

Toward the Prediction of Organic Hydrate Crystal Structures

Ashley T. Hulme[†] and Sarah L. Price*

*Department of Chemistry, University College London, 20 Gordon Street,
London WC1H 0AJ, United Kingdom*

Received February 22, 2007

Abstract: Lattice energy minimization studies on four ordered crystal structures of ice and 22 hydrates of approximately rigid organic molecules (along with 11 corresponding anhydrate structures) were used to establish a model potential scheme, based on the use of a distributed multipole electrostatic model, that can reasonably reproduce the crystal structures. Transferring the empirical repulsion–dispersion potentials for organic oxygen and polar hydrogen atoms to water appears more successful for modeling ice phases than using common water potentials derived from liquid properties. Lattice energy differences are reasonable but quite sensitive to the exact conformation of water and the organic molecule used in the rigid molecule modeling. This potential scheme was used to test a new approach of predicting the crystal structure of 5-azauracil monohydrate (an isolated site hydrate) based on seeking dense crystal packings of 66 5-azauracil···water hydrogen-bonded clusters, derived from an analysis of hydrate hydrogen bond geometries involving the carbonyl- and aza-group acceptors in the Cambridge Structural Database. The known structure was found within 5 kJ mol^{−1} of the global minimum in static lattice energy and as the third most stable structure, within 1 kJ mol^{−1}, when thermal effects at ambient temperature were considered. Thus, although the computational prediction of whether an organic molecule will crystallize in a hydrated form poses many challenges, the prediction of plausible structures for hydrogen-bonded monohydrates is now possible.

1. Introduction

Hydrate formation is common for organic molecules, with estimates varying from a third of organic molecules^{1,2} to perhaps three-quarters of pharmaceutical compounds forming hydrates.³ Many manufacturing processes provide an opportunity for hydrates to form,² and the state of hydration can be changed with environmental humidity and time.^{2,4} The state of hydration of an active pharmaceutical ingredient can significantly affect the solubility and dissolution rate and therefore its bioavailability.⁴ Thus intimate control of the hydration state of an active pharmaceutical ingredient is required. Such control can only follow detailed investigations into the existence of hydrated states, their formation, and

their dehydration products.⁵ The aim of this study is to take the first steps toward the use of computational crystal structure prediction to aid investigations into hydrate as well as polymorph screening.⁶

The roles of the water molecules in hydrates have been classified into three categories:² those in which each water molecule is isolated in the lattice (isolated site hydrate) and only hydrogen bonded to the organic molecule; those where water occupies channels in the crystal structure in which the water content is invariant (stoichiometric channel hydrate) or can readily vary in stoichiometry with the environmental humidity (nonstoichiometric channel hydrate); and those where the water is associated with ionic species (ion associated hydrates). Attempts have also been made to classify the extended water hydrogen-bonded substructures in hydrates,^{3,7} in a manner analogous to graph set analysis,⁸ defining chains, rings, tapes, and layers of water molecules.

* Corresponding author e-mail: s.l.price@ucl.ac.uk.

[†] Current address: Pharmorphix Ltd., 250 Cambridge Science Park, Milton Rd., Cambridge CB4 0WE, U.K.

Well-defined hydrate crystal structures in the Cambridge Structural Database (CSD)⁹ show a range of possible water hydrogen-bonding geometries¹⁰ from a single hydrogen bond up to the maximum of four. It has been contended that hydrate formation is more prevalent for those organic molecules in which there is a hydrogen bond donor/acceptor imbalance,¹¹ particularly when there are fewer donors than acceptors, with the inclusion of the water addressing this imbalance. This contention has been investigated by a statistical analysis of hydrate structures in the CSD,¹² which concluded that the sum of donors and acceptors of a molecule, rather than their ratio, influences hydrate formation, which is also increased with the polarity of the surface of the organic molecule. Hydrates can also be formed with more extensive, disordered water filling channels and voids in the crystal structure, extending to highly solvated protein structures where only a minority of waters in close proximity to the protein are in fixed positions.¹³ While a computational method of predicting hydrate formation would be a useful complement to the formulation and process design for pharmaceuticals, it is clearly a major challenge which will be very dependent on the molecule involved: a comprehensive method of hydrate prediction would have to search through a range of stoichiometries for possible hydrate crystal structures and consider entropic effects for disordered water as well as incorporating conformational flexibility. More organic hydrate crystal structures are known than all other solvates combined,¹⁴ emphasizing water's unique role in solvation and crystallization and being a solvent of choice in industrial processes.

The most fundamental requirement for hydrate prediction is the accurate modeling of the balance of organic molecule...organic molecule, organic molecule...water, and water...water intermolecular interactions. This study has investigated whether current models for intermolecular forces that are used for the crystal structure prediction of rigid organic molecules¹⁵ are suitable for hydrate prediction. Most simple water potentials have been parametrized against a wide range of liquid properties,¹⁶ with only one explicit attempt to modify a water potential for use with ice,¹⁷ which reparameterized the TIP4P potential to reproduce the density of several forms of ice. A wide range of computationally inexpensive models for water...water interactions were tested for their ability to reproduce four of the ordered crystal structures of ice, to assess their ability to model water...water interactions in the crystalline state. The repulsion–dispersion model potentials tested range from several that have been developed and are widely used for simulating liquid water to those derived from the empirical model potentials commonly used in organic crystal structure prediction (CSP).¹⁸ Both atomic charge and atomic multipole¹⁹ descriptions of the dominant electrostatic interactions were tested in conjunction with each repulsion–dispersion potential. Given the computational expense required for crystal structure prediction, more realistic but more complex model intermolecular pair potentials for water, which include further terms such as flexibility, anisotropic repulsion, and polarization,^{16,20,21} were not considered. The most promising water model was then tested, in conjunction with a commonly used

intermolecular potential for organic molecules, for its ability to reproduce the organic molecule...water interactions in the crystal structures of a range of 22 hydrate structures of rigid organic molecules. Corresponding anhydrate crystal structures were also available for seven of these compounds, allowing the lattice energies of the hydrates to be compared with those of the corresponding anhydrates and ice.

