hydrophobic the aqueous solubility may become very low resulting in poor absorption after oral administration. In the early development stage when the bioactivity of the drug molecules is to be evaluated in animal studies, it is important to prepare an appropriate formulation. The aqueous solubility of a drug molecule is one of the important properties in the selection of a suitable formulation for animal studies. Therefore, a reliable, robust, and fast method for the prediction of the aqueous solubility of drugs would be of great value. One possibility is, of course, to measure the aqueous solubility experimentally, but this is a difficult and time-consuming task, especially when the solubility is very low. Other difficulties are that the amount of the drug substance available in the discovery phase is usually very small and that knowledge of the solid state of the drug is limited.

One promising alternative to experiments is then to use theoretical and numerical methods to predict the aqueous solubility instead. A common method used in drug design in the pharmaceutical industry during the last 25 years is the quantitative structure−property relations (QSPR) method.[2] In this type of approach one usually tries to find a correlation between a relevant physiochemical property, e.g., in our case the aqueous solubility, and a number of basic descriptors, e.g., the molecular surface area, the molecular volume, and the logarithm of the octanol−water partition coefficient. Weighting factors for the different descriptors used in the QSPR relations are optimized against experimental data from a required training set consisting of an appropriate number of molecules. The main advantages of such methods are that they are easy to use, computationally fast, and able to produce reasonable results for molecules structurally similar to those included in the training set. The main disadvantages are the usually poor results for molecules that deviate too much in molecular structure from the molecules in the training set and the limited physical understanding contained in and generated by such statistical correlations.

The intrinsic water solubility S_0 is defined as the water solubility of a neutral substance in its stable crystalline form. It is, however, common that a drug can take on different crystalline forms (polymorphs), and it might be both difficult and time-consuming to find the most stable polymorph by experiment. Finding stable crystalline forms of compounds by calculations or simulations is an active field of research,[3] but, to our knowledge, no one has yet presented a reliable method to predict such results.[4] Because of these difficulties and given that most organic substances have several polymorphs, we have instead focused on the aqueous solubility of the supercooled liquid (amorphous hypothetical solid state), S_{amorph}. Approximate values of S_0 can be obtained from the calculated values of S_{amorph} and via a thermodynamic cycle for the crystal → supercooled liquid process according to the method discussed by Neau et al.[5]

Our ambition in this paper as in the previous work[6] is to propose approximate theories that are robust, relatively fast, and

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

[†] Göteborg University.

[‡] AstraZeneca R&D, Mölndal.

10.1021/jp0642239 CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/01/2007

able to produce reliable results for more complex systems. The theories should be physically sound and based on simply accessible quantities such as interaction energies, molecular size, and shape. These properties can easily be calculated from simulation data and employed in predictions of free energy changes, which can then be compared to the free energies obtained from more time-consuming simulations both for the hydration process and the process of bringing the solute molecule from the gas phase into the pure supercooled liquid (amorphous) phase. Unfortunately, as most drugs at room temperature form glasses rather than liquids this will lead to difficulties in the free energy simulations because of the rigidity of such a system. To reduce these difficulties, the free energy simulations are here instead performed at 673.15 K (400 °C) to guarantee that the molecules form melts, i.e., pure drug liquids. The free energy difference between the melt at 400 °C and the glass at 25 °C will be the subject of a forthcoming paper. We are aware that we have not solved the problem of simulating the pure melts at room temperature. But our ambition is to find a reliable theory that can replace or complement the free energy simulations at low temperatures.

In this work, results calculated by approximate theory are referred to as "predicted results", and free energy changes calculated by free energy simulations are referred to as "simulated free energy perturbation (FEP) results".[7] The reasons for comparing the predicted results to simulated free energy data are twofold: (i) Direct experimental results on the solubility of melts are lacking for all molecules studied in this work, and (ii) we want at this stage to test the accuracy and robustness of the proposed theory.

For the simulation of thermodynamic properties we have chosen the Metropolis Monte Carlo (MC) method and the isobaric−isothermal (NPT) ensemble. This is an efficient method commonly used for treating systems containing many molecules within classical statistical mechanics. Thus, the MC method implemented with accurate molecular interaction potentials is a powerful methodology for modeling molecular liquids. We have used the Biochemical and Organic Simulation System (BOSS) MC program developed by Jorgensen et al.[8,9] Included in the program is a potential function library with parameters from the all-atom optimized potential for liquid simulations (OPLS-AA) force field developed by Jorgensen et al.[10] The atomic parameters contained in the OPLS-AA force field have been developed to simultaneously yield accurate descriptions of molecular structure, conformational energetics, heats of vaporization, free energies of solvation, etc. Accurate values of properties such as energies, molecular areas, and volumes can easily be obtained by a MC simulation. In contrast, free energies and entropies are much more difficult to calculate and require longer simulations. We will compare free energies obtained by our theories with the free energies obtained from FEP simulations using the BOSS program.

2. Theory

2.1. Equilibrium between Supercooled Liquid and Solution. The aqueous solubility of a supercooled liquid can be calculated from the relation[6]

$$c_w = c_l \exp(-\Delta G^\bullet_{lw}/RT) \tag{1}$$

where R is the gas constant, T is the absolute temperature, and c_l and c_w are the number densities of the drug of pure liquid and in aqueous solution, respectively. $\Delta G^\bullet_{lw}$ is the change in free energy upon transfer of one drug molecule from its pure liquid to aqueous solution. The subscript lw stands for the liquid−water process, and the • notation will here and below mean that the free energy change is described according to the Ben-Naim[11] definition, which takes the center of mass of the molecule to be stationary in each phase. The total process can be divided into two subprocesses: (i) the free energy of hydration of the drug molecule ($\Delta G^\bullet_{vw}$) discussed in our previous paper[6] where the subscript vw stands for the vapor−water process and (ii) the free energy of solvation ($\Delta G^\bullet_{vl}$) of a drug molecule in pure liquid. We have

$$\Delta G^\bullet_{lw} = -\Delta G^\bullet_{vl} + \Delta G^\bullet_{vw} \tag{2}$$

where vl denotes the vapor−liquid process.

The Ben-Naim[11] pseudo-chemical potential (PCP) is defined as

$$\mu^\bullet = \mu - RT \ln c\Lambda^3 \tag{3}$$

where μ is the (ordinary) chemical potential, R is the gas constant, c is the number density, $\Lambda = \sqrt{h^2/(2\pi m k_B T)}$ is the thermal de Broglie wavelength, h is Planck's constant, m is the molecular mass, and k_B is the Boltzmann constant. Thus, the PCP is equal to the change in free energy when adding one solute molecule to a system and then keeping its center of mass fixed.

We want to determine the solubility at $T = 298.15$ K and $P = 1$ atm, but due to the difficulties in the estimation of accurate $\Delta G^\bullet_{vl}$ values from free energy simulations of the supercooled liquid at room temperature we will instead describe simulations at 673.15 K where all drugs are liquids. $\Delta G^\bullet_{vl}$ values at room temperature can then be calculated from the sum of the results obtained at high temperature and the change in free energy upon cooling the melt to room temperature. The latter process will be discussed in a future paper. Therefore the results, except for cyclohexane, refer to the conditions at $T =$ 673.15 K and atmospheric pressure.

