In Silico Prediction of Drug Solubility: 4. Will Simple Potentials Suffice?

Kai Lüder,[*]Lennart Lindfors,[†]
Jan Westergren,[‡]Sture Nordholm,[§]
Rasmus Persson,[¶]Mikaela Pedersen[‖]

October 29, 2008

Abstract

In view of the extreme importance of reliable computational prediction of aqueous drug solubility, we have established a Monte Carlo simulation procedure which appears, in principle, to yield reliable solubilities even for complex drug molecules. A theory based on judicious application of linear response and mean field approximations has been found to reproduce the computationally demanding free energy determinations by simulation while at the same time offering mechanistic insight. The focus here is on the suitability of the model of both drug and solvent, *i. e.*, the force fields. The optimized potentials for liquid simulations all atom (OPLS-AA) force field, either intact or combined with partial charges determined either by semiempirical AM1/CM1A calculations or taken from the condensed-phase optimized molecular potentials for atomistic simulation studies (COMPASS) force field has been used. The results illustrate the crucial role of the force field in determining drug solubilities. The errors in interaction energies obtained by the simple force fields tested here are still found to be too large for our purpose but if a component of this error is systematic and readily removed by empirical adjustment the results are significantly improved. In fact, consistent use of the OPLS-AA Lennard-Jones force field parameters with partial charges from the COMPASS force field will in this way produce good predictions of amorphous drug solubility within 1 day on a standard desktop PC. This is shown here by the results of extensive new simulations for a total of 47 drug molecules which were also improved by increasing the water box in the hydration simulations from 500 to 2000 water molecules.

Keywords: solubility, drug molecule, free energy, Monte-Carlo simulation, crystal energy calculation, amorphous phase, aqueous solution, OPLS-AA, AM1/CM1A, COMPASS ∎

[*]Department of Chemistry, University of Gothenburg, SE-412 96 Göteborg, Sweden
[†]Pharmaceutical and Analytical R&D, AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden
[‡]Wendelsbergs beräkningskemi AB, SE-435 35 Mölnlycke, Sweden
[§]Department of Chemistry, University of Gothenburg, SE-412 96 Göteborg, Sweden
[¶]Department of Chemistry, University of Gothenburg, SE-412 96 Göteborg, Sweden
[‖]Department of Chemistry & Biotechnology, Chalmers University of Technology SE-412 96 Göteborg, Sweden

INTRODUCTION

In this series of studies, we aim to develop a computationally tractable yet reliable method of determining the aqueous solubility of increasingly complex and generally sparingly soluble drug molecules. The only computational method available for this purpose at the moment is the quantitative structure property relations (QSPR)[1,2] by which an attempt is made to find and exploit statistical correlations between structural features of the drug molecules and their solubility. Although this method is commonly used in the pharmaceutical industry today it is not sufficiently reliable.[3] Particularly when the solubility is to be predicted for a new molecule not well represented in the training set of molecules of known solubility used to establish the statistical correlations the prediction can be far from realistic. Such errors are generally difficult to analyze and interpret because there is no underlying theory to reexamine.

The ambition here has instead been to develop a computational tool which in an automatic and robust manner can predict the amorphous solubility of a newly developed drug candidate from the chemical structure. This tool is of general importance and crucial in the formulation work when deciding how the drug candidate should be administered in the early animal studies to obtain the desired effect. It is desirable to replace existing QSPR methods by this new computational tool. The prediction of the amorphous solubility of a modern drug candidate should be computable within 1 day with the new proposed tool on a standard PC of today.

We believe a theory is needed both for the short- and long-term resolution of solubility by computation. In the short term, given the enormous complexity of the problem of drug molecule solubility, our hope is to develop a sound theoretical basis for a computational prediction with a minimal empirical input in which any parametrization is limited to a form which is physically interpretable. In the longer term, we look forward to eliminate all empirical input and arrive at a truly *a priori* prediction of solubility from first principles.

Our prediction of aqueous drug solubility is based on a two-pronged approach where we use both a brute force Monte Carlo (MC) simulation method combined with the free energy perturbation (FEP) method and an approximate theory to generate the free energy

differences required to determine the solubility. Our MC simulations have been done with the program package BOSS (Biochemical and Organic Simulation System) of Jorgensen[4,5] in which the required potential model for the system of drug in water or pure drug is provided. The use of the FEP method entails long sequences of simulations in which various interactions, Lennard-Jones or electrostatic pairwise atomic interactions in our case, are turned on or off. This is computationally demanding and gives the calculation a brute force character. For this reason, we have been developing and testing a simpler approximate theory which we shall call SR theory where "SR" stands for "simplified response." The SR theory is based on the assumption that the response of the system to electrostatic interactions can be described by a linear response theory and the response to dispersion or other induction and correlation interactions by mean field theory. The main and immediate advantage of the SR theory is that the MC simulations need to produce only system energies which are the forte of standard simulation methods. The entropies required are generated by the simplified response estimates. There are, however, further advantages in that the simplified mechanistic analysis of SR theory provides the possibility of entering empirical input in a physically interpretable manner. Thus it allows the type of physical empiricism that we may need to produce useful results in the short term. Moreover, the explicit mechanistic analyses provide the possibility of more insightful error analysis when difficulties are encountered.

In earlier work, we have sequentially studied the insertion of a drug molecule initially in vacuum into water[6] and into liquid drug at 400 °C.[7] Finally, we have studied the insertion of a drug molecule into a pure amorphous drug phase at 25 °C by cooling the pure liquid drug from 400 °C to 25 °C.[8] Thus we have generated the free energies required to predict the aqueous solubility of amorphous drug. We have also noted the possibility of estimating the solubility of crystalline drug by use of a thermodynamic cycle requiring the entropy of melting and the melting temperature of the crystalline drug. However, only a few preliminary comparisons of solubility predictions with experiment have been made so far. The focus has instead been on establishing the simulation procedures and developing and testing the SR theory. This work has been successful and we now have well-established simulation procedures and a functional SR theory which has proven able to reproduce the brute force simulation results for a large ensemble of drug molecules.

In this work, we shall report the first comprehensive comparison of predicted and experimental aqueous drug solubilities. Successful computational predictions require not only the adequate simulation method and theory that we believe we have developed but also accurate but computationally tractable potential models describing the crucial intermolecular interactions determining solubility. For this reason much of the attention will in the present report be focused on the adequacy of the optimized potentials for liquid simulations all atom (OPLS-AA)[9] force field combined with the AM1/CM1A[10,11] partial atomic charges which we have so far used in our calculations. We are well aware that this is an extremely simplified force field which cannot be relied on to deliver the accuracy of ± 6 kJ/mol that we would need to compute solubilities within an order of magnitude of experiment. As the errors are almost certainly going to be larger than this we want to find out here (i) how large they might be, (ii) whether we can reduce them by readily accessible variation of the previously chosen force field, or (iii) find it possible to empirically correct a systematic bias appearing in the solubilities due to a systematic error in the force field.

In view of the uncertainty about the adequacy of available force fields and the evidence from the testing of our SR theory that accurate energy estimation is the key requirement we shall here complement the solubility calculations with calculations of crystal energies. For this purpose, we have picked out an ensemble of 43 organic molecules for which both crystal structures and sublimation enthalpies are known from experiment. Thus we are able to get an independent picture of the accuracy of our tested interaction potentials by calculating these crystal sublimation enthalpies. We should note, however, that this is done with a given experimental crystal structure so the error we see may be both direct and indirect through a shifted equilibrium geometry. We shall also report calculations of the enthalpy of melting of 24 molecules in our set of drug molecules which can be compared with corresponding experimental values. Finally we shall report results for aqueous drug solubilities for a total of 47 drug molecules including six modern drug molecules for which the amorphous solubility has been determined.[12]

The force fields that we shall explore here are OPLS-AA[9] with (i) AM1/CM1A[10,11] partial charges or (ii) with OPLS–AA partial charges and (iii) the condensed-phase optimized molecular potentials for atomistic simulation studies (COMPASS)[13] force field which comes

with its own method of determining partial charges. We shall treat the electrostatic interactions represented in terms of partial atomic charges and the steric repulsion, dispersion attractions, and other induced interaction are represented by Lennard-Jones pair-potentials acting between atoms as independent physical mechanisms that can be obtained from different sources and joined into a complete force field. We are aware, however, that when simplified models are used for the interactions the distinction between terms of different physical origin becomes blurred and their respective parameters in the potentials interdependent. Bluntly phrased the performance of a good but oversimplified force field will rely on some error cancellation which is likely to be lost when the two types of interactions are independently improved. That this may be the case is illustrated by a recent study of Mobley et al.[14] who calculated hydration free energies for an ensemble of small molecules using a variety of different methods for estimating atomic partial charges and found that the nominally most accurate method did not necessarily produce the best results. Clearly the AMBER GAFF[15] force field with either TIP3P[16] or TIP4P-Ew[17] water and the different partial charges that they used requires a balanced and interdependent parameter determination to produce the best results. This will quite likely be so also in our study below but we shall nevertheless proceed.

Our earlier work was done with a force field combining OPLS-AA Lennard-Jones parameters with AM1/CM1A partial charges determined by independent quantum mechanical calculations. Thus we have already suffered the lack of interdependency of potential parameters and can hope to improve our force field both in terms of mechanistic accuracy and by achieving a greater consistency between electrostatic and nonelectrostatic interactions.