This model potential scheme is then used in a proof-of-concept test to predict the crystal structure of 5-azauracil monohydrate with the foreknowledge that 5-azauracil does indeed form a monohydrate. In order to avoid searching through the whole multidimensional space for crystals with two independent molecules in the unit cell, we developed a specific search strategy, analogous to one recently applied to diastereomeric salts,²² based on the assumption that the water would be hydrogen bonded to the organic molecule, and using an analysis of the CSD to determine the likely water hydrogen-bonding geometries. This approach, assuming the approximate hydrogen-bonding geometry and dense packing, significantly reduces the number of structures that have to be considered compared with more mathematically complete search methods appropriate for two independent molecules in the asymmetric unit cell.^{23,24} The random search method used in the only previously published work on the prediction of hydrate structures²⁵ found the experimental structure for only five of the nine polyalcohol and carbohydrate monohydrates considered, though the difficulty of these searches was considerably increased by the flexibility of the molecules.

The crystal structure prediction of 5-azauracil monohydrate tests the ability of our search strategy to find the known structure and that of the model potential scheme to successfully model the energy of the known structure relative to the alternative structures. Thus, we can assess the current possibilities of hydrate prediction on the basis of static lattice energy minimization.

2. Method

2.1. Testing Intermolecular Potentials for Their Ability To Model Ice. The phase diagram of ice presently contains 14 distinct crystalline phases. Phases I_h (common ice), III, IV, V, VI, VII, and XII are disordered and are unsuitable for testing the suitability of potentials for modeling crystalline phases through static lattice energy minimization. However, ices II^{26,27} and VIII,²⁸ which are high pressure ordered phases with no disordered analogues, ice IX,²⁹ a nearly ordered modification of ice III, and ice XI,^{30,31} a low temperature, ambient pressure proton-ordered modification of ice I_h, are suitable for this task. Recently, after the completion of the current study, an ordered version of ice V and a partially ordered version of ice XII have been reported.³² The structural properties of ices II, VIII, IX, and XI are reported in Table 1 and were used to test a range of water intermolecular potentials. These crystal structures were determined for D₂O using neutron diffraction and the accurate location of the hydrogen atoms allowed the experimental water geometries to be used. However, the variations in these molecular geometries (Table 1) from the isolated molecule³³ bond length (0.9572 Å) and bond angle (104.52°) demon-

Table 1. Summary of the Ice Structures Used To Test the Water Intermolecular

| structure | space group (SG for energy minimization) | Z | intramolecular geometry | | hydrogen bonds | |
|------------------------|--|----------------|----------------------------|-----------------------|------------------------|-------------------------|
| | | | O—H length (Å) | H—O—H angle (°) | O...O length (Å) | O—H...O angle (°) |
| ice II ²⁷ | $R\bar{3}$ | 2 | 0.958 | 103 | 2.805 | 166 |
| | | | 0.972 | 107 | 2.767 | 167 |
| | | | 0.942 | - | 2.779 | 178 |
| | | | 1.014 | - | 2.845 | 168 |
| ice VIII ²⁸ | $I4_1/amd$ ($C112_1$) | 0.5 (1) | 0.968 | 106 | 2.879 | 178 |
| | | | | | | |
| ice IX ²⁹ | $P4_12_12$ ($P2_1$) | 1.5 | 0.977 | 106 | 2.75 | 167 |
| | | (6) | 0.971 | 105 | 2.797 | 175 |
| | | | 0.979 | - | 2.763 | 165 |
| ice XI ³¹ | $Cmc2_1$ ($P\bar{1}$) | 1 ^a | 0.976 | 108 | 2.74 | 177 |
| | | (4) | 1.054 | 114 | 2.803 | 178 |
| | | | 0.947 | - | 2.737 | 176 |

^a Asymmetric unit contains two-half molecules — one located on a 2-fold axis and one on a mirror plane.

strate the flexibility of the molecular structure to distort within the tetrahedral hydrogen-bonding coordination upon crystallization and highlights the inherent limitations of our approximation of modeling the water as rigid.

Each ice structure was lattice energy minimized using DMAREL,³⁴ allowing for rigid body rotation, translation, and cell changes within the symmetry constraints of the subgroup of the experimental space group which gave whole molecules in the asymmetric unit (Table 1). After energy minimization, the retention of the higher experimental symmetry by the energy minimized structures was confirmed by the PLATON³⁵ ADDSYM algorithm. Consideration was limited to model potentials where the repulsion–dispersion potential was of the Lennard-Jones 12–6 form, specifically SPC,³⁶ its derivatives SPC/E³⁷ and MSPC/E,³⁸ and the two related potentials TIP3P³⁹ and TIP4P³⁹ or the Buckingham exp-6 form using the NSPC/E⁴⁰ model. Two other Buckingham potentials which had been derived by empirical fitting to organic crystal structures rather than water were also considered: the FIT potential¹⁵ where the oxygen potential was derived from oxohydrocarbons⁴¹ and the polar hydrogen from N–H groups mainly hydrogen bonded to carbonyl groups¹⁵ and a variation of this, FIT(COOH), where the polar hydrogen parameters were optimized to carboxylic acids.⁴² These published water potentials were tested with their own charge model (denoted STD) and in combination with both the CHELPG potential derived charges (ESP)⁴³ and distributed multipoles (DMA)¹⁹ derived from the MP2/6-31G(d,p) charge density calculated for the molecular structure for each ice. Only potential models with interaction sites on the atomic nuclei were considered, thus excluding using the full TIP4P model which has the oxygen partial charge at a non-nuclear position. The electrostatic contribution to the lattice energy was evaluated by Ewald summation for all charge–charge, charge–dipole, and dipole–dipole terms, with the repulsion–dispersion potential summed to a 15 Å atom–atom cutoff, and the higher multipole–multipole interactions in the DMA models up to R^{-5} summed to a 15

Å center of mass cutoff. All minima were confirmed to be stable by considering their second derivative rigid body properties. Thus a total of 21 intermolecular potential combinations for water were tested for their ability to reproduce the four ordered ice structures, by static lattice energy minimization. While the reproduction of the lattice parameters and density of ice XI is of most interest, as its stability domain of under 73 K at ambient pressure most closely matches the static lattice energy approximation of zero temperature and pressure, the lattice parameters of the other forms should be only underestimated by a few percent with the structures substantially unchanged. The overall structure reproduction for both ice structures and organic hydrates was quantified by the usual weighted F -value ‘figure of shame’⁴⁴

$$F = \left(\frac{100\Delta a}{a}\right)^2 + \left(\frac{100\Delta b}{b}\right)^2 + \left(\frac{100\Delta c}{c}\right)^2 + \Delta\alpha^2 + \Delta\beta^2 + \Delta\gamma^2 + (10\text{rms}\Delta x)^2 + \left(\frac{\text{rms}\Delta\theta}{2}\right)^2$$

where Δa is the error in cell length a (Å), $\Delta\alpha$ is the error in cell angle α (°), and the root-mean-square (rms) values of the space-group symmetry-allowed rigid-body center of mass translations Δx (Å) and rotations $\Delta\theta$ (°) are calculated over all the molecules in the unit cell. Since there are more symmetry-unconstrained molecular translations with two (or more) independent molecules in the asymmetric unit, F -values are expected to be larger for hydrates than for $Z' = 1$ anhydrate crystal structures.