2.2. Solvation Analysis in Pure Melt. The transfer of a solute molecule from vapor to a pure melt can be separated into a number of steps, which we shall resolve to improve our chances of predicting the experimental solubilities of drug molecules. The three steps that we consider are analogous to those discussed in our previous work:[6] (i) the creation of a cavity in the pure melt; (ii) turning on the dispersion interactions; (iii) turning on the Coulomb interactions. The free energy changes of these steps are denoted ΔG_{cav}, $\Delta G^\bullet_{insert}$, and $\Delta G^\bullet_{charge}$, respectively. The Ben-Naim notation for ΔG_{cav} can be omitted because the solute molecule is not involved in the cavity creation step. We note that the cavity should have the volume and shape that eliminate most of the repulsive interaction of the Lennard-Jones potentials acting between the solute and the surrounding molecules. Note also that the nonpolar solute molecule inserted into the cavity in step ii feels the full intramolecular potential including the intramolecular electrostatic interactions present. The intermolecular electrostatic interactions between the solute and the surrounding molecules are set to zero in this step but then increased from zero to full interaction in step iii.

To improve efficiency of computation and physical understanding of the process we shall analyze the individual free energy changes above in terms of a semiempirical theory. In this theory ΔG_{cav} will be related to an appropriate equation of state discussed more in detail in the Results and Discussion section. The insertion free energy will be assumed to be equal to the interaction energy due to the Lennard-Jones interaction.

This amounts to the use of a mean field approximation since no response of the surrounding molecules is accounted for. Finally, the electrostatic interactions are assumed to produce a linear response that includes an entropy compensation that generally reduces the free energy change to half of the electrostatic interaction energy. To clarify the energies that we compute, they are defined as solute energies, i.e., the total interaction energies between the solute and all of the surrounding molecules in the melt. However, it should be noted that the solute energy of interaction with the surrounding molecules amounts to twice the average energy per molecule in the drug melt. According to our simplified theory we then have

$$\Delta G^{\bullet}_{\text{insert}} \cong E_{\text{LJ}} \tag{4}$$

$$\Delta G^{\bullet}_{\text{charge}} \cong E_{\text{C}}/2 \tag{5}$$

Here E_{LJ} is the Lennard-Jones interaction potential, and E_{C} is the electrostatic potential energy between nonbonded atoms in different molecules. The Ben-Naim notation for the interaction energies can be left out due to their lack of dependence on the location of the center of mass. The division by 2 in eq 5 is the energy−entropy compensation effect according to linear response theory in its simplest form where only the linear term is accounted for. A minor complication arises in this analysis since we assume that intramolecular interactions are completely separable from the intermolecular interactions. This is not completely the case, but our simulations indicate that the coupling between inter- and intramolecular energies is small.

In our simulations we have actually used a procedure different from the stepwise analysis discussed above. First of all, the simulations are performed in the reverse direction by removing rather than adding molecules, which should not make any difference more than that the sign of the ΔG value will be changed. Thus we begin by identifying one of the drug molecules in the melt as the solute molecule and proceed to remove it from the pure liquid phase. In the first step the solute electrostatic interactions are scaled from full to zero. The accompanying change in intramolecular energy is subtracted by simulating the same process in vacuum. In the second step, the Lennard-Jones interactions are gradually turned off while reducing the molecule to 10% of its normal size. This is effectively the combination of steps i and ii above in the reverse direction and is denoted by $\Delta G^{\bullet}_{\text{cav+insert}}$. The change in intramolecular van der Waals energy upon the change in size of the solute caused an undesired change in free energy. To estimate this quantity, the same process is simulated again in vacuum, and the corresponding change in intramolecular van der Waals energy in vacuum is subtracted from the combined estimate of $\Delta G^{\bullet}_{\text{cav+insert}}$. In summary, the total solvation free energy for the melt is given by the sum

$$\Delta G^{\bullet}_{\text{vl}} \cong \Delta G_{\text{cav}} + \Delta G^{\bullet}_{\text{insert}} + \Delta G^{\bullet}_{\text{charge}} \tag{6}$$

3. Simulation Methodology

From the MC simulations in the NPT ensemble we obtain energies, free energies, molar volumes, molecular volumes, and molecular surface areas. The molecular surface area and the molecular volume are calculated by rolling a pointlike sphere on the molecular surface defined as the surface of the union of spheres representing all atoms, with radii set to the vdW radii multiplied by $2^{1/6}$. For the simulations we have used the BOSS MC program[8,9] (version 4.6). The initial three-dimensional molecular structures were obtained from so-called *z*-matrices supplied with the program. The inter- and intramolecular nonbonded interactions between atoms were described by potential functions consisting of Coulomb and Lennard-Jones interactions. The intramolecular through-bond interactions are modeled by harmonic bond stretches and bond angles while a Fourier series is used for modeling the dihedral angles in a molecule. Parameters for the different potential functions are taken from the OPLS-AA library,[10] which is considered to be a good force field for the description of organic liquids. The partial charges of the atoms in the molecule are located at the center of the corresponding Lennard-Jones spheres and are calculated from the Austin Model 1 (AM1) (a semiempirical quantum chemical method)[12] wave function by the Charged Method 1A (CM1A)[13] procedure. A geometry optimization of the molecule in vacuum is done, and the partial charges are recalculated in a single-point calculation for the optimized molecular structure. We have used the AM1, CM1A, and geometry optimization methods implemented in the BOSS program for all 46 molecules.

In all simulations cubic boxes were used with periodic boundary conditions. The cutoff distance r_c for the interactions was set to 10 Å. To verify that two molecules were within r_c, they were chosen in the following manner: (i) The first atom labeled, the center atom, was the one selected closest to the center of the molecule; (ii) the second atom was the one most distant from the center atom; (iii) the third atom was the one chosen most distant from the second atom; (iv) the fourth atom was the one most distant from the second and the third atom. This is of course a restriction and may lead to configurations where atom pairs in different molecules are closer in distance than the cutoff distance and which then are not included in the average data calculations. For three different molecules we have tested several configurations generated during the sampling and thereby compared the intermolecular energies calculated in two different ways: (i) using the four-atom cutoff method mentioned above and (ii) using the all-atom cutoff distance. We found that the energies are ∼1% lower when the all-atom cutoff method is used for all three tested pure melts. For all studied configurations we do not find any overlap problems, and the minimum distance between atom pairs of excluded molecules is found to be larger than 4 Å. For interactions beyond the cutoff distance 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 smoothly brought to zero in a quadratic manner in the range from $r_c - 0.5$ Å to r_c.

A pure liquid MC simulation starts from an equilibrated argon box where the 128 argon atoms are replaced by identical drug molecules. One identical molecule is added to the system, which is referred to as the solute molecule despite the fact that the solute and the rest of the molecules are of the same type. The reason for this is technical and serves only to distinguish between the molecule to be removed in the FEP simulations and the remaining molecules. Then, to save computation time, the system is reduced by removing 65 "solvent" molecules with the largest interactions energies. The simulation itself starts with a short equilibration in the canonical ensemble (NVT) to avoid undesirable volume changes due to some unphysical short interatomic separation. This is followed by an equilibration at constant pressure and temperature.

In the second half of the equilibration, the ranges for translation, rotations, and volume moves are adjusted to obtain a suitable acceptance rate of 10%. An attempt of molecule displacement 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 are incorporated when the solute molecule is rebuilt from the coordinated structure file. A volume displacement attempt is periodically performed after 375 molecule displacement attempts. It was revealed by exploratory simulations that for systems with large molecules it was difficult to achieve acceptance rates greater than 12−13% for the translational and rotational displacements. Therefore the acceptance rate of 10% was set as the default in all systems studied. Allen and Tildesley discuss such low acceptance rates in their textbook.[14] During the sampling of physical and thermodynamical properties the maximal translation, rotation, and volume changes are kept constant.

