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

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

Department of Physics, University of Strathclyde, John Anderson Building, 107 Rottenrow, Glasgow G4 0NG, UK, Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, DE-04103 Leipzig, Germany, and Biomedical Sciences Research Complex and EaStCHEM School of Chemistry, Purdie Building, University of St Andrews, North Haugh, St Andrews, Scotland KY16 9ST, UK

E-mail: jbom@st-andrews.ac.uk; maxim.fedorov@strath.ac.uk

[*]To whom correspondence should be addressed
[†]University of Strathclyde
[‡]Max Planck Institute for Mathematics in the Sciences
[¶]University of St Andrews

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 (3DRISM) of the Integral Equation Theory of Molecular Liquids (IET). The solubilities of 25 crystalline druglike molecules are predicted by the model with $R = 0.85$ and $RMSE = 1.45$ $\log_{10}$ S units, which is significantly more accurate than results obtained using implicit continuum solvent models. The method is unique in that it is not parameterized against experimental solubility data and it offers a full computational characterization of the thermodynamics of transfer of 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 unionized molecule in saturated aqueous solution at thermodynamic equilibrium at a given temperature.[1,2] It is related to both pH-dependent solubility and dissolution rate by models such as the Henderson-Hasselbalch equation[3,4] and Noyes-Whitney equation,[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 (QSPR), 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 rationalise or predict different solubilities for different polymorphs of a molecule.

A more satisfactory approach to predicting intrinsic aqueous solubility would be to calculate it directly from molecular simulation. Up until now, however, few such approaches have been published, even though a large number of similar methods have been proposed to calculate other pharmacokinetic 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 moleular 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 semi-flexible organic molecules to be calculated with reasonable confidence,[25–29] with some recent successes also 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 here take from experiment, but which might in future work be obtained by crystal structure prediction. Since saturated aqueous solutions of druglike molecules are difficult to simulate from crystalline molecules, 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 X). Although the solution free energy 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 crystal to supercooled liquid to solution is problematic because the Gibbs free energy change for transfer from crystal to supercooled liquid is not easily accessible by either experiment or computation. Luder et al. have developed Monte Carlo simulations to predict $\Delta G_{liq-water}$. However, Monte Carlo simulations are computationally expensive and, hence, the method is not applicable to high-throughput drug discovery. Also, the supercooled liquid state of most drugs at room temperature is not accessible and so it is necessary to carry out simulations at elevated temperatures. The most successful method for prediction of solubility from this thermodynamic cycle is the general solubility equation (GSE), which relates log S to melting point and the logarithm of the octanol-water partition coefficient (log P). It can be derived (if some assumptions are made about the entropy of melting) from the thermodynamic cycle of gas to supercooled liquid to solution. Dannenfelser et al. have provided models for the prediction of ΔS_m, and Wassvik et al. have demonstrated that the GSE is more accurate if experimental values of ΔS_m are used. However, the best models for predicting melting point still give 40-50 degrees centigrade predictive errors and so the GSE is not usually applicable to as yet unsynthesized molecules.

The thermodynamic cycle for transfer from crystal to vapor to solution has been the subject of both experimental and computational studies. Reinwald et al. predicted aqueous solubility of drugs from experimental enthalpies of sublimation and calculated hydration energies. Unfortunately, this method was only accurate for one molecule from a data set of 12, and no computational procedure was suggested for the calculation of ΔH_{sub}. Perlovich et al. have published a series of papers that investigate the thermodynamic properties of drugs by experiment and computation, but they have not provided any methods for the prediction of solubility from structure alone without empirical parameterization. The most successful application of a thermodynamic cycle via the vapor has been the prediction of the solubility of liquids (and a small number of low molecular weight solids) from both experimental and calculated vaporisation and hydration energies by Thompson et al. The authors report predictive mean unsigned errors in the range of 0.4-0.6 in log solubility for a data set comprising simple low molecular weight compounds.

In previous work, some of the current authors (DSP and JBOM) attempted to predict solubility from calculation of sublimation and hydration free energies.[18] In this work, where ΔG_{sol} was calculated using an implicit solvent model based upon 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[31] and hydration free energies. Moreover, additional experimental data to benchmark these calculations has become available.[32] In particular, motivated by our earlier results, some of the current authors (DSP, MVF) 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).[33–37] This 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. Furthermore, new improved continuum solvent models for quantum mechanics calculations 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 parameterization 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 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_o) and the overall change in Gibbs free energy is

$$\Delta G^*_{sol} = \Delta G^*_{sub} + \Delta G^*_{hyd} = -RTln(S_o V_m) \tag{1}$$

where ΔG^*_{sol} is the Gibbs free energy for solution, ΔG^*_{sub} is the Gibbs free energy for sublimation, ΔG^*_{hyd} is the Gibbs free energy for hydration, R is the molar gas constant, T is the temperature (298 K), V_m is the molar volume of the crystal, S_o is the intrinsic solubility in moles per liter, and the superscript * denotes that we are using the Ben-Naim terminology, 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).

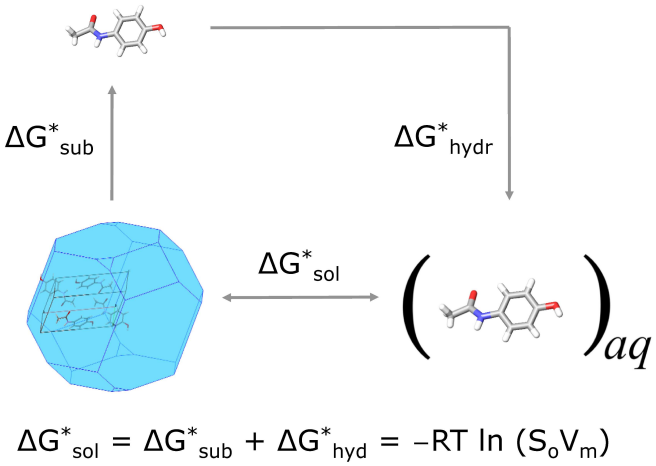


Figure 1: Thermodynamic cycle for transfer from crystal to gas and then to aqueous solution. This figure is based on Figure 1 from our earlier work.[18]