2.2. Testing Model Intermolecular Potentials for Organic Hydrates. A test set of hydrate structures was constructed by searching the CSD (May 2004) for suitable hydrate crystal structures containing only the atomic species C, H, N, O, and F and excluding structures containing ions, polymers, disordered structures, structures without three-dimensional coordinates determined, and structures solved from powder X-ray diffraction data. Each structure in the subsequent set was examined manually to eliminate those with undetermined water hydrogen positions and those in which the parent molecule was deemed too flexible, i.e., contained groups with greater conformational flexibility than methyl, nitro, and amino substituents. Of the resulting test set of 22 hydrates, seven were found to have one or more corresponding anhydrous crystal structures which were also modeled for comparison.

The repulsion–dispersion potential tested was the FIT potential with the exp-6 form

$$U = \sum_{i \in M, k \in N} (A_u A_{kk})^{1/2} \exp(-(B_u + B_{kk})R_{ik}/2) - (C_u C_{kk})^{1/2}/R_{ik}^6$$

where atom i of molecule M is of type u and separated by an intermolecular distance R_{ik} from atom k of molecule N of type κ . Following the success of this model for reproducing the ice structures (section 3.1), the parameters used for C, N, and H_{np} (bonded to C) were those that had originally been fitted to azahydrocarbons,⁴⁵ for F to perfluorocarbons,⁴⁶ and for O (in the organic molecule and water) to oxohydrocar-

bonds.⁴¹ The parameters used for H_p, bonded to either N, O or in water, had been fitted to a range of polar and hydrogen bonded organic crystal structures¹⁵ in combination with the same C, H_{np}, N, and O parameters and a DMA electrostatic model. This model potential combined with a DMA model has been widely used in organic crystal structure prediction⁴⁷ including fluorinated compounds.^{48,49} This validation against hydrate crystal structures was to assess the ability of the geometric combining rules to correctly model organic molecule...water interactions by extrapolating from models for water...water and organic molecule...organic molecule crystal structures. This empirical model effectively represents all contributions to the intermolecular potential except the electrostatic contribution, which was represented by the distributed multipole representation of the MP2/6-31G(d,p) wavefunctions of the isolated molecules.

In order to test sensitivity to the molecular structure, for each hydrate crystal structure the lattice energy minimum using both the experimental molecular conformation (denoted ExpMinExp) and the ab initio optimized molecular conformation (denoted ExpMinOpt) were determined using the same DMAREL methodology described in 2.1. To allow for the systematic error in X-ray location of hydrogens, the ExpMinExp molecular structures had all bond lengths to hydrogen extended to neutron values,⁵⁰ with the standard SPC length for water (1.0 Å).³⁶ For the ExpMinOpt structures, the organic molecules were optimized at the MP2/6-31G(d,p) level, and a standard water geometry was used with the MP2/6-31G(d,p) optimized water molecule (angle 103.8° and bond length of 0.961 Å) (cf. Table 1). For three molecules, the crystal structure reproduction was sufficiently sensitive to certain flexible torsion angles that additional studies were performed with molecular structures in which just these torsion angles were constrained to experimental values and all other molecular parameters optimized (ExpMinConOpt). The ExpMinConOpt set of calculations constrained the NO₂ torsion for anhydrous 5-nitouracil and its monohydrate, one hydroxyl group proton torsion in dialuric acid monohydrate and the ring atom positions in anhydrous 5-fluorocytosine form 1. All ab initio calculations were performed using GAUSSIAN03,⁵¹ and the distributed multipoles for the specific molecular charge density were obtained using GDMA1.0.⁵²

2.3. Crystal Structure Prediction for 5-Azuracil Monohydrate. The model potential, as validated in the previous sections, was then tested for its ability to predict a monohydrate in which the water is hydrogen bonded to the organic molecule, by generating densely packed crystals from rigid molecule...water bimolecular clusters and then minimizing the lattice energy, allowing the water and molecule to independently adjust their relative orientation and position within the crystal lattice. The test system chosen was 5-azauracil monohydrate as both the anhydrate and monohydrate crystal structures are both reproduced with typical accuracy by the model intermolecular potential scheme (sections 3.1 and 3.2). The anhydrate structure has already been predicted as the global minimum in its lattice energy⁵³ and is well reproduced by Molecular Dynamics simulations at 310 K⁵⁴ with the same model potential. However, the main

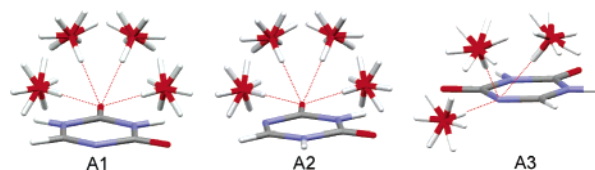


Figure 1. The 66 5-azauracil-water clusters used to generate initial monohydrate crystal structure, with notation for acceptors A1 and A2 and A3.

reason for choosing 5-azauracil is that this rigid molecule has three hydrogen bond donors and two acceptors, providing considerable variety of possible 5-azauracil...water, 5-azauracil...5-azauracil, and water...water hydrogen bonds in the crystals.

