JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

First-Principles Calculation of the Intrinsic Aqueous Solubility of Crystalline Druglike Molecules

David S. Palmer,[†,‡] James L. McDonagh,[¶] John B. O. Mitchell,*[,¶] Tanja van Mourik,[¶] and Maxim V. Fedorov*[,†,‡]

[†]Department of Physics, University of Strathclyde, John Anderson Building, 107 Rottenrow, Glasgow, Scotland G4 0NG, United Kingdom

[‡]Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, DE-04103 Leipzig, Germany

[¶]Biomedical Sciences Research Complex and EaStCHEM School of Chemistry, University of St. Andrews, Purdie Building, North Haugh, St. Andrews, Scotland KY16 9ST, United Kingdom

Supporting Information

ABSTRACT: We demonstrate that the intrinsic aqueous solubility of crystalline druglike molecules can be estimated with reasonable accuracy from sublimation free energies calculated using crystal lattice simulations and hydration free energies calculated using the 3D Reference Interaction Site Model (3D-RISM) of the Integral Equation Theory of Molecular Liquids (IET). The solubilities of 25 crystalline druglike molecules taken from different chemical classes are predicted by the model with a correlation coefficient of $R = 0.85$ and a root mean square error (RMSE) equal to 1.45 $\log_{10} S$ units, which is significantly more accurate than results obtained using implicit continuum solvent models. The method is not directly parametrized against experimental solubility data, and it offers a full computational characterization of the thermodynamics of transfer of the drug molecule from crystal phase to gas phase to dilute aqueous solution.

1. INTRODUCTION

The intrinsic aqueous solubility of an ionizable molecule is defined as the concentration of the un-ionized molecule in saturated aqueous solution at thermodynamic equilibrium at a given temperature.[1,2] It is used to calculate dissolution rate and pH-dependent solubility in models such as the Noyes−Whitney equation[3] and the Henderson−Hasselbalch equation,[4,5] respectively. Prediction of the intrinsic aqueous solubility of bioactive molecules is of great importance in the biochemical sciences, because it is a key determinant in the bioavailability of novel pharmaceuticals[6−11] and the environmental fate of potential pollutants.[12,13]

Over the last two decades, more than 100 different computational methods to predict the solubility of organic molecules in water have been published.[14−16] The vast majority of these are quantitative structure−property relationships (QSPRs), which use experimental data to learn a statistical relationship between the physical property of interest (e.g., solubility) and molecular descriptors calculable from a simple computational representation of the molecule.[6] QSPRs have been widely used because they are computationally inexpensive and may offer reasonably accurate predictions for molecules similar to those in the training set.[17] It is well-known, however, that QSPR models are unreliable for molecules dissimilar to those in the training set. Furthermore, since QSPRs are not based on any fundamental physical theory, they provide little information about the underlying physical chemistry. In all but a few cases,[18−20] QSPR models predict solubility from molecular rather than crystal structure, which means they are not able to rationalize or predict different solubilities for different polymorphs of a molecule.

A more rigorous approach to predicting intrinsic aqueous solubility would be to calculate it directly from molecular simulation. Until now, however, few such approaches have been published, even though a large number of similar methods have been proposed to calculate other physicochemical properties, such as octanol−water partition coefficients,[21] acid−base dissociation coefficients,[22] and protein−ligand binding free energies.[23,24] One reason for this observation is that the crystalline polymorphic form of organic molecules has traditionally been difficult to predict from molecular structure. However, there has been significant progress in this field in recent years. The current state of the art allows the polymorphic landscape of rigid and semiflexible organic molecules to be calculated with reasonable confidence,[25−29] with some recent successes also being reported for crystal structure prediction of molecules with multiple rotatable bonds.[30]

The aim of this work is to propose and test several methods to calculate the intrinsic aqueous solubility of druglike molecules starting from a known crystal structure, which we take from experiment here, but which might, in future work, be obtained by crystal structure prediction. The most tractable approach to calculating intrinsic aqueous solubility from molecular simulation is via computation of the free energy of solution (ΔG_{sol}), which is the free-energy change associated with transfer of the molecule from the crystalline phase to aqueous solution under standardized conditions (see Figure 1, presented later in this work). Although the solution free energy

Received: May 2, 2012
Published: July 25, 2012

cannot easily be calculated from a single simulation, it may, in principle, be decomposed into terms that can be computed in separate simulations, via a thermodynamic cycle.

The thermodynamic cycle of solid drug to supercooled liquid to aqeuous solution has been the basis of many different methods to predict intrinsic aqueous solubility. Although the supercooled liquid state is not accessible by experiments, it is introduced into the thermodynamic cycle to allow the process of breaking down the crystal lattice (step 1: crystal to supercooled liquid) to be decoupled from the process of hydration (step 2: supercooled liquid to aqueous solution). The most successful method for prediction of solubility from this thermodynamic cycle is the general solubility equation (GSE), which relates log S to the melting point and the logarithm of the octanol−water partition coefficient (log P). It can be derived from the thermodynamic cycle of gas to supercooled liquid to solution (if some assumptions are made about the entropy of melting, ΔS_m). The GSE provides useful estimates of solubility when experimental melting point and ΔS_m[31] data are available. However, the GSE is not usually applicable to unsynthesized molecules, because the best empirical methods for predicting the melting point give predictive errors of 40−50 °C.[32,33] A thermodynamic cycle similar to that discussed in this paragraph has also been used to predict the solubility of amorphous druglike molecules, using Monte Carlo simulations with additional empirical parameters employed to remove systematic errors.[34]

The thermodynamic cycle for transfer from crystal to vapor to solution has been the subject of both experimental and computational studies. The most successful such study has been the prediction of the solubility of liquids (and a small number of low-molecular-weight solids) from both experimental and calculated vaporization and hydration energies by Thompson et al.[35] The authors report predictive mean unsigned errors in the range of 0.4−0.6 in log solubility (log S) for a dataset comprising simple low-molecular-weight compounds. Schnieders et al. have recently proposed a method to predict the solubility of organic crystals from molecular dynamics simulations.[36] For a small dataset comprised of four low-molecular-weight n-alkylamides (acetamide, butanamide, pentanamide, hexanamide), solubility was predicted with a root mean square error (RMSE) of 0.83 log S units (calculated from data presented in Table 8 of the original manuscript[36]). Reinwald et al. predicted the aqueous solubility of a more diverse selection of drugs from the experimental enthalpies of sublimation (ΔH_{sub}) and calculated hydration energies.[37] Unfortunately, this method was only accurate for one molecule from a dataset of 12, and no computational procedure was suggested for the calculation of ΔH_{sub}. Perlovich et al. published a series of papers that investigate the thermodynamic properties of a wide variety of structurally diverse drugs, by both experiment and computation, but they have not provided any methods for the prediction of solubility from structure alone without empirical parametrization.[38,39]

In previous work, some of the current authors (D.S.P. and J.B.O.M.) attempted to predict solubility from calculation of sublimation and hydration free energies.[18] In that work, where ΔG_{sol} was calculated using an implicit solvent model based on the Poisson−Boltzmann equation, ab initio results were not found to deliver the required accuracy, but after the introduction of a small number of empirical corrections, accurate predictions of a druglike test set were obtained. Since then, however, there has been significant progress in the development of methods to calculate sublimation[40] and hydration free energies. Moreover, additional experimental data to benchmark these calculations have become available.[41] In particular, motivated by our earlier results, some of the current authors (D.S.P., M.V.F.) have developed a set of free-energy functionals that allow hydration free energies to be calculated accurately using the 3D Reference Interaction Site Model (3D-RISM).[42−46] The 3D-RISM/UC solvation free-energy functional is easily implemented using existing computational software and allows in silico screening of druglike molecules at significantly lower computational expense than explicit solvent simulations. (Here, UC stands for universal correction.) Furthermore, new improved continuum solvent models for quantum mechanics calculations[47] and new density functionals for thermochemistry[48] have recently been developed. In the current work, we assess how accurately the intrinsic aqueous solubility of crystalline druglike molecules can be calculated without empirical parametrization of the computational methods against experimental solubility data. In particular, we consider three different model potentials used to calculate sublimation free energies, and four different methods for calculating hydration free energies, taken from both implicit continuum solvation theory and the Integral Equation Theory of molecular liquids.

2. THEORY

2.1. Calculation of Intrinsic Aqueous Solubility from Sublimation and Hydration Free Energies. The intrinsic aqueous solubility of a crystalline solute is measured at thermodynamic equilibrium between the undissolved crystalline form of the molecule and the neutral form of the molecule in solution, which can be written $X_s \rightleftharpoons X_{aq}$. If the activity coefficient for the solute in solution is assumed to be unity, then the relationship between intrinsic solubility (S_0) and the overall change in Gibbs free energy is[35]

$$\Delta G^*_{sol} = \Delta G^*_{sub} + \Delta G^*_{hyd} = -RT \ln(S_0 V_m) \quad (1)$$

where ΔG^*_{sol} is the Gibbs free energy for solution, ΔG^*_{sub} the Gibbs free energy for sublimation, ΔG^*_{hyd} the Gibbs free energy for hydration, R the molar gas constant, T the temperature (298 K), V_m the molar volume of the crystal, and S_0 the intrinsic solubility (in moles per liter, mol/L). The superscript asterisk (*) denotes that we are using the Ben−Naim terminology,[49,50] which refers to the Gibbs free energy for transfer of a molecule between two phases at a fixed center of mass in each phase (see Figure 1).

The intrinsic aqueous solubility of a solute depends on the crystal form of the precipitate at thermodynamic equilibrium during the solubility experiment. For neutral druglike molecules, it is important to consider the influence that crystalline polymorphism has on solubility. The average difference in molar solubilities between polymorphs has been estimated to be ∼2-fold.[51] Since this value is larger than the average error in models to predict solubility (5-fold to 10-fold molar solubility),[14−16,52] it has previously been suggested that a simulated low-energy crystal structure may be a useful starting point from which to predict solubility, even if small errors exist in the calculated polymorphic landscape.[18,19]

2.2. Calculation of Hydration Free Energy Using 3D-RISM-UC. *2.2.1. Background.* The 3D Reference Interaction Site Model (3D-RISM)[53−56] is a theoretical method for modeling solution phase systems based on classical statistical

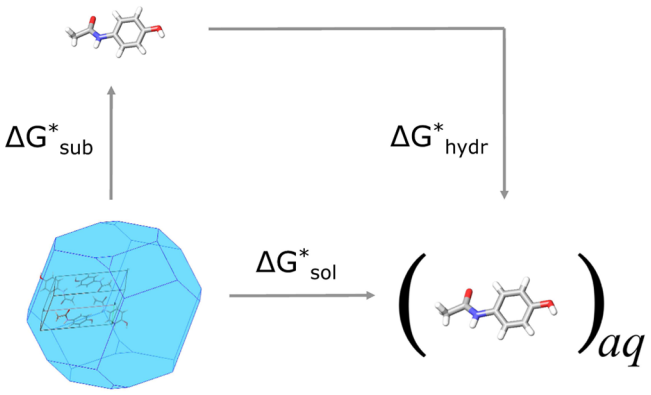

Figure 1. Thermodynamic cycle for transfer from crystal to gas and then to aqueous solution. This figure is based on Figure 1 from our earlier work.[18]

mechanics. The 3D-RISM equations relate 3D intermolecular *solvent site−solute* total correlation functions ($h_\alpha(\mathbf{r})$), and direct correlation functions ($c_\alpha(\mathbf{r})$) (index α corresponds to the solvent sites):[55,53]

$$h_\alpha(\mathbf{r}) = \sum_{\xi=1}^{N_{\text{solvent}}} \int_{R^3} c_\xi(\mathbf{r}-\mathbf{r}')\chi_{\xi\alpha}(|\mathbf{r}'|)\,d\mathbf{r}' \tag{2}$$

where $\chi_{\xi\alpha}(r)$ is the bulk solvent susceptibility function, and N_{solvent} is the number of sites in a solvent molecule (see Figure 2). The solvent susceptibility function $\chi_{\xi\alpha}(r)$ describes the mutual correlations of sites ξ and α in solvent molecules in the bulk solvent. It can be obtained from the solvent intramolecular correlation function ($\omega^{\text{solv}}_{\xi\alpha}(r)$), site−site radial total correlation functions ($h^{\text{solv}}_{\xi\alpha}(r)$) and the solvent site number density (ρ_α): $\chi_{\xi\alpha}(r) = \omega^{\text{solv}}_{\xi\alpha}(r) + \rho_\alpha h^{\text{solv}}_{\xi\alpha}(r)$ (from here onward, we imply that each site is unique in the molecule, so that $\rho_\alpha = \rho$ for all α).[55] In this work, these functions were obtained by solution of the RISM equations of the pure solvent.[55,57]

In order to calculate $h_\alpha(\mathbf{r})$ and $c_\alpha(\mathbf{r})$, N_{solvent} *closure* relations are introduced:

$$h_\alpha(\mathbf{r}) = \exp(-\beta u_\alpha(\mathbf{r}) + h_\alpha(\mathbf{r}) - c_\alpha(\mathbf{r}) + B_\alpha(\mathbf{r})) - 1 \quad (\alpha = 1, \ldots, N_{\text{solvent}}) \tag{3}$$

where $u_\alpha(\mathbf{r})$ is the 3D interaction potential between the solute molecule and α solvent site, $B_\alpha(\mathbf{r})$ are bridge functionals, $\beta = 1/(k_B T)$ (here, k_B is the Boltzmann constant, and T is the temperature).