When the pure liquid simulation is finished, the free energy simulation for the removal of one molecule starts from the equilibrated pure liquid box. Our FEP simulation procedure decouples the solute molecule from the surrounding molecules using two subprocesses. The first process involves the decoupling of the electrostatic interactions by gradually scaling down the atomic partial charges by a coupling parameter λ in the Coulomb potential

$$E_C(\lambda) = \lambda \sum_{i\in\text{solute}} \sum_{j\in\text{bulk}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{7}$$

Here q_i and q_j denote the partial charges of atom i in the solute molecule and atom j in the surrounding bulk molecules, ϵ_0 is the permittivity of vacuum, and r_{ij} is the distance between atom i and j.

When $\lambda = 1$ the solute molecule is fully interacting, and when $\lambda = 0$ the surrounding molecules experience the solute as an alkane-like molecule. In the second process the Lennard-Jones interaction of the uncharged molecule is turned off by reducing its coupling parameter λ to zero, and simultaneously the size of the solute molecule is reduced to 10% of its normal size by use of a modified Lennard-Jones potential

$$E_{LJ}(\lambda) = \lambda \sum_{i\in\text{solute}} \sum_{j\in\text{bulk}} 4\epsilon_{ij}\left(\left(\frac{\sigma_{ij}(0.9\lambda + 0.1)}{r_{ij} + (1-\lambda)r_{\text{shift}}}\right)^{12} - \left(\frac{\sigma_{ij}(0.9\lambda + 0.1)}{r_{ij} + (1-\lambda)r_{\text{shift}}}\right)^{6}\right) \tag{8}$$

Here ϵ_{ij} and σ_{ij} are calculated from $\epsilon_{ij} = (\epsilon_i\epsilon_j)^{1/2}$ and $\sigma_{ij} = (\sigma_i\sigma_j)^{1/2}$. In eq 8 r_{shift} is a constant set to 1.5 Å that allows the surrounding molecules to spatially overlap with the solute atoms when $\lambda < 1$.

Both the decouplings of the electrostatic and the Lennard-Jones interactions are performed in 10 consecutive windows of coupling strength where the different windows are linearly distributed in the range from $\lambda = 0.05$ to $\lambda = 0.95$. To make the FEP simulations more effective we use double-wide sampling,[7] which means that both of the energy changes $\Delta E = E_{\lambda+0.05} - E_\lambda$ and $\Delta E = E_\lambda - E_{\lambda-0.05}$ are calculated for each window. Corresponding free energy increments $\Delta G = G_{\lambda+0.05} - G_\lambda$ and $\Delta G = G_\lambda - G_{\lambda-0.05}$ are then obtained from the energy changes by the FEP equation

$$\Delta G = -RT \ln \langle \exp -(\Delta E)/RT \rangle_\lambda \tag{9}$$

where $\langle \cdots \rangle_\lambda$ represents the ensemble average at coupling parameter λ. The total change in free energy is then obtained by summing the increments of ΔG for the different perturbation steps. We have studied the same molecules here as in the previous work[6] except for pseudoephedrine and benzene, which do not form stable liquids at 673.15 K. The different molecules studied and their names are listed in the Supporting Information.

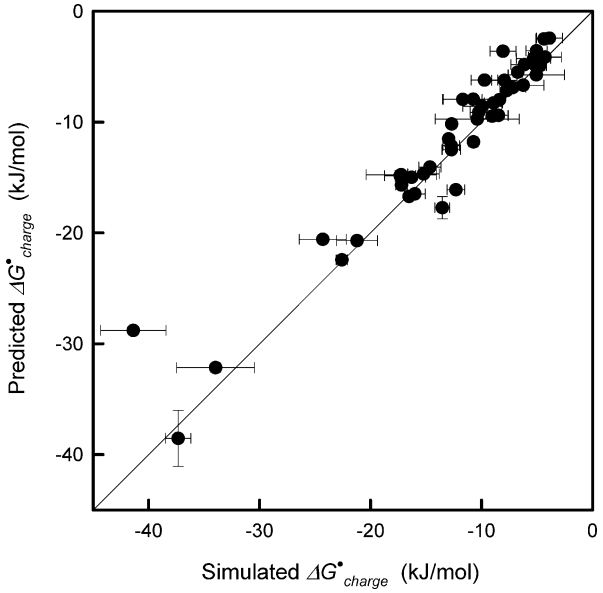

Figure 1. Predicted free energy for the electrostatic decoupling of the solute molecule from the surrounding molecules, $\Delta G^{\bullet}_{\text{charge}}$, as obtained using the linear response theory, eq 5. These results are compared to the simulated $\Delta G^{\bullet}_{\text{charge}}$ obtained from FEP simulation. The error bars represent the standard error of the mean.

It is not easy to simulate pure systems of large flexible molecules to analyze the errors involved in our obtained simulation results. We have added some additional information in this respect in the Supporting Information where we present the convergence in our simulation results for the Lennard-Jones and the charging decoupling of the solute molecule, respectively. There we also present results for investigating the hysteresis effects for the Lennard-Jones decoupling of three different solute molecules. We find that the deviations are minor for all of the investigated molecules. The charge decoupling of the solute molecule is a simpler simulation procedure compared to the Lennard-Jones decoupling step, and for this step we know that the hysteresis effects are small.

3. Results and Discussion

3.1. Results for the Drug Molecules. To be sure that we obtain reliable results when simulating small systems consisting of 64 drug molecules, interaction energies and molar volumes are evaluated for systems of 260 molecules, and the results are compared. The deviations were found to be on the average 2.9% for the intermolecular interaction energy and 1.4% for the molar volume. Hence, the smaller box is expected to produce reliable results.

In the process of removing the solute molecule from its pure liquid phase the electrostatic interactions of the solute molecule are turned off. In Figure 1 the correlation between the predicted $\Delta G^{\bullet}_{\text{charge}}$ values from eq 5 and the corresponding FEP results for this process is shown. The error bars in the results presented in Figure 1 and the following figures represent standard errors of the mean obtained for each drug from three independent simulations with different initial seeds generated by a random number generator. The results appear to be in good agreement with the linear response theory, which may be understood in

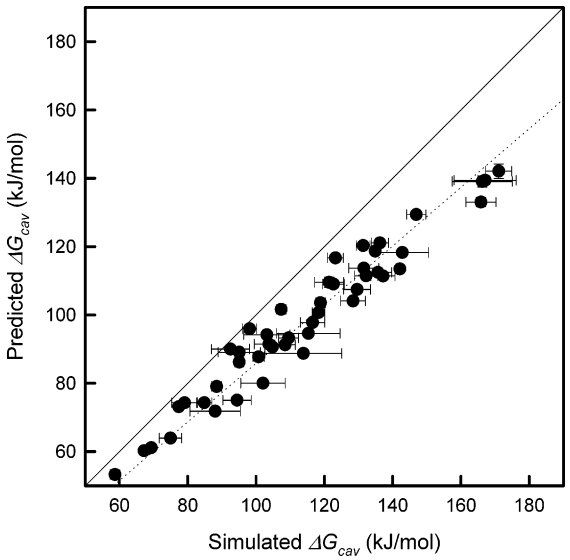

Figure 2. Predicted free energy of cavity creation, ΔG_{cav}, as calculated by eq 12 with the volume fraction estimated by the modified Carnahan−Starling equation of state, eq 11. These results are compared to simulated ΔG_{cav} data calculated by the relation $\Delta G_{cav} = \Delta G^{\bullet}_{cav+insert} - E_{LJ}$, where $\Delta G^{\bullet}_{cav+insert}$ and E_{LJ} data are taken from FEP simulations. The dotted line is a linear regression (intercept forced to zero) with a slope of 0.86.

terms of the, relatively speaking, weak electrostatic interactions in the pure drug melt.