The search strategy employed required sufficient initial crude monohydrate crystal structures to be generated to ensure that all plausible hydrogen-bonding geometries could be sampled during the lattice energy optimization. Since the lattice energy minimization procedure was unlikely to significantly move the water molecules when optimizing from a densely packed crystal structure, it was necessary to establish a range for common hydrogen-bonded geometries for water (H_w—O_w—H_w) around the carbonyl and nitrogen acceptor groups by analysis of hydrate structures present in the CSD as described in the Supporting Information. The distributions for the H_w...O and H_w...N bond lengths and O_w—H_w...O and O_w—H_w...N angles were sufficiently sharp (Supporting Information, Figures S2 and S3) that clusters were defined with these parameters set to 1.9 Å, 1.9 Å, 135°, and 120°, respectively, and the hydrogen bonds were fixed to be linear. However, the torsion angles required to define the positions of the water molecule relative to the 5-azauracil molecular plane and the orientation of the non-hydrogen-bonded H_w atom were found to be fairly evenly distributed in both cases (Supporting Information, Figures S2 and S3). Hence clusters were defined at each acceptor (shown in Figure 1) so that the O_w was positioned by a torsion angle with values of 0°, 60°, 120°, or 180° (excluding one geometry where the water molecule physically overlapped the 5-azauracil molecule), and then the non-hydrogen-bonded H_w was defined by its torsion angle being either 0°, 60°, 120°, 180°, 240°, or 300°. Figure 1 shows all the resulting 66 starting point clusters which were built from the optimized molecular conformations. The implicit assumption that these clusters would automatically generate hydrogen bonds to the two N—H acceptors in the densely packed crystal structures was carefully monitored during the results analysis and appeared to be appropriate in this specific case because of the proximity of the donor and acceptor sites.

Each rigid 5-azauracil...water cluster was used in MOLPAK⁵⁵ to generate monohydrate crystal structures with one 5-azauracil...water cluster in the asymmetric unit cell in 37 MOLPAK packing types covering 18 space groups. The MOLPAK packing types were defined to represent the most common modes of packing of organic molecules, and this remains appropriate since the space group distribution of hydrate crystal structures is similar to that for organic structures generally.² The densest 125 of the approximately 5000 structures generated in each packing type were passed

Table 2. Summary of the Average F -Value for Each Repulsion–Dispersion Plus Electrostatic Model Combination Averaged over the Four Ordered Ice Structures

| repulsion–dispersion potential | electrostatic model | average F -value |
|--------------------------------|---------------------|--------------------|
| FIT | DMA | 21 |
| SPC/E | DMA | 56 |
| SPC | DMA | 57 |
| TIP4P | DMA | 68 |
| TIP3P | DMA | 75 |
| TIP3P | STD | 94 |
| SPC/E | STD | 95 |
| MSPC/E | ESP | 96 |
| SPC | STD | 99 |
| MSPC/E | STD | 101 |
| TIP3P | ESP | 105 |
| TIP4P | ESP | 109 |
| SPC | ESP | 117 |
| SPC/E | ESP | 118 |
| MSPC/E | DMA | 130 |
| NSPC/E | DMA | 173 |
| FIT(COOH) | ESP | 175 |
| FIT(COOH) | DMA | 192 |
| NSPC/E | ESP | 204 |
| NSPC/E | STD | 240 |
| FIT | ESP | 285 |

to DMAREL for lattice energy minimization, using the same model potential and methodology as described in 2.1 and 2.2. Thus approximately 300 000 lattice energy minimizations were carried out. All structures that minimized to saddle points were discarded. The optimized structures were compared using COMPACK⁵⁶ to overlay the 15 molecule coordination spheres and powder patterns^{57,58} to determine the unique low-energy structures in each search and then from the combination of all 66 searches. The elastic constants and $k = 0$ phonons for the unique crystal structures were calculated within the rigid-body harmonic approximation^{59,60} and used to estimate the zero-point energy, entropy, and the Helmholtz free energy⁶¹ at 298 K.

3. Results

3.1. The Testing of Water Potentials To Reproduce the Ordered Structures of Ice. The minima in the lattice energy for each of the eight different repulsion–dispersion models in combination with up to three different electrostatic models are compared with the experimentally known ice structures used as the starting point for the minimization in the Supporting Information (Tables S1–S4 for ice II, VIII, IX, and XI, respectively). Although many potentials reproduce one or more structures satisfactorily within the limits of static lattice energy minimization (errors in cell dimensions $< 5\%$, $F < \sim 50$), many minimizations result in grossly distorted structures, for example, with one cell parameter changing by over 10%. The average of the F -values for all four structures for each intermolecular potential model are summarized in Table 2, in order of increasing average F -value.

From this ranked list, it is clear that for the majority of potentials, the electrostatic model is of principal importance in determining the ability to reproduce the ice structures, with the theoretically more accurate distributed multipole model giving superior results compared to atomic charge models. This is in accord with the ability of the distributed multipole model to model the orientation dependence of the dominant term for hydrogen-bonding directionality.^{18,62,63} The NSPC/E and FIT(COOH) dispersion–repulsion potentials performed poorly, irrespective of the electrostatic model. It can be concluded from the poor performance of the FIT(COOH) that the hydrogen repulsion in water is closer to that for an N–H hydrogen than a carboxylic acid O–H hydrogen. The FIT dispersion–repulsion potential, combined with the distributed multipole electrostatic model, gives the overall best reproduction of the four test ice structures (Table 2). For this potential, all of the lattice energy minima are up to 5% denser than the corresponding experimental structure, consistent with the neglect of thermal effects, and with all of the structural variables quantitatively reproduced to reasonably good agreement (Table 3). The sensitivity to the intramolecular geometry was also investigated by contrasting (Table 3) the lattice energy minima for all four ice structures found with their specific molecular structure (ExpMinExp) with those (ExpMinOpt) minima found using the MP2/6-31G(d,p) optimized rigid water geometry, as used in the hydrate modeling. This made little difference to the structural reproductions, though it did destabilize the lattice energy in all cases by up to 2.5 kJ mol^{−1}. The lattice energies for the four polymorphs fell into a small range, −55.53 to −53.27 kJ mol^{−1} (ExpMinExp), which is typical of the relative energy differences between polymorphs, and compares reasonably well⁶⁴ with the sublimation enthalpy of ice I_h at 0 K, calculated to be 47.34(2) kJ mol^{−1}.⁶⁵

It is perhaps surprising that transferring the repulsion–dispersion model from organic functional groups rather than a specific water model appears more successful; perhaps this reflects that the errors in transferring a simple model from the liquid to the idealized crystalline state are larger. A significant component of the errors in the reproductions of the ice structures are likely to be due to the approximations of lattice energy minimization to model these ice phases, most of which are only stable at high pressures. The success of this relatively simple model for ice structures arises from the realistic modeling of the dominant electrostatic contribution. For comparison we note that an eight-site intermolecular potential for water with 77 fitted-parameters reproduces²¹ 1077 calculated points with negative (stable) energies for the water dimer with an rms error of 0.4 kJ mol^{−1}. The results in Table 3 lead to the conclusion that further empirical refinement of specific water parameters for lattice energy minimization studies by fitting to the ordered ice structures was not warranted.