Originally we felt that the nonelectrostatic interactions represented in our force fields by Lennard-Jones pairwise atomic interactions might be the weak link in the force fields but statistical analyses of the sensitivity of the results to intermolecular interaction parameters showed little support for this. We have instead found the evidence consistent with the partial charges being more critical for the performance of the force field. Thus most of the attention will be focused on the three different methods of determining partial charges employed in the AM1/CM1A scheme based on semiempirical quantum chemistry and the largely empirical parameters used in the OPLS-AA and COMPASS force fields. We also consider a suggested

upscaling of the AM1/CM1A charges in the aqueous phase due to its highly polar nature. As it turns out we will find in the end that the consistent use of either COMPASS or OPLS-AA partial charges in all phases is to be preferred and the combination of COMPASS charges with OPLS-AA Lennard-Jones parameters produces interaction energies and solubilities which, even though not accurate enough *a priori*, show a sufficiently strong correlation with reality that a simple empirical adjustment produces useful solubilities for our large set of drug molecules. The water model used throughout is TIP4P (Transferable Interaction Potential 4 Points) which was developed by Jorgensen et al.[16] and follows the same pattern of Lennard-Jones interaction combined with electrostatic interactions represented in terms of partial charges.

THE METHODOLOGY OF "IN SILICIO PREDICTION"

Approximate Theory

The solvation of a poorly soluble drug molecule in aqueous phase can be modeled by the process of transferring one molecule from pure phase into the aqueous phase. The free energy for such a process can advantageously be obtained from a thermodynamic cycle involving the free energy change, $-\Delta G^{\odot}_{\mathrm{va}}$, on transfer of one molecule from pure amorphous phase to vacuum and the corresponding change, $\Delta G^{\odot}_{\mathrm{va}}$, on going from vacuum to aqueous solution. Here the superscripted circled dot designates Ben Naim quantities.[18] The total free energy change is the sum of these two individual terms. In the SR theory proposed,[6–8] the same *Ansatz* is used for the predictions of the two individual free energy increments and we have

$$\Delta G^{\odot}_{\mathrm{aw}} = -\Delta G^{\odot}_{\mathrm{va}} + \Delta G^{\odot}_{\mathrm{vw}} = -(\Delta G_{\mathrm{cav,a}} + E_{\mathrm{LJ,a}} + E_{\mathrm{C,a}}/2) + (\Delta G_{\mathrm{cav,w}} + E_{\mathrm{LJ,w}} + E_{\mathrm{C,w}}/2) \quad (1)$$

Here $\Delta G_{\mathrm{cav,a}}$ and $\Delta G_{\mathrm{cav,w}}$ are the work (Gibbs free energy changes) required required to create a cavity in the amorphous pure and aqueous phases, respectively. indices "va" and "vw" denote the phase transfers vapor-to-amorphous and vapor-to-water, respectively. The free energy changes associated with solute-solvent interactions are evaluated in the SR theory with the aid of the mean field and linear response approximations in the second line of

eq. (1). The interaction between the solute molecule and all other molecules in the system is represented by the Lennard-Jones energies $E_{\mathrm{LJ,a}}$ and $E_{\mathrm{LJ,w}}$ and Coulomb energies $E_{\mathrm{C,a}}$ and $E_{\mathrm{C,w}}$ in the respective phases. Here the mean field approximation implies that there is no response to the Lennard-Jones attractions so that the free energy change is equal to the intermolecular interaction energy change. The assumption of a linear response implies that the electrostatic free energy is precisely half the corresponding electrostatic intermolecular interaction energy. The introduction of electrostatic interaction is of quantitative importance for drug molecules in contrast to the unpolar alkanes. it is common to use linear response as a first approximation for the response of water due to the coupling of the electrostatic interactions of a solute molecule. Åqvist and coworkers[19,20] have shown that this is a good approximation for ions and less accurate for neutral small organic molecules. This was confirmed in our previous work on studies of neutral molecules. The work required to create a cavity in water to fit the drug molecule is estimated from the product of the surface tension of water and the molecular surface area of the drug molecule as described in previous work.[6] There it was found that the molecular surface area calculated from the atomic radii defined by the minimum of the Lennard-Jones potential yielded the best agreement with FEP[21] simulation data combined with the TIP4P surface tension of 63.5 mN/m.[22] For the pure systems, the surface tension is usually unknown and the work required for the cavity creation in pure system has to be estimated in a different way. An equation of state approach for hard nonspherical particles[23] was used instead as described more in detail in previous work.[7,8] The physicochemical data need in the SR theory can easily be obtained from standard MC or molecular dynamics simulations on molecular systems. In contrast to QSPR methods, an advantage of the SR theory is the mechanistic insight and physical clarification this description might provide. Failures are more easily detected and resolved than with QSPR methods where their origin in most cases is impossible to recognize. Another advantage is the possibility to insert more accurate experimental data to adjust for errors in the predictions due to inaccuracies in the simulation data.

The MC simulation by BOSS

In previous work and in this work, the BOSS program was used for the statistical MC simulations of flexible drug molecules in aqueous and pure systems. BOSS is a molecular modeling program which can perform molecular mechanics calculation, semiempirical quantum chemical calculations, and statistical MC simulations. It is tailored for simulations of biochemical molecular systems in gas and liquid phase. It is based on the traditional Metropolis algorithm and contains many different features like preferential sampling and an algorithm for the change of the intramolecular structure of flexible molecules. Constant volume (NVT) and pressure (NPT) simulations with cubic periodic boundary conditions can be carried out in a standard manner. In the previous work, the statistical FEP[21] method implemented in BOSS was used to calculate free energy changes and to verify the accuracy of the predictions obtained by the SR theory. These simulations are significantly more time-consuming than the standard energy MC simulations required when using the SR theory. For a detailed description of the different simulations on the aqueous and the pure amorphous systems of drug-like molecules, we refer to our previous work.[6–8]

Calculation Method for Crystal Energies

To verify the accuracy of the force field used for modeling, the pure phase intermolecular crystal energies were calculated and compared with the corresponding experimental data. Only the non-bonded interactions described by Lennard-Jones (12-6) and Coulomb potentials were considered as we are interested in the interaction energy between molecules for a given experimentally known crystal structure. No tail corrections fro truncations were applied for the long distance atomic interactions. Given the minor computational effort required for such calculations, an extended cut-off distance of 30 Å was employed instead. This extension of the cut-off distances was deemed necessary on physical grounds. The high space symmetry of the crystal phase means that the Coulomb interaction is much longer ranged: the additive contributions from many small dipoles add up quickly and it would constitute a great neglect to truncate these prematurely. This is contrary to the aqueous phase, where the much higher

degree of thermal randomness, means that the electrostatic potential decays more rapidly, as partial cancellation occurs from imperfectly aligned dipoles.

Furthermore, the same criterion was used in the crystal energy calculations as in the pure amorphous MC simulations[7] for the determination of whether two molecules were within or beyond the cut-off distance. If determined to be within range, all atomic interactions between the molecules were included. The total interaction energy per molecule, which is equal to the lattice energy per molecule multiplied by a factor of two, was obtained by considering one of the crystal unit cells to be located in the center and interacting with surrounding unit cells. If there were more than one molecule in a unit cell the interactions between all these molecules were included. The number of surrounding unit cells is increased until the total interaction energy has converged.

The Modeling of Drug and Solvent by Simple Force Fields

The atomic interactions within a solute molecule, between the solute and the solvent molecules, and between the solvent molecules are described by simple pairwise interaction potentials. The force fields implemented in BOSS are the OPLS-AA[9] and the OPLS-UA[24] (OPLS – United Atom). They are similar in form to the force fields AMBER[25] and CHARMM[26] and correspond to the same level of theory. The simplicity of these force fields is suitable for modeling molecular systems of 1000−−100 000 atoms. The intramolecular conformation energy is calculated from bond stretches and bending angles represented by simple harmonic spring potentials. The dihedral interactions between atoms separated by three bonds are calculated by a four-term Fourier expansion whereas the nonbonded pairwise interactions fro atoms separated by more than three bonds are represented by Coulomb and Lennard-Jones (12-6) potentials. The nonbonded interactions for atoms separated by three covalent bonds are scaled by a factor [1/2] to compensate for the contribution already accounted for in the dihedral interactions within a molecule. Additional potentials for improper angles can be specified by dihedral or harmonic constraints. The same nonbonded potentials are used for the intermolecular atomic interactions between different molecules. As reported elsewhere,[9] the OPLS-AA force field parameters have been optimized against experimental data from

an ensemble of common organic liquids.

In previous and present work, the TIP4P[16] model was used for modeling the aqueous phase. It is a rigid water model with four interaction sites where the spherical Lennard-Jones site is placed at the origin of the sphere representing the center of the oxygen atom. Two equal, positive charges are placed at the centers of the hydrogen atoms and the negative charge is shifted by 0.15 Å from the center of the oxygen atom in the direction along the molecular axis. There exists a wide range of different water models proposed in the literature and some of them are implemented in BOSS: SPC,[27] TIP3P,[16] TIP4P,[16] and TIP5P.[28] The TIP4P model was chosen because it yields good agreement with the experimental density and cohesive energy data of water at 25 °C. The structural information obtained with TIP4P also agrees well with experimental neutron diffraction data for liquid water.[16] As an alternative, the TIP5P water model was evaluated because this model is thought to be improved compared with the TIP4P and generally yields even better agreement with experimental data for pure water. However, unrealistically large absolute hydration energies were found using this model, and it was conjectured that the two extra charges on the TIP5P water molecule in some cases interacted too strongly with positively charged atoms in the solute molecule. If one combines a drug force field that has no hydrogen Lennard-Jones parameters on acidic hydrogens with TIP5P water, collapse of an acidic hydrogen onto the TIP5P "lone pair" can apparently occur. Extreme interaction energies are obtained and a TIP5P water molecule becomes stuck to the acidic hydrogen of the solute molecule for the remainder of the Markov chain. These artifacts do not occur when using TIP4P water making it more suitable in simulations of aqueous systems of drug molecules. The TIP3P and TIP4P are well-known water potentials widely used and discussed in the literature where thermodynamic properties of aqueous solutions have been investigated.[29]

Further artifacts arise when periodic boundary conditions are applied to systems that are too limited in their spatial extension. Because of the size and complexity in terms of chemical structure of the drug molecules, a large water box is required so that the bulk conditions of water can be realized far away from the solute molecule. As six modern drug molecules of considerable complexity were to be included in these studies, the simulations were performed with 2000 TIP4P water molecules instead of 500 as in previously reported studies.[6] A more

detailed investigation of the size effects on both larger and smaller molecules indicated that water systems of 500 TIP4P water molecules are, in fact, too small also for some of the smaller solute molecules investigated. The dependence of the results for $\Delta_{\mathrm{vw}}^{\odot}$ on water box size for two selected molecules (famotidine and nitrobenzene) is shown in the Supporting Information. Thus, to avoid this problem, the water box was increased to 2000 molecules in all simulations also in the case of smaller solute molecules like benzene. Except for the increased system size and the fact that the simulation time was extended, the same simulation parameters were applied as in the previous investigated aqueous systems consisting of 500 water molecules. A cut-off distance of 10 Å was applied with a standard tail correction for the Lennard-Jones interactions where the radial distribution function is assumed to be equal to unity beyond the cut-off. No tail corrections were added to the Coulomb interactions. Instead these interactions are smoothly brought to zero in a quadratic manner within the range 9.5−10 Å.