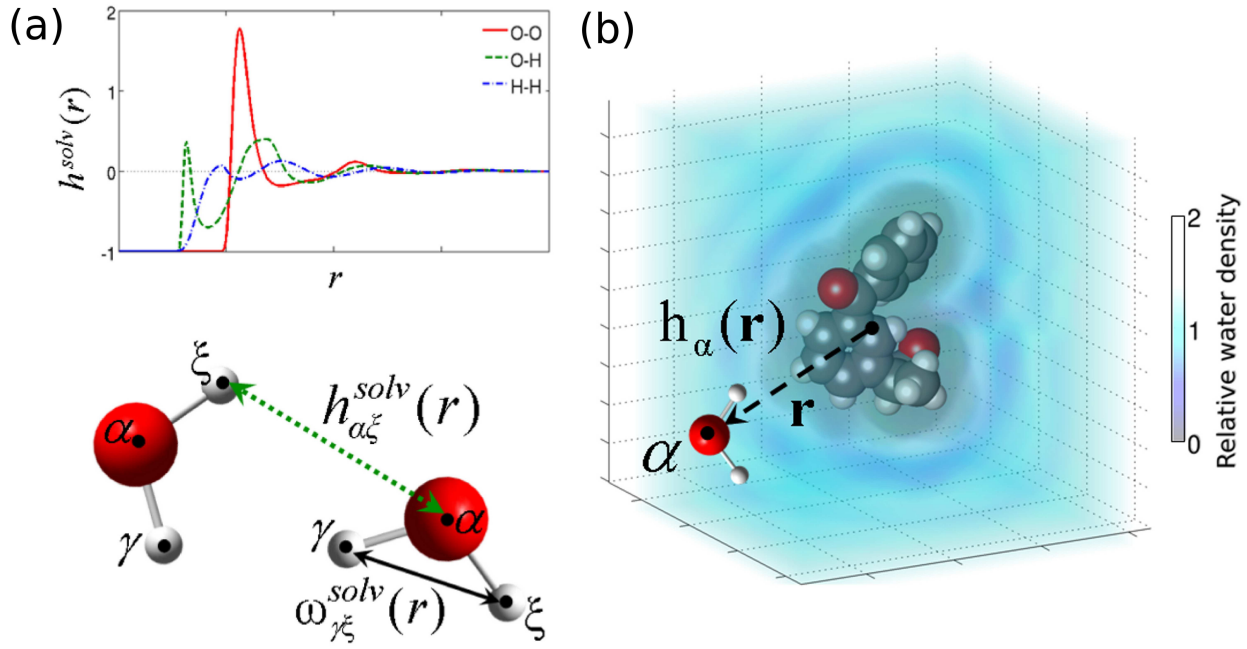


Figure 2: Correlation functions in the 3D RISM approach. (a) Site-site intramolecular ($\omega^{solv}_{\gamma\xi}(r)$) and intermolecular ($h^{solv}_{\alpha\xi}(r)$) correlation functions between sites of solvent molecules. The inset shows the radial projections of water solvent site-site density correlation functions: oxygen-oxygen (OO, red solid), oxygen-hydrogen (OH, green dashed) and hydrogen-hydrogen (HH, blue dash-dotted); (b) Three-dimensional intermolecular solute-solvent correlation function $h_\alpha(\mathbf{r})$ around a model solute. This figure is based on Figure 1 from our earlier work.[34]

2.2 Calculation of Hydration Free Energy using 3DRISM-UC

2.2.1 Background

The 3D Reference Interaction Site Model (3D RISM)[38–41] is a theoretical method for modelling solution phase systems based on classical statistical 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):[38,40]

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

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 $(\omega^{solv}_{\xi\alpha}(r))$, site-site radial total correlation functions $(h^{solv}_{\xi\alpha}(r))$ and the solvent site number density (ρ_α): $\chi_{\xi\alpha}(r) = \omega^{solv}_{\xi\alpha}(r) + \rho h^{solv}_{\xi\alpha}(r)$.[40] In this work, these functions were obtained by solution of the RISM equations of the pure solvent.[40,42]

To make Eq. (2) complete, $N_{solvent}$ *closure* relations are introduced:

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

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, $\beta = 1/k_BT$, 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 are therefore practically incomputable, which makes it necessary to incorporate some approximations.[40,43,44] In the current work, we use a closure relationship proposed by Kovalenko and Hirata (the KH closure),[45] which was designed to improve convergence rates and to prevent possible divergence of the numerical solution of the RISM equations:[45]

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

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).[33]

Within the framework of the RISM theory there exist several approximate functionals that allow one to analytically obtain values of the HFE from the total $h_\alpha(\mathbf{r})$ and direct $c_\alpha(\mathbf{r})$ correlation functions.[24,46,47] Although these functionals have been extensively used to *qualitatively* model thermodynamics of different chemical systems[24,48,49] they generally give HFE values that are strongly biased from experimental data with a large standard deviation error.[24,33,46,47,50,51] In recent work, DSP and MVF have developed a new free energy functional (3D RISM/UC) that allows the hydration free energies (HFE) of molecules ranging from simple alkanes to pharmaceuticals to be calculated accurately in the scope of the 3D RISM.

2.3 3D RISM/UC functional

The Gaussian fluctuations (GF) HFE functional was initially developed for by Chandler, Singh and Richardson, for 1D RISM, and adopted by Kovalenko and Hirata for the 3D RISM case:[40,52]

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

where ρ_α is the number density of a solvent sites α. Unfortunately, HFEs calculated using GF free energy functional have only a *qualitative* agreement with experiment. We (DSP and MVF) have recently shown that 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.[34–36] 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 (intercept):[34]

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

where the scaling coefficient a and intercept b values are obtained by linear regression against the experimental data for the simple organic molecule dataset. For the combination of methods used here (e.g. KH closure, GF free energy functional, molecular geometries optimized at the AM1 level of theory, AM1-BCC partial charges, and Lennard-Jones parameters taken from the AMBER GAFF forcefield), the coefficients have the values $a = -3.2217$ 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:[53,54]

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

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

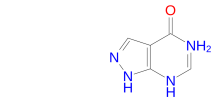
The 3D RISM/UC method has been shown to give accurate calculations of hydration free energies for both simple organic molecules and bioactive (druglike) molecules).[34–36]

3 Methods

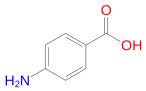
3.1 Datasets

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 (CSD) refcode of the polymorph used in the calculations . The experimental solubility, sublimation and hydration data including references are given in Table 1, Table 3, and Table 4, respectively. Sublimation free energy data could only be

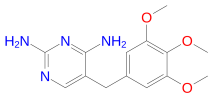
found in the published literature for four molecules in the dataset. For this reason, the sublimation free energy calculations were also benchmarked against sublimation free energy data obtained from experimental intrinsic aqueous solubility and hydration free energy data using Eq. (1). The benefit of 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.[55,56] (By "well-documented" we mean that both the methodology and the experimental conditions must be clearly reported)