In general, the exact bridge functions $B_\alpha(\mathbf{r})$ in eq 3 are represented as an infinite series of integrals over high-order correlation functions and, therefore, are practically incomputable, which makes it necessary to incorporate some approximations.[55,58,59] In the current work, we use a closure relationship proposed by Kovalenko and Hirata (the KH closure),[60] which was designed to improve convergence rates and to prevent possible divergence of the numerical solution of the RISM equations:[60]

$$h_\alpha(\mathbf{r}) = \begin{cases} \exp(\Xi_\alpha(\mathbf{r})) - 1 & (\text{when } \Xi_\alpha(\mathbf{r}) \le 0) \\ \Xi_\alpha(\mathbf{r}) & (\text{when } \Xi_\alpha(\mathbf{r}) > 0) \end{cases} \tag{4}$$

where $\Xi_\alpha(\mathbf{r}) = -\beta u_\alpha(\mathbf{r}) + h_\alpha(\mathbf{r}) - c_\alpha(\mathbf{r})$.

The 3D interaction potential between the solute molecule and α site of solvent ($u_\alpha(\mathbf{r})$, eq 3) is estimated as a superposition of the site−site interaction potentials between solute sites and the particular solvent site, which depend only on the absolute distance between the two sites. We use the common form of the site−site interaction potential represented by the long-range electrostatic interaction term and the short-range term (Lennard-Jones potential).[42]

Within the framework of the RISM theory, there exist several approximate functionals that allow one to obtain values of the hydration free energy (HFE) from the total $h_\alpha(\mathbf{r})$ and direct

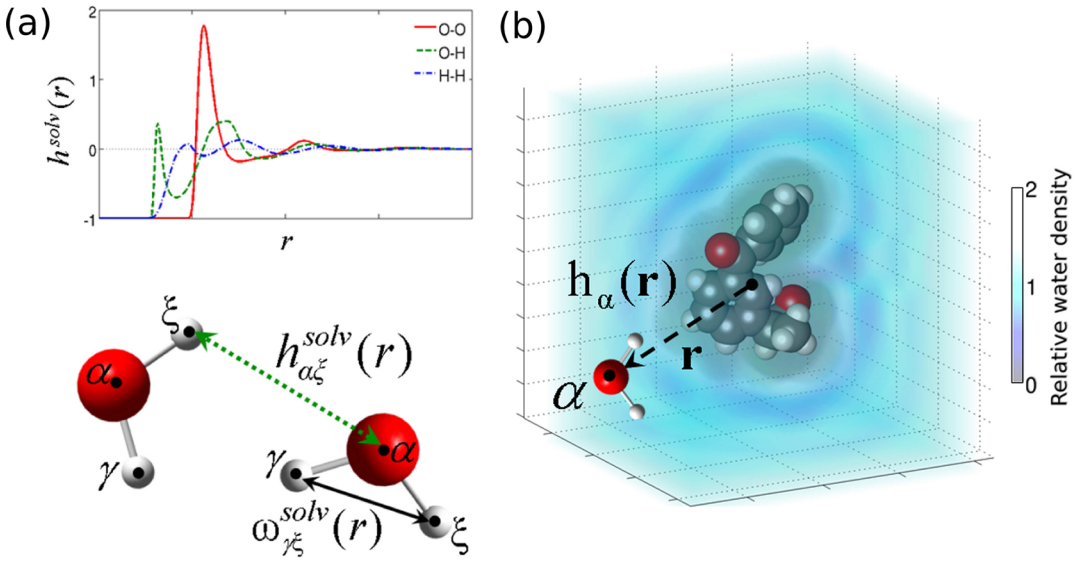

Figure 2. Correlation functions in the 3D-RISM approach. (a) Site−site intramolecular ($\omega^{\text{solv}}_{\gamma\xi}(r)$) and intermolecular ($h^{\text{solv}}_{\alpha\xi}(r)$) correlation functions between sites of solvent molecules. The graph shows the radial projections of water solvent site−site density correlation functions: oxygen−oxygen (OO, solid red curve), oxygen−hydrogen (OH, dashed green curve), and hydrogen−hydrogen (HH, dash-dotted blue line). (b) Three-dimensional (3D) intermolecular solute−solvent correlation function $h_\alpha(\mathbf{r})$ around a model solute (diclofenac). This figure is based in Figure 1 from our earlier work.[43].

dx.doi.org/10.1021/ct300345m | J. Chem. Theory Comput. 2012, 8, 3322−3337

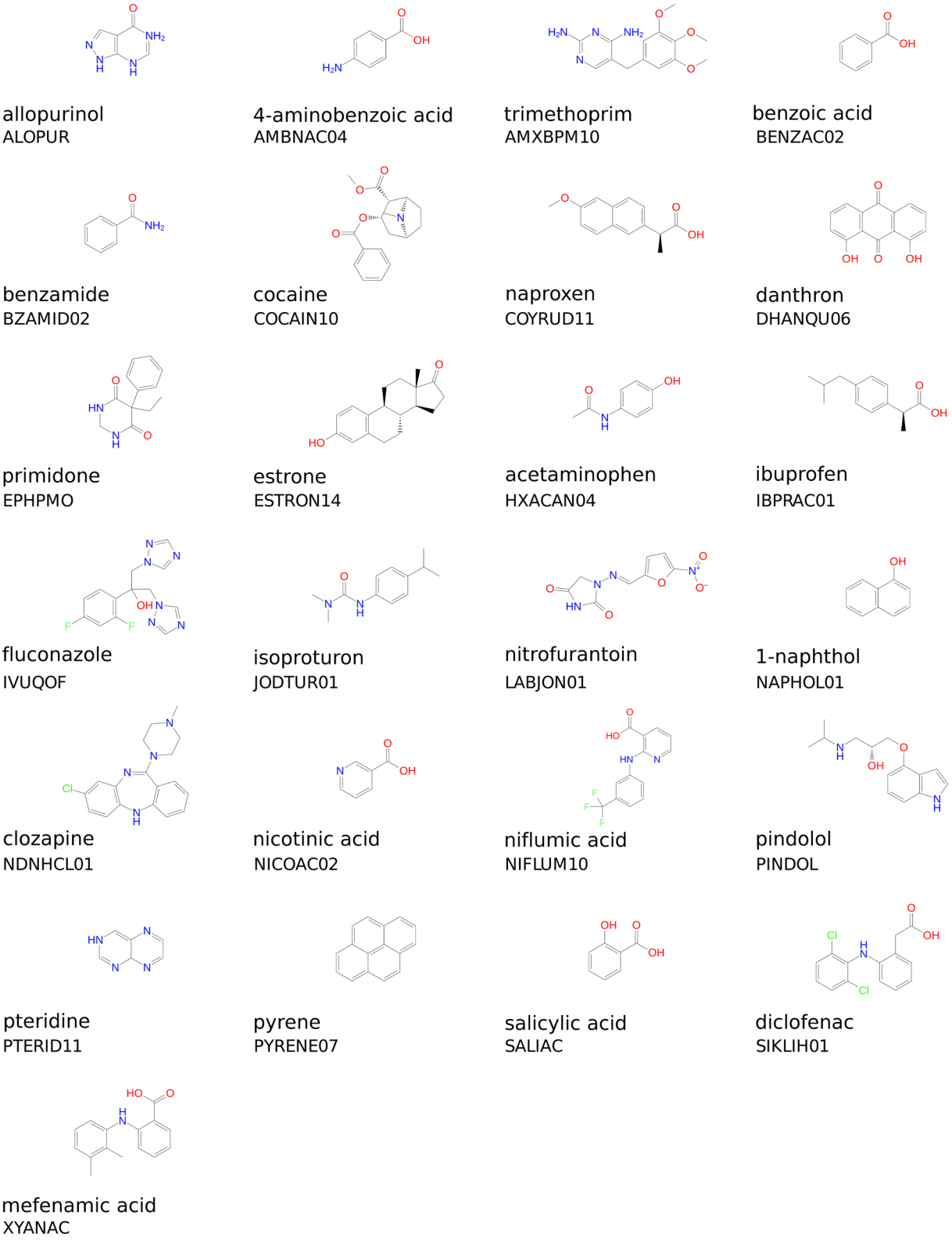

Figure 3. Chemical structures, common names, and selected crystalline polymorphic form (as Cambridge Structural Database refcodes) for the 25 molecules in the dataset.

$c_\alpha(\mathbf{r})$ correlation functions analytically.[24,61,62] Although these functionals have been extensively used to *qualitatively* model thermodynamics of different chemical systems,[24,63,64] they generally give HFE values that are strongly biased from experimental data with a large standard deviation error.[24,42,61,62,65,66] In recent work, some of the current authors (D.S.P. and M.V.F.) developed a new free-energy functional (3D-RISM/UC)[43] that allows the hydration free energies of molecules ranging from simple alkanes to pharmaceuticals to be calculated accurately in the scope of the 3D-RISM.[42−46]

dx.doi.org/10.1021/ct300345m | J. Chem. Theory Comput. 2012, 8, 3322−3337

2.3. 3D-RISM/UC Functional. The Gaussian fluctuations (GF) HFE functional was initially developed by Chandler, Singh, and Richardson, for 1D-RISM,[67] and adapted by Kovalenko and Hirata for the 3D-RISM case:[55]

$$\Delta G_{hyd}^{GF} = k_B T \sum_{\alpha=1}^{N_{solvent}} \rho_\alpha \int_{R^3} \left[-c_\alpha(\mathbf{r}) - \frac{1}{2} c_\alpha(\mathbf{r}) h_\alpha(\mathbf{r}) \right] d\mathbf{r} \quad (5)$$

where ρ_α is the number density of solvent sites α. Unfortunately, HFEs calculated using the GF free-energy functional have only a *qualitative* agreement with experiment. The error in hydration free energies calculated by the GF functional in 3D-RISM is strongly correlated with the partial molar volume calculated by 3D-RISM.[43−45] The 3D-RISM/UC free energy functional developed from this observation is a linear combination of the ΔG_{hyd}^{GF}, the dimensionless partial molar contribution (ρV), and a bias correction (b, the intercept):[43]

$$\Delta G_{hyd}^{3D\text{-}RISM/UC} = \Delta G_{hyd}^{GF} + a(\rho V) + b \quad (6)$$

where the values of the scaling coefficient a and intercept b are obtained by linear regression against experimental data for simple organic molecules. For the combination of methods used here, the coefficients have values of $a = -3.2217$ kcal/mol and $b = 0.5783$ kcal/mol.