In contrast to the results obtained in our previous work on the free energy of hydration,[6] where the electrostatic interactions of the solute with its aqueous environment are much more dominant compared to the Lennard-Jones interactions, the reverse is the case for the pure organic liquids studied here. The nature of an aqueous solution with its stronger electrostatic interactions and its ability to form hydrogen bonds is quite different from that of the systems studied here.

When the Coulomb interactions have been turned off the system has one alkane-like solute molecule, which interacts only via Lennard-Jones potentials with the surrounding, fully interacting molecules. In the second simulation step, the Lennard-Jones parameters of the solute molecule are gradually turned to zero, and simultaneously the solute molecule is reduced to 10% of its normal size. Consequently the reversal of this step accounts for the work needed to both create the cavity and then to scale up the van der Waals interactions of the solute molecule located inside the cavity. In our previous work on the free energy of hydration[6] the free energy of cavity formation, ΔG_{cav}, was calculated from the surface tension of water−vacuum γ and the molecular surface area A_{MS} of the solute by the expression

$$\Delta G_{cav} = A_{MS}\gamma \tag{10}$$

Due to the lack of surface tension data for the various pure drug melts an alternative approach for the calculation of the work needed to create a cavity inside a pure melt has to be used here. If one assumes that the drug molecules can be described as spheres, then one can try to model the pure melts by the Carnahan−Starling[15] equation of state and add pressure contributions due to Lennard-Jones and electrostatic interaction according to the relation

$$\frac{PV_m}{RT} = \frac{1 + \phi + \phi^2 - \phi^3}{(1-\phi)^3} + \frac{E_{LJ} + E_C/2}{2RT} \tag{11}$$

Here P is pressure, V_m is the molar volume, R is the gas constant, and T is the absolute temperature. The first term on the right-hand side of the equation of state above represents the hard sphere repulsive contribution to the pressure as modeled by the Carnahan−Starling equation of state for hard spheres where ϕ is the drug volume fraction given by the molecular volume, v_{drug}, and the molar volume V_m of the liquid as $\phi = v_{drug}N_A/V_m$. Here N_A is Avogadro's constant. Note that for a fluid of hard spheres of diameter d the volume fraction is $\phi = (\pi d^3/6)N_A/V_m$. The numerator of the second term on the right-hand side of eq 11 consists of the solute Lennard-Jones energy E_{LJ} and half of the electrostatic energy E_C, respectively. Recall that the solute energy is defined as the total interaction energy between the solute and all of the surrounding molecules in the bulk liquid and amounts to twice the average interaction energy in the melt. Thus the factor of 2 in the denominator of the last term in eq 11 arises because of the use of solute energies. We assume in this approximate treatment that we can apply mean field theory for the van der Waals interactions and linear response theory for the electrostatic interactions of the pure liquids. In Appendix A we have presented a derivation of the equation of state presented in eq 11 based on our assumptions mentioned above. The expression for the free energy of cavity formation, ΔG_{cav}, for hard sphere fluids modeled by the Carnahan−Starling equation of state (EOS)[16] can be written as

$$\Delta G_{cav} = RT\frac{8\phi - 9\phi^2 + 3\phi^3}{(1-\phi)^3} \tag{12}$$

By inserting intermolecular energies and molar volumes V_m from simulations into eq 11 a volume fraction ϕ can be calculated. Inserting the calculated ϕ values together with intermolecular energies from pure liquid simulations in eq 12, ΔG_{cav} can be predicted for the different solute molecules. We will compare these ΔG_{cav} values with corresponding estimates obtained from the difference $\Delta G^{\bullet}_{cav+insert} - E_{LJ}$ where the $\Delta G^{\bullet}_{cav+insert}$ data are obtained from FEP simulations and the E_{LJ} values are obtained from pure liquid simulations where the partial charges are fully turned on. In the previous work for the hydration free energies we introduced in an ad hoc manner a contribution $\Delta E_{LJ} = E_{LJ}(q = 1) - E_{LJ}(q = 0)$ to improve the correlation between simulated and predicted results from theories. Here, this contribution is almost zero due to weak electrostatic interactions and is not included. In Figure 2 the predicted ΔG_{cav} values versus the simulated ΔG_{cav} values are shown. The results are promising, but there appears to be a slight but systematic underestimation of the free energy of cavity formation of approximately 15% by the theory.

To assume that all of the drug molecules studied can be treated as spheres is of course a crude simplification, and an interesting alternative would be to describe them in a more realistic shape. As most of the drug molecules studied here contain planar aromatic ring structures, a more realistic shape to investigate would be the oblate spheroid, i.e., a flattened sphere. The question is then how to determine the short and the long axes of the oblate. If we assume that all molecules have the shape of oblate spheroids, then we can derive an expression that relates the molecular volume V_{MS} and the molecular surface area (MSA) A_{MS} to the oblate axial ratio a_r by the geometric functions of oblates.[17] We then obtain the

equation

$$\frac{A_{MS}}{V_{MS}^{2/3}} = \frac{1}{2}\left[1 + \frac{a_r^2}{\sqrt{1-a_r^2}} \ln \frac{1+\sqrt{1-a_r^2}}{a_r}\right] \pi^{1/3} \left(\frac{6}{a_r}\right)^{2/3} \quad (13)$$

From this equation the axial ratio a_r for a solute molecule can then be calculated from the estimated molecular volume and surface area. For the most spherical molecule, salicylic acid, an axial ratio of 0.38 is obtained while for the least spherical molecule, prostaglandin E2, an axial ratio of 0.19 is calculated.

We are well aware that the molecules studied are neither spheres nor oblate spheroids. Nevertheless, it is worthwhile to investigate whether the results obtained by eq 12 can be improved by taking into consideration the nonspherical shapes of the molecules and replacing the Carnahan−Starling equation of state by an equation of state for oblates instead. Boublík et al.[17] have developed such an equation of state for hard convex bodies, and the total equation of state would then take the following form

$$\frac{PV_m}{RT} = \frac{1}{1-\phi} + \frac{3\alpha\phi}{(1-\phi)^2} + \frac{(3\alpha^2\phi^2) - \alpha(6\alpha-5)\phi^3}{(1-\phi)^3} + \frac{E_{LJ} + E_C/2}{2RT} \quad (14)$$

Here α is a parameter describing the degree of sphericity as defined by $\alpha = R_c A_{oblate}/V_{oblate}$,[17] where for the oblate spheroid R_c is the mean curvature integral divided by 4π, A_{oblate} is the oblate surface area, and V_{oblate} is the corresponding oblate volume. If the geometric functions[17] for R_c, A_{oblate}, and V_{oblate} are inserted into the expression for α, then one obtains the relation between α and a_r as

$$\alpha = \frac{1}{4a_r}\left(a_r + \frac{\arccos(a_r)}{\sqrt{1-a_r^2}}\right)\left(1 + \frac{a_r^2}{\sqrt{1-a_r^2}} \ln\left(\frac{1+\sqrt{1-a_r^2}}{a_r}\right)\right) \quad (15)$$