The limited availability of appropriate atomic partial charges in the OPLS–AA force field for many atomic groups, *e. g.*, cyclic aromatic nitrogen compounds, makes it unsuitable for describing newly discovered drug candidates. Instead, these charges can simply be obtained from semiempirical molecular orbital calculations. Two such methods, AM1[10] (Austin Model 1) and PM3[30,31] (parametric Method 3) are implemented in BOSS with the charge calculation methods CM1A (Charge Model 1A) and CM3A (Charge Model 3A) developed by Storer et al.[11] From rapidly calculated wave functions, appropriate atomic partial charges in a molecule in gas phase can be estimated. Combining these charges with the OPLS–AA nonelectrostatic part provides a fast procedure to provide a new molecule with suitable nonbonded interaction parameters.

In the work on hydration free energies of organic and drug molecules reported by Udier-Blagović et al.,[32] they combined the OPLS–AA Lennard-Jones parameters with partial charges given by AM1/CM1A. The procedure is to calculate the initial atomic charges from a single-point calculation for a given molecular structure. The structure is then energy-minimized *in vacuo* under these charges before they are recalculated for the optimized structure. To take the polarization effects of water into account, it was reported[32] that increasing the atomic charges by 14% yielded improved agreement between experimental and simulated

hydration free energy data for this charging methodology. The virtue of this method is that it is defined for a very wide variety of molecules. It is only compounds containing elements not parameterized in AM1 that cannot be treated. However, the accuracy of the partial charges thus obtained might not always be satisfactory.

As an alternative, we have evaluated the accuracy achieved by using atomic charges from the COMPASS[13] force field together with the OPLS–AA nonelectrostatic potential parameters, which were assigned with complete fidelity to the specifications of the OPLS–AA parameter descriptions. These charge assignments were carried out in an automatic manner from tabulated bond charge increments and the molecular connectivity, in complete accord with the COMPASS bond increment specifications. Our reasons for testing these force field parameters are first that the parameters are tuned to reproduce thermodynamical properties of solid-state phases and second that the range of derived charge types available is broader than that available in the OPLS–AA force field. The valence parameters and the atomic partial charges were derived by tuning with respect to *ab initio* data whereas the Lennard-Jones parameters were obtained by conducting MD simulations on molecular liquids where the cohesive energies and equilibrium densities were adjusted to experimental data. The empirical potentials describing the intermolecular interactions in COMPASS were more sophisticated than those in the OPLS–AA description with additional anharmonic and mixing potentials implemented. The nonbonded atomic interactions are described by a Coulomb and a Mie (9-6) potential with sixth order combination laws for unlike atom pairs expressed by

$$\sigma_{ij} = \left(\frac{\sigma_i^6 + \sigma_j^6}{2} \right)^{1/6} \tag{2}$$

$$\varepsilon_{ij} = 2\sqrt{\varepsilon_i \varepsilon_j} \left(\frac{\sigma_i^3 \sigma_j^3}{\sigma_i^6 + \sigma_j^6} \right) \tag{3}$$

Justification for these somewhat unusual mixing rules can be found in the literature.[33]

COMPUTATIONAL PREDICTIONS IN COMPARISON WITH EXPERIMENT

Crystal Energies

A major question raised in our computations of drug solubility is whether the force field is sufficiently accurate to describe the interactions. Although the ultimate answer will be found in the solubilities themselves, we shall calculate crystal energies and compare them with experiment to obtain an independent and more direct test of the accuracy with which interaction energies of drug-like molecules are predicted. We have evaluated all force field combinations discussed hitherto in modeling the pure crystalline phase of an ensemble (referred to as the sublimation data set) of 43 organic molecules for which experimentally determined crystal structures and enthalpies of sublimation were available. The crystal structures were taken from the Cambridge Crystallographic Data Center (CCDC). References to the experimental data for the different compounds are listed in Table A in the Supporting Information. The molecules in the sublimation data set consist of alcohols, carboxylic acids, aldehyde, amides and a few actual drug molecules.

Rather than relaxing the crystal structures under the different force field regimes, we opted to use the experimentally determined structure in each case. This was done because relaxing the molecular structure was prone to yield large and difficult-to-determine conformation changes, particularly so for the simple force field combinations studied here. Moreover, Osborn and York[34] showed that when comparing calculated crystal energies with experimental data, the agreement is not necessarily improved when relaxing the structures of the calculations.

The calculated intermolecular crystal energies were compared with experimental data obtained from the approximate relation[35,36]

$$E_{\text{inter}} = 2E_{\text{lattice}} \approx -2(\Delta H_{\text{sub}} + 2RT) \tag{4}$$

where E_{lattice} is the lattice energy, ΔH_{sub} is the experimental sublimation enthalpy, R is the gas constant, and T is the absolute temperature. Note that E_{inter} is twice the lattice energy

which is also the total interaction energy between any one molecule and all others in the crystal.

Significant deviations between prediction and experiment were found in the crystal energies of six dicarboxylic acids of varying carbon chain lengths (1, 2, 4, 6, 8, or 10 carbons), as shown in Figure 1, when calculated with OPLS–AA Lennard-Jones parameters and COMPASS charges. Part of the problem is the uncertainties in the precise determination of the hydrogen atom positions from X-ray diffraction experiments, particularly so for highly polarized bonds: for the O–H separations, the uncertainty in bond distance was on the order of 0.109 Å, on a mean bond length of 1.081 Å in this data set.

From more precise neutron scattering measurements, it has been recommended by Allen et al.[37] that all terminal X–H bonds with X equal to C, O, or N should be adjusted to the atom-atom separations $r^{\mathrm{Norm}}_{\mathrm{C-\!-H}} = 1.083$ Å, $r^{\mathrm{Norm}}_{\mathrm{O-\!-H}} = 0.983$ Å, and $r^{\mathrm{Norm}}_{\mathrm{N-\!-H}} = 1.009$ Å, respectively. With this correction, improved agreement between the calculated and the experimental results is obtained as shown in Figure 1, except for dodecandioic acid. In all further crystal calculations, the X–H bonds have therefore been adjusted.

In Figure 2, intermolecular crystal energies, E_{inter}, from OPLS–AA–AM1/CM1A parameters for the molecules of the sublimation ensemble are compared with experimental data obtained using eq. (4). The square symbols (both open and filled) are included to illustrate the uncertainty in the experimental sublimation enthalpy results for the molecules cocaine and ibuprofen from two different literature sources (experimental results are listed in Table A in the Supp. Info.). The RMS deviation of 31 kJ/mol for the 45 data points shown in Figure 2 is considerable.

In Figure 3, the effect of replacing the AM1/CMA1A atomic charges with the OPLS-AA atomic charges is shown. For two cyclic amine compounds, benimidazole and 1,2,4,-triazole, there are no suitable charge parameters available in the OPLS–AA force field and they were thus excluded. The RMS deviation for the 43 data points in Figure 3 is 24 kJ/mol representing an improvement compared with results obtained with OPLS–AA force field and AM1/CM1A partial charges. It is, however, important to emphasize that suitable parameters for describing certain functional groups of new molecules are missing in this force field. The OPLS-AA charges are, unlike the AM1/CM1A charges, independent of the

molecular conformation and depend only on the connectivity to other atoms which makes them more robust.

For flexible molecules, quantum mechanical charges obtained for a single conformation, in a single medium, may not be representative of all conformations in the ensuing Markov chain of the simulation. This problem is alleviated somewhat when using charges fitted against experimental data, or averaged over several conformational quantum mechanical calculations, as they in a mean-value sense incorporate these variations.

One such example of the latter is the partial charges of the COMPASS force field.[13,38–41] Like the OPLS-AA partial charges, they are independent of the conformation of the molecule and depend only on the connectivity of an atom with its neighbors. The results are shown in Figure 4 where the RMS energy deviation in this case is similar, 24 kJ/mol, to the results obtained with the OPLS-AA charges. Crystal energy calculations, as discussed earlier, have also been carried out using the Materials Studio[42] software package version 31. using the full COMPASS force field.[13] In these calculations, the crystal structures were not normalized. Instead the Materials Studio program adjusted nonphysical bond separations found in the molecular crystalline structure in an automatic and *ad hoc* manner. A cut-off of 12.5 Å was used with a cubic spline tail correction for both the Lennard-Jones and the Coulomb interactions. For 34 of the molecules in the sublimation data set, the Lennard-Jones interaction energy $E_{\mathrm{inter,LJ}}$ calculated with the COMPASS and OPLS-AA force fields, using the (9-6) and (12-6) potentials, respectively, are compared (see Fig. 5a). The results are similar despite the two different functional forms and sets of atomic parameters applied in COMPASS and OPLS-AA, respectively, and the different tail corrections. Corresponding results for the Coulomb interactions are shown in Figure 5b and the agreement between the two different sets of partial charges is good. These two force fields seem to yield approximately the same accuracy on average in the calculation of crystalline interaction energies for the investigated molecules in the sublimation data set despite the different repulsion potentials used, *i. e.*, the inverse ninth power repulsion in the COMPASS force field and the inverse twelfth power repulsion in the OPLS-AA force field. The combination laws for interactions between different atom types are more complex in the COMPASS force field compared with those applied in the OPLS-AA force field. The close correlation between the Lennard-Jones interactions

in the two different force fields suggests that we might combine Lennard-Jones interactions modeled by the OPLS-AA force field with electrostatics modeled by the COMPASS force field without losing much accuracy. Clearly these results indicate that the prediction of crystalline interaction energies is improved by replacing the AM1/CM1A partial charges with either OPLS-AA or COMPASS partial charges.