We estimate the solute partial molar volume via *solute−solvent site* correlation functions, using the standard 3D-RISM theory expression:[68,69]

$$V = k_B T \eta \left(1 - \rho_\alpha \sum_{\alpha=1}^{N_{solvent}} \int_{R^3} c_\alpha(\mathbf{r})\, d\mathbf{r} \right) \quad (7)$$

where η is the pure solvent isothermal compressibility and ρ_α is the number density of solute sites α.

The 3D-RISM/UC method has been shown to give accurate hydration free energies for both simple organic molecules and bioactive (druglike) molecules.[43−45]

3. METHODS

3.1. Datasets. The dataset used in this work contains 25 druglike molecules with experimental data taken from the published literature. The chemical structures and common names of these molecules are illustrated in Figure 3, along with the Cambridge Structural Database refcode of the polymorph used in the calculations. The experimental sublimation free energy, hydration free energy, and intrinsic aqueous solubility data are given later in Tables 1, 3, and 4, respectively. Solubility data are expressed in terms of log S, using a decadic logarithm with units referred to mol/L, and free energies are given in units of kJ/mol. Sublimation free energy data could only be found in the published literature for four molecules in the dataset. For this reason, the sublimation free-energy calculations were also benchmarked against sublimation free-energy data obtained from experimental intrinsic aqueous solubility and hydration free-energy data using eq 1. The benefit of this approach is that it ensures that both sublimation free energy and solubility are given for the same polymorph, but it may also cause an amplification of the experimental error. The lack of accurate and well-documented experimental thermodynamic data for druglike molecules in the published literature has previously been recognized by other authors as a significant stumbling block in the development of new computational models.[70,71] [By "well-documented", we mean that both the methodology and the experimental conditions must be clearly reported.]

3.2. Calculation of ΔG_{hyd} Using 3D-RISM/UC. *3.2.1. RISM Calculations.* RISM calculations were performed assuming infinitely dilute solution. We used the Lue and Blankschtein version of the SPC/E model of water (MSPC/E).[72] This differs from the original SPC/E water model[73] by the addition of modified Lennard-Jones (LJ) potential parameters for the water hydrogen, which were altered to prevent possible divergence of the algorithm.[74−77] The Lorentz−Berthelot mixing rules were used to generate the solute−water LJ potential parameters.[78] The following LJ parameters (for water hydrogen) were used to calculate the interactions between solute sites and water hydrogens: $\sigma_{H_w}^{LJ} = 1.1657$ Å and $\varepsilon_{H_w}^{LJ} = 0.0155$ kcal/mol.

3.2.2. 3D-RISM Calculations. The 3D-RISM calculations were performed using the NAB simulation package[57] in the AmberTools 1.4 set of routines.[79] The 3D-grid around a solute was generated such that the minimum distance between any solute atom and the edge of the solvent box (*buffer* in NAB notation) was equal to 30 Å. The linear grid spacing in each of the three directions was 0.3 Å. We employed the MDIIS iterative scheme,[80] where we used five MDIIS vectors, an MDIIS step size of 0.7, and a residual tolerance of 10^{-10}. The KH closure was used for solution of the 3D-RISM equations. Solvent susceptibility functions were taken from the 1D RISM calculations.

3.2.3. Solvent Susceptibility Functions. Solvent susceptibility functions were calculated with the 1D RISM method present in AmberTools 1.4. The dielectrically consistent RISM was employed,[81] using the KH closure. The grid size for 1D functions was 0.025 Å, which gave a total of 16 384 grid points. We employed the MDIIS iterative scheme, where we used 20 MDIIS vectors, an MDIIS step size of 0.3, and a residual tolerance of 10^{-12}. The solvent was considered to be pure water with a number density of 0.0333 Å^{-3}, and a dielectric constant of 78.497. The final solvent susceptibility site−site functions were stored and then used as input for the 3D-RISM calculations. The solvent isothermal compressibility evaluated from the 1D RISM calculation was $k_B T\eta = 1.949459$ Å^3.

3.2.4. Input Structures and potential parameters. The following data are needed for 3D-RISM calculations in the NAB simulation package: (1) atomic coordinates, (2) partial charges of atoms, and (3) atom−atom potential parameters representing the van der Waals interactions. These parameters were assigned to each molecule using programs distributed with the AmberTools 1.4 package.[79,82,83]

(1) The coordinates of each molecule were optimized using the AM1 Hamiltonian[84] via the Antechamber[85] suite, which uses the SQM[79] program for semiempirical QM calculations. The input coordinates of each solute were taken from the crystal structure used in the sublimation free-energy calculation.
(2) Atomic partial charges were calculated using the AM1-BCC method,[86−88] as implemented in Antechamber from the AmberTools 1.4 package.[79] The BCC parameters were taken from Jakalian et al.[87]
(3) For all compounds, the GAFF LJ parameters[88] were assigned to solute atoms with the Antechamber and Tleap programs.[85]

The mean time required to calculate the hydration free energy of a single solute using 3D-RISM/UC was ∼45 min,

while the minimum and maximum values were ~30 min and ~75 min, respectively (on a single Intel Core 2 Duo CPU E8600 3.33 GHz machine). The time required for a single calculation could be significantly reduced by using advanced numerical algorithms[89,90] or by performing the simulations using parallel computation.[57]

3.3. Calculation of ΔG_{hyd} Using Implicit Continuum Models. HFEs were calculated using three commonly used continuum solvent models in the scope of quantum mechanics: (1) HF/6-31G(d) PCM - Hartree−Fock theory with the 6-31G(d) basis set and the integral equation formalism of the polarizable continuum model,[91] as implemented in Gaussian03[92] (UAHF atomic radii were used to define the molecular cavity); (2) HF/6-31G(d) SMD - Hartree−Fock theory with the 6-31G(d) basis set and the SMD solvent model,[47] as implemented in Gaussian09;[93] (3) M06−2X/6-31G(d) SMD - the M06−2X density functional[48] with the 6-31G(d) basis set and the SMD solvent model, as implemented in Gaussian09. Ultrafine integral grids (as defined in Gaussian03/09) were used since these are expected to be required for M06−2X calculations.[94] These combinations of QM theory and solvent model were selected because they performed well in a recent blind challenge for HFE calculation.[95−97] We note that these are also the recommended methods for HFE calculation in Gaussian03 and Gaussian09, respectively. The calculations assume infinite dilution of the solute in pure aqueous solvent at 298 K.

3.4. Calculation of ΔG_{sub}. The Gibbs free energy for sublimation was calculated assuming a 1 atm standard state in the gas (denoted by the superscript symbol °). The Gibbs−Helmholtz equation was used to calculate ΔG°_{sub}, where ΔH_{sub} was computed from a calculated lattice energy and ΔS_{sub} was considered to be the difference between the entropy of an ideal gas and the entropy of the crystal at 298 K (where the latter was estimated from the calculated phonon modes of the crystal). ΔG°_{sub} can be converted to ΔG^{*}_{sub} using the following equation, which is derived considering the work for isothermal expansion of an ideal gas:[18,98]

$$\Delta G^{*}_{sub} = \Delta G^{\circ}_{sub} - RT \ln\left(\frac{V_m p_0}{RT}\right) \tag{8}$$

V_m is the molar volume of the crystal and p_0 is standard atmospheric pressure (1 atm = 101.325 kPa). By substituting eq 8 into eq 1, solubility can be defined in terms of ΔG°_{sub} and ΔG^{*}_{hyd}, to eliminate V_m:

$$S = \frac{p_0}{RT} \exp\left(\frac{\Delta G^{\circ}_{sub} + \Delta G^{*}_{hyd}}{-RT}\right) \tag{9}$$

This convention is useful because sublimation free energies are almost exclusively given as ΔG°_{sub} in the literature, while hydration free energies are more commonly given as ΔG^{*}_{hyd}. [We note that converting experimental values of ΔG°_{sub} to ΔG^{*}_{hyd} requires knowledge of V_m, which is not always available if the polymorphic form is not accurately reported.] It is also convenient because, by default, all of the computational methods to calculate sublimation and hydration free energies tested here produce values in these standard states. Therefore, in what follows, thermodynamic data will be tabulated as ΔG°_{sub} and ΔG^{*}_{hyd}, and eq 9 will be used to calculate solubility.

3.4.1. Calculation of ΔH_{sub}. The enthalpy of sublimation ΔH_{sub} can be approximated from the crystal lattice energy, U_{latt}, by $\Delta H^{\circ}_{sub} = -U_{latt} - 2RT$. The "$-2RT$" term arises because the lattice energy does not include lattice vibrational energies (which can be approximated by $6RT$ for crystals of rigid molecules oscillating in a harmonic potential), the energy of the vapor is $3RT$ and a $PV = RT$ correction is necessary to change energies into enthalpies, thus yielding $-6RT + 3RT + RT = -2RT$.[99]

Crystal lattice energies were minimized with DMACRYS.[40] An initial crystalline structure for each solute was selected from the Cambridge Structural Database using the algorithm discussed in Section 3.4.3. Hydrogen atom coordinates were modified to standardize heavy-atom-to-hydrogen bond lengths to standard values obtained from neutron spectroscopy, using routines implemented in DMACRYS. Crystal lattice energies were minimized treating each molecule as a rigid body using the intermolecular potential described below.

The repulsion−dispersion contributions to the intermolecular potential were evaluated as

$$U_{rep-disp} = \sum_{M,N}^{N_{mol}} \left[\sum_{i \in M < k \in N} \left(A_{\iota\kappa} e^{-B_{\iota\kappa} R_{ik}} - \frac{C_{\iota\kappa}}{R_{ik}^{6}} \right) \right] \tag{10}$$

where atoms i and k in molecules M and N are of types ι and κ, respectively, and the parameters $A_{\iota\kappa}$, $B_{\iota\kappa}$, and $C_{\iota\kappa}$ are characteristic of the atom types. The atom−atom potential parameters were taken from Williams and Houpt (C−C, H_c−H_c, N−N, O−O, F−F),[100] Coombes et al. (H_p−H_p),[101] Hsu and Williams (Cl−Cl),[102] and Filippini and Gavezzotti (S−S);[103] here, H_c are hydrogen atoms bonded to carbon and H_p are polar hydrogen atoms (bonded to either oxygen or nitrogen). These atom−atom potentials are often referred to collectively as the FIT potential.[40] Potential parameters for interactions between different atoms were constructed as geometric averages for parameters A and C and arithmetic averages for parameter B. Repulsion−dispersion interactions were evaluated up to a 15 Å cutoff.

Electrostatic contributions to the intermolecular potential were calculated using a distributed multipole representation of the charge distribution,[104] using multipoles up to the hexadecapole, which were computed at three different levels of theory: (i) MP2/6-31G(d,p),[105] (ii) B3LYP/6-31G(d,p),[106,107] and (iii) HF/6-31G(d,p). Ewald summation was used for charge−charge, charge−dipole, and dipole−dipole interactions, while all higher order electrostatic terms (up to R^{-5}) were summed to a 15 Å cutoff between molecular centers of mass.

We note that the intermolecular potentials used in the crystal lattice simulation differ from those used in the 3D-RISM calculations. This is, in part, due to the limitations of the available software, but is also a reflection of the fact that different calculation methods have different sensitivities to the intermolecular potential. Thus, for modeling organic crystals, where an anisotropic description of molecular interactions is required, we have used an intermolecular potential operating with a distributed multipole expansion of the molecular charge distribution. By contrast, for 3D-RISM calculations, which are less sensitive to the intermolecular potential, an atom-based partial charge model has been used to represent the static charge distribution. The selected intermolecular potentials are those recommended by the developers of the various computational methods,[40,43] which have been proven to perform well in previous studies.[45,108]

3.4.2. Calculation of ΔS_{sub}. The molar entropy change for sublimation was calculated as $\Delta S^{o}_{sub} = S_{rot,gas} + S_{trans,gas} - S_{ext,cryst}$, where $S_{rot,gas}$ and $S_{trans,gas}$ are the rotational and translational contributions to the entropy of the gas at 298 K, respectively, and $S_{ext,cryst}$ is the intermolecular vibrational contribution to the entropy of the crystal at 298 K. The change in electronic entropy was assumed to be zero. The intramolecular and intermolecular contributions to the entropy of the crystal were considered to be decoupled, such that the change in intramolecular vibrational entropy for transfer from crystal to gas was taken to be zero. The gain in conformational entropy for flexible molecules was initially set to zero. The use of a correction of between 0.5*R* and 1.5*R* per rotatable bond was also tested.