One can easily show that eq 14 reduces to eq 11 when $\alpha = 1$. By inserting $P = 1$ atm, the molar volume, and the intermolecular interaction energies obtained from simulations in eq 14 with an α value calculated from eq 15, the (predicted) volume fraction ϕ for oblates can be obtained. In Figure 3 the ϕ values for hard spheres and oblates are compared with simulation results for ϕ estimated as $\phi = \phi_{BOSS}/2^{1/2}$. The reason for the division of the BOSS value by a factor of $2^{1/2}$ is that in the BOSS simulation the molecular volumes and the MSAs are computed using an atomic diameter corresponding to the energy minimum of the Lennard-Jones interaction potential (the Lennard-Jones parameter σ multiplied by $2^{1/6}$). The ϕ values obtained by eq 11 are in general overestimated compared to the simulated ϕ values, while in contrast the ϕ values obtained by eq 14 are in general too low compared to the simulated values. However, the results predicted by eq 14 are significantly less scattered, and the linear regression slope represented by the dotted line is close to 1. This result of our theory based on Boublík's equation of state, i.e., eq 14, is therefore an improvement compared to our corresponding result using the Carnahan−Starling theory, eq 11.

This is, of course, an inspiring finding leading us to believe that the results for ΔG_{cav} will also improve. In a similar way as described above, the free energy of cavity formation in systems

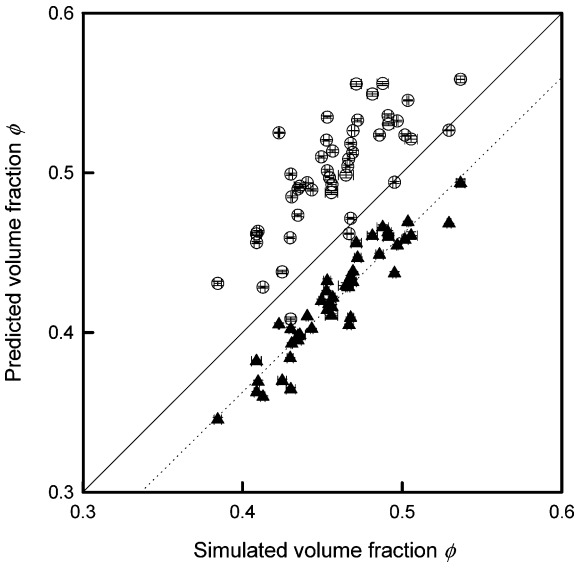

Figure 3. Two different predicted volume fractions ϕ, the first calculated by eq 11 (Carnahan−Starling theory[15]) and represented by the open ring symbols and the second calculated by eq 16 (Boublík et al. hard oblate theory[17]) and represented by the filled triangles, compared to the simulated volume fractions ϕ_{BOSS} obtained from pure liquid simulations. Note that the input data required to calculate the volume fractions by eqs 11 and 16 are taken from pure liquid simulations.

of oblate spheroids can be derived from eq 14 obtaining (Appendix B)

$$\frac{\Delta G_{cav}}{RT} = \frac{\phi(2+2\alpha+12\alpha^2-4\phi+7\alpha\phi-21\alpha^2\phi+2\phi^2+\alpha\phi^2+3\alpha^2\phi^2)}{2(1-\phi)^3} + (\alpha-1)(1+6\alpha)\ln(1-\phi) \quad (16)$$

In Figure 4 ΔG_{cav} predicted by eq 16 is plotted versus ΔG_{cav} obtained from the FEP simulations. The slope of a linear regression, with the intercept forced to zero, of the predicted data against the simulated results is closer to unity when the molecules are described as oblate spheroids instead of spheres.

Still we are aware that treating all molecules as oblates is a rough simplification, and we have carried out some further work where molecules were classified as either oblates or prolates. It can be shown that eqs 14 and 16 are applicable to both oblates and prolates and the only difference is how the parameter α depends on the axial ratio ($a_r < 1$ for oblates and $a_r > 1$ for prolates).[17] The classification is carried out by assuming that the atoms in a molecule are solid spheres with a volume given by their respective Lennard-Jones σ values. Then we calculate the moment of inertia along the principal axes and the corresponding radii of gyration, K_1, K_2, and K_3, for one typical conformation. Since we are interested in the form of the molecule we set the mass of each atom to be proportional to the atomic volume instead of equal to the proper atomic mass. For a true solid homogeneous ellipsoid the radii of gyration are related to the semiaxes a (shortest), b, and c (longest)

$$a = \sqrt{\frac{5}{2}(K_2^2 + K_3^2 - K_1^2)}$$

$$b = \sqrt{\frac{5}{2}(K_1^2 + K_3^2 - K_2^2)}$$

$$c = \sqrt{\frac{5}{2}(K_1^2 + K_2^2 - K_3^2)} \quad (17)$$

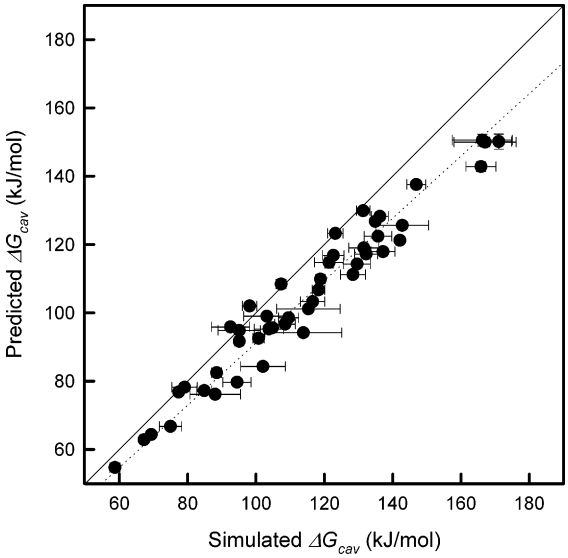

Figure 4. Predicted free energy of cavity creation, ΔG_{cav}, as calculated by eq 16 with the volume fraction and an α value calculated as described in the text. These results are compared with simulated ΔG_{cav} calculated by the relation $\Delta G_{cav} = \Delta G^{\bullet}_{cav+insert} - E_{LJ}$, where the data for $\Delta G^{\bullet}_{cav+insert}$ and E_{LJ} data are taken from FEP simulations. The dotted line is a linear regression (intercept forced to zero) with a slope of 0.91.

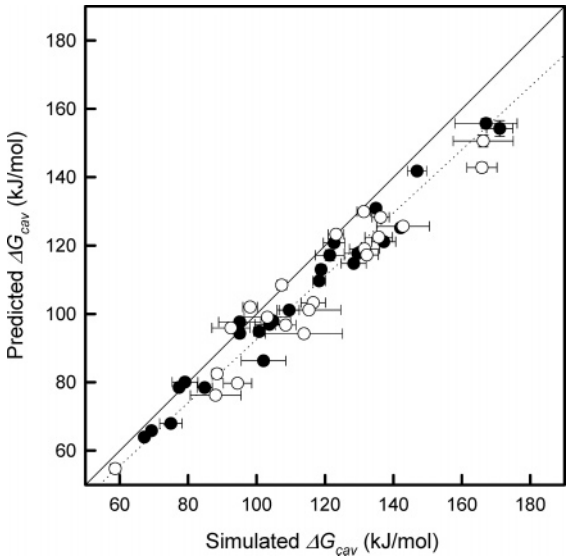

Figure 5. Predicted free energy of cavity creation, ΔG_{cav}, when some of the molecules are modeled as oblates (filled symbols) and the other ones are modeled as prolates (unfilled symbols), as described in the text. These results are compared with simulated ΔG_{cav} calculated by the relation $\Delta G_{cav} = \Delta G^{\bullet}_{cav+insert} - E_{LJ}$. The dotted line is a linear regression (intercept forced to zero) to all data with a slope of 0.93.