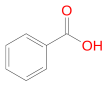
allopurinol
ALOPUR



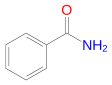
4-aminobenzoic acid
AMBNAC04



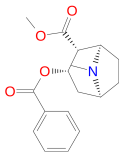
trimethoprim
AMXBPM10



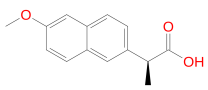
benzoic acid
BENZAC02



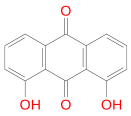
benzamide
BZAMID02



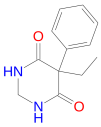
cocaine
COCAIN10



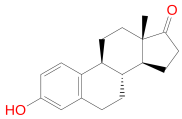
naproxen
COYRUD11



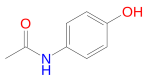
danthron
DHANQU06



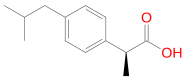
primidone
EPHPMO



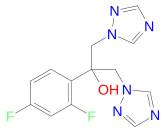
estrone
ESTRON14



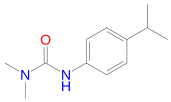
acetaminophen
HXACAN04



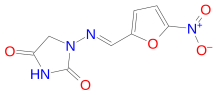
ibuprofen
IBPRAC01



fluconazole
IVUQOF

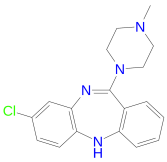


isoproturon
JODTUR01

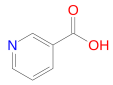


nitrofurantoin
LABJON01

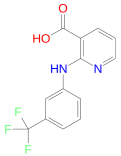
1-naphthol
NAPHOL01



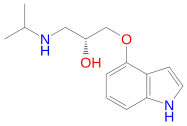
clozapine
NDNHCL01



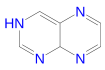
nicotinic acid
NICOAC02



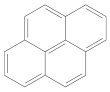
niflumic acid
NIFLUM10



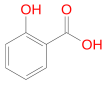
pindolol
PINDOL



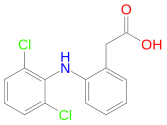
pteridine
PTERID11



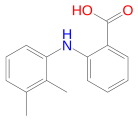
pyrene
PYRENE07



salicylic acid
SALIAC



diclofenac
SIKLIH01



mefenamic acid
XYANAC

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

3.2 Calculation of ΔG_{hyd} using 3D RISM/UC

3.2.1 RISM calculations

RISM calculations were performed assuming infinitely diluted solution. We used the Lue and Blankshtein version of the SPC/E model of water (MSPC/E).[57] This differs from the original SPC/E water model[58] by the addition of modified LJ potential parameters for the water hydrogen, which were altered to prevent possible divergence of the algorithm.[59–62] The Lorentz-Berthelot mixing rules were used to generate the solute-water LJ potential parameters.[63] 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[42] in the AmberTools 1.4 set of routines.[64] The 3D-grid around a solute was generated such that the minimum distance between any solute atom and the edge of 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,[65] where we used 5 MDIIS vectors, MDIIS step size - 0.7, residual tolerance - 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,[66] using the KH closure. The grid size for 1D-functions was 0.025 Å, which gave a total of 16384 grid points. We employed the MDIIS iterative scheme, where we used 20 MDIIS vectors, MDIIS step size - 0.3, and residual tolerance - 10^{-12}. The solvent was considered to be pure water with a number density 0.0333 Å^{-3}, a dielectric constant of 78.497. The final susceptibility solvent 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.[64,67,68]

(1) The coordinates of each molecule were optimized using the AM1 Hamiltonian[69] via the *antechamber*[70] suite, which uses the *sqm*[64] 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,[71–73] where BCC stands for bond charge correction, as implemented in Antechamber from the AmberTools 1.4 package.[64] The BCC parameters were taken from Jakalian et al.[72]

(3) For all compounds, the LJ parameters from the General Amber Force Field (GAFF)[73] were assigned to solute atoms with the *antechamber* and *tleap* programs.[70]

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 polarizable continuum model, as implemented in Gaussian03. 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, as implemented in Gaussian09; (3) HF/6-31G(d) SMD - the M06-2X density functional with the 6-31G(d) basis set and the SMD solvent model, as implemented in Gaussian09. These combinations of QM theory and solvent model were selected because they performed well in a recent blind challenge for HFE calculation.[74–76] 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 298K.

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 ^{o}). The Gibbs-Helmholtz equation was used to calculate ΔG^o_{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^o_{sub} can be converted to ΔG^*_{sub} using the following equation, which is derived considering the work for isothermal expansion of an ideal gas:

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

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

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

This convention is useful because sublimation free energies are almost exclusively given as ΔG^o_{sub} in the literature, while hydration free energies are more commonly given as ΔG^*_{hyd}. (We note that converting experimental values of ΔG^o_{sub} to ΔG^*_{sub} 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^o_{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 $\Delta H^o_{sub} = -U_{latt} - 2RT$. 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

Crystal lattice energies were calculated with DMACRYS from the energy-minimized crystal structures. The repulsion-dispersion contributions to the intermolecular potential were evaluated as

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

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), Coombes et al. (H_P-H_P), Hsu and Williams (Cl-Cl), and Filippini and Gavezzotti (S-S); here, H_C are hydrogen atoms bonded to carbon and H_P are polar hydrogen atoms (bonded to either oxygen or nitrogen). 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 Åcutoff.

Electrostatic contributions to the intermolecular potential were calculated from a distributed multipole representation of the electron distribution, which was evaluated by single point calculation, using MP2 and the 6-31G(d,p) basis set in Gaussian03, including multipoles up to the hexadecapole. 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 Åcutoff between molecular centers of mass.

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 intra- 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 $1/2RT$ and $3/2RT$ 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.

$S_{ext,crystal}$. 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.[77] In these calculations, it is assumed that vibrations are harmonic and coupling between inter 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 (CSD) using the following algorithm:

1. Extract all entries from the Cambridge Structural Database (CSD) that have 3D coordinates for the single molecule (no salts, solvates, cocrystals, 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 four 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.[78–80] For those molecules in the dataset for which the solubility data is 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. Nevertheless, as has previously been suggested by other authors, a simulated crystal structure may be sufficient to improve models to predict solubility, even if small errors exist in the simulated polymorphic landscape.,[18,19] since the average differences in experimental molar solubilities between polymorphs (2-fold)[81] are considerably lower than the average errors in models to predict solubility (6- to 10-fold molar solubility),.[14–16]

3.5 Statistical Analysis

To compare calculated and experimental results for different computational models, a correlation coefficient and the root mean squared deviation (*RMSD*) were evaluated:

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

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

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 the two parts: bias (or mean displacement, M) and standard deviation (SD), which are calculated by the formulae:

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

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

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 formulae:

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

From inspection of Eq. (14) and Eq. (15), it is clear that models reporting *RMSE* 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 analysed in terms of three different categories, some of which have previously been adopted by other authors:[16] (i) accurate predictions - molecules whose solubilities are calculated with an absolute error of less than 0.5 log S units; (ii) satisfactory predictions - molecules whose solubilities are calculated with an absolute error of less than 1 log S units; (iii) outliers - molecules for which the absolute error in the calculated solubility is more than two times the standard deviation of the experimental data ($1.79 \cdot 2 = 3.58$ $\log_{10}$S, referred to units of mol/l).

Statistical analyses were carried out in the R Statistical Computing Environment.[82] Python scripts were used to manipulate raw data files.

3.6 Computational Expense

The calculations discussed in this paper were performed in duplicate on different computing clusters, at the University of Strathclyde (by JM) and at the Max Planck Institute (MPI) for Mathematics in the Sciences (by DSP). Here, we report timings for computations performed on a single machine at the MPI, an Intel(R) Core (TM)2 Duo CPU E8600 3.33 GHz processor. The mean time required to calculate the hydration free energy of a single solute using 3D RISM/UC was $\sim$ 45 min, while the minimum and maximum values were $\sim$ 30 min and $\sim$ 75 min, respectively. The time required for a single calculation could be significantly reduced by using advanced numerical algorithms[83] or by performing the simulations using parallel computation.[42] The most time-consuming step in the sublimation free energy calculations is the single-point calculation at the MP2/6-31G(d,p) level, which required between 1 and 11 hours on a single CPU depending on the size of the molecule. The remaining steps in the calculation of sublimation free energy require minimal computational expense (10 to 20 mins on a single CPU).

4 Results

The aim of the current work is to assess how accurately the intrinsic aqueous solubility of crystalline druglike molecules can be estimated from calculated sublimation and hydration free energies based on a thermodynamic cycle via the vapour (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); (ii) B3LYP/6-31G(d,p); (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 ($bias_{MP2} = 0.16$ kJ/mol and $bias_{B3LYP} = 0.70$ kJ/mol). Both models provided a better estimate of the data than its mean since the root-mean-square errors ($RMSE_{MP2} = 5.64$ kJ/mol and $RMSE_{B3LYP} = 5.67$ 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} ($R_{HF} = 0.82$, $RMSE_{HF} = 11.50$ kJ/mol, $bias_{B3LYP} = -8.90$ kJ/mol). The correlation between ΔG_{sub}^{exp} and ΔG_{sub}^{MP2} is plotted in Figure 4 (similar graphs for ΔG_{sub}^{B3LYP} 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 $\Delta G_{sub}^{MP2} = 54.89$ kJ/mol.