As far as we are aware, among the most accurate calculations of crystal sublimation enthalpies are those reported by Feng and Li.[43] In their work, they used density functional theory (DFT) with the B3LYP exchange-correlation functional and the 6-21G** basis set and counterpoise corrections with an extra correction term to account for the long-range dispersion interactions. The predicted internal crystal energies for 34 organic crystals yielded an RMS deviation of 16 kJ/mol vs. experimental results. It seems unlikely that we would obtain as accurate results using much simpler force field but it would be of great interest for us to improve our accuracy beyond the 24 kJ/mol uncertainty. As the Lennard-Jones interaction energies are greater in magnitude than the electrostatic interactions in the pure drug phase, we have investigated the possibility of optimizing the Lennard-Jones parameters for fixed COMPASS charges. We tried to optimize the Lennard-Jones ε parameters of the OPLS-AA force field for certain atom types by a simulated annealing approach. Thus minimizing the RMS deviation between experimental and calculated crystal energies for the sublimation data set, we found, however, that the relative change of ε values dwarfed the marginal improvements in energy. This indicated to us that we were dealing with error cancellation rather than improvement in the representation of the nonelectrostatic interactions. As small, acceptable changes for the chosen atom types hardly affected the RMS deviation we did no make any further investigations in this direction.

A further test of the force field parameters in the pure drug phase is to compute the enthalpy difference between the crystalline and amorphous phases, ΔH_{ca}. For 24 molecules in the solubility data set, experimental data, *i. e.*, crystal structures and enthalpies of melting, are available in the literature (for references and data see Table B in the Supp. Info.). The enthalpy change, ΔH_{ca}, can be computed from the relation

$$\Delta H_{\mathrm{ca}} = \Delta H_{\mathrm{va}} - \Delta H_{\mathrm{vc}} \approx \Delta E_{\mathrm{inter,ca}}/2 - 3RT + p\Delta V_{\mathrm{ca}} \tag{5}$$

where the subscript "ca" denotes the difference between crystalline and amorphous phases and "vc" denotes the difference between vapor and crystal. $\Delta E_{\mathrm{inter,ca}}$ is the intermolecular interaction energy difference and ΔV_{ca} is the molar volume difference, between the two condensed phases. In eq. (5), it is assumed that the intramolecular energies of all phases are equal. The amorphous phase is modeled as a liquid where the enthalpy change for the vapor to amorphous phase transfer is then approximated by $\Delta H_{\mathrm{va}} \approx E_{\mathrm{inter,a}}/2 - RT + PV_{\mathrm{a}}$. The interaction energy of the crystalline state is calculated in the same manner as discussed previously. the crystal molar volume is taken from the single-crystal X-ray crystallographic data. The corresponding energy and volume data for the amorphous phase are obtained from MC simulations at 25 °C and 1 atm as discussed in earlier work.[8]

A thermodynamic cycle was applied to adjust the experimental melting enthalpy ΔH_{m} obtained at the melting temperature T_{m} to the temperature of interest, *i. e.*, 25 °C. The relation proposed by Neau et al.,[44] where the difference between the heat capacities of the amorphous and crystalline states is approximated by the use of the melting entropy ΔS_{m}, yields

$$\Delta H_{\mathrm{ca}}(T) \approx \Delta H_{\mathrm{m}}(T_{\mathrm{m}}) - \Delta S_{\mathrm{m}}(T_{\mathrm{m}} - T) \tag{6}$$

In Figure 6, experimental enthalpies of melting, extrapolated to 25 °C using eq. (6) to yield ΔH_{ca}, are compared with predicted results obtained from eq. (5). The RMS deviation of 10 kJ/mol is comparable in magnitude to the deviation obtained previously for the intermolecular crystal energies, taking into account that ΔH_{ca} is proportional to the average energy. It should be pointed out that ΔH_{ca} is obtained as the difference between two rather large negative numbers, and it is therefore encouraging that the RMS deviation appears rather small. For chloramphenicol, however, the results, become nonphysical, *i. e.*, a negative enthalpy difference is predicted. This failure might possibly be related to the uncertainty in the crystal structure determination because even more negative values were obtained using another structure in the CCDC database (CLMPCL01). Nevertheless, it could also be due to the inaccuracy of the force field. Except for this particular molecule, the results obtained indicate that the pure systems, both in crystalline and in amorphous states, are reasonably well described by the OPLS-AAA force field with COMPASS atomic charges.

Drug Solubilities

The predicted free energy changes, $\Delta G^{\odot}_{\mathrm{aw}}$, for the total process are obtained from eq. (1) and thermodynamic properties obtained from standard MC simulations both in aqueous and in pure drug phase. To verify these predictions, they are compared with the corresponding experimental data for 47 molecules. The experimental data are derived from the expression

$$S_{\mathrm{a}} = \exp(\Delta G^{\odot}_{\mathrm{aw}}/RT - \ln V_{\mathrm{m,a}}) \tag{7}$$

where S_{a} is the amorphous solubility and $V_{\mathrm{m,a}}$ is the molar volume. As shown previously,[6,8] the amorphous solubility can be approximated from the intrinsic crystal solubility S_0 by the relation

$$S_{\mathrm{a}} \approx S_0 \exp\left(\frac{\Delta S_{\mathrm{m}}}{R} \ln(T_{\mathrm{m}}/T)\right) \tag{8}$$

Thus, by combining eqs. (7) and (8), we obtained $\Delta G^{\odot}_{\mathrm{aw}}$ at $T = 298.15$ K from experimental data on the crystal solubility S_0, the entropy of melting ΔS_{m} and the melting temperature T_{m} together with standard MC simulation data for the estimation of $V_{\mathrm{m,a}}$. Although this means that $\Delta G^{\odot}_{\mathrm{aw}}$ is not strictly experimental, it is nevertheless an instructive comparison. Furthermore, for the six more modern molecules, the experimental amorphous solubility is known[12] and in those cases the approximation in eq. (8) is not necessary. The chemical structures of these molecules are shown in Figure 7 and the corresponding experimental data can be found in the Supporting Information. However, before we go further into the results for the total process, we will first investigate the hydration free energy, $\Delta G^{\odot}_{\mathrm{vw}}$, in more detail.

In Figure 8, results are shown for the free energy of hydration $\Delta G^{\odot}_{\mathrm{vw}}$, obtained using the AM1/CM1A charges scaled by 1.14 the TIP4P surface tension (63.5 ± 5 mN/m)[22] and our SR theory,[6] as described in eq. (4) above. The predictions are compared with experimental data for 12 solute molecules (acetaminophen, acetylsalicylic acid, ethyl-p-hydroxy, flurbiprofen, ibuprofen, ketoprofen, naproxen, phenacetin, salicylic acid, benzene, nitrobenzene, and famotidine). The experimental hydration free energies have been computed from the crystalline solubility, S_0, and the vapor pressure, P_{vap}, at 25 °C according to

$$\Delta G^{\odot}_{\mathrm{vw,exp}} = -RT \ln\left(\frac{S_0 RT}{P_{\mathrm{vap}}}\right) \tag{9}$$

As can be observed in Figure 8, the predictions obtained using scaled AM1/CM1A partial charges are too low by $\sim 11\%$ for 11 of the 12 solute molecules. For famotidine, shown in the lower left corner, there is an error of more than 200 kJ/mol between the predicted value $\Delta G^{\odot}_{\rm vw} = -298$ kJ/mol and the experimental value $\Delta G^{\odot}_{\rm vw} = -80$ kJ/mol. The deviation for nitrobenzene (experimental $\Delta G^{\odot}_{\rm vw} = -17$ kJ/mol)[45] is also quite significant considering its small size. We draw the conclusion that AM1/CM1A charges, when scaled by 1.14, may be too large, particularly so for molecules containing nitro or sulfone groups, and possibly also sulfoxides as discussed elsewhere.[46] It should be noted that famotidine ($C_8H_{15}N_7O_2S_3$) is a rather polar molecule containing one sulfone group, two other sulfurs, and as many as 7 nitrogen atoms and hence most likely represents an extreme. The errors in the partial charges of famotidine contribute to the overestimation of the Coulomb attractions with the surrounding water molecules. Nitrobenzene in TIP4P water has been studied by Carlson and Jorgensen.[45] In their work, partial charges were derived by fitting to the electrostatic potential surface (EPS) obtained from *ab initio* HF/6-31G* wave functions.[46] The EPS partial charges were found to accurately reproduce the experimental data for the free energy of hydration of nitrobenzene. Furthermore, the EPS charges gave a nitrobenzene-water Coulomb interaction energy of around -60 kJ/mol,[45] significantly different from -89 kJ/mol obtained using AM1/CM1A charges scaled up by the factor 1.14.