$S_{rot,gas}$ and $S_{trans,gas}$. The rotational and translational entropies of the gas were calculated from statistical thermodynamics, assuming an ideal gas at 298 K, using the routines implemented in Gaussian09.[93]

$S_{ext,crystal}$. From the Third Law of Thermodynamics, the entropy of all perfect crystalline substances is zero at $T = 0$ K. At 298 K, it is necessary to consider intermolecular and intramolecular vibrations. Translational and rotational entropies are assumed to be negligible, and the crystal lattice is considered to be infinite and perfect. The vibrational terms arise from the intramolecular vibrations and from the phonon modes of the crystal. The latter were calculated using the rigid molecule lattice dynamics implemented in DMACRYS, with the same model potential used for the lattice energy minimizations. Only the $6N - 3$ (where N is the number of molecules in the unit cell) optical zone-center ($k = 0$) phonons were calculated; the remaining three acoustic modes have zero frequency at $k = 0$. The density of states was calculated using a hybrid Debye−Einstein approximation for $k \neq 0$, where the frequencies of the optical phonons were assumed to be independent of k and the acoustic contribution was modeled by the Debye approximation, with the Debye cutoff frequency estimated by extrapolating the acoustic modes to the zone boundary, using sound velocities calculated from the elastic stiffness tensor. The resulting free energy expression is given in Anghel et al.[109] In these calculations, it is assumed that vibrations are harmonic and coupling between intermolecular and intramolecular vibrations is ignored.

3.4.3. Selection of a Crystal Polymorph. For the lattice energy calculations, a single-crystal structure for each solute was selected for analysis from the Cambridge Structural Database, using the following algorithm:

(1) Extract from the Cambridge Structural Database all entries for the given compound that have 3D coordinates with only one type of molecule in the unit cell from the Cambridge Structural Database (i.e., no salts, solvates, co-crystals, etc.)
(2) Calculate the lattice energy for each entry
(3) Select the crystal structure that has the lowest calculated lattice energy

The majority of experimental data in the published literature is reported without characterization of the crystalline polymorphic form that is observed at thermodynamic equilibrium in the solubility experiment, which makes it difficult to compile an accurate database of polymorph solubility. For the eight molecules in the dataset for which polymorph information was available, the crystalline form used in the calculations was the same as that observed in experiment. This is not unexpected, since both the experimental and computational methodologies will, on average, select more thermodynamically stable polymorphs. In the case of the computations, it is clear that the algorithm discussed above will explicitly select the most thermodynamically stable polymorph (as defined by the model potential). In the case of the experiments, the repeated dissolution and reprecipitation of the solute that occurs during a single solubility measurement often promotes changes in crystal polymorph, from less to more thermodynamically stable forms in accordance with Ostwald's law of stages.[110−112] This is especially true for solubilities measured by the CheqSol method.[111] For those molecules in the dataset for which the solubility data are reported without characterization of the polymorphic form of the precipitate, it is not possible to assess whether the polymorphic form used in the computations is the same as that used in the experiment.

3.5. Statistical Analysis. To compare calculated and experimental results for different computational models, a correlation coefficient (R) and the root mean square deviation (RMSD) were evaluated:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y^i - y^i_{exp})^2}{\sum_{i=1}^{n} (y^i_{exp} - M(y^i_{exp}))^2} \tag{11}$$

$$\mathrm{RSMD}(y, y_{exp}) = \sqrt{\frac{1}{N} \sum_i (y^i - y^i_{exp})^2} \tag{12}$$

where index i runs through the set of N selected molecules, and y^i and y^i_{exp} are the calculated and the experimental values, respectively, for molecule i for a given property (i.e., ΔG_{sub}, ΔG_{hyd}, or $\log_{10} S$). The total deviation can be split into two parts: bias (or mean displacement, M) and standard deviation (σ), which are calculated by the formulas

$$\mathrm{bias} = M(y - y_{exp}) = \frac{1}{N} \sum_{i \in S} (y^i - y^i_{exp}) \tag{13}$$

$$\sigma(y - y_{exp}) = \sqrt{\frac{1}{N} \sum_{i \in S} [y^{(i)} - y^{(i)}_{exp} - M(y - y_{exp})]^2} \tag{14}$$

The bias gives the systematic error, which can be corrected by a simple constant term. The standard deviation gives the random error that is not explained by the model. One can see the connection between these three formulas:

$$\mathrm{RSMD}(y, y_{exp})^2 = M(y - y_{exp})^2 + \sigma(y - y_{exp})^2 \tag{15}$$

Models reporting RMSE values greater than the standard deviation of the experimental data (1.79 log S units) offer less-accurate predictions of solubility than the null model provided by the mean of the experimental data. Here, the standard error of the RMSE is estimated using 1000 bootstrap samples of 25 molecules taken with replacement from the 25 molecules in the original dataset.

The statistics defined previously give measures of the prediction error for the complete dataset. To further validate our results, the predicted solubility data is analyzed in terms of three different categories, some of which have previously been adopted by other authors:[16] (i) accurate predictions, which involve molecules whose solubilities are calculated with an absolute error of <0.5 log S units; (ii) satisfactory predictions, which involve molecules whose solubilities are calculated with

an absolute error of <1 log S unit; (iii) outliers, which involve molecules for which the absolute error in the calculated solubility is more than twice the SD of the experimental data ($1.79 \times 2 = 3.58$ log S, referred to units of mol/L).

Statistical analyses were carried out in the R Statistical Computing Environment.[113] Python scripts were used to manipulate raw data files.

The calculations discussed in this paper were performed on different computing clusters, at the EaStCHEM Research Computing Facility at the University of St. Andrews (by J.L.McD.) and at the Max Planck Institute (MPI) for Mathematics in the Sciences (by D.S.P.).

4. RESULTS

The objective of the current work is to assess how accurately the intrinsic aqueous solubility of crystalline druglike molecules can be computed from calculated sublimation and hydration free energies based on a thermodynamic cycle via the vapor (eq 1). We begin by comparing the calculated and experimental data for sublimation and hydration free energies.

4.1. Sublimation Free Energy. The model potential used to calculate sublimation free energies comprises two terms: a repulsion−dispersion term and an electrostatic term. The repulsion−dispersion term was evaluated using the Buckingham potential and empirical parameters obtained from the FIT potential. The electrostatic term was calculated using a distributed multipole representation of the charge distribution using multipoles up to the hexadecapole, which were computed at three different levels of theory: (i) MP2/6-31G(d,p),[105] (ii) B3LYP/6-31G(d,p),[106,107] and (iii) HF/6-31G(d,p). The sublimation free energies calculated by these methods will be referred to as ΔG_{sub}^{MP2}, ΔG_{sub}^{B3LYP} and ΔG_{sub}^{HF}, respectively, with ΔG_{sub}^{exp} used to refer to the experimental sublimation free-energy data.

A significantly better correlation was observed between ΔG_{sub}^{exp} and either ΔG_{sub}^{MP2} or ΔG_{sub}^{B3LYP}, than between ΔG_{sub}^{exp} and ΔG_{sub}^{HF}. The statistics reported in Table 1 indicate that both ΔG_{sub}^{MP2} and ΔG_{sub}^{B3LYP} explain much of the variance in the experimental sublimation free-energy data ($R_{MP2} = 0.87$ and $R_{B3LYP} = 0.87$) without a significant systematic error ($bias_{MP2} = 0.17$ kJ/mol and $bias_{B3LYP} = 0.71$ kJ/mol). Both models provided a better estimate of the data than the mean of the experimental data, since the root-mean-square errors ($RMSE_{MP2} = 5.63$ kJ/mol and $RMSE_{B3LYP} = 5.66$ kJ/mol) were lower than the standard deviation of the experimental sublimation free-energy data ($\sigma = 10.34$ kJ/mol). By contrast, ΔG_{sub}^{HF} does not provide a good estimate of ΔG_{sub}^{exp} ($R_{HF} = 0.82$, $RMSE_{HF} = 11.64$ kJ/mol, $bias_{B3LYP} = -9.30$ kJ/mol). The correlation between ΔG_{sub}^{exp} and ΔG_{sub}^{B3LYP} is plotted in Figure 4 (similar graphs for ΔG_{sub}^{MP2} and ΔG_{sub}^{HF} are provided in the Supporting Information). The large outlier in this graph is ibuprofen, for which $\Delta G_{sub}^{exp} = 42.06$ kJ/mol and $\Delta G_{sub}^{B3LYP} = 54.60$ kJ/mol.

Table 1. Sublimation Free Energies at 298 K from Experiment and Calculated Using the FIT Potential Parameters and Distributed Multipoles Evaluated at Different Levels of Theory Using the 6-31G(d,p) Basis Set

		ΔG_{sub}^{calc} (kJ/mol)		
molecule	ΔG_{sub}^{exp} (kJ/mol)	(MP2)	(B3LYP)	(HF)
BENZAC02	34.23[114,115]	35.08	34.59	48.91
BZAMID02	43.14[114,116]	36.41	36.98	48.76
COCAIN10	54.90[114,115]	56.23	56.55	61.07
COYRUD11	61.06[45,41]	65.62	64.84	77.84
HXACAN04	59.95[45,41]	54.91	53.93	69.13
IBPRAC01	42.06[45,115]	54.89	54.60	65.80
JODTUR01	59.45[114,116]	60.15	59.75	68.22
NAPHOL01	35.38[114,41]	35.82	33.23	39.73
PYRENE07	46.25[114,116]	41.88	41.77	43.81
SALIAC	40.31[114,41]	34.04	33.40	42.41
R		0.87	0.87	0.82
RMSE		5.63	5.66	11.64
σ		5.63	5.62	7.00
bias		0.17	0.71	−9.30

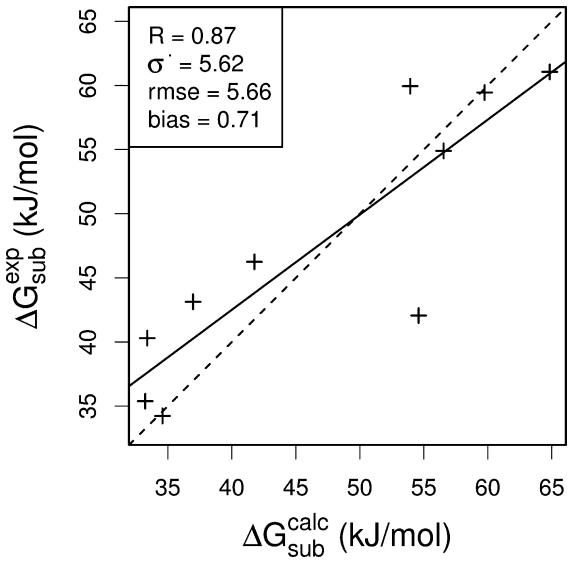

Figure 4. Correlation between experimental and calculated sublimation free energy, where the calculations were performed using the FIT potential parameters and distributed multipoles evaluated at the B3LYP/6-31G(d,p) level of theory. The solid black line indicates the line of best fit through the data. The dashed line is the identity line.