3.2. Results for the Reproduction of Hydrate Crystal Structures. The reproductions of the hydrate and anhydrate crystal structures on lattice energy minimization vary considerably. The F -values are summarized in Tables 4 and 5, and the detailed results, including an analysis of the most poorly reproduced hydrogen bonds, are included in the

Table 3. Reproduction of the Crystal Structures of the Ordered Polymorphs of Ice by the FIT+DMA Model Potential Used in the Hydrate Modeling

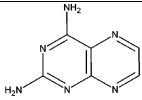
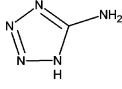
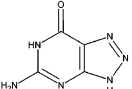
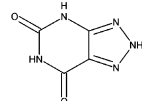
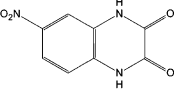
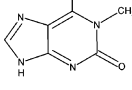
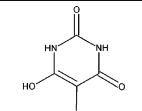
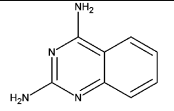
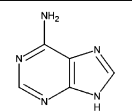
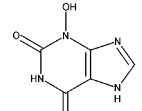
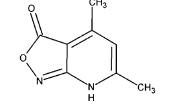
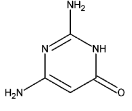
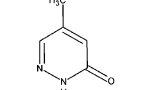
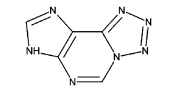
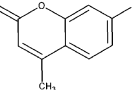
| | <i>a</i> (Å) | % error | <i>b</i> (Å) | % error | <i>c</i> (Å) | % error | density (g cm ⁻³) | % error | energy (kJ mol ⁻¹) | <i>F</i> |
|----------------------------|--------------|---------|--------------|---------|--------------|---------|-------------------------------|---------|--------------------------------|----------|
| Ice II | | | | | | | | | | |
| experimental ²⁷ | 12.983 | | 12.983 | | 6.254 | | 1.18 | | | |
| ExpMinExp | 12.773 | -1.6 | 12.773 | -1.6 | 6.190 | -1.0 | 1.23 | 4.4 | -53.61 | 14 |
| ExpMinOpt | 12.709 | -2.1 | 12.709 | -2.1 | 6.226 | -0.5 | 1.24 | 4.8 | -53.13 | 18 |
| Ice VIII | | | | | | | | | | |
| experimental ²⁸ | 4.656 | | 4.656 | | 6.775 | | 1.63 | | | |
| ExpMinExp | 4.583 | -1.6 | 4.583 | -1.6 | 6.635 | -2.1 | 1.72 | 5.4 | -55.53 | 10 |
| ExpMinOpt | 4.551 | -2.3 | 4.551 | -2.3 | 6.732 | -0.6 | 1.72 | 5.4 | -54.81 | 12 |
| Ice IX | | | | | | | | | | |
| experimental ²⁹ | 6.73 | | 6.83 | | 6.73 | | 1.16 | | | |
| ExpMinExp | 6.639 | -1.4 | 6.701 | -1.9 | 6.639 | -1.4 | 1.22 | 4.7 | -53.27 | 16 |
| ExpMinOpt | 6.64 | -1.3 | 6.696 | -2.0 | 6.64 | -1.3 | 1.22 | 4.8 | -52.59 | 17 |
| Ice XI | | | | | | | | | | |
| experimental ³¹ | 4.5019 | | 7.7979 | | 7.328 | | 0.93 | | | |
| ExpMinExp | 4.696 | 4.3 | 7.484 | -4.0 | 7.194 | -1.8 | 0.95 | 1.7 | -55.47 | 44 |
| ExpMinOpt | 4.483 | -0.4 | 7.760 | -0.5 | 7.320 | -0.1 | 0.94 | 1.0 | -53.07 | 6 |

Supporting Information, Tables S5 and S6. Examination of the lattice energy minimized structures using the standard hydrogen bond criteria within PLUTO⁶⁶ led to the conclusion that in only 5 of the 22 hydrate structures had the hydrogen bonding been altered upon energy minimization. The worst result of these five (THYMMH, CYTOSM, AZGUA, CIMMEQ, DIALAC02) structures is shown in Figure 2, from which it is clear that the huge F_{opt} value of 1633 arises from large changes in the relative water positions, with corresponding large cell deviations, but the thymine hydrogen-bonded ribbon remains intact. This contrasts with the majority of the monohydrate minimizations, which resulted in $F_{\text{opt}} \leq 120$ and the hydrogen-bonding motif being well reproduced, as exemplified by the overlay in Figure 3 for 5-azauracil monohydrate.

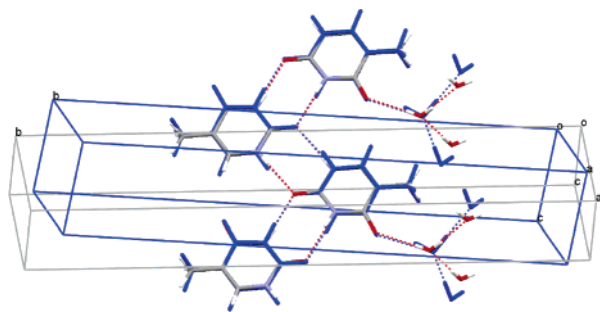
From the successful reproduction of the majority of hydrates structures it was concluded that the FIT potential would be adequate for use in a crystal structure prediction search for monohydrates of a small organic molecule, but, as generally found in organic crystal structure modeling, there are some pathological cases. The seven cases where both the hydrate and the anhydrate are modeled allowed comparison of the hydrate lattice energies with the weighted mean of the anhydrate and ice energy calculated with the same computational model (Table 5). Crudely, at 0 K there would be no thermodynamic reason for the monohydrate to form if its lattice energy was less stable than the sum of the lattice energies of the anhydrate and ice. However, the results show that the calculated hydrate lattice energies are commonly close to this value: i.e. the calculations do not show any great thermodynamic driving force for hydrate formation, with the differences generally comparable with the differences in lattice energy between ExpMinExp and ExpMin(Con)Opt energy minimizations. The former are generally more stable, implying that the very small conformational distortions probably arise from the crystal packing, but the energies are very sensitive to the exact positions of the protons in this rigid body modeling.⁶⁴