In Figure8, the hydration free energy results obtained when using the COMPASS charges with scale factor 1.0 in combination with the OPLS-AA parameters for the Lennard-Jones interactions in our simulations are also shown. As can be observed the predictions in this case are on average too high by $\sim 25\%$ compared with experimental results. This was previously also reported by other groups[47,48] where they reported hydration free energies evaluated with the full OPLS-AA interaction potentials for a set of smaller organic molecules and obtained predictions too high by 25%. As previously illustrated, the partial charges are similar in magnitude to the COMPASS partial charges. The reason may be that the OPLS-AA and COMPASS force field parameters were both tuned to experimental thermodynamic properties of pure organic phases. However, a significant improvement was obtained for famotidine with an error of only 5 kJ/mol in comparison with experiment. The major reason for this particular improvement is that the sulfone group is significantly less charged

in the COMPASS description (which also applied to the sulfoxide group). It is important to point out that sulfone and sulfoxide groups are common in modern drugs and therefore require an accurate description. The error in the prediction for nitrobenzene is reduced to -5 kJ/mol with the COMPASS charges because the nitro group is slightly less polar than in the AM1/CM1Ax1.14 description. Still the nitro group is too polar also within the COMPASS description if one considers that the predictions for the remaining molecules are too high. The results obtained suggest that we may use the COMPASS charges also in the aqueous phase as an alternative to the AM1/CM1Ax1.14 charges. With these results together with data obtained for the pure drug phase presented above we will now analyze the total process and the prediction of the amorphous drug solubility using simple force fields.

In Figure 9a, the predictions of the free energy changes ($\Delta G^{\odot}_{\mathrm{aw}}$) calculated by eq. (1) for transferring a molecule from pure amorphous drug phase to aqueous solution are compared with experimental data. The AM1/CM1Ax1.14 partial charges were used for the description of the interactions in aqueous solution and the COMPASS partial charges were used for the description of the interactions in the pure phase as those yielded the best agreement with experimental results for the crystal energies. In the SR theory, the TIP4P surface tension of 63.5 ± 5 mN/m[22] and the molecular surface area obtained from simulations were used to estimate the required work to create a cavity in aqueous solution. The results are quite promising and distributed around the diagonal with an RMS deviation of 10.2 kJ/mol. The large error bars representing the standard error of the mean are mainly due to the uncertainty in the simulated surface tension (63.5 ± 5 mN/m).[2]2 However, for four of the molecules we obtain $\Delta G^{\odot}_{\mathrm{aw}} < 0$ which means that the solute molecule would prefer the aqueous phase to the pure phase. Although not strictly nonphysical, except in the case of the crystalline phase, it is nevertheless an indication of error.

In Figure 9b, the corresponding results obtained instead by the OPLS-AA/COMPASS combination for both phases are compared with experimental results. However, using COMPASS charges in both phases seems to overestimate $\Delta G^{\odot}_{\mathrm{aw}}$ by a factor of 2 which means that predictions will yield too low solubilities (dotted line shows linear regression $\Delta G^{\odot}_{\mathrm{aw,predicted}} = 2.04 \Delta G^{\odot}_{\mathrm{aw,experimental}}$). This might suggest that we should slightly increase the electrostatic interactions in the aqueous phase, which also is illustrated for the 12 solute molecules in

Figure 8. Therefore, we performed simulations for the same molecules as shown in Figure 8 with the COMPASS charges increased by using scale factors 1.10 and 1.20. The predicted free energy of hydration becomes slightly more negative as expected. The slope of the linear correlation between predicted and experimental $\Delta G^{\odot}_{\mathrm{aw}}$ is reduced from 2.04 to 1.62 using a scale factor of 1.10 and to 1.12 with a scale factor of 1.20. However, it should be noted that the scattering of the $\Delta G^{\odot}_{\mathrm{aw}}$ values is more pronounced with the increased charges introduced by a linear scaling. This may indicate that the partial charges should not be scaled linearly for all atoms as proposed by Udier-Blagović et al.[32]

In view of these results, we feel that the overestimation of $\Delta G^{\odot}_{\mathrm{aw}}$ in Figure 9b cannot be corrected using simple charge scaling, without increasing the uncertainty. A crude an completely empirical approach would be to simply correct $\Delta G^{\odot}_{\mathrm{aw}}$ by an added constant, *e. g.*, the constant obtained in the linear regression in Figure 9b. With this adjustment, the scattering of the data is reduced and RMS deviations of less than 6 kJ/mol between the predictions and the experimental data are obtained. This is similar to what is done in conventional QSPR analysis, and thus, requires a "training set." The correlation obtained here, however, contains only one parameter, $\Delta G^{\odot}_{\mathrm{aw}}$, that should capture the basic physical process of solvation: it is a correction due to the inaccuracy of the force fields used. In Figure 10, the result of such a "correction" is shown in terms of the prediction of the amorphous solubility using eq. (8).

The predictions in Figure 10 show a correlation with a coefficient of $r^2 = 0.65$ which does not impress when compared with those obtained in many QSPR studies reported in the literature.[2,49] However, the root mean square deviation between experimental and our corrected data for the free energy of the amorphous to water transfer is less than 6 kJ/mol, which translates into approximately one order of magnitude in terms of solubility (RMS deviation of 1.0 logarithmic units). Compared with conventional QSPR, we expect that our approach will be significantly more robust when predictions are carried out on new molecules, which is not always the case for QSPR.[49] Moreover, our aim will be to take advantage of the mechanistic clarity of our approach to find the cause of the shift required and in course eliminate it.

DISCUSSION AND CONCLUSION

With the use of the proposed SR theory and thermodynamic properties determined by conventional MC simulations, we have shown here that apparently robust and for our use sufficiently accurate aqueous amorphous solubilities for a wide range of larger and flexible organic molecules can be obtained. To our knowledge, this has not been done before for such a wide set of complex molecules with the level of mechanistic analysis proposed and discussed here. Commonly QSPR methods are used for predictions of physicochemical properties like aqueous solubility and $\log P$ data, where P is the octanol-water partition coefficient of organic and more complex substances. Our approach is an important alternative to these methods where the improved physical insight obtained with our method is an advantage compared with the QSPR methods which are more purely based on statistics. The evaluation of the semiempirical SR theory both with brute force free energy simulation data and experimental solubility data showed that robust results were obtained despite the fact that one adjustment parameter had to be used for the corrections of inaccuracies in the force fields to yield a better agreement with the experimental data. Such calculations can be performed within 1 day on a dual core standard PC of today for a drug molecule of $\sim$ 70 atoms. The corresponding brute force free energy simulation would take 1 week or more for a complex molecule. In the pharmaceutical industry, computational speed is an important factor and it is desirable that the prediction of accurate aqueous solubility of a newly developed drug does not take more than one day. Otherwise experimental solubility measurement would be a more attractive choice. Brute force free energy simulations are therefore at present not an attractive alternative.

The force field modeling plays a dominant role in the computational prediction of accurate drug solubilities in aqueous phase by MC simulation. We need to find or generate a tractable model for both the interaction of drug molecules in pure phase and in dilute water solution. The SR theory has given us an important possibility to efficiently evaluate the accuracy of the force field used to describe the molecular interactions in the pure and aqueous phases, respectively. Two important questions here were (i) the reliability of the simple force field used in our earlier work and (ii) whether this force field could be improved without the loss of

simplicity. With the help of the CCDC data on crystal structures and enthalpies of complex molecules, we have tested the reliability of the Lennard-Jones modeling of the dispersive attractions used in the OPLS-AA and COMPASS force fields. The calculations soon showed that despite a dominant contribution to the total interaction energies in the condensed drug phase these corresponding Lennard-Jones parameters in our original OPLS-AA force field were not readily improved by statistical means. This result serves as a reminder of the difficulty of replacing mechanistic analysis by statistics.

Focus was then shifted to the partial charges and the corresponding electrostatic interactions which are dominant in the aqueous phase and found to be more variable between the AM1/CM1A method and the OPLS-AA and COMPASS force fields. In reality, the charges in a molecule or molecular fluid are distributed in accord with the corresponding quantum mechanical ground-state structures. The partial charges are no more than a numerically convenient way of representing far more complex electrostatic interactions. Thus we are not surprised that they vary according to the procedure by which they are fitted and are responsible for a large share of the systematic deviations and scatter in our results. Although the quantum chemical AM1/CM1A partial charges at first appeared to have an advantage in their first principles determination for a molecule in vacuum our results indicate that they are not optimal for our applications in condensed phases. When these charges are used in combination with a linear scale procedure to incorporate the polarization of liquid water, a scale constant of 1.14 seems to yield an overestimation of the polarization effects. In the pure phase, the results with the unscaled AM1/CM1A charges are on average reasonable in magnitude but the predicted results for the crystal energies show significant scatter. A more empirical procedure such as used in the original OPLS-AA scheme or in the empirical and *ab initio* COMPASS scheme is able to significantly reduce the scatter and thereby open the way for improved solubility prediction as demonstrated in this work. Clearly, the polarization effects in water are significantly underestimated with both our sets of force field charge parameters. This has previously been noted by, among others, Oostenbrink et al.[50] who found that optimized parameters for solvents of different dielectric constants differ considerably. To compensate for this polarization effect, the COMPASS charge parameters were linearly increased by scale factors of 1.10 or 1.20. The hydration free energies were then strength-

ened as desired, but the scatter was increased by this procedure. The results presented here illustrate the importance of using partial charges which are consistent and appropriate for both phases, pure drug or water solution, rather than imbalanced in favor of either one. It appears that until they can be placed on a more firm footing scaling of the charges should be avoided if possible or applied uniformly in both phases. The simpler procedure employed in the COMPASS force field which is independent of molecular configuration and environment seems also to be the most robust in our present application.

Further progress in computational solubility prediction relies heavily on the accuracy of the force field used. Although our potentials evaluated in the present work is able to produce an RMS deviation of roughly 10 kJ/mol for the needed $\Delta G^{\odot}_{\mathrm{aw}}$ value of transfer from pure solute to aqueous phase the aim is to get this value down to less than 6 kJ/mol without any empirical correction as applied here. The simple force fields used in this article therefore need to be improved further before sufficiently accurate *a priori* predictions can be obtained. We feel that such activities to improve the modeling of the interactions are of utmost importance. Significant progress in force field generation would also require more high quality experimental data for complex molecules, *e. g.*, crystal vapor pressures.[51] Such experimental work is, unfortunately, limited to a few groups in the world.[51–53]

A potentially fruitful alternative to scaling the charges, or the Lennard-Jones parameters, of the solute, is exemplified by Shirts and Pande who optimized the Lennard-Jones parameters of TIP3P water to more accurately reproduce the free energies of hydration for some small organic molecules.[29] We have ourselves begun investigations in this direction, which we hope will be advantageous also for computing the free energies of hydration of larger, bioactive molecules.