The sublimation free-energy data calculated by the three model potentials may be further validated against experimental intrinsic aqueous solubility data by using the experimental hydration free-energy data to complete the thermodynamic cycle. Solubility is predicted more accurately when MP2/6-31G(d,p) or B3LYP/6-31G(d,p) multipoles are used in the model potential ($R_{MP2} = 0.90$, $RMSE_{MP2} = 0.99$ log S units, $\sigma_{MP2} = 0.99$ log S units and $bias_{MP2} = -0.03$ log S units; $R_{B3LYP} = 0.90$, $RMSE_{B3LYP} = 0.99$ log S units, $\sigma_{B3LYP} = 0.99$ log S units and $bias_{B3LYP} = -0.12$ log S units) than when HF/6-31G(d,p) multipoles are used ($R_{HF} = 0.78$, $RMSE_{HF} = 2.04$ log S units, $\sigma_{HF} = 1.23$ log S units and $bias_{HF} = 1.63$ log S units). For the calculations using HF/6-31G(d,p) multipoles, the RMSE value is larger than the SD of the experimental solubility data (1.79 log S units, referred to units of mol/L), which indicates that the model gives less-accurate predictions than the mean of the experimental data. The correlation between experimental and calculated solubility obtained using multipoles evaluated at the B3LYP/6-31G(d,p) level of theory is plotted in Figure 5 (similar graphs for ΔG_{sub}^{MP2} and ΔG_{sub}^{HF} are provided in the Supporting Information). The sublimation thermodynamic data calculated using the FIT potential parameters and distributed

dx.doi.org/10.1021/ct300345m | J. Chem. Theory Comput. 2012, 8, 3322−3337

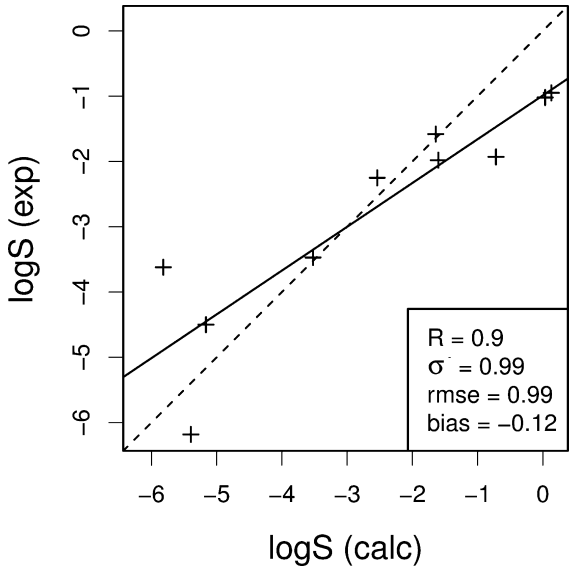

Figure 5. Correlation between experimental and calculated solubility, where the latter was computed from experimental hydration free energies and calculated sublimation free energies, and the calculations were performed using the FIT potential parameters and distributed multipoles evaluated at the B3LYP/6-31G(d,p) level of theory. The solid black line indicates the line of best fit through the data. The dashed line is the identity line.

multipoles evaluated at the B3LYP/6-31G(d,p) level of theory is provided in Table 2.

The results show that distributed multipoles calculated at the MP2/6-31G(d,p) or B3LYP/6-31G(d,p) level of theory provide a more accurate description of the electrostatic interaction energy in the crystal than do multipoles calculated at the HF/6-31G(d,p) level. Furthermore, the observed correlation between ΔG_{sub}^{MP2} and ΔG_{sub}^{B3LYP} (R = 0.99, σ = 0.86 kJ/mol, bias = 0.54 kJ/mol) suggests that the density functional method is a useful alternative to MP2 for the molecules and calculations considered here.

4.2. Hydration Free Energy. Four different methods were used to calculate hydration free energies: (1) 3D-RISM/UC; (2) HF/6-31G(d) PCM - Hartree−Fock theory with the 6-31G(d) basis set and the polarizable continuum model; (3) HF/6-31G(d) SMD - Hartree−Fock theory with the 6-31G(d) basis set and the SMD solvent model, as implemented in Gaussian09; and (4) M06-2X/6-31G(d) SMD - the M06-2X density functional with the 6-31G(d) basis set and the SMD solvent model. The hydration free energies calculated using these methods will be referred to as $\Delta G_{hyd}^{3D\text{-}RISM/UC}$, ΔG_{hyd}^{PCM}, $\Delta G_{hyd}^{SMD(HF)}$, and $\Delta G_{hyd}^{SMD(M06\text{-}2X)}$. The calculated hydration free-energy data are presented in Table 3 and are plotted in Figure 6. It is clear that, for this dataset, the 3D-RISM/UC and HF/6-31G(d) SMD methods provide significantly more accurate estimates of hydration free energies than the other methods tested here. The correlation statistics for the 3D-RISM/UC method are R = 0.93, σ = 4.49 kJ/mol, RMSE = 4.85 kJ/mol, and bias = 1.82 kJ/mol, whereas for the HF/6-31G(d) SMD method, the statistics are R = 0.97, σ = 2.81 kJ/mol, RMSE = 2.91, and bias = −0.72 kJ/mol. The errors in prediction using SMD are lower than those previously reported by other authors for similar molecules,[47,95,117] which may indicate that a small cancellation of errors has occurred for the molecules considered here. The RMSE values for both of the 3D-RISM/UC and SMD methods are below the value of 5.7 kJ/mol, which would equate to a 1 $\log_{10}$ unit error in the related equilibrium property

Table 2. Sublimation Thermodynamics (n = 25) Calculated Using the FIT Potential Parameters and Distributed Multipoles Evaluated at the B3LYP/6-31G(d,p) Level of Theory

molecule	U_{latt} (kJ/mol)	ΔH_{sub} (kJ/mol)	S_{trans} (J/(mol K))	S_{rot} (J/(mol K))	S_{ext} (J/(mol K))	ΔS_{sub} (J/(mol K))	ΔG_{sub} (kJ/mol)
ALOPUR	−129.58	124.62	170.02	120.90	95.30	195.61	66.33
AMBNAC04	−115.42	110.46	170.11	123.11	89.11	204.11	49.64
AMXBPM10	−176.85	171.89	179.46	144.23	99.83	223.87	105.18
BENZAC02	−95.68	90.72	168.67	119.80	100.10	188.36	34.59
BZAMID02	−98.09	93.13	168.56	119.90	100.02	188.44	36.98
COCAIN10	−124.18	119.23	180.01	144.39	114.09	210.32	56.55
COYRUD11	−131.73	126.78	176.57	138.31	107.05	207.83	64.84
DHANQU06	−127.69	122.73	177.10	137.08	111.52	202.66	62.34
EPHPMO	−136.95	132.00	175.91	133.82	106.57	203.16	71.46
ESTRON14	−142.34	137.39	178.58	141.18	107.77	211.99	74.21
HXACAN04	−118.76	113.80	171.33	126.29	96.72	200.89	53.93
IBPRAC01	−121.14	116.19	175.20	136.11	104.64	206.68	54.60
IVUQOF	−155.62	150.67	180.13	144.13	107.11	217.15	85.95
JODTUR01	−125.78	120.83	175.20	135.93	106.17	204.96	59.75
LABJON01	−158.23	153.26	177.00	139.31	107.26	209.05	90.97
NAPHOL01	−95.41	90.45	170.73	124.12	102.84	192.01	33.23
NDNHCL01	−142.56	137.61	180.92	147.57	115.62	212.87	74.17
NICOAC02	−102.07	97.12	168.77	119.57	93.97	194.37	39.20
NIFLUM10	−142.58	137.62	179.11	142.94	124.28	197.77	78.69
PINDOL	−156.18	151.23	177.51	142.14	97.92	221.74	85.15
PTERID11	−83.93	78.97	169.65	119.86	118.12	171.39	27.90
PYRENE07	−104.77	99.81	174.95	132.69	112.88	194.77	41.77
SALIAC	−96.16	91.21	170.20	122.57	98.77	194.00	33.40
SIKLIH01	−137.15	132.19	179.67	142.37	112.19	209.85	69.66
XYANAC	−137.05	132.09	177.15	139.27	104.08	212.35	68.81

Table 3. Hydration Free Energies (n = 25) from Experiment and Calculated Using Four Different Computational Methods: (1) 3D-RISM/UC; (2) HF/6-31G(d) PCM - Hartree−Fock Theory with the 6-31G(d) Basis Set and the Polarizable Continuum Model; (3) HF/6-31G(d) SMD - Hartree−Fock Theory with the 6-31G(d) Basis Set and the SMD Solvent Model; and (4) M06-2X/6-31G(d) SMD - the M06-2X Density Functional with the 6-31G(d) Basis Set and the SMD Solvent Model

molecule	ΔG_{hyd}^{exp} (kJ/mol)	$\rho V^{3D\text{-}RISM}$[a]	$\Delta G_{hyd}^{3D\text{-}RISM/UC}$ (kJ/mol)	$\Delta G_{hyd}^{SMD(M06\text{-}2X)}$ (kJ/mol)	$\Delta G_{hyd}^{SMD(HF)}$ (kJ/mol)	ΔG_{hyd}^{PCM} (kJ/mol)
ALOPUR		4.36	−73.88	−63.96	−76.63	−68.32
AMBNAC04		5.21	−56.29	−46.01	−51.70	−44.18
AMXBPM10		11.21	−92.41	−69.63	−79.84	−53.47
BENZAC02	−33.14[114]	4.80	−36.67	−25.10	−31.66	−27.95
BZAMID02	−45.64[114]	5.15	−49.86	−37.66	−44.33	−35.73
COCAIN10	−49.98[114]	11.83	−59.30	−40.59	−52.52	−27.87
COYRUD11	−43.30[45]	9.23	−44.44	−36.04	−45.42	−36.61
DHANQU06		8.18	−39.47	−26.68	−38.26	−21.42
EPHPMO		8.40	−70.40	−63.67	−73.65	−64.39
ESTRON14		11.66	−42.51	−42.78	−52.89	−47.40
HXACAN04	−62.05[45]	5.97	−60.31	−55.53	−63.63	−53.35
IBPRAC01	−29.33[45]	9.83	−31.74	−23.02	−30.56	−18.24
IVUQOF		10.63	−70.75	−78.77	−93.96	−52.13
JODTUR01	−47.58[114]	9.07	−48.01	−33.09	−40.51	−28.79
LABJON01		7.55	−92.77	−75.46	−100.63	−96.48
NAPHOL01	−32.01[114]	5.81	−24.70	−28.15	−30.64	−28.24
NDNHCL01		12.21	−49.05	−53.04	−55.02	−36.99
NICOAC02		4.65	−43.74	−35.98	−44.28	−38.16
NIFLUM10		9.60	−65.73	−25.72	−34.33	−22.43
PINDOL		10.31	−67.78	−52.92	−57.67	−45.73
PTERID11		4.74	−52.06	−55.90	−64.98	−37.49
PYRENE07	−18.91[114]	8.01	−26.18	−15.72	−19.15	−11.30
SALIAC	−37.21[114]	5.02	−36.13	−27.12	−33.49	−29.62
SIKLIH01		10.31	−41.77	−33.80	−42.77	−19.79
XYANAC		9.76	−35.80	−23.63	−26.65	−16.65
R			0.93	0.97	0.97	0.88
RMSE			4.85	8.3	2.91	11.58
σ			4.49	3.06	2.81	5.58
bias			1.82	−7.71	−0.72	−10.15

[a]The term $\rho V^{3D\text{-}RISM}$ is the unitless partial molar volume of the solute in infinitely dilute solution as calculated by 3D-RISM theory.

(e.g., intrinsic aqueous solubility). The RMSE values for the other methods are all above 8 kJ/mol.

4.3. Solubility. Twelve different predictions of intrinsic aqueous solubility can be made from the calculated thermodynamic data by considering a thermodynamic cycle via the vapor (three different methods for calculating sublimation free energies, combined with four different methods for calculating hydration free energies).