3.3. Crystal Structure Prediction Applied to 5-Azauracil Monohydrate. Comparison of the low-energy monohydrate structures found in the search with the ExpMinOpt minimized experimental 5-azauracil monohydrate structure showed that it was found as the 23rd ranked structure, 4.3 kJ mol⁻¹ above the global minimum in the lattice energy. All the more stable predicted structures are shown in Table 6. Inspection of these structures showed that the majority corresponded to sheets with hydrogen bonds solely between the 5-azauracil and water molecules, with none between pairs of 5-azauracil molecules. The experimental structure has the water hydrogen bonded only within the sheet (Figure 3) and only weak van der Waals interactions between sheets, whereas the majority of lower energy structures contained one of two sheet motifs (Figure 4) with the out-of-plane water proton hydrogen bonding to an adjacent sheet. The hypothetical sheet 1 can be derived from the experimental sheet structure (Figure 5), by the rotation of the water molecule out of the plane and a compensating adjustment of the 5-azauracil molecules to give a closer contact between the carbonyl oxygens. The more stable sheet 2 differs significantly from sheet 1 (and the experimental sheet) in that alternating 5-azauracil molecules have to be rotated by about 120° in the plane of the molecule (Figure 4), which is unlikely to occur as a solid-state transformation. The lowest energy nonsheet structure (A3_1_c ad/9) has a water molecule doubly hydrogen bonded to N1-H3 and O2 in a ring motif, and A3_1_f ad/31 has the water hydrogen bonded to the acceptor atoms N3 and O3, both being considerable distortions from the initial bimolecular cluster hydrogen-bonding motifs. However, the two other possible doubly hydrogen-bonded water...5-azauracil motifs do not appear to pack sufficiently well to appear in this low-energy region. The other nonsheet structures only have single hydrogen bonds to water. Thus only the experimental sheet structure shows the most common hydrogen-bonding motif for water in hydrates¹² (donor-donor-acceptor with the water ac-

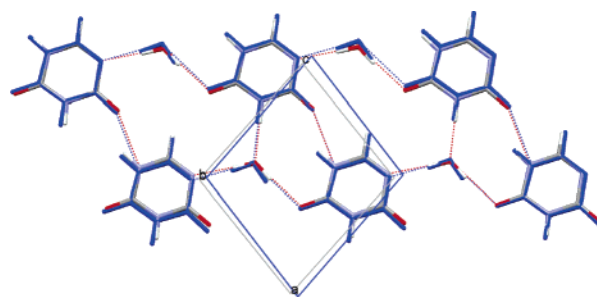
Table 4. Some Hydrate Structures Used To Test the Intermolecular Potential^a

| | | |
|--|---|--|
|  2,4-Diaminopteridine monohydrate AMPTRA10 $F_{\text{exp}} = 96$; $F_{\text{opt}} = 56$ |  5-Aminotetrazole monohydrate AMTETZ01 $F_{\text{exp}} = 32$; $F_{\text{opt}} = 115$ |  8-Azaguanine monohydrate AZGUAN $F_{\text{exp}} = 458$; $F_{\text{opt}} = 295$ |
|  Xanthazole monohydrate XANAZH01 $F_{\text{exp}} = 114$; $F_{\text{opt}} = 95$ |  6-Nitro-2,3-dihydroxy-quinoxaline monohydrate BAKGOJ01 $F_{\text{exp}} = 27$; $F_{\text{opt}} = 57$ |  1-Methyl-isoguanine dihydrate CIMMEQ $F_{\text{exp}} = 504$; $F_{\text{opt}} = 1366$ |
|  Dialuric acid monohydrate DIALAC02 $F_{\text{exp}} = 198$; $F_{\text{opt}} = 237$ |  2,4-Diaminoquinazoline monohydrate DUPYIW $F_{\text{exp}} = 16$; $F_{\text{opt}} = 43$ |  Adenine trihydrate FUSVAQ01 $F_{\text{exp}} = 111$; $F_{\text{opt}} = 106$ |
|  3-Hydroxyxanthine dihydrate HXANTH10 $F_{\text{exp}} = 37$; $F_{\text{opt}} = 84$ |  4,6-Dimethyl-isoxazolo-(3,4-b)pyridin-3-one monohydrate MIOZPO $F_{\text{exp}} = 106$; $F_{\text{opt}} = 208$ |  2,6-Dimano-4-pyrimidinone monohydrate SEYDIJ $F_{\text{exp}} = 134$; $F_{\text{opt}} = 184$ |
|  5-Methylpyridazine-3-one monohydrate TEKVI0 $F_{\text{exp}} = 58$; $F_{\text{opt}} = 46$ |  Tetrazolo(5,1)-purine monohydrate TRZPUR $F_{\text{exp}} = 136$; $F_{\text{opt}} = 118$ |  7-Hydroxy-4-methyl-chromen-2-one monohydrate WIKDAV $F_{\text{exp}} = 25$; $F_{\text{opt}} = 36$ |

^a The CSD refcode for the specific crystal structure used is given, with the F -values for the ExpMinExp (F_{exp}) and ExpMinOpt (F_{opt}) or ExpMinConOpt (F_{copt}) lattice energy minimizations with the different rigid molecule structures.

**Figure 2.** Overlay of the experimental crystal structure (colored by element) with that of the ExpMinOpt energy minimized structure (colored blue) for thymine monohydrate, the worst reproduction.

cepting one hydrogen bond and acting as a hydrogen bond donor to two independent acceptors).

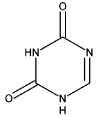
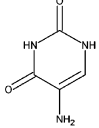
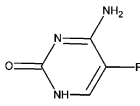
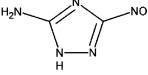
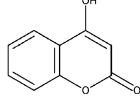

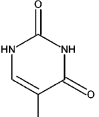
**Figure 3.** Overlay of the experimental crystal structure (colored by element) with that of the ExpMinOpt energy minimized structure (colored blue) for 5-azauracil monohydrate.