Table 1: Sublimation free energies at 298K 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

Molecule	ΔG_{sub}^{exp} kJ/mol	ΔG_{sub}^{calc} kJ/mol (MP2)	ΔG_{sub}^{calc} kJ/mol (B3LYP)	ΔG_{sub}^{calc} kJ/mol (HF)
BENZAC02	34.23	35.08	34.59	48.91
BZAMID02	43.14	36.41	36.98	48.76
COCAIN10	54.90	56.23	56.55	61.07
COYRUD11	61.06	65.62	64.84	77.84
HXACAN04	59.95	54.91	53.93	69.13
IBPRAC01	42.06	54.89	54.60	65.80
JODTUR01	59.45	60.15	59.75	68.22
NAPHOL01	35.38	35.82	33.23	39.73
PYRENE07	46.25	41.88	41.77	43.81
SALIAC	40.31	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

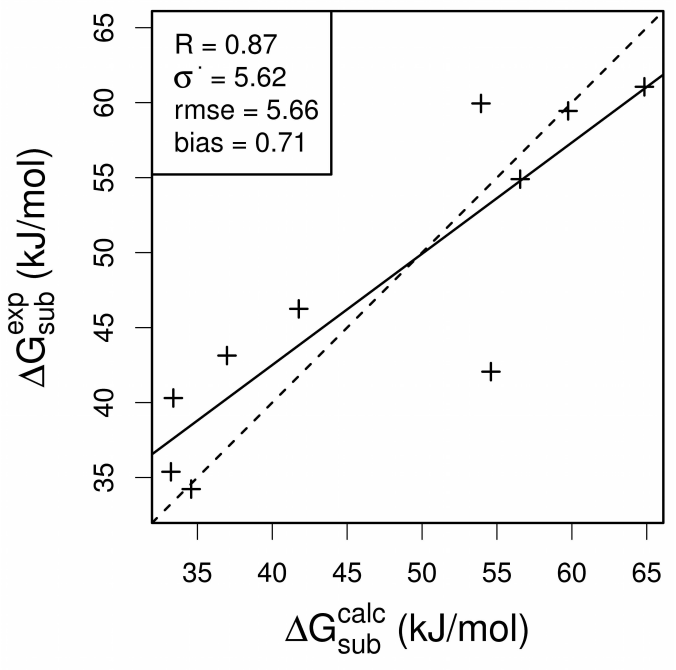


Figure 4: Correlation between experimental and calculated sublimation free energy, where the calculations were performed using the FIT potential parameters and distributed multipoles evaluated at the B3LYP/6-31G(d,p) level of theory

The 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 (see Table 1 $R_{MP2} = 0.90$, $RMSE_{MP2} = 0.99$ 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 value of the RMSE is larger than the standard deviation 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 MP2/6-31G(d,p) level of theory is plotted in Figure 4 (similar graphs for ΔG_{sub}^{B3LYP} and ΔG_{sub}^{HF}

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

are provided in the supporting information).

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). Furthermore, the observed correlation between ΔG_{sub}^{MP2} and ΔG_{sub}^{B3LYP} ($R = 0.998$, $\sigma = 0.863$ kJ/mol, $bias = 0.539$ kJ/mol) suggests that the density functional method is a useful alternative to MP2 for the molecules and calculations considered here.

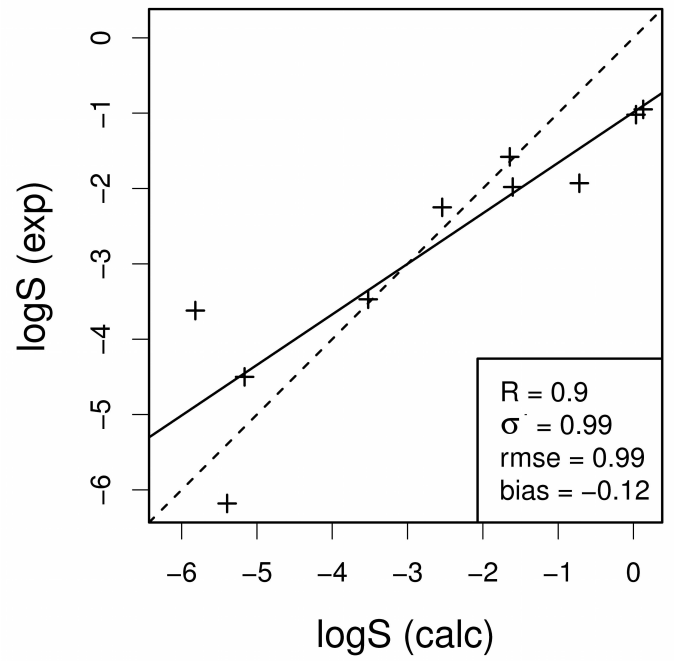


Figure 5: Correlation between experimental and calculated solubility, where the latter was computed from experimental hydration free energies and calculated sublimation free energies, and the calculations were performed using the FIT potential parameters and distributed multipoles evaluated at the B3LYP/6-31G(d,p) level of theory

4.2 Enthalpy-entropy compensation

A linear correlation is observed between the calculated enthalpy and entropy of sublimation, which is illustrated in Figure 6, where ΔH_{sub} is plotted against $-T\Delta S_{sub}$ with both quantities given in units of kJ/mol. The line of best fit is $\Delta H_{sub} = 0.13T\Delta S_{sub} - 44.61$ with $R = -0.88$. The idea of enthalpy-entropy compensation has been used to describe phenomena observed in a wide variety of different chemical systems including proteins,[84] protein-ligand complexes,,[85] etc. It is normally invoked to model relationships between enthalpy and entropy of homologous series of compounds, or of measurements on a small number of chemical systems carried out at a range of different temperatures. In our opinion, the results presented here do not imply a mechanism for enthalpy-entropy compensation in the sublimation of real organic crystals, since we work with a simplified computational model (infinite, perfect crystal and simplified entropy expression). Nevertheless, it is interesting to note that, of the terms comprising ΔS_{sub} (see **??**), ΔH_{sub} is more strongly correlated