Although hopeful that more accurate force fields can be found, we feel confident about the continued utility of our approximate SR theory which is based on a mechanistic analysis of the interactions and varying response of the system to the different components of the total interaction. With this tool backed up by our full simulation method, we can now understand, and perhaps in the future anticipate and rectify, failure of our solubility prediction in individual cases. Furthermore, systematic errors may be also revealed and corrected either by empirical means or eventually by improved theoretical analysis and refined calculations.

Such understanding is not easily generated by QSPR analysis and it lays a foundation for further progress in this area.

ACKNOWLEDGMENTS

This work has been financially supported by the Swedish Science Research Council (VR) and AstraZeneca. Special gratitude is expressed to Professor Per Svensson at AstraZeneca R&D Södertälje for the valuable discussions and suggestions on the accuracy of crystal structures. Accelrys Software Inc. is also acknowledged for letting us evaluate the Materials Studio software package.

References

1. Jain, N., J. Pharm. Sci., **2001**, 90,234.

2. Jorgensen, W. L., Duffy, E. M., Adv. Drug Deliv. Rev., **2002**, 54, 355.

3. Johnson, S. R., J. Chem. Inf. Model., **2008**, 48, 25.

4. Jorgensen, W. L. In The Encyclopedia of Computational Chemistry; Schleyer, P. v. R., Ed.; Wiley: New York, USA, **1998**, p. 3281.

5. Jorgensen, W. L., Tirado-Rives, J., J. Comput. Chem., **2005**, 26, 1689.

6. Westergren, J., Lindfors, L., Höglund, T.; Lüder, K., Nordholm, S.; Kjellander, R., J. Phys. Chem. B. **2007**, 111, 1872.

7. Lüder, K., Lindfors, L., Westergren J., Nordholm, S., Kjellander, R., J. Phys. Chem. B. **2007**, 111, 1883.

8. Lüder, K., Lindfors, L., Westergren, J., Nordhom, S., Kjellander, R., J. Phys. Chem. B **2007**, 111, 7303.

9. Jorgensen, W. L., Maxwell, D. S., Tirado-Rives, J., J. Am. Chem. Soc. **1996**, 118, 11225.

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

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

12. Lindfors, L., Forssén, S., Skantze, P., Skantze, U., Zackrisson, A., Olsson, U., Langmuir **2006**, 22, 911.

13. Sun, H., J. Phys. Chem. B **1998**, 102, 7338.

14. Mobley, D. L., Dumont, E., Chodera, J. D., Dill, K. A., J. Phys. Chem. B **2007**, 111, 2242.

15. Wang, J., Wolf, R. M., Caldwill, J. W., Kollman, P. A., Case, D. J., J. Comput. Chem. **2004**, 25, 1157.

16. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L., J. Chem. Phys. **1983**, 79, 926.

17. Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L., Head-Gordon, T., J. Chem. Phys. **2004**, 120, 9665.

18. Ben-Naim, A., Solvation Thermodynamics: Plenum Press: New York, **1987**.

19. Åqvist, J., Hansson, T., J. Phys. Chem. **1996**, 100, 9512.

20. Almlöf, M., Carlsson, J., Åqvist, J., J. Chem. Theory Comput. **2007**, 3, 2162.

21. Perlman, D. A., Rao, B. G. In The Encyclopedia of Computational Chemistry; Schleyer, P. v. R., Ed., Wiley: Athens, USA, **1998**, p. 1036.

22. Pohorille, A., Wilson, M., J. Mol. Struct. (Theochem.) **1993**, 284, 271.

23. Boublik, T., Coll. Czech Chem. **1986**, 51, 2301.

24. Jorgensen, W. L., Madura, J. D., Swenson, C. J., J. Am. Chem. Soc. **1984**, 106, 6638.

25. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, K. M., Merz Jr., K. M., Fergusson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A., J. Am. Chem. Soc. **1995**, 117, 5179.

26. MacKerell Jr., A. D., Bashrofrd, D., Bellot, W., Dunbrack Jr., R. L., Evanseck, J. D., Field, M. J., Fisher, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuzera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reither III, W. E., Raux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuzera, J., Yin, D., Karplus M., J. Phys. Chem. B **1998**, 102, 3586.

27. Toukan, K., Rahman, A., Phys. Rev. B **1985**, 31, 2643.

28. Mahoney, M. W., Jorgensen, W. L., J. Chem. Phys. **2000**, 112, 8910.

29. Shirts, M. R., Pande, V. S., J. Chem. Phys. **2005**, 122, 134508.

30. Stewart, J. J. P., J. Comput. Chem. **1989**, 10, 209.

31. Stewart, J. J. P., J. Comput. Chem. **1989**, 10, 221.

32. Udier-Blagović, M., De Tirado, P. M., Pearlman, S. A., Jorgensen, W. L., J. Comput. Chem. **2004**, 25, 1322.

33. Waldman, M., Hagler, A. T., J. Comput. Chem. **1993**, 14, 1077.

34. Osborn, J. C., York, P., J. Mol. Struct. **1999**, 43, 17.

35. Perstin, A. J., Kitaigorodsky, A. I., The Atom-Atom Potential Method: Application to Organic MOlecular Solids; Springer: New York, **1987**.

36. Warshel, A., Lifson S., J. Chem. Phys. **1970**, 53, 582.

37. Allen, F. H., Kennard, O., Watson, D. G., Brammer, L., Orpen, A. G., Taylor, R., J. Chem. Soc. Perkin. Trans. II **1985**, 12, S1.

38. Bunte, S. W., Sun, H., J. Phys. Chem. B **2000**, 104, 2477.

39. Rigby, D., Sun, H., Eichinger, B. E., Polym. Int. **1997**, 44, 311.

40. Sun, H., Rigby, D., Spect. Acta A **1997**, 53, 1301.

41. Sun, P., Ren, P., Fried, J. R., Comp. Theor. Polym. Sci. **1998**, 8, 229.

42. Materials Studio 3.1. Accelrys Software Inc.: San Diego, CA.

43. Feng, S., Li, T., J. Chem. Theory Comput. **2006**, 2, 149. R., J. Chem. Soc. Perkin. Trans. II **1985**, 12, S1.

44. Neau, S. H., Bhandarkar, S. V., Hellmuth, E. H., Pharm. Res. **1997**, 14, 601.

45. Carlson, H. A., Jorgensen, W. L., J. Phys. Chem. **1995**, 99, 10667.

46. Carlson, H. A., Nguyen, T. B., Orozco, M., Jorgensen, W. L., J. Comput. Chem. **1993**, 14, 1240.

47. Shirts, M. R., Pitera, J. W., Swope, W. C., Pande, V. S., J. Chem. Phys. **2003**, 119, 5740.

48. Maccallum, J. L., Thielman, D. P., J. Comput. Chem. **2003**, 24, 1930.

49. Faller, B., Ertl, P., Adv. Drug Deliv. Rev. **2007**, 59, 533.

50. Oostenbrink, C., Villa, A., Mark, A. E., van Gunsteren, W. F., J. Comput. Chem. **2004**, 25, 1656.

51. Perlovich, G. L., Volkova, T. V., Bauer-Brandl, A., J. Pharm. Sci. **2006**, 95, 2158.

52. Elder, J. P., J. Thermal Anal. **2997**, 49, 897.

53. Wright, S. F., Dollimore, D., Dunn, J. G., Alexander, K., Therm. Acta **2004**, 421, 25.

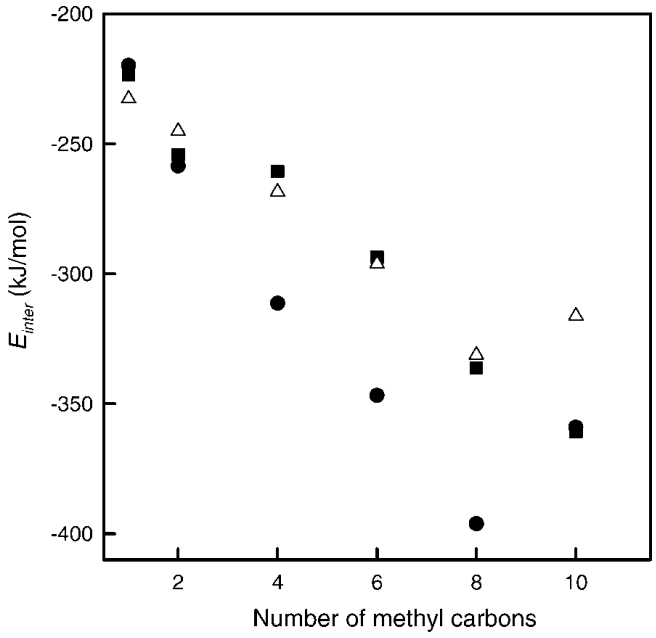

Figure 1: Intermolecular crystal energies for six dicarboxylic acids comparing experimental sublimation energy data (open triangles) with calculated results using OPLS-AA Lennard-Jones interaction parameters and COMPASS charges. Filled circles represent calculated results obtained using the experimental crystal structures directly from the CCDC database, whereas filled squares represent calculated eneriges where all terminal H positions have been normalized.[37]

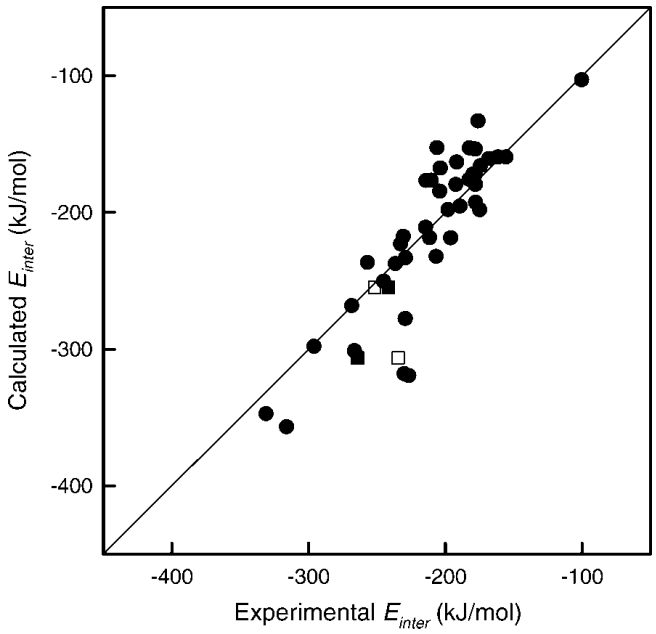

Figure 2: Calculated intermolecular crystal energies vs. experimental sublimation energy data (filled circles). The interaction energies are calculated using OPLS-AA Lennard-Jones interaction parameters and AM1/CM1A charges for crystal structures obtained from the CCDC database with normalized terminal H positions.[37] The filled and open squares represent results for crystals from two different literature sources of experimental sublimation energy data illustrating the experimental uncertainty in such experiments.