Thus, the search strategy was certainly successful in locating the known monohydrate structure and a range of plausible alternatives. Indeed the frequency with which all the sheet structures were found suggests a high level of redundancy in generation of these structures, though this has to be balanced by two of the nonsheet structures being found only once. Detailed analysis (Supporting Information, Table S7) shows that one of the sheet 1 structures could be found starting from all 66 clusters and that the experimental structure was found starting from 50 cluster geometries. Thus the exploration of possible hydrogen-bonding geometries appears reasonably complete for this system.

The known structure is not the lowest in static lattice energy, though the energy gap to the global minimum structure of less than 5 kJ mol⁻¹ is small. Its relative stability is improved by considering entropic effects, as the experimental structure is the third most stable according to the estimated Helmholtz free energy at 298 K (Table 6). While a sheet 2 structure remains the most stable, the difference is reduced to only 0.8 kJ mol⁻¹, though certain sheet 1 structures and nonsheet structures remain thermodynamically competitive. It is notable that there are many different ways of stacking both hypothetical sheet structures with very similar lattice energies but different susceptibility to shearing forces. The known structure is particularly susceptible to shear, as might be expected from the lack of hydrogen bonding between the sheets.

The energy differences between the known and the other low-energy structures in Table 6 are small compared with the many approximations in the computational model. In this case, the most important approximation is probably that the conformation of water and 5-azauracil has not been optimized within each crystal structure, particularly since there is an 18 kJ mol⁻¹ difference in lattice energy between the ExpMinExp and ExpMinOpt minimizations of the experimental 5-azauracil monohydrate crystal structure (Table 5). Figure 3 shows how these small molecular structure differences lead to minor changes in the hydrogen-bonding geometry and yet significant energy differences. Also, it is notable that all hypothetical crystal structures based on sheets (Figure 4), other than the experimental structure, have head-to-head carbonyl O...O distances less than 3 Å, which are rather short according to a recent estimate of the oxygen van der Waals radius⁶⁸ of 1.58 Å, though within previous estimates.⁶⁴ Since such head-to-head interactions differ from

Table 5. Some Hydrates with Corresponding Anhydrate Structures and Results of the Lattice Energy Minimization Calculations of the Intermolecular Potential^a

| Compound | Experimental information | | ExpMinExp Minimizations | | | ExpMin(Con)Opt Minimizations | | |
|---|--|---|-------------------------|-------------|---------|------------------------------|-------------|---------|
| | Anhydrate | Monohydrate | Anhydrate | Monohydrate | Stab. E | Anhydrate | Monohydrate | Stab. E |
| 5-azauracil  | XERBEB $F_{\text{exp}} = 38$; $F_{\text{opt}} = 72$ | HOQHAW $F_{\text{exp}} = 86$; $F_{\text{opt}} = 99$ | -117.41 | -178.26 | -5.38 | -109.3 | -160.7 | 1.68 |
| 5-nitrouracil  | NIMFOE $F_{\text{exp}} = 45$; $F_{\text{opt}} = 51$ | NURMAH $F_{\text{exp}} = 29$; $F_{\text{opt}} = 72$ | -124.92 | -183.7 | -3.30 | -116.6 | -163.8 | 5.88 |
| | NIMFOE01 $F_{\text{exp}} = 75$; $F_{\text{opt}} = 98$ | | -124.26 | -183.7 | -3.96 | -115.54 | -163.8 | 4.78 |
| | NIMFOE01 $F_{\text{exp}} = 12$; $F_{\text{opt}} = 20$ | | -129.94 | -183.7 | 1.72 | -119.34 | -163.8 | 8.62 |
| 5-fluorocytosine  | MEBQE01 $F_{\text{exp}} = 4$; $F_{\text{opt}} = 7$ | BIRMEU01 $F_{\text{exp}} = 41$; $F_{\text{opt}} = 60$ | -127.59 | -181.2 | 1.88 | -123.68 | -163.9 | 12.86 |
| | MEBQE0 $F_{\text{exp}} = 4$; $F_{\text{opt}} = 11$ | | -132.84 | -181.2 | 7.12 | -117.12 | -163.9 | 6.29 |
| 3-amino-5-nitro-1,2,4-triazole  | JOWWIB $F_{\text{exp}} = 58$; $F_{\text{opt}} = 187$ | JYIWET $F_{\text{exp}} = 187$; $F_{\text{opt}} = 109$ | -137.06 | -193.16 | -0.62 | -117.82 | -169.84 | 1.06 |
| 4-hydroxycoumarin  | Anhydrate 2 ⁶⁷ $F_{\text{exp}} = 131$; $F_{\text{opt}} = 178$ | HOXCUM01 $F_{\text{exp}} = 73$; $F_{\text{opt}} = 89$ | -112.5 | -156.54 | 11.43 | -106.85 | -146.44 | 13.48 |
| | Anhydrate 3 ⁶⁷ $F_{\text{exp}} = 38$; $F_{\text{opt}} = 55$ | | -110.24 | -156.54 | 9.17 | -105.98 | -146.44 | 12.61 |
| Cytosine  | CYTSIN $F_{\text{exp}} = 67$; $F_{\text{opt}} = 10$ | CYTOSM $F_{\text{exp}} = 183$; $F_{\text{opt}} = 812$ | -135.75 | -182.46 | 8.76 | -121.54 | -161.3 | 13.32 |
| Thymine  | THYMIN01 $F_{\text{exp}} = 28$; $F_{\text{opt}} = 16$ | THYMMH $F_{\text{exp}} = 1308$; $F_{\text{opt}} = 1633$ | -110.15 | -162.32 | 3.30 | -104.1 | -155.22 | 1.95 |

^a The CSD refcode for the specific crystal structure used is given, with the F -values for the ExpMinExp (F_{exp}) and ExpMinOpt (F_{opt}) or ExpMinConOpt (F_{cop}) lattice energy minimization with the different rigid molecule structures. Stab. E is the difference between the calculated hydrate lattice energy and the sum of the anhydrate and ice XI lattice energies, with a negative sign implying that the monohydrate is more stable than this extrapolated value. Note that the lattice energies of ice XI from Table 3 are used, i.e., the water geometry differs between the monohydrates and ice XI for the ExpMinExp values. Values in italics are derived using ExpMinConOpt lattice energies, with the torsions specified in section 2.2 fixed.

commonly found attractive carbonyl...carbonyl geometries⁶⁹ (which are well modeled by this potential in cases such as alloxan),⁷⁰ there could well be an underestimate of the repulsion in this contact in the two unobserved sheet motifs.