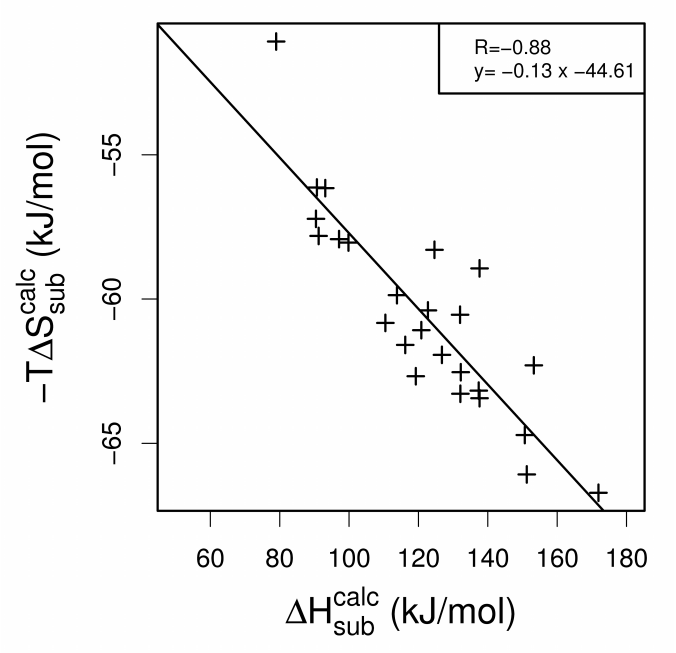


Figure 6: Correlation between sublimation enthalpies and entropies calculated using the FIT potential parameters and distributed multipoles evaluated at the B3LYP/6-31G(d,p) level of theory

with both $S_{trans}(gas)$ and $S_{rot}(gas)$ ($R = 0.7$ and $R = 0.7$, respectively), than with $S_{ext}(cryst)$ ($R = 0.04$), which might suggest that the observed correlation between ΔH_{sub} and ΔS_{sub} is not simply an artifact due to use of the same model-potential in the calculation of these quantities. The outlier in Figure 6 is pteridine (PTERID11), which is heterocycle containing two ring systems, four nitrogens and no rotatable bonds.

4.3 Hydration Free Energy

Four different methods were used to calculate hydration free energies: (1) 3DRISM/UC; (2) HF/6-31G(d) with the 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; (4) HF/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 $\Delta G_{hyd}^{3DRISM/UC}$, ΔG_{hyd}^{PCM}, $\Delta G_{hyd}^{SMD(HF)}$, and $\Delta G_{hyd}^{SMD(M06-2X)}$. The calculated hydration free energy data are presented in Table 3 and plotted in Figure 7. It is clear that for this dataset the 3DRISM/UC and HF/6-31G(d) SMD methods provide significantly more accurate estimates of hydration free energies than the other four methods tested here.

The correlation statistics for the 3DRISM/UC method are $R = 0.929$, $\sigma = 4.491$ kJ/mol, $RMSE = 4.896$ and $bias = 1.821$ kJ/mol, whereas for the HF/6-31G(d) SMD method they are $R = 0.973$, $\sigma = 2.815$ kJ/mol, $RMSE = 2.906$ and $bias = -0.721$ kJ/mol. It is noteworthy that the root-mean-square-errors for both of these methods are below the value of 5.7 kJ/mol that would equate to a 1 log_{10} unit error in the related equilibrium property (e.g. intrinsic aqueous solubility). The root-mean-square errors for the other four methods are all above 8 kJ/mol.

Table 3: Hydration free energies (n=25) from experiment and calculated using four different computational methods: (1) 3DRISM/UC; (2) HF/6-31G(d) with the 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; (4) HF/6-31G(d) SMD - the M06-2X density functional with the 6-31G(d) basis set and the SMD solvent model. $\rho V^{3D-RISM}$ is the unit-less partial molar volume of the solute in infinitely dilute solution as calculated by 3D RISM theory.

Molecule	ΔG_{hyd}^{exp} kJ/mol	$\rho V^{3D-RISM}$	$\Delta G_{hyd}^{3D-RISM/UC}$ kJ/mol	$\Delta G_{hyd}^{SMD(M06-2X)}$ 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 [a]	4.80	-36.67	-25.10	-31.66	-27.95
BZAMID02	-45.64 [a]	5.15	-49.86	-37.66	-44.33	-35.73
COCAIN10	-49.98 [a]	11.83	-59.30	-40.59	-52.52	-27.87
COYRUD11	-43.30 [b]	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 [b]	5.97	-60.31	-55.53	-63.63	-53.35
IBPRAC01	-29.33 [b]	9.83	-31.74	-23.02	-30.56	-18.24
IVUQOF		10.63	-70.75	-78.77	-93.96	-52.13
JODTUR01	-47.58 [a]	9.07	-48.01	-33.09	-40.51	-28.79
LABJON01		7.55	-92.77	-75.46	-100.63	-96.48
NAPHOL01	-32.01 [a]	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	8.01	-26.18	-15.72	-19.15	-11.30
SALIAC	-37.21 [a]	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] Ref. [86]
[b] Ref. [36]

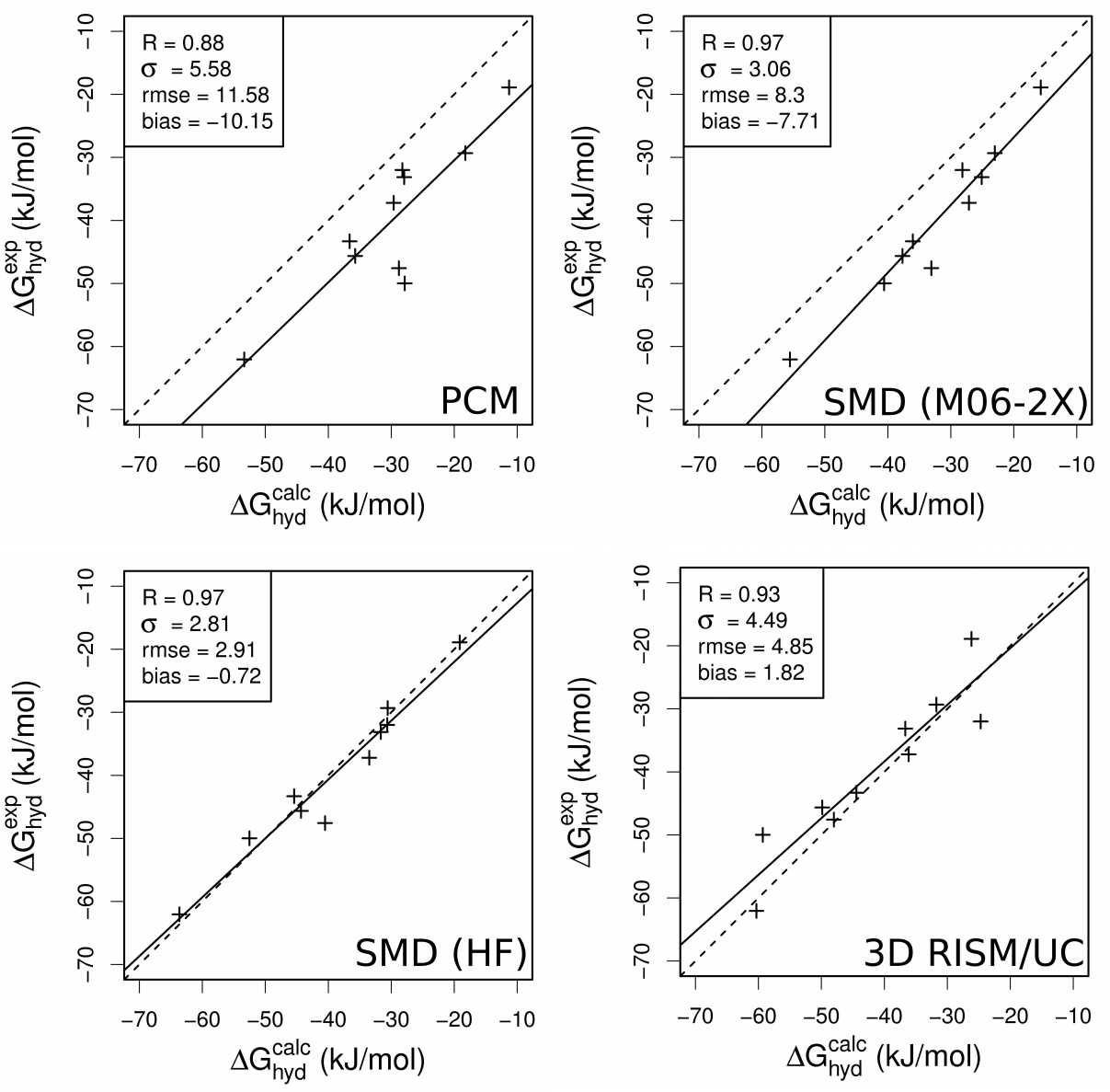


Figure 7: Calculation of hydration free energy made using four different solvation models plotted against experimental data for 10 solutes