By describing the molecule as an ellipsoid, the semiaxes of a molecule can be calculated. The shapes of the molecules are generally neither exactly oblate nor prolate, but we classify them as oblate if $(c/b)/(b/a) < 1$ and otherwise prolate. In Figure 5 the results for the predicted ΔG_{cav} values when the molecules are treated as oblates represented by the filled symbols or prolates represented by the unfilled symbols are compared to simulation results. One can see that the correlation between the predicted and the simulation results is very little improved by such classification so we will model all molecules as oblates instead.

Finally, predicted results for the free energy of solvation in pure melts $\Delta G^{\bullet}_{vl}$ are calculated by the relation

$$\Delta G^{\bullet}_{vl} \cong \Delta G_{cav} + E_{LJ} + \frac{E_C}{2} \tag{18}$$

The required work of cavity formation ΔG_{cav} is calculated by the hard oblate theory, eq 16. The van der Waals interactions are calculated by mean field theory in eq 4, and the electrostatic interactions are calculated by linear response theory in eq 5. In Figure 6 the predicted results for $\Delta G^{\bullet}_{vl}$ are shown versus the corresponding simulation results from FEP simulations. At first sight the results do not look promising because of the large standard errors in the FEP simulation results in comparison to the small absolute values of $\Delta G^{\bullet}_{vl}$. This is caused by the common phenomenon of subtracting two numbers of similar sizes resulting in a small difference and where the uncertainties are summed. The work required to introduce a cavity into the pure melt contributes to a large positive value of ΔG_{cav}, and in contrast, turning on the Lennard-Jones interactions contributes to a large negative energy comparable in magnitude to the cavity energy. Thus these two quantities will almost cancel each other to yield a small absolute number for the free energy of solvation. The contribution of the electrostatics is small for most systems studied.

In a future paper we will address the process of going from $\Delta G^{\bullet}_{vl}$ (673.15 K) to $\Delta G^{\bullet}_{vl}$ (298.15 K). Our preliminary results indicate that the relatively large uncertainties observed here reduces compared to mean values upon cooling.

3.2. Results for Cyclohexane. In contrast to the case of free energy of hydration of organic compounds, there are not many attempts in the literature addressing the, from the solubility point of view, equally important solvation free energies of druglike or organic molecules in their pure liquid state. However, the work by Graziano, e.g., on benzene[18] and cyclohexane,[19] are exceptions. In the case of cyclohexane,[19] where a simple comparison can be carried out, predicted values of $\Delta G^{\bullet}_{vl}$ for the process of transferring one cyclohexane molecule from an ideal gas into a pure liquid were obtained by an approach that is very similar to ours; Graziano used the approximation

$$\Delta G^{\bullet}_{vl} \cong \Delta G_{cav} + E_{LJ} \tag{19}$$

where the free energy of cavity formation was obtained from scaled particle theory and Pierotti's formula[18,20] was used for E_{LJ}. At $T = 298.15$ K he obtained $\Delta G^{\bullet}_{vl} = -19.0$ kJ/mol, which is in good agreement with the experimental value of −18.5 kJ/mol. To investigate this system further we have run a liquid simulation at this temperature. We obtain a molar volume $V_m = 111$ cm^3/mol and an enthalpy of vaporization of 33.4 kJ/mol as calculated from

$$\Delta H_{vap} \approx \Delta E_{intra} - (E_{LJ} + E_C)/2 + RT \tag{20}$$

where ΔE_{intra} is the change in the intramolecular energy from the liquid to the vapor phase and we have assumed that the vapor phase can be treated as ideal and that $RT \gg PV_m$ in the liquid phase. These results compare favorably with the experimental data for molar volume $V_m = 109$ cm^3/mol[18] and enthalpy of vaporization of 33.0 kJ/mol,[18] respectively. The results from our simulations are $\Delta E_{intra} = -0.6$ kJ/mol, $E_{LJ} = -63.2$ kJ/mol, and $E_C = -0.15$ kJ/mol. Using these data we obtain $\Delta G^{\bullet}_{vl} = -21.9$ kJ/mol by eqs 12 and 18 in our Carnahan−Starling description, while the use of hard oblate theory in eq 16 together with eq 18 instead yields $\Delta G^{\bullet}_{vl} = -20.6$ kJ/mol. Despite the

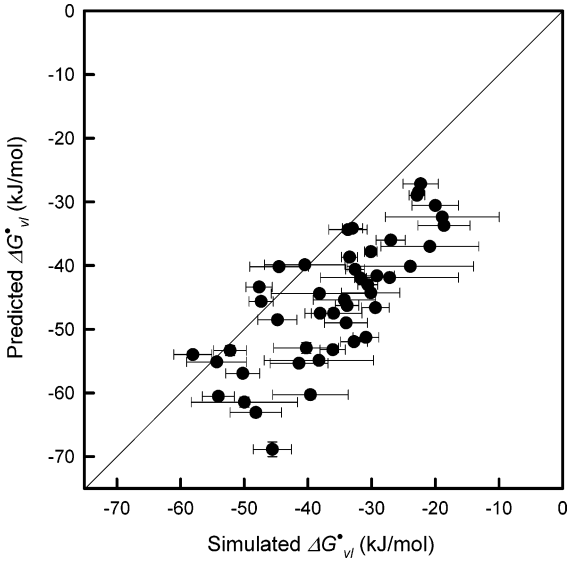

Figure 6. Predicted free energy of solvation, $\Delta G^{\bullet}_{vl}$, as calculated by eq 18, where ΔG_{cav} is substituted by the hard oblate theory, eq 16, and where the interaction energies are taken from pure liquid simulations. These results are compared to simulated $\Delta G^{\bullet}_{vl}$ obtained from FEP simulations.

TABLE 1: Predicted Free Energy Change, $\Delta G^{\bullet}_{vl}$, for the Vacuum → Liquid Transfer of One Cyclohexane Molecule at Different Temperatures Compared to Experimental Data[a]

T (K)	E_{LJ} (kJ/mol)	ΔG_{cav} (kJ/mol)	$\Delta G^{\bullet}_{vl}$ (predicted) (kJ/mol)	$\Delta G^{\bullet}_{vl}$ (experimental)[b] (kJ/mol)
278.15	−64.6	43.4	−21.2	−19.5
298.15	−62.2	42.0	−20.3	−18.5
323.15	−59.4	40.1	−19.3	−17.5
358.15	−56.4	38.3	−18.4	−16.5
373.15	−53.8	36.3	−17.5	−15.7

[a] The predicted $\Delta G^{\bullet}_{vl}$ values are calculated by eq 19 where ΔG_{cav} is calculated from the hard oblate theory, eq 16, and the energy E_{LJ} is extracted from the experimental enthalpy of vaporization ΔH_{vap}[18] by eq 20. [b] Reference 18.

good agreement between our results for those $\Delta G^{\bullet}_{vl}$ results of Graziano the individual terms of ΔG_{cav} and E_{LJ} calculated by our models differ significantly. The free energy of cavity formation is larger in both our implementations than that found by Graziano using scaled particle theory.[18] Furthermore, the dispersion interactions are more negative in our model. It should be noted that the simulations yield very small electrostatic interactions for cyclohexane. Thus, E_C in the second term on the right-hand side in eq 20 is zero to a good approximation, which facilitates the extraction of solute Lennard-Jones interaction energies from experimental data for the enthalpy of vaporization. In the paper by Graziano, the Pierotti formula gives $E_{LJ} = -49.5$ kJ/mol at $T = 298.15$ K, which is significantly lower than the value obtained from our simulations, which appear to be consistent with the experimental data for the enthalpy of vaporization. However, the use of experimental ΔH_{vap} for the calculation of E_{LJ} combined with scale particle theory appear to yield estimates of the free energy of solvation that deviate further from the experimental result.