4. Discussion

The computational search strategy and lattice energy modeling developed here are certainly successful at generating a range of plausible crystal structures for 5-azauracil monohydrate, including the known one, within a small energy range. It is tempting to assume that crystallographic experience might lead to the discounting of the alternative structures based on either the observed preferences for

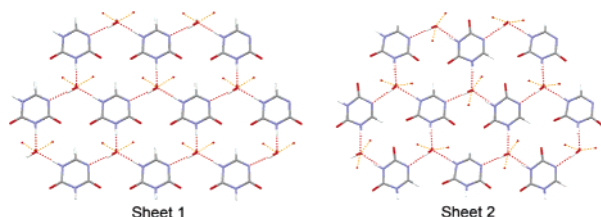
hydrate hydrogen-bonding geometries¹² or the unusual carbonyl interactions in the alternative sheets and so favor the observed sheet structure. However, crystallographic intuition in selecting plausible structures is far from reliable.⁷¹ While this study did not predict the known monohydrate structure as the most thermodynamically stable in free energy, it may well have been selected as one of the three most plausible structures on this basis and hence would have been a successful prediction by the rules of the international blind tests.⁷²

It is highly likely that the observed structure is the thermodynamically most stable at ambient conditions, though it is possible that the observed structure is kinetically favored,

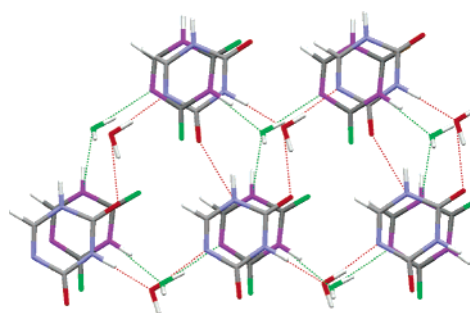
Table 6. Summary of the Low-Energy 5-Azauracil Monohydrate Predicted Structures, in Order of Lattice Energy Stability

| structure ^a | N_{find} | lattice energy (kJ mol ⁻¹) | free energy ^b | free energy rank | space group | a (Å) | b (Å) | c (Å) | α (°) | β (°) | γ (°) | shear ^c (GPa) | hydrogen bond motif |
|------------------------|-------------------|--|--------------------------|------------------|--------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------------------|---------------------|
| A1_3_c ad/32 | 38 | -165.0 | -166.1 | 1 | <i>Pc</i> | 5.064 | 7.084 | 7.275 | 90 | 90.61 | 90 | 0.4 | sheet 2 |
| A2_3_b ad/1 | 31 | -164.2 | -165.7 | 2 | <i>Pc</i> | 5.499 | 6.972 | 6.844 | 90 | 93.71 | 90 | 0.3 | sheet 2 |
| A1_2_d da/92 | 29 | -164.1 | -164.2 | 10 | <i>Cc</i> | 7.337 | 7.039 | 10.04 | 90 | 90.31 | 90 | 1.5 | sheet 2 |
| A3_1_c ad/9 | 1 | -162.5 | -165.3 | 4 | <i>Pc</i> | 3.683 | 5.706 | 12.646 | 90 | 100.22 | 90 | 1.1 | not sheet |
| A2_1_d aa/82 | 65 | -162.4 | -165.1 | 6 | <i>P1</i> | 3.701 | 5.762 | 6.172 | 102.43 | 94.7 | 91.35 | 3.8 | sheet 1 |
| A2_2_e da/110 | 46 | -162.3 | -164.7 | 8 | <i>Cc</i> | 11.934 | 7.403 | 10.183 | 90 | 144.497 | 90 | 1.5 | sheet 1 |
| A3_1_f ad/17 | 25 | -162.3 | -164.7 | 7 | <i>Pc</i> | 5.04 | 7.413 | 7.007 | 90 | 93.38 | 90 | 1.1 | sheet 1 |
| A3_1_d aa/93 | 66 | -162.2 | -163.8 | 13 | <i>P1</i> | 4.895 | 4.995 | 5.361 | 92.44 | 92.01 | 90.78 | 5.5 | sheet 1 |
| A1_2_b ah/65 | 64 | -162.1 | -165.2 | 5 | <i>P2₁</i> | 3.625 | 12.149 | 5.802 | 90 | 91.11 | 90 | 1.7 | not sheet |
| A3_1_f ad/31 | 1 | -162.0 | -163.3 | 16 | <i>P2₁</i> | 4.716 | 5.261 | 10.623 | 90 | 96.65 | 90 | 7.5 | not sheet |
| A3_1_c da/66 | 17 | -161.9 | -163.2 | 17 | <i>Cc</i> | 6.998 | 7.406 | 10.113 | 90 | 93.09 | 90 | 1.2 | sheet 1 |
| A2_4_b da/75 | 51 | -161.9 | -163.0 | 18 | <i>Cc</i> | 12.858 | 6.906 | 6.789 | 90 | 119.13 | 90 | 1.7 | sheet 1 |
| A2_2_d da/97 | 29 | -161.4 | -162.7 | 22 | <i>Cc</i> | 7.186 | 7.012 | 10.349 | 90 | 90.88 | 90 | 1.5 | sheet 2 |
| A2_2_b da/84 | 41 | -161.3 | -162.4 | 23 | <i>Cc</i> | 12.66 | 6.796 | 7.194 | 90 | 123.67 | 90 | 1.1 | sheet 1 |
| A2_2_b ab/123 | 63 | -161.1 | -163.9 | 11 | <i>P-1</i> | 6.92 | 6.871 | 6.911 | 104.06 | 109.32 | 113.49 | 3.0 | sheet 1 |
| A3_1_e ad/104 | 11 | -161.1 | -163.4 | 14 | <i>Pc</i> | 3.811 | 7.475 | 9.066 | 90 | 91.11 | 90 | 2.8 | sheet 1 |
| A2_3_e de/8 | 21 | -160.9 | -164.3 | 9 | <i>C2/c</i> | 11.535 | 7.57 | 15.747 | 90 | 131.81 | 90 | 0.6 | sheet 1 |
| A2_4_b ai/7 | 15 | -160.8 | -163.9 | 12 | <i>P2₁/c</i> | 7.227 | 6.946 | 10.125 | 90 | 92.25 | 90 | 1.8 | not sheet |
| A1_3_d dc/13 | 40 | -160.8 | -163.3 | 15 | <i>C