4.4 Solubility

Twelve different predictions of intrinsic aqueous solubility can be made from the calculated thermodynamic data by considering a thermodynamic cycle via the vapour (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 Section 4.1 and Section 4.3. However, some caution must be excercised in making these inferences. Firstly, 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 vapour pressures near room temperature are more difficult to measure than saturated solution concentrations). Secondly, experimental sublimation and hydration free energy data are only available for lower molecular weight, less drug-like molecules (again due to the problems of measuring partial vapour pressures for druglike molecules near room temperature). Thus, the 10 molecules considered in Section 4.1 and Section 4.3 are a biased sample of the chemical space represented by the full dataset of 25 molecules. Thirdly, since the hydration free energy data for the 10 molecule dataset are freely available in the published literature, they may have been used in the parameterization of some of the implicit continuum methods used to calculate hydration free energies (they were *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 Section 4.1 and Section 4.3 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 solubility than ΔG_{sub}^{MP2}, but only by relatively small margins (Δ RMSE <

0.25 , referred to units of mol/l).

For the complete dataset of 25 molecules, the most accurate predictions of solubility were made using $\Delta G_{hyd}^{3DRISM/UC}$ and ΔG_{sub}^{B3LYP}, where $R = 0.85$, $RMSE = 1.45$ log S, $\sigma = 1.43$ 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 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 $\Delta G_{hyd}^{3DRISM/UC}$. See Figure 7). The accuracy of the predictions of solubility obtained with $\Delta G_{hyd}^{SMD(HF)}$ ($RMSE = 2.14$ log S, referred to units of mol/l) are relatively poor, which is in part due to two large outliers: NIFLUM10, Δlog S = 4.58 ; PTERID11 Δlog S = -5.09. NIFLUM contains 3 flourine atoms which may be a contributing factor to the prediction errors (IVUQOF is the only other molecular crystal that contains fluorine atoms).

Table 4: Prediction of solubility (logS) 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

Molecule	logS (Exp.)	Polymorph (Exp.)	logS (3D RISM/UC)	Error	logS (SMD-M062X)	Error	logS (SMD-HF)	Error	logS (PCM)	Error
NDNHCL01	-3.24 [e]	NDNHCL01	-5.79	2.55	-5.09	1.85	-4.75	1.51	-7.91	4.67
IVUQOF	-1.80 [a]		-4.05	2.25	-2.65	0.85	0.02	-1.82	-7.32	5.52
IBPRAC01	-3.62 [a]		-5.40	1.78	-6.92	3.30	-5.60	1.98	-7.76	4.14
ESTRON14	-5.32 [b]		-6.94	1.62	-6.90	1.58	-5.13	-0.19	-6.09	0.77
NAPHOL01	-1.98 [c]	NAPHOL01	-2.88	0.90	-2.28	0.30	-1.84	-0.14	-2.26	0.28
SIKLIH01	-5.46 [c]	SIKLIH01	-6.28	0.82	-7.67	2.21	-6.10	0.64	-10.13	4.67
AMXBPM10	-2.95 [c]	AMXBPM10	-3.63	0.68	-7.62	4.67	-5.83	2.88	-10.45	7.50
PINDOL	-3.79 [c]		-4.43	0.64	-7.04	3.25	-6.21	2.42	-8.30	4.51
COYRUD11	-4.50 [c]	COYRUD	-4.96	0.46	-6.44	1.94	-4.79	0.29	-6.34	1.84
XYANAC	-6.74 [c]	XYANAC	-7.17	0.43	-9.31	2.57	-8.78	2.04	-10.53	3.79
DHANQU06	-5.19 [d]		-5.40	0.21	-7.64	2.45	-5.61	0.42	-8.56	3.37
JODTUR01	-3.47 [d]		-3.45	-0.02	-6.06	2.59	-4.76	1.29	-6.82	3.35
NICOAC02	-0.85 [d]		-0.59	-0.26	-1.95	1.10	-0.50	-0.35	-1.57	0.72
BENZAC02	-1.58 [a]		-1.02	-0.56	-3.05	1.47	-1.90	0.32	-2.55	0.97
HXACAN04	-1.02 [c]	HXACAN01	-0.27	-0.75	-1.11	0.09	0.31	-1.33	-1.49	0.47
NIFLUM10	-4.58 [c]		-3.66	-0.92	-10.67	6.09	-9.16	4.58	-11.25	6.67
SALIAC	-1.93 [c]	SALIAC03	-0.91	-1.02	-2.49	0.56	-1.37	-0.56	-2.05	0.12
EPHPMO	-2.64 [d]		-1.57	-1.07	-2.75	0.11	-1.00	-1.64	-2.63	-0.01
AMBNAC04	-1.37 [d]		-0.22	-1.15	-2.02	0.65	-1.03	-0.34	-2.34	0.97
COCAIN10	-2.25 [a]		-0.91	-1.34	-4.19	1.94	-2.10	-0.15	-6.42	4.17
BZAMID02	-0.95 [d]		0.87	-1.82	-1.27	0.32	-0.10	-0.85	-1.61	0.66
PYRENE07	-6.18 [d]		-4.12	-2.06	-5.95	-0.23	-5.35	-0.83	-6.73	0.55
LABJON01	-3.26 [c]		-1.07	-2.19	-4.11	0.85	0.31	-3.57	-0.42	-2.84
ALOPUR	-2.26 [a]		-0.06	-2.20	-1.80	-0.46	0.42	-2.68	-1.04	-1.22
PTERID11	0.02 [c]		2.85	-2.83	3.52	-3.50	5.11	-5.09	0.29	-0.27

[e] Ref. [16]
[a] Ref. [87]
[b] Ref. [88]
[c] Ref. [32]
[d] Ref. [89]

Table 5: Calculation of intrinsic aqueous solubility (S) for a dataset of 25 molecules (Figure 3) using sublimation and hydration free energies calculated by different methods. Solubilities are expressed as $\log_{10}$S with units referred to mol/l.