To evaluate our own proposed model further, we have used experimental data for molar volume and enthalpy of vaporization, combined with molecular volume and surface area obtained from simulations, to predict $\Delta G^{\bullet}_{vl}$ values for cyclohexane in the temperature range 273.15−373.15 K. Here we assume that the physical size of the molecule does not change in this temperature range and that ΔE_{intra} is constant, $\Delta E_{intra} = -0.6$ kJ/mol, i.e., the value that we obtain at $T = 298.15$ K. The results obtained are compared with experimental data in Table 1. It seems that our proposed model can predict the temperature dependence of $\Delta G^{\bullet}_{vl}$ in reasonable quantitative agreement with experimental data. The average deviation from the experimental $\Delta G^{\bullet}_{vl}$ values at different temperatures is −1.8 kJ/mol.

To summarize, the use of eq 18 combined with an equation of state using Boublík's hard body theory works satisfactorily for the prediction of both $\Delta G^{\bullet}_{vl}$ values of the 46 pure drug liquids at 673.15 K and for the temperature dependence of $\Delta G^{\bullet}_{vl}$ for cyclohexane. This is promising for our future work on the aqueous solubility of supercooled liquids at room temperature.

Appendix A

Equation of State for Pure Drug Fluids. We shall use the generalized van der Waals approach to the equation of state of the pure drug fluids. Recall that the partition function for an ideal gas of spherical particles is

$$Q = \Lambda^{-3N} V^N / N! \approx (v/\Lambda^3)^N \, e^N \tag{A1}$$

where $v = V/N$ is the volume per particle. For a liquid if the molecules are treated as hard spheres with no soft interactions, then the volume v is replaced by the free volume $v_f = v(1 - \phi)$, where the volume fraction for a hard sphere with diameter d is $\phi = (\pi d^3/6)\, N/V$. We then obtain

$$Q = [(v(1 - \phi))/\Lambda^3]^N \, e^N \tag{A2}$$

and the free energy is

$$A = -k_B T \ln Q = -N k_B T [\ln(v(1 - \phi))/\Lambda^3 + 1] \tag{A3}$$

The pressure is given by

$$P = -\frac{\partial A}{\partial V} = \frac{\phi}{V}\frac{\partial A}{\partial \phi} = \frac{k_B T}{v(1 - \phi)} \tag{A4}$$

where we have assumed that the volume of the particle is independent of v. For a realistic molecule there will also be soft interactions, e.g., dispersion interactions and other forms of van der Waals interactions. We shall first assume that these interactions are acting pairwise between the molecules and sufficiently weakly that they can be treated by the mean field approximation. This means that they do not induce any response in the fluid. We can then express their contribution to the free energy as a pure energy

$$\Delta A_{LJ} = E_{LJ} = -N a_{LJ} n = -N a_{LJ}/v \tag{A5}$$

where a_{LJ} is the binding Lennard-Jones energy per molecule at unit density in the fluid and n is the number density. We assume that a_{LJ} is independent of n. The Lennard-Jones energy per molecule E_{LJ}/N is then proportional to n, which is the inverse of the volume fraction per molecule v. The free energy is then

$$A = -N k_B T [\ln[(v(1 - \phi))/\Lambda^3] + 1] - N a_{LJ}/v \tag{A6}$$

and the pressure is

$$P = \frac{k_B T}{v(1 - \phi)} - \frac{a_{LJ}}{v^2} \tag{A7}$$

The equation of state above is the van der Waals EOS. We could apply it to the drug fluid, but in these fluids substantial interactions of electrostatic origin are present for which the mean field approximation is inaccurate due to the long-range nature of the electrostatic interactions. We shall assume instead that these interactions can be treated by linear response theory. This means that they induce a response in the fluid such that the corresponding correlation reduces the free energy contribution to half of the interaction energy

$$\Delta A_C = \frac{1}{2} E_C = -N \frac{1}{2} a_C n = -N \frac{1}{2} a_C / \upsilon \qquad \text{(A8)}$$

where a_C is the Coulomb binding energy per molecule, which is assumed to be independent of n. The free energy then becomes

$$A = -Nk_BT[\ln[(\upsilon(1-\phi))/\Lambda^3] + 1] - Na_{LJ}/\upsilon - \frac{1}{2} Na_C/\upsilon \qquad \text{(A9)}$$

and the pressure is

$$P = \frac{k_BT}{\upsilon(1-\phi)} - \frac{a_{LJ}}{\upsilon^2} - \frac{1}{2}\frac{a_C}{\upsilon^2} \qquad \text{(A10)}$$

If we now multiply by υ/k_BT we obtain

$$\frac{P\upsilon}{k_BT} = \frac{1}{1-\phi} - \frac{a_{LJ} + a_C/2}{\upsilon k_BT} \qquad \text{(A11)}$$

The average Lennard-Jones and Coulomb energies per molecule are given by

$$\bar{e}_{LJ} = -a_{LJ}/\upsilon = -a_{LJ} n \qquad \text{(A12)}$$

and

$$\bar{e}_C = -a_C/\upsilon = -a_C n \qquad \text{(A13)}$$

Finally we note that given the pairwise interactions of the molecules the total energy of interaction of one molecule with all others is twice the corresponding average energy where each pairwise energy is divided equally between the two interacting molecules. Thus we have

$$e_{LJ} = 2\bar{e}_{LJ} = -2a_{LJ} n \qquad \text{(A14)}$$

$$e_C = 2\bar{e}_C = -2a_C n \qquad \text{(A15)}$$

We can then finally write the equation of state in the form

$$\frac{P\upsilon}{k_BT} = \frac{1}{1-\phi} + \frac{e_{LJ} + e_C/2}{2k_BT} \qquad \text{(A16)}$$

which upon change to molar variables becomes

$$\frac{PV_m}{RT} = \frac{1}{1-\phi} + \frac{E_{LJ} + E_C/2}{2RT} \qquad \text{(A17)}$$

Now we can replace the hard sphere pressure term by the Carnahan−Starling or Boublík hard body equation of state to obtain the equation of state that we have used in our solvation theory in this publication.