Since the intrinsic solubility is clearly defined in terms of the sublimation and hydration free energies (see eq 1), it is reasonable to assume that the accuracy of the calculated solubilities can be inferred directly for different methods from the accuracy with which these methods calculate sublimation and hydration free energies, as discussed in Sections 4.1 and 4.2. However, some caution must be exercised in making these inferences. First, the experimental errors associated with measurements of sublimation and hydration free energies are significantly larger than those with intrinsic aqueous solubilities (because, for druglike molecules, partial vapor pressures near room temperature are more difficult to measure than saturated solution concentrations). Second, experimental sublimation and hydration free energy data are only available for lower-molecular-weight, less-druglike molecules (again due to the problems of measuring partial vapor pressures for druglike molecules near room temperature). Thus, the 10 molecules considered in Sections 4.1 and 4.2 are a biased sample of the chemical space represented by the full data set of 25 molecules. Third, since the 10-molecule data set is freely available in the published literature, it may have been used during parametrization of some of the implicit continuum methods to calculate hydration free energies (it was *not* used in the development of the 3D-RISM/UC free-energy functional).

Despite these caveats, the majority of the trends observed for the calculation of sublimation and hydration free energies in Sections 4.1 and 4.2 hold true for the calculation of solubility. For example, the predictions of solubility made using ΔG_{sub}^{MP2} or ΔG_{sub}^{B3LYP} are significantly more accurate than those made using ΔG_{sub}^{HF}, regardless of which estimate of ΔG_{hyd} is used. For each of the four different estimates of hydration free energy considered here, ΔG_{sub}^{B3LYP} gives slightly more-accurate predictions of log S than ΔG_{sub}^{MP2}, but only by relatively small margins (ΔRMSE $\leq$ 0.2, referred to units of mol/L).

For the complete dataset of 25 molecules, the most-accurate predictions of solubility were made using $\Delta G_{hyd}^{3D\text{-}RISM/UC}$ and ΔG_{sub}^{B3LYP}, where R = 0.85, RMSE = 1.45 log S, σ = 1.43 log S, and bias = −0.23 log S (units referred to mol/L). The solubilities of 5 molecules were predicted with absolute errors <0.5 log S units, while 12 molecules were predicted with absolute errors <1 log S units; there were no outliers (based on the categories defined earlier). It is perhaps surprising that the

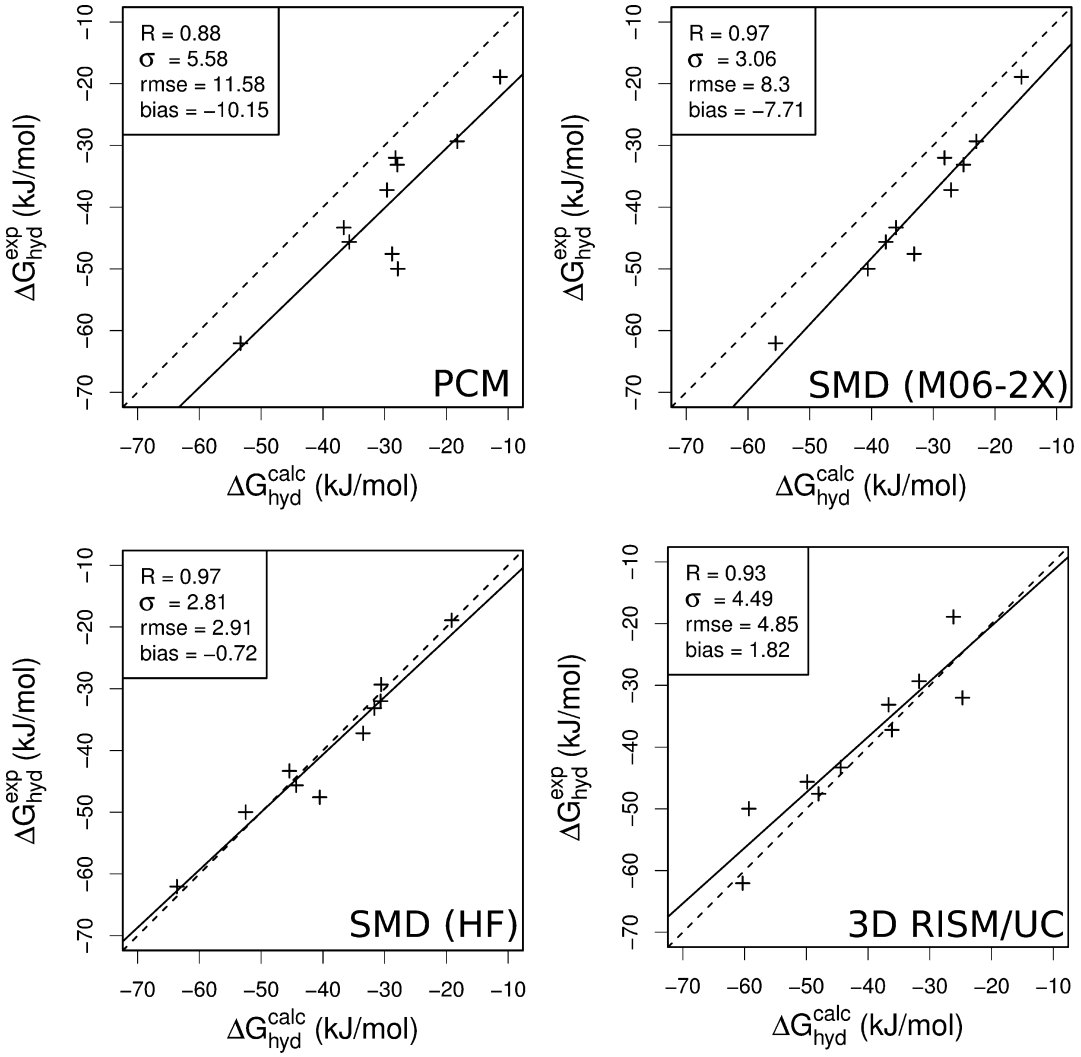

Figure 6. Calculation of hydration free energy made using four different solvation models (PCM, SMD (M06-2X), SMD (HF), and 3D-RISM/UC), plotted against experimental data for 10 solutes. The solid black line indicates the line of best fit through the data. The dashed line is the identity line.

most-accurate prediction of solubility was not obtained using $\Delta G_{hyd}^{SMD(HF)}$, since this method gave the most-accurate estimates of hydration free energies for the 10 molecules for which experimental data were available (RMSE = 2.91 kJ/mol, compared to RMSE = 4.85 kJ/mol for $\Delta G_{hyd}^{3D\text{-}RISM/UC}$; see Figure 6). The accuracy of the predictions of solubility obtained with $\Delta G_{hyd}^{SMD(HF)}$ (RMSE = 2.03 log S, referred to units of mol/L) is relatively poor, which is partially due to two large outliers: NIFLUM10 (Δ(log S) = 4.58) and PTERID11 (Δ(log S) = −5.09). NIFLUM contains three fluorine (F) atoms, which may be a contributing factor to the prediction errors (IVUQOF is the only other molecular crystal that contains F atoms). (See data listed in Tables 4 and 5, as well as that plotted in Figure 7.)

5. DISCUSSION

Prediction of the solubility of bioactive molecules is of great importance in the biochemical sciences, because solubility is a key physicochemical property in estimating the bioavailability of novel pharmaceuticals and the environmental fate of potential pollutants. The objective of the current work has been to assess how accurately the intrinsic aqueous solubility of crystalline druglike molecules can be calculated using existing computational methods. Since the solution free energy is not amenable to direct computation, it has been decomposed into sublimation and hydration free energies by a thermodynamic cycle via the vapor. Sublimation free energies have been calculated using model potential-based crystal lattice simulations (in DMACRYS), while hydration free energies have been computed using both implicit continuum solvent approaches and molecular integral equation theory. It should be noted that other decompositions of the solution free energy are possible.

The results presented here support many conclusions. First, distributed multipoles calculated at the MP2/6-31G(d,p) or B3LYP/6-31G(d,p) level of theory provide a significantly more-accurate description of the electrostatic interaction energy in the crystal than do multipoles calculated at the HF/6-31G(d,p) level. The difference between the sublimation free energies (and solubilities) obtained using the MP2 or B3LYP multipoles is small, which suggests that B3LYP can be used as a less computationally expensive alternative to the MP2 calculations.

The most-accurate calculations of intrinsic aqueous solubility were obtained using the 3D-RISM/UC method of molecular integral equation theory. When the hydration free energies calculated by 3D-RISM/UC were combined with sublimation free energies computed using multipoles evaluated at the

Table 4. Prediction of Solubility (log *S*) for a Dataset of 25 Molecules Using Sublimation Free Energies Calculated Using the FIT Potential Parameters and Distributed Multipoles Evaluated at the B3LYP/6-31G(d,p) Level of Theory, Combined with Hydration Free Energies Calculated by Different Methods[a]

	Experiment		3D-RISM/UC		SMD M06-2X		SMD-HF		PCM	
molecule	log *S* (Exp.)	polymorph (Exp.)	log *S* (3D-RISM/UC)	error	log *S* (SMD M06-2X)	error	log *S* (SMD-HF)	error	log *S* (PCM)	error
NDNHCL01	−3.24[16]	NDNHCL01	−5.79	2.55	−5.09	1.85	−4.75	1.51	−7.91	4.67
IVUQOF	−1.80[115]		−4.05	2.25	−2.65	0.85	0.02	−1.82	−7.32	5.52
IBPRAC01	−3.62[115]		−5.40	1.78	−6.92	3.30	−5.60	1.98	−7.76	4.14
ESTRON14	−5.32[118]		−6.94	1.62	−6.90	1.58	−5.13	−0.19	−6.09	0.77
NAPHOL01	−1.98[41]	NAPHOL01	−2.88	0.90	−2.28	0.30	−1.84	−0.14	−2.26	0.28
SIKLIH01	−5.46[41]	SIKLIH01	−6.28	0.82	−7.67	2.21	−6.10	0.64	−10.13	4.67
AMXBPM10	−2.95[41]	AMXBPM10	−3.63	0.68	−7.62	4.67	−5.83	2.88	−10.45	7.50
PINDOL	−3.79[41]		−4.43	0.64	−7.04	3.25	−6.21	2.42	−8.30	4.51
COYRUD11	−4.50[41]	COYRUD11	−4.96	0.46	−6.44	1.94	−4.79	0.29	−6.34	1.84
XYANAC	−6.74[41]	XYANAC	−7.17	0.43	−9.31	2.57	−8.78	2.04	−10.53	3.79
DHANQU06	−5.19[116]		−5.40	0.21	−7.64	2.45	−5.61	0.42	−8.56	3.37
JODTUR01	−3.47[116]		−3.45	−0.02	−6.06	2.59	−4.76	1.29	−6.82	3.35
NICOAC02	−0.85[116]		−0.59	−0.26	−1.95	1.10	−0.50	−0.35	−1.57	0.72
BENZAC02	−1.58[115]		−1.02	−0.56	−3.05	1.47	−1.90	0.32	−2.55	0.97
HXACAN04	−1.02[41]	HXACAN04	−0.27	−0.75	−1.11	0.09	0.31	−1.33	−1.49	0.47
NIFLUM10	−4.58[41]		−3.66	−0.92	−10.67	6.09	−9.16	4.58	−11.25	6.67
SALIAC	−1.93[41]	SALIAC	−0.91	−1.02	−2.49	0.56	−1.37	−0.56	−2.05	0.12
EPHPMO	−2.64[116]		−1.57	−1.07	−2.75	0.11	−1.00	−1.64	−2.63	−0.01
AMBNAC04	−1.37[116]		−0.22	−1.15	−2.02	0.65	−1.03	−0.34	−2.34	0.97
COCAIN10	−2.25[115]		−0.91	−1.34	−4.19	1.94	−2.10	−0.15	−6.42	4.17
BZAMID02	−0.95[116]		0.87	−1.82	−1.27	0.32	−0.10	−0.85	−1.61	0.66
PYRENE07	−6.18[116]		−4.12	−2.06	−5.95	−0.23	−5.35	−0.83	−6.73	0.55
LABJON01	−3.26[41]		−1.07	−2.19	−4.11	0.85	0.31	−3.57	−0.42	−2.84
ALOPUR	−2.26[115]		−0.06	−2.20	−1.80	−0.46	0.42	−2.68	−1.04	−1.22
PTERID11	0.02[41]		2.85	−2.83	3.52	−3.50	5.11	−5.09	0.29	−0.27

[a]Solubilities are expressed in terms of $\log_{10} S$ with units referred to mol/L. The polymorphic form of the undissolved precipitate, as observed at thermodynamic equilibrium during the solubility experiment, is given in the third column for molecules for which it was reported.