ΔG_{hyd}	ΔG_{sub}	*R*	*RMSE*	σ	*bias*	|error| < 0.5	|error| < 1	outliers
3DRISM/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	MP2	0.81	2.14	2.13	0.14	8 (32%)	12 (48%)	4 (16%)
(HF)	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	MP2	0.84	2.49	1.87	1.65	6 (24%)	8 (32%)	3 (12%)
(M062X)	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	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%)

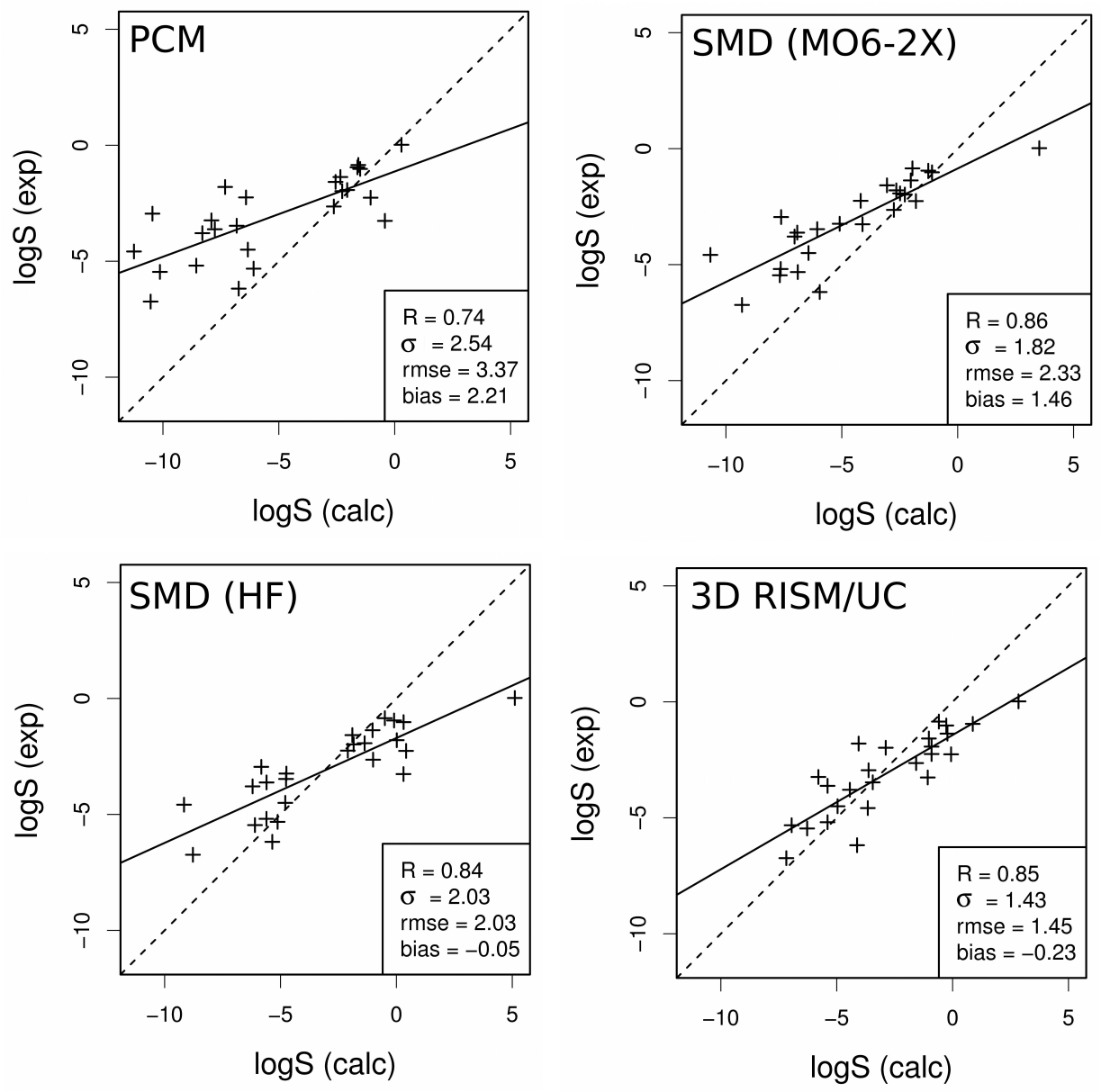


Figure 8: Prediction of solubility (logS) for a dataset of 25 molecules using sublimation free energies calculated using the FIT potential parameters and distributed multipoles evaluated at the B3LYP/6-31G(d,p) level of theory, combined with hydration free energies calculated by different methods

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. Over the last two decades, more than 100 different methods to predict the solubility of organic molecules in water have been published. The vast majority of these computational methods are Quantitative Structure-Property Relationships (QSPR), which use experimental data to learn a statistical relationship between the physical property of interest (solubility) and molecular descriptors calculable from a simple computational representation of the molecule. Up until now, very few theoretical approaches to calculate solublity have been published, even though a large number of similar methods have been proposed to calculate other pharmacokinetic properties, such as octanol-water partition coefficients, acid-base dissociation coefficients and protein-ligand binding free energies.

The computation of intrinsic aqueous solubility from molecular simulation may be decomposed into two seperate steps: (i) the prediction of crystal structure; (ii) the calculation of the solution free energy (from which intrinsic aqueous solubility can be estimated). The aim of the current work has been to assess how accurately step (ii) can be performed 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 vapour. 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 a number of conclusions. Firstly, distributed multipoles calculated at the MP2/6-31G(d,p) or B3LYP/6-31G(d,p) level of theory provide a far more accurate description of the electrostatic interaction energy in the crystal than do multipoles calculated at the HF/6-31G(d,p). 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 3DRISM/UC method of molecular integral equation theory. When the hydration free energies calculated by 3DRISM/UC were combined with sublimation free energies computed using multipoles evaluated at the 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. Although this level of accuracy is higher than that normally reported by QSPR models,[90,91] it is still lower than the standard deviation of the experimental solubility data ($\sigma = 1.79$ log S units), which indicates that the method provides useful predictions (better than the null prediction provided by the mean of the experimental data). Furthermore, unlike the QSPR models, the method proposed here is not parameterized on solubility data. 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 by, for example, considering more accurate intermolecular potentials, more advanced models from molecular integral equation theory, etc. Alternatively, a different approach might be taken, by combining the computational calculations with QSPR methods to obtain a hybrid approach. Both of these approaches would benefit significantly from the measurement of new accurate thermodynamic data for druglike molecules.

The lack of accurate and well-documented experimental sublimation and hydration free energy data for druglike molecules in the published literature is a signficant stumbling block in the development of new computational models.[55,56] (By "well-documented" we mean that both the methodology and the experimental conditions must be clearly reported.) Intrinisic aqueous solubility data was recently published for a large dataset of druglike molecules. We note that the 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

It has been commonly reported that calculation of the intrinsic aqueous solubility of crystalline druglike molecules is not possible with standard computational methods without incorporating specific empirical parameters. 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. Whilst 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.

Acknowledgement

We acknowledge the partial financial support of the Deutsche Forschungsgemeinschaft (DFG) - German Research Foundation, Research Grant FE 1156/2-1. We acknowledge the use of the computing facilities at the John von Neumann-Institut für Computing (NIC), Juelich Supercomputing Centre (JSC), Forschungszentrum Juelich GmbH, Germany - Project IDs: HLZ18 and HLZ16. DSP is funded by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (FP7-PEOPLE-2010-IEF).

Supporting Information Available

The 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.

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, N. J., 2003.

(3) Hendersen, L. J. *Am. J. Physiol.* **1908**, *21*, 173–179.

(4) Hasselbalch, K. A. *Biochemische Zeitschrift* **1917**, *78*, 112–144.