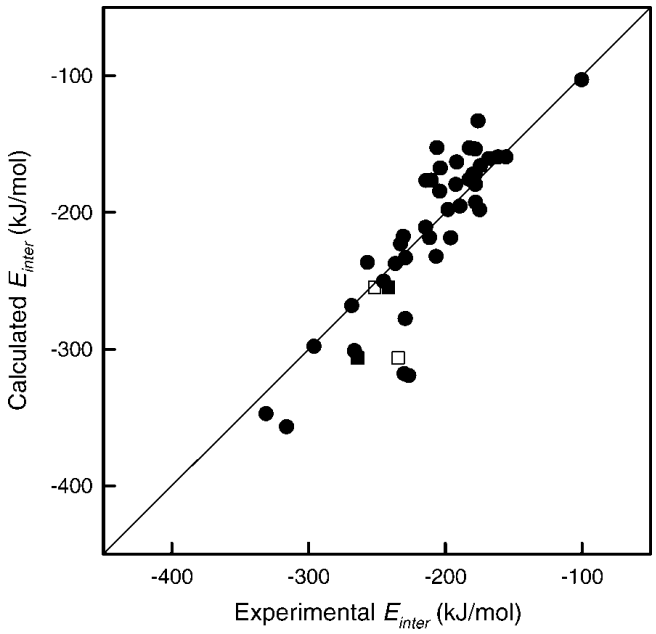

Figure 3: Calculated intermolecular crystal energies vs. experimental sublimation energy data (filled circles). The interaction energies are calculated using OPLS-AA Lennard-Jones interaction parameters and OPLS-AA charges for crystal structures obtained from the CCDC database with normalized terminal H positions.[37] The filled and open squares represent results for crystals from two different literature sources of experimental sublimation energy data illustrating the experimental uncertainty in such experiments.

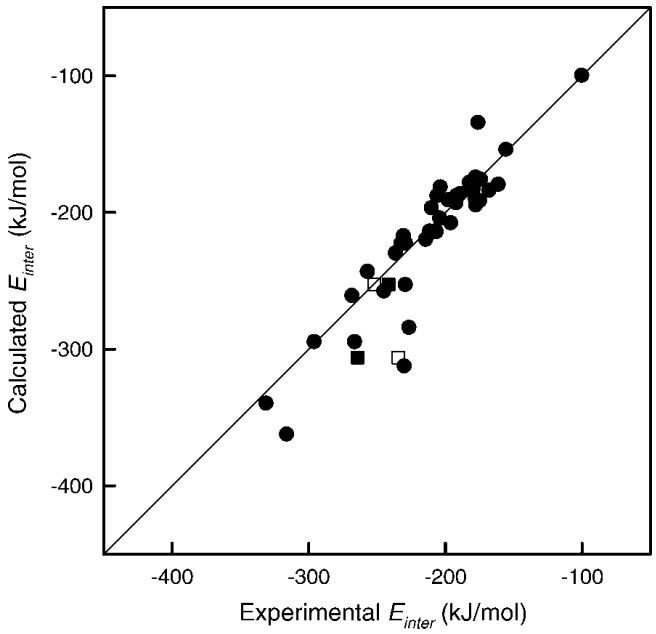

Figure 4: Calculated intermolecular crystal energies vs. experimental sublimation energy data (filled circles). The interaction energies are calculated using OPLS-AA Lennard-Jones interaction parameters and COMPASS charges for crystal structures obtained from the CCDC database with normalized terminal H positions.[37] The filled and open squares represent results for crystals with two different literature sources of experimental sublimation energy data illustrating the experimental uncertainty in such experiments.

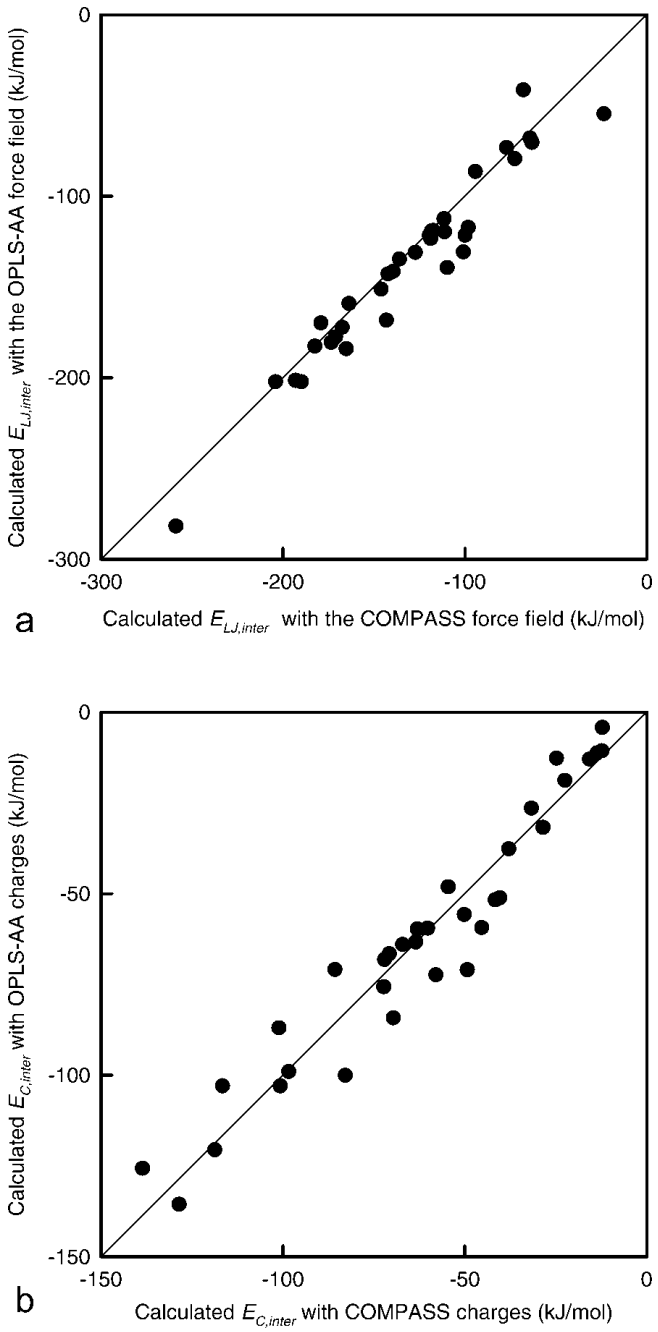

Figure 5: (a) Calculated Lennard-Jones intermolecular interaction energies using the OPLS-AA force field vs. results obtained using the COMPASS force field for experimental crystal structures from the CCDC database (no normalization of terminal H positions). (b) Calculated Coulomb intermolecular interaction energies using the OPLS-AA charges vs. results obtained using COMPASS charges for experimental crystal structures from the CCDC database (no normalization of terminal H positions).

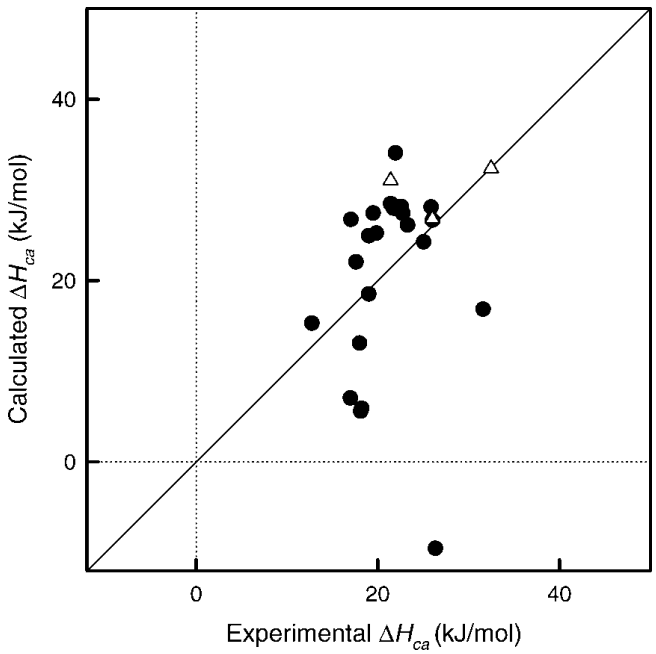

Figure 6: Calculated changes in enthalpies for the process, crystal to amorphous phase vs. experimental data where the calculated data are obtained from crystal structures in the CCDC (with normalized terminal H positions) and Monte Carlo simulations of amorphous drugs using OPLS-AA Lennard-Jones interaction parameters and COMPASS charges. The filled circles are results for molecles from the BOSS data set and the open triangles are results for felodipine, nifedipine and bicalutamide. The dotted lines are guide lines.