Table 5. Calculation of Intrinsic Aqueous Solubility (S) for a Dataset of 25 Molecules (See Figure 3) Using Sublimation and Hydration Free Energies Calculated by Different Methods[a]

$\lvert\Delta G_{hyd}$	ΔG_{sub}	*R*	RMSE	σ	bias	|error| < 0.5	|error| < 1	outliers
3D-RISM/UC	MP2	0.81	1.58	1.58	−0.05	5 (20%)	10 (40%)	0 (0%)
	B3LYP	0.85	1.45	1.43	−0.23	5 (20%)	12 (48%)	0 (0%)
	HF	0.75	2.51	1.64	1.90	1 (4%)	2 (8%)	5 (20%)
SMD (HF)	MP2	0.81	2.14	2.13	0.14	8 (32%)	12 (48%)	4 (16%)
	B3LYP	0.84	2.03	2.03	−0.05	8 (32%)	12 (48%)	2 (8%)
	HF	0.75	3.02	2.19	2.08	2 (8%)	5 (20%)	6 (24%)
SMD (M06-2X)	MP2	0.84	2.49	1.87	1.65	6 (24%)	8 (32%)	3 (12%)
	B3LYP	0.86	2.33	1.82	1.46	6 (24%)	10 (40%)	2 (8%)
	HF	0.74	4.2	2.17	3.59	1 (4%)	1 (4%)	13 (52%)
PCM (HF)	MP2	0.71	3.57	2.65	2.40	3 (12%)	9 (36%)	9 (36%)
	B3LYP	0.74	3.37	2.54	2.21	5 (20%)	11 (44%)	9 (36%)
	HF	0.65	5.11	2.69	4.35	0 (0%)	2 (8%)	13 (52%)

[a]Solubilities are expressed as $\log_{10} S$ with units referred to mol/L.

B3LYP/6-31G(d,p) level of theory, the intrinsic aqueous solubilities of 25 druglike molecules could be calculated with RMSE = 1.45 log *S* units. The fact that the RMSE value is lower than the standard deviation of the experimental solubility data (σ = 1.79 log *S* units) indicates that the method provides useful predictions (better than the null prediction provided by the mean of the experimental data). Although this level of accuracy is worse than those reported by QSPR models in a recent blind challenge to predict solubility,[16,41] there is clearly great scope to develop the methods presented here. From one side, the computational methods used to calculate sublimation and hydration free energies might be systematically improved, for example, by considering more-accurate intermolecular potentials, more-advanced models from molecular integral equation

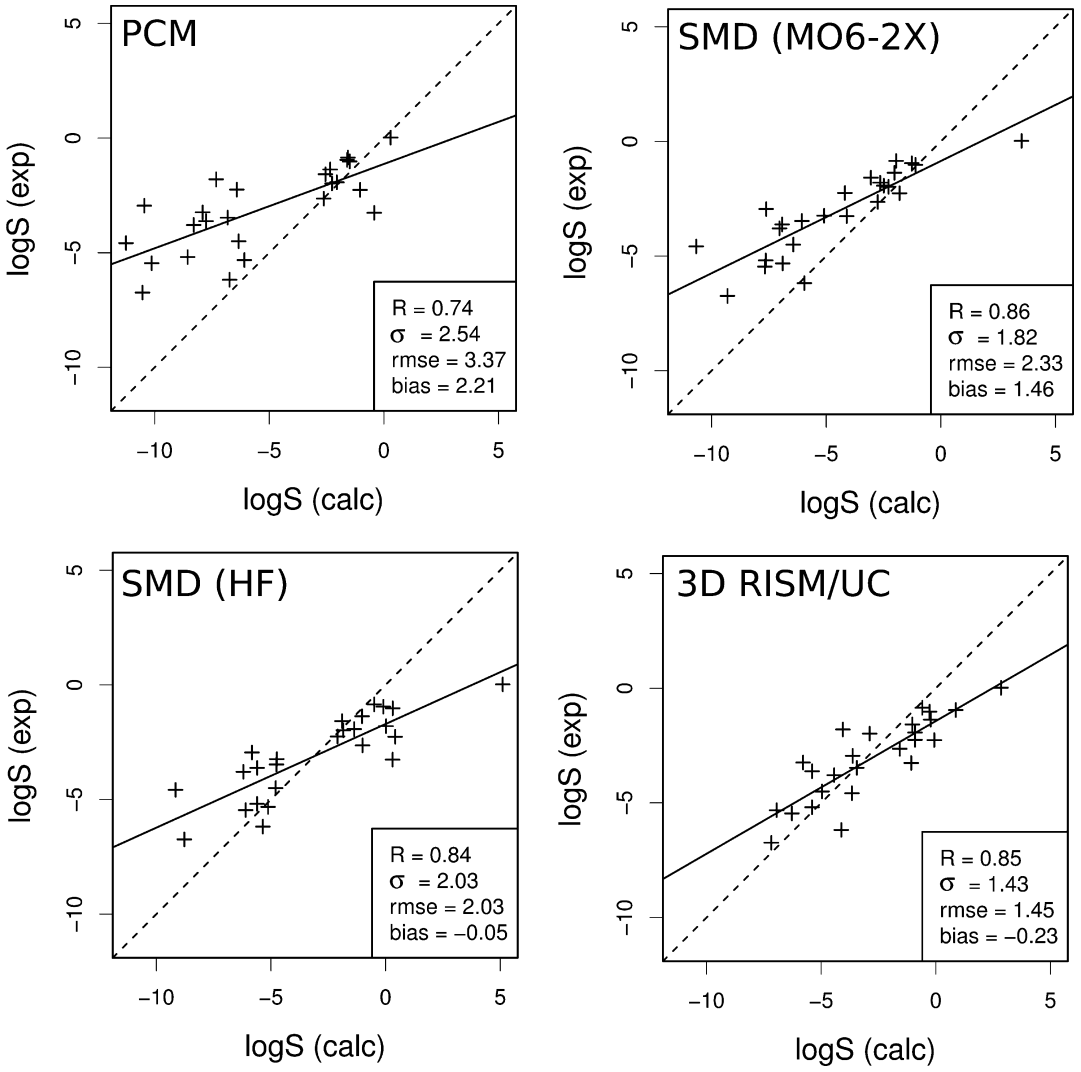

Figure 7. Prediction of solubility (log *S*) for a dataset of 25 molecules using sublimation free energies calculated using the FIT potential parameters and distributed multipoles evaluated at the B3LYP/6-31G(d,p) level of theory, combined with hydration free energies calculated by different methods. The solid black line indicates the line of best fit through the data. The dashed line is the identity line.

theory, etc. Alternatively, a different approach might be taken, by combining the computational calculations with QSPR methods to obtain a hybrid approach.

As well as predicting solubility, the method presented here provides a complete characterization of the thermodynamics of transfer of solute from crystal to gas phase to aqueous solution. Since the solubility of a crystalline solute depends on the properties of the undissolved crystalline precipitate as well as the properties of the solution, the thermodynamic data provides valuable information in understanding not only which of two molecules is more soluble, but also why the selected molecule is more soluble. By contrast, QSPR models, which are statistical rather than first-principles approaches, provide only limited statistical information about the underlying physicochemical processes. Moreover, since most QSPR models predict solubility from molecular rather than crystal structure, they are not able to rationalize or predict different solubilities for different polymorphs of a molecule.

The development of computational methods such as those described here is currently hindered by the lack of accurate and well-documented experimental sublimation and hydration free-energy data for druglike molecules.[70,71] Intrinisic aqueous solubility data were recently published for a large dataset of druglike molecules. We note that measurement of accurate thermodynamic data (including, but not limited to, sublimation and hydration free energies) for these molecules would significantly benefit the development of new computational solvent models.

6. CONCLUSIONS

The results presented here show that, by combining model potential-based crystal lattice simulations to calculate sublimation free energies with a statistical mechanics approach to calculating hydration free energies, the solubility of crystalline druglike molecules can be predicted with reasonable accuracy. While these proof-of-concept results are not yet as accurate as those reported by purely empirical approaches, there is clearly a very wide scope for systematic improvement. The method is not directly parametrized against experimental solubility data, and it offers a full computational characterization of the thermodynamics of transfer of the drug molecule from crystal phase to gas phase to dilute aqueous solution.

■ ASSOCIATED CONTENT

Supporting Information

Complete datasets, including all experimental and calculated data, and details of the FIT potential parameters. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jbom@st-andrews.ac.uk (J.B.O.M.), maxim.fedorov@strath.ac.uk (M.V.F.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

D.S.P. and M.V.F. thank the Scottish Universities Physics Alliance (SUPA) for funding through M.V.F.'s SUPA2 start-up funds. D.S.P. and M.V.F. are also grateful for partial financial support from the Deutsche Forschungsgemeinschaft (DFG)—German Research Foundation (Research Grant No. FE 1156/2-1) and for the use of the computing facilities at the John von Neumann-Institut für Computing (NIC), Juelich Supercomputing Centre (JSC), Forschungszentrum Juelich GmbH, Germany (Project ID Nos. HLZ18 and HLZ16). D.S.P. is grateful for funding from the European Commission, through a Marie Curie Intra-European Fellowship within the seventh European Community Framework Programme (No. FP7-PEOPLE-2010-IEF). J.B.O.M. and J.L.McD. thank the Scottish Universities Life Sciences Alliance (SULSA) for financial support. T.v.M., J.B.O.M., and J.L.McD. thank EaStCHEM for access to the EaStCHEM Research Computing Facility. We are grateful to Dr. Graeme Day (University of Cambridge) for providing a script to calculate the entropies of crystal structures.

■ REFERENCES

(1) Yalkowsky, S. H. *Solubility and Solubilization in Aqueous Media*; Oxford University Press: New York, 1999.

(2) Avdeef, A. *Absorption and Drug Development: Solubility, Permeability, and Charge State*; Wiley−Interscience: Hoboken, NJ, 2003.

(3) Noyes, A. A.; Whitney, W. R. *J. Am. Chem. Soc.* **1897**, *19*, 930−934.

(4) Hendersen, L. J. *Am. J. Physiol.* **1908**, *21*, 173−179.

(5) Hasselbalch, K. A. *Biochem. Z.* **1917**, *78*, 112−144.

(6) van de Waterbeemd, H.; Gifford, E. *Nat. Rev. Drug Discovery* **2003**, *2*, 192−204.

(7) Kubinyi, H. *Nat. Rev. Drug Discovery* **2003**, *2*, 665−668.

(8) Paolini, G. V.; Shapland, R. H. B.; van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L. *Nat. Biotechnol.* **2006**, *24*, 805−815.

(9) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.

(10) Jorgensen, W. L. *Science* **2004**, *303*, 1813−1818.

(11) Walters, W. P.; Namchuk, M. *Nat. Rev. Drug Discovery* **2003**, *2*, 259−266.

(12) Chiou, C. T.; Peters, L. J.; Freed, V. H. *Science* **1979**, *206*, 831−832.

(13) Mackay, D.; Shiu, W. Y.; Ma, K. C. *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals, Vol. 2: Polynuclear Aromatic Hydrocarbons, Polychlorinated Dioxins, and Dibenzofurans*; Lewis Publishers: Boca Raton, FL, 1992.

(14) Dearden, J. C. *Expert Opin. Drug Discovery* **2006**, *1*, 31−52.

(15) Balakin, K. V.; Savchuk, N. P.; Tetko, I. V. *Curr. Med. Chem.* **2006**, *13*, 223−241.

(16) Hopfinger, A. J.; Esposito, E. X.; Llinas, A.; Glen, R. C.; Goodman, J. M. *J. Chem. Inf. Model.* **2009**, *49*, 1−5.