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

(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, W. P., R. H. B.; van Hoorn *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, Volume 2, Polynuclear Aromatic Hydrocarbons, Polychlorinated Dioxins, and Dibenzofurans*; Lewis Publishers, 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) Bayorath, F.

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

(19) Johnson, S. R.; Chen, X. Q.; Murphy, D.; Gudmundsson, O. *Mol. Pharmaceutics* **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., 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 Cryst. B* **2000**, *56*, 697–714.

(26) Motherwell, W. D. S. et al. *Acta Cryst. B* **2002**, *58*, 647–661.

(27) Day, G. M. et al. *Acta Cryst. B* **2005**, *61*, 511–527.

(28) Day, G. M. et al. *Acta Cryst. B* **2009**, *65*, 107–125.

(29) Bardwell, D. A. et al. *Acta Cryst. B* **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) 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.

(32) Llinas, A.; Glen, R. C.; Goodman, J. M. *J. Chem. Inf. Model.* **2008**, *48*, 1289–1303.

(33) Palmer, D. S.; Sergiievskyi, V. P.; Jensen, F.; Fedorov, M. V. *J. Chem. Phys.* **2010**, *133*, 044104.

(34) Palmer, S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. *J. Phys. Cond. Matt.* **2010**, *22*, 492101.

(35) Palmer, D. S.; Chuev, G. N.; Ratkova, E. L.; Fedorov, M. V. *Curr. Pharm. Des.* **2011**, *17*, 1695–1708.

(36) Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. *Mol. Pharmaceutics* **2011**, *8*, 1423–1429.

(37) Frolov, A. I.; Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. *J. Phys. Chem. B* **2011**, *115*, 6011–6022.

(38) Beglov, D.; Roux, B. *J. Phys. Chem.* **1997**, *101*, 7821–7826.

(39) Du, Q. H.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **2000**, *104*, 796–805.

(40) Hirata, F., Ed. *Molecular theory of solvation*; Kluwer Academic Publishers, Dordrecht, Netherlands, 2003.

(41) Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. *J. Am. Chem. Soc.* **2009**, *131*, 12430–12440.

(42) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. *J. Chem. Theory Comput.* **2010**, *6*, 607–624.

(43) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids, 4th ed*; Elsevier Academic Press, Amsterdam, The Netherlands, 2000.

(44) Duh, D. M.; Haymet, A. D. J. *J. Chem. Phys.* **1995**, *103*, 2625–2633.

(45) Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **1999**, *103*, 7942–7957.

(46) Ratkova, E. L.; Chuev, G. N.; Sergiievskyi, V. P.; Fedorov, M. V. *J. Phys. Chem. B* **2010**, *114*, 12068–12079.

(47) Ten-no, S.; Jung, J.; Chuman, H.; Kawashima, Y. *Mol. Phys.* **2010**, *108*, 327–332.

(48) Drabik, P.; Gusarov, S.; Kovalenko, A. *Biophys. J.* **2007**, *92*, 394–403.

(49) Blinov, N.; Dorosh, L.; Wishart, D.; Kovalenko, A. *Biophys. J.* **2010**, *98*, 282–296.

(50) Ten-no, S. *J. Chem. Phys.* **2001**, *115*, 3724–3731.

(51) Ratkova, E. L.; Fedorov, M. V. *J. Chem. Theory Comput.* **2011**, *7*, 1450–1457.

(52) Chandler, D.; Singh, Y.; Richardson, D. M. *J. Chem. Phys.* **1984**, *81*, 1975–1982.

(53) Harano, Y.; Imai, T.; Kovalenko, A.; Kinoshita, M.; Hirata, F. *J. Chem. Phys.* **2001**, *114*, 9506–9511.

(54) Imai, T.; Harano, Y.; Kovalenko, A.; Hirata, F. *Biopolymers* **2001**, *59*, 512–519.

(55) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. *J. Comput. Aided Mol. Des.* **2010**, *24*, 259–279.

(56) 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.

(57) Lue, L.; Blankschtein, D. *J. Phys. Chem.* **1992**, *96*, 8582–8594.

(58) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(59) Hirata, F.; Rossky, P. J. *Chem. Phys. Lett.* **1981**, *83*, 329–334.

(60) Lee, P. H.; Maggiora, G. M. *J. Phys. Chem.* **1993**, *97*, 10175–10185.

(61) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **2000**, *113*, 2793–2805.

(62) Chuev, G.; Fedorov, M.; Crain, J. *Chem. Phys. Lett.* **2007**, *448*, 198–202.

(63) Allen, M. P., Tildesley, D. J., Eds. *Computer Simulation of Liquids*; Clarendon Press, Oxford, 1987.

(64) 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.

(65) Kovalenko, A.; Ten-No, S.; Hirata, F. *J. Comput. Chem.* **1999**, *20*, 928–936.

(66) Perkyns, J. S.; Pettitt, B. M. *Chem. Phys. Lett.* **1992**, *190*, 626–630.

(67) Case, D. A. et al. Amber Version 11. WWW page, 2010; `http://ambermd.org`.

(68) 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.

(69) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(70) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Model.* **2006**, *25*, 247–260.

(71) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146.

(72) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(73) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(74) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 4538–4543.

(75) Nicholls, A.; Wlodek, S.; Grant, J. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 293–306.

(76) Meunier, A.; Truchon, J.-F. *J. Comput-Aided Mol. Des.* **2010**, *24*, 361–372.

(77) Anghel, A. T.; Day, G. M.; Price, S. L. *Crystengcomm* **2002**, 348–355.

(78) Llinas, A.; Burley, J. C.; Box, K. J.; Glen, R. C.; Goodman, J. M. *J. Med. Chem.* **2007**, *50*, 979–983.

(79) Llinas, A.; Box, K. J.; Burley, J. C.; Glen, R. C.; Goodman, J. M. *J. Appl. Crystallogr.* **2007**, *40*, 379–381.

(80) Llinas, A.; Burley, J. C.; Prior, T. J.; Glen, R. C.; Goodman, J. M. *Crystal Growth & Design* **2008**, *8*, 114–118.

(81) Pudipeddi, M.; Serajuddin, A. T. M. *J. Pharm. Sci.* **2005**, *94*, 929–939.

(82) 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.

(83) Sergiievskyi, V. P.; Hackbusch, W.; Fedorov, M. V. *J. Comput. Chem.* **2011**, *32*, 1982–1992.

(84) Cavanagh, J.; Akke, M. *Nat. Struct. Biol.* **2000**, *7*, 11–13.

(85) Gilli, P.; Ferretti, V.; Gilli, G.; Borea, P. A. *J. Phys. Chem.* **1994**, *98*, 1515–1518.

(86) US EPA. 2011, EPISUITE for Microsoft Windows, v 4.10, United States Environmental Protection Agency, Washington, DC, USA.

(87) Bergstrom, C. A. S.; Wassvik, C. M.; Norinder, U.; Luthman, K.; Artursson, P. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1477–1488.

(88) Shareef, A.; Angove, M. J.; Wells, J. D.; B., J. B. *J. Chem. Eng. Data* **2006**, *51*, 879–881.

(89) Rytting, E.; Lentz, K. A.; Chen, X. Q.; Qian, F.; Venkatesh, S. *AAPS Journal* **2005**, *7*, E78–E105.

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

(91) Palmer, D. S.; O'Boyle, N. M.; Glen, R. C.; Mitchell, J. B. O. *J. Chem. Inf. Model.* **2007**, *47*, 150–158.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.