Appendix B

If we replace the first term in the right-hand side of eq A17 with the equation of state for oblates proposed by Boublík,[17] then one obtains with the substitution $\phi = \upsilon_0/\upsilon$

$$P = k_BT\left(\frac{1}{\upsilon - \upsilon_0} + \frac{3\alpha\upsilon_0}{(\upsilon-\upsilon_0)^2} + \frac{3\alpha^2\upsilon_0^2 - \alpha(6\alpha-5)\upsilon_0^3}{(\upsilon-\upsilon_0)^3}\right) + \frac{b}{\upsilon^2} = P(\upsilon) \qquad \text{(B1)}$$

where υ_0 is the volume of an oblate particle and υ is again the volume per particle. The second term on the right-hand side of eq B1 is an energy density (b) that is volume-independent and defined by

$$-\frac{2b}{\upsilon} = \frac{1}{N_A}\left(E_{LJ} + \frac{E_{Coul}}{2}\right) \qquad \text{(B2)}$$

The thermodynamic relation of the pressure of N particles and at temperature T is

$$P = -\left(\frac{\partial A}{\partial V}\right)_{N,T} = -\left(\frac{\partial f}{\partial \upsilon}\right)_{N,T} \qquad \text{(B3)}$$

where $f = A/N$ is the Helmholtz free energy per particle. From the EOS in eq B1 and the thermodynamic relation in eq B3 follows

$$\begin{aligned} f &= -\int_{\upsilon_0+\Delta}^{\upsilon} d\upsilon' P(\upsilon') + f(\upsilon_0+\Delta) \\ &= -\int_{\upsilon_0+\Delta}^{\upsilon} d\upsilon' \left[k_BT\left(\frac{1}{\upsilon'-\upsilon_0} + \frac{3\alpha\upsilon_0}{(\upsilon'-\upsilon_0)^2} + \frac{3\alpha^2\upsilon_0^2 - \alpha(6\alpha-5)\upsilon_0^3}{(\upsilon'-\upsilon_0)^3}\right) + \frac{b}{\upsilon'^2}\right] + f(\upsilon_0+\Delta) \\ &= -k_BT\left[\ln(\upsilon-\upsilon_0) - \frac{3\alpha\upsilon_0}{\upsilon-\upsilon_0} - \frac{3\alpha^2\upsilon_0^2}{2(\upsilon-\upsilon_0)^2} + \left(\frac{6\alpha^2-5\alpha}{2}\right)\left(\frac{\upsilon_0(2\upsilon-3\upsilon_0)}{(\upsilon-\upsilon_0)^2} + 2\ln\frac{\upsilon-\upsilon_0}{\upsilon}\right) + \frac{b}{k_BT\upsilon}\right] + k_BT\left[\ln(\Delta) - \frac{3\alpha\upsilon_0}{\Delta} - \frac{3\alpha^2\upsilon_0^2}{2\Delta^2} + \left(\frac{6\alpha^2-5\alpha}{2}\right)\left(\frac{\upsilon_0(2(\upsilon_0+\Delta)-3\upsilon_0)}{\Delta^2} + 2\ln\frac{\Delta}{\upsilon_0+\Delta}\right) + \frac{b}{k_BT(\upsilon_0+\Delta)}\right] + f(\upsilon_0+\Delta) \end{aligned} \qquad \text{(B4)}$$

Here Δ is a small constant used to avoid singularities in the integration. The chemical potential is defined as

$$\mu = \frac{\partial}{\partial N} Nf = f + N\frac{\partial}{\partial N} f = f + P\upsilon \qquad \text{(B5)}$$

and the Gibbs free energy is given by

$$G = A + PV = N(f + P\upsilon) \qquad \text{(B6)}$$

Note that if we obtain f from the EOS there will be undetermined constants in the result. If we let υ be very large, then we expect the ideal gas law to apply and yield

$$Q = \frac{V^N}{\lambda^{3N} N!}$$

$$A = -k_BT\ln(Q) = -Nk_BT[\ln\upsilon - \ln\lambda^3 + 1] \qquad \text{(B7)}$$

For eq B7 to agree with eq B4 in the limit $v \rightarrow \infty$ we must have

$$f = -k_B T\left[\ln(v - v_0) - \ln \lambda^3 + 1 - \frac{3\alpha v_0}{v - v_0} - \frac{3\alpha^2 {v_0}^2}{2(v - v_0)^2} + \left(\frac{6\alpha^2 - 5\alpha}{2}\right)\left(\frac{v_0(2v - 3v_0)}{(v - v_0)^2} + 2\ln\frac{v - v_0}{v}\right) + \frac{b}{k_B T v}\right] \quad \text{(B8)}$$

Thus we obtain

$$\mu = -k_B T\left(\ln(v - v_0) - \ln \lambda^3 + 1 - \frac{3\alpha v_0}{v - v_0} - \frac{3\alpha^2 {v_0}^2}{2(v - v_0)^2} + \left(\frac{6\alpha^2 - 5\alpha}{2}\right)\left(\frac{v_0(2v - 3v_0)}{(v - v_0)^2} + 2\ln\frac{v - v_0}{v}\right) + \frac{b}{k_B T v}\right) + Pv \quad \text{(B9)}$$

Note that for the ideal gas we have

$$\mu^{id} = -k_B T(\ln v - \ln \lambda^3) \quad \text{(B10)}$$

Thus we can write

$$\mu = \mu^{id} + \mu^{ex} \quad \text{(B11)}$$

where μ^{ex} is the excess chemical potential. By substituting $\phi = v_0/v$ and inserting eq B2 into eq B9 we then obtain the final equation for the predictions of changes in excess Gibbs free energy in molar variables

$$\Delta G^{\bullet} = RT\left(\frac{\phi(2 + 2\alpha + 12\alpha^2 - 4\phi + 7\alpha\phi - 21\alpha^2\phi + 2\phi^2 + \alpha\phi^2 + 3\alpha^2\phi^2)}{2(1 - \phi)^3} + (\alpha - 1)(1 + 6\alpha)\ln(1 - \phi)\right) + E_{LJ} + \frac{E_{Coul}}{2} = \Delta G_{cav} + E_{LJ} + \frac{E_{Coul}}{2} \quad \text{(B12)}$$

Due to the assumption that the intramolecular structure of the extra molecule is unaffected during the transfer from the ideal gas to the pure melt we have that $\Delta G^{\bullet} = \Delta G^{ex}$.

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

Supporting Information Available: A summary of all important simulation results for 46 molecules. and information about the convergence and hysteresis effects in the simulation results. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Lipinski, C. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235.
(2) Jorgensen, W. L.; Duffy, E. M. *Adv. Drug Delivery Rev.* **2002**, *54*, 355.
(3) Gavezotti, A. *CrystEngComm* **2002**, *4*, 343.
(4) 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.
(5) Neau, S. H.; Bhandarkar, S. V.; Hellmuth, E. H. *Pharm. Res.* **1997**, *14*, 601.
(6) Westergren, J.; Lindfors, L.; Höglund, T.; Lüder, K.; Nordholm, S.; Kjellander, R. *J. Phys. Chem. B* **2007**, *111*, 1872.
(7) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.
(8) Jorgensen, W. L. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley: New York, 1998; Vol. 5, p 3281.
(9) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689.
(10) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(11) Ben-Naim, A. *Solvation Thermodynamics*, 1st ed.; Plenum Press: New York, 1987.
(12) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(13) Storer, J. W.; Giesen, D. J.; Cramer C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.
(14) Allen M. P.; Tildesley D. J. *Computer Simulation of Liquids*, 1st ed.; Oxford University Press: New York, 1987.
(15) Carnahan, N. F.; Starling, K. E. *J. Phys. Chem.* **1969**, *51*, 635.
(16) Mansoori, G. A.; Carnahan, N. F.; Starling, K. E.; Leland, T. W., Jr. *J. Chem. Phys.* **1971**, *54*, 1523.
(17) Boublík, T.; Nezbeda, I. *Collect. Czech. Chem. Commun.* **1986**, *51*, 2301.
(18) Graziano, G. *Biophys. Chem.* **1999**, *82*, 69.
(19) Graziano, G. *Can. J. Chem.* **2000**, *78*, 1233.
(20) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717.