(17) Bajorath, F. *Nat. Rev. Drug Discovery* **2002**, *1*, 882−894.

(18) Palmer, D. S.; Llinas, A.; Morao, I.; Day, G. M.; Goodman, J. M.; Glen, R. C.; Mitchell, J. B. O. *Mol. Pharm.* **2008**, *5*, 266−279.

(19) Johnson, S. R.; Chen, X. Q.; Murphy, D.; Gudmundsson, O. *Mol. Pharm.* **2007**, *4*, 513−523.

(20) Abramov, Y.; Pencheva, K. In *Chemical Engineering in the Pharmaceutical Industry: From R and D to Manufacturing*; am Ende, D. J., Ed.; Wiley: New York, 2010; pp 491−501.

(21) Garrido, N. M.; Queimada, A. J.; Jorge, M.; Macedo, E. A.; Economou, I. G. *J. Chem. Theory Comput.* **2009**, *5*, 2436−2446.

(22) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A. *J. Phys. Chem. A* **2005**, *109*, 6634−6643.

(23) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67−74.

(24) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. *J. Phys. Chem. B* **2010**, *114*, 8505−8516.

(25) Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooij, W. T. M.; Price, S. L.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. *Acta Crystallogr., Sect. B: Struct. Sci.* **2000**, *56*, 697−714.

(26) Motherwell, W. D. S.; Ammon, H.; Dunitz, J.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Lommerse, J. P. M.; Mooij, W. T. M.; Price, S. L.; Scheraga, H.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 647−661.

(27) Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, 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.; Verwer, P. *Acta Crystallogr., Sect. B: Struct. Sci.* **2005**, *61*, 511−527.

(28) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J. S.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M. A.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A. K.; Schweizer, B. *Acta Crystallogr., Sect. B: Struct. Sci.* **2009**, *65*, 107−125.

(29) Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X. M.; Braun, D. E.; Cruz-Cabeza, A.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M.; Hofmann, D. W. M.; Hofmann, F.; Jose, K. V. J.; Karamertzanis, P. G.; Kazantsev, A. V.; Kendrick, J.; Kuleshova, L. N.; Leusen, F. J. J.; Maleev, A. V.; Misquitta, A. J.; Mohamed, S.; Needs, R. J.; Neumann, M.; Nikylov, D.; Orendt, A. M.; Pal, R.; Pantelides, C. C.; Pickard, C. J.; Price, L. S.; Price, S. L.; Scheraga, H. A.; van de Streek, J.; Thakur, T. S.; Tiwari, S.; Venuti, E.; Zhitkov, I. K. *Acta Crystallogr., Sect. B: Struct. Sci.* **2011**, *67*, 535−551.

(30) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L.; Galek, P. T. A.; Day, G. M.; Cruz-Cabeza, A. J. *Int. J. Pharm.* **2011**, *418*, 168−178.

(31) Wassvik, C. M.; Holmen, A. G.; Bergstrom, C. A. S.; Zamora, I.; Artursson, P. *Eur. J. Pharm. Sci.* **2006**, *29*, 294−305.

(32) Hughes, L. D.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. O. *J. Chem. Inf. Model.* **2008**, *48*, 220−232.

(33) O'Boyle, N. M.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. O. *Chem. Cent. J.* **2008**, *2*, 21.

(34) Luder, K.; Lindfors, L.; Westergren, J.; Nordholm, S.; Persson, R.; Pedersen, M. *J. Comput. Chem.* **2009**, *30*, 1859−1871.

(35) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **2003**, *119*, 1661−1670.

(36) Schnieders, M. J.; Baltrusaitis, J.; Shi, Y.; Chattree, G.; Zheng, L.; Yang, W.; Ren, P. *J. Chem. Theory Comput.* **2012**, *8*, 1721−1736.

(37) Reinwald, G.; Zimmermann, I. *J. Pharm. Sci.* **1998**, *87*, 745−750.

(38) Perlovich, G. L.; Volkova, T. V.; Manin, A. N.; Bauer-Brandl, A. *AAPS PharmSciTech* **2008**, *9*, 205−216.
(39) Perlovich, G. L.; Volkova, T. V.; Manin, A. N.; Bauer-Brandl, A. *J. Pharm. Sci.* **2008**, *97*, 3883−3896.
(40) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478−8490.
(41) Llinas, A.; Glen, R. C.; Goodman, J. M. *J. Chem. Inf. Model.* **2008**, *48*, 1289−1303.
(42) Palmer, D. S.; Sergiievskyi, V. P.; Jensen, F.; Fedorov, M. V. *J. Chem. Phys.* **2010**, *133*, 044104.
(43) Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. *J. Phys. Condens. Matter* **2010**, *22*, 492101.
(44) Palmer, D. S.; Chuev, G. N.; Ratkova, E. L.; Fedorov, M. V. *Curr. Pharm. Des.* **2011**, *17*, 1695−1708.
(45) Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. *Mol. Pharm.* **2011**, *8*, 1423−1429.
(46) Frolov, A. I.; Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. *J. Phys. Chem. B* **2011**, *115*, 6011−6022.
(47) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.
(48) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(49) Ben-Naim, A. *J. Phys. Chem.* **1978**, *82*, 792−803.
(50) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016−2027.
(51) Pudipeddi, M.; Serajuddin, A. T. M. *J. Pharm. Sci.* **2005**, *94*, 929−939.
(52) Palmer, D. S.; O'Boyle, N. M.; Glen, R. C.; Mitchell, J. B. O. *J. Chem. Inf. Model.* **2007**, *47*, 150−158.
(53) Beglov, D.; Roux, B. *J. Phys. Chem.* **1997**, *101*, 7821−7826.
(54) Du, Q. H.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **2000**, *104*, 796−805.
(55) Hirata, F., Ed. *Molecular Theory of Solvation*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003.
(56) Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. *J. Am. Chem. Soc.* **2009**, *131*, 12430−12440.
(57) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. *J. Chem. Theory Comput.* **2010**, *6*, 607−624.
(58) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 4th Edition; Elsevier Academic Press: Amsterdam, 2000.
(59) Duh, D. M.; Haymet, A. D. J. *J. Chem. Phys.* **1995**, *103*, 2625−2633.
(60) Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **1999**, *103*, 7942−7957.
(61) Ratkova, E. L.; Chuev, G. N.; Sergiievskyi, V. P.; Fedorov, M. V. *J. Phys. Chem. B* **2010**, *114*, 12068−12079.
(62) Ten-no, S.; Jung, J.; Chuman, H.; Kawashima, Y. *Mol. Phys.* **2010**, *108*, 327−332.
(63) Drabik, P.; Gusarov, S.; Kovalenko, A. *Biophys. J.* **2007**, *92*, 394−403.
(64) Blinov, N.; Dorosh, L.; Wishart, D.; Kovalenko, A. *Biophys. J.* **2010**, *98*, 282−296.
(65) Ten-no, S. *J. Chem. Phys.* **2001**, *115*, 3724−3731.
(66) Ratkova, E. L.; Fedorov, M. V. *J. Chem. Theory Comput.* **2011**, *7*, 1450−1457.
(67) Chandler, D.; Singh, Y.; Richardson, D. M. *J. Chem. Phys.* **1984**, *81*, 1975−1982.
(68) Harano, Y.; Imai, T.; Kovalenko, A.; Kinoshita, M.; Hirata, F. *J. Chem. Phys.* **2001**, *114*, 9506−9511.
(69) Imai, T.; Harano, Y.; Kovalenko, A.; Hirata, F. *Biopolymers* **2001**, *59*, 512−519.
(70) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 259−279.
(71) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *J. Med. Chem.* **2008**, *51*, 769−779.
(72) Lue, L.; Blankschtein, D. *J. Phys. Chem.* **1992**, *96*, 8582−8594.
(73) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(74) Hirata, F.; Rossky, P. J. *Chem. Phys. Lett.* **1981**, *83*, 329−334.
(75) Lee, P. H.; Maggiora, G. M. *J. Phys. Chem.* **1993**, *97*, 10175−10185.
(76) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **2000**, *113*, 2793−2805.
(77) Chuev, G.; Fedorov, M.; Crain, J. *Chem. Phys. Lett.* **2007**, *448*, 198−202.
(78) Allen, M. P., Tildesley, D. J., Eds. *Computer Simulation of Liquids*; Clarendon Press: Oxford, U.K., 1987.
(79) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(80) Kovalenko, A.; Ten-No, S.; Hirata, F. *J. Comput. Chem.* **1999**, *20*, 928−936.
(81) Perkyns, J. S.; Pettitt, B. M. *Chem. Phys. Lett.* **1992**, *190*, 626−630.
(82) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. Amber Version 11. Available via the Internet at http://ambermd.org, 2010.
(83) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(84) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(85) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(86) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132−146.
(87) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(88) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(89) Sergiievskyi, V. P.; Hackbusch, W.; Fedorov, M. V. *J. Comput. Chem.* **2011**, *32*, 1982−1992.
(90) Sergiievskyi, V. P.; Fedorov, M. V. *J. Chem. Theory Comput.* **2012**, *8*, 2062−2070.
(91) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117−129.
(92) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*. Gaussian, Inc., Wallingford, CT, 2004.
(93) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.;

Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision C.1*; Gaussian, Inc.: Wallingford, CT, 2009.
(94) Wheeler, S. E.; Houk, K. N. *J. Chem. Theory Comput.* **2010**, *6*, 395−404.
(95) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 4538−4543.
(96) Nicholls, A.; Wlodek, S.; Grant, J. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 293−306.
(97) Meunier, A.; Truchon, J.-F. *J. Comput-Aided Mol. Des.* **2010**, *24*, 361−372.
(98) Atkins, P.; de Paula, J. *Physical Chemistry*, 9th ed.; Oxford University Press: Oxford, U.K., 2009.
(99) Gavezzotti, A.; Filippini, G. *Theoretical Aspects and Computer Modeling of the Molecular Solid State*; Wiley and Sons: Chichester, U.K., 1997; Chapter 3, pp 61−99.
(100) Williams, D. E.; Houpt, D. J. *Acta Crystallogr., Sect. B: Struct. Sci.* **1986**, *42*, 286−295.
(101) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352−7360.
(102) Hsu, L. Y.; Williams, D. E. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1980**, *A36*, 277−281.
(103) Filippini, G.; Gavezzotti, A. *Acta Crystallogr., Sect. B: Struct. Sci.* **1993**, *49*, 868−880.
(104) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233−239.
(105) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−622.
(106) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(107) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(108) Day, G. M.; Motherwell, W. D. S.; Jones, W. *Cryst. Growth Des.* **2005**, *5*, 1023−1033.
(109) Anghel, A. T.; Day, G. M.; Price, S. L. *CrystEngComm* **2002**, 348−355.
(110) Llinas, A.; Burley, J. C.; Box, K. J.; Glen, R. C.; Goodman, J. M. *J. Med. Chem.* **2007**, *50*, 979−983.
(111) Llinas, A.; Box, K. J.; Burley, J. C.; Glen, R. C.; Goodman, J. M. *J. Appl. Crystallogr.* **2007**, *40*, 379−381.
(112) Llinas, A.; Burley, J. C.; Prior, T. J.; Glen, R. C.; Goodman, J. M. *Cryst. Growth Des.* **2008**, *8*, 114−118.
(113) R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2011; ISBN 3-900051-07-0.
(114) USEPA. Estimation Programs Interface Suite for Microsoft Windows, v 4.10, United States Environmental Protection Agency (USEPA): Washington, DC, 2011.
(115) Bergstrom, C. A. S.; Wassvik, C. M.; Norinder, U.; Luthman, K.; Artursson, P. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1477−1488.
(116) Rytting, E.; Lentz, K. A.; Chen, X. Q.; Qian, F.; Venkatesh, S. *AAPS J.* **2005**, *7*, E78−E105.
(117) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 317−333.
(118) Shareef, A.; Angove, M. J.; Wells, J. D.; Johnson, B. B. *J. Chem. Eng. Data* **2006**, *51*, 879−881.