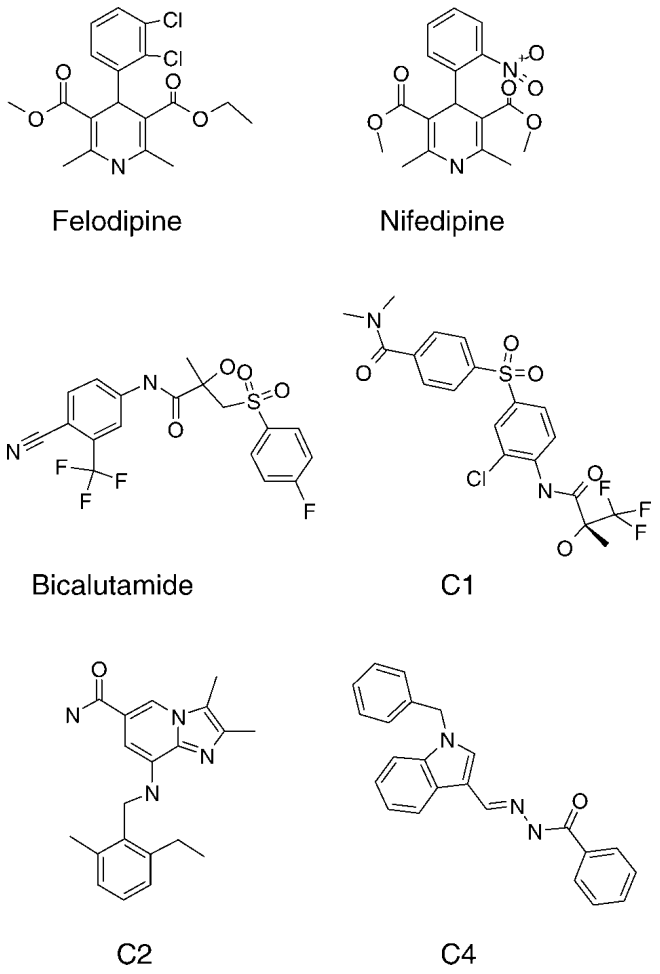

Figure 7: Molecular structures of the six drug compounds here refereed to as more modern molecules where the amorphous solubility is experimentally determined.[12]

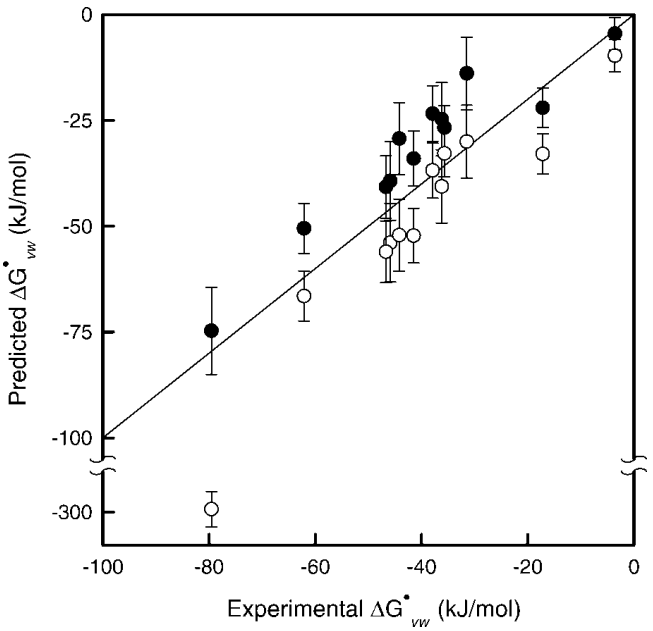

Figure 8: Predicted results for the free energy of hydration vs. experimental data (nitrobenzene $\Delta G^{\odot}_{\rm vw} = -17$ kJ/mol[45] and famotidine $\Delta G^{\odot}_{\rm vw} = -80$ kJ/mol). Theoretical predictions are carried out using data from simulations obtained with the OPLS-AA Lennard-Jones interaction parameters, and AM1/CM1A×1.14 (open circles) or COMPASS charges (filled circles) in TIP4P water simulations using the TIP4P surface tension ($\gamma_{\rm TIP4P} = 63.5 \pm 5$ mN/m).[22]

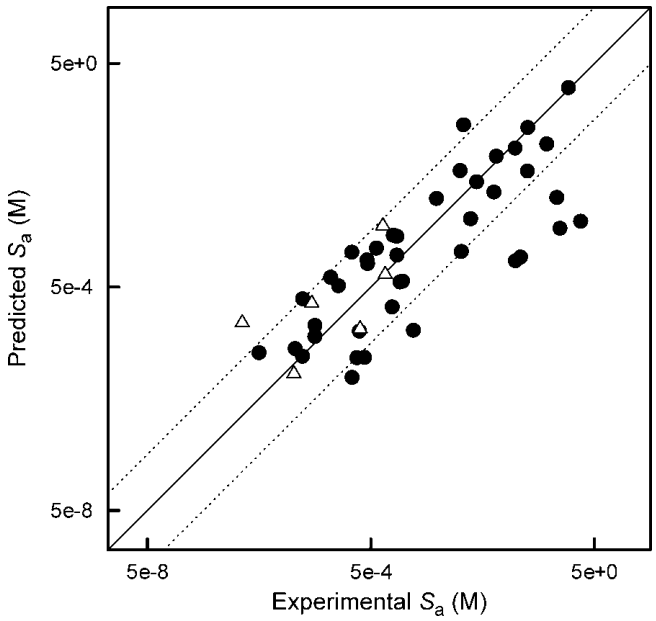

Figure 9: (a) Predicted results for the free energy of transfer amorphous to water vs. data calculated from experimental crystal solubility data, crystal melting data, and molar volumes from simualtions of amorphous drugs (filled circles). Theoretical predictions are carried out using data from simulations obtained with the OPLS-AA Lennard-Jones interaction parameters, AM1/CM1A scaled by a factor of 1.14 and COMPASS charges in water and pure drug systems, respectively. Predicted results for the six more modern durgs for which the amorphous solubility has been experimentally determined (open triangles).[12] The full line represents a guide line with slope = 1. (b) Predicted results for the free energy of transfer amorphous to water vs. experimental data for molecules in the solubility data set (filled circles). The open triangles represent results for the six drug molecules with known amorphous solubility.[12] Theoretical predictions are carried out using data from simulations obtained with the OPLS-AA Lennard-Jones interaction parameters, COMPASS charges inw ater and pure drug systems, respectively. The dotted line is a best fit to experimental results for the molecules in the solubility data set givin $\Delta G^{\odot}_{\mathrm{aw,predicted}} = 2.04 \Delta G^{\odot}_{\mathrm{aw,experimental}}$.

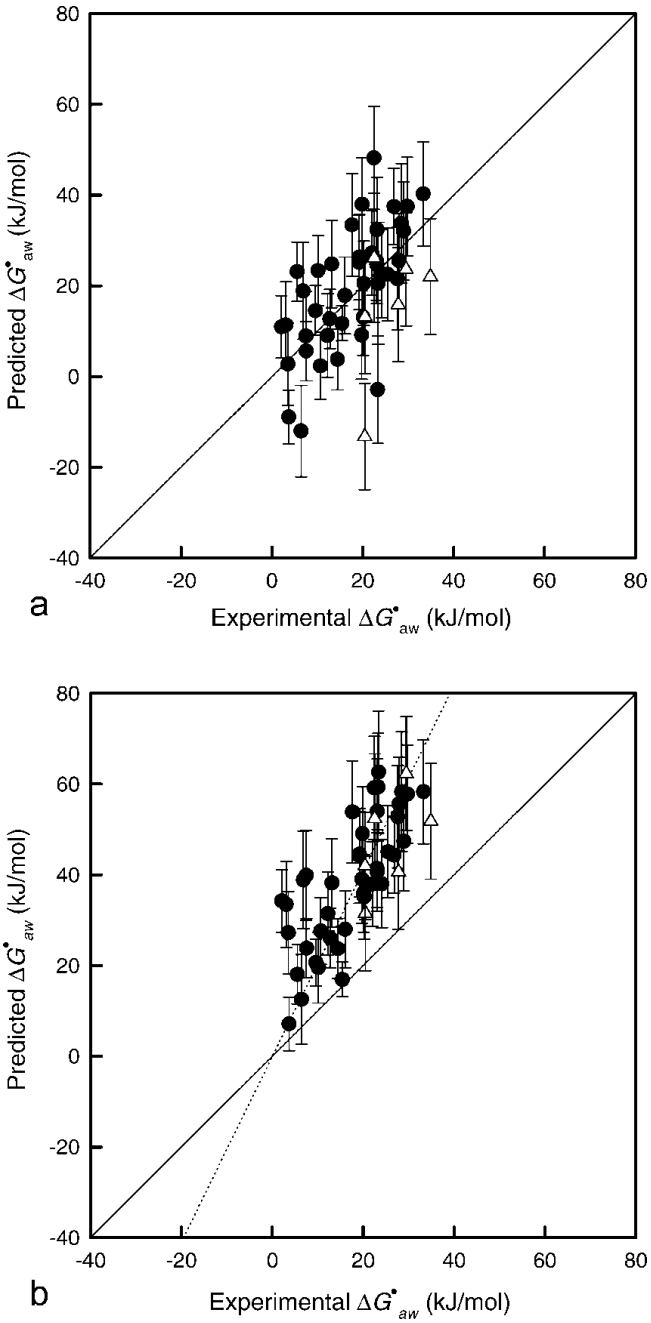

Figure 10: Predicted amorphous solubilities vs. experimental data for 41 drug molecules (filled circles) and six drug molecules with known amorphous solubility (open triangles).[12] The predicted results are obtained from Monte Carlo simulations using the OPLS-AA Lennard-Jones interaction parameters with COMPASS charges in TIP4P water and pure drug systems, respectively. The dotted lines illustrate the ± 1 unit in $\log(S_{\mathrm{a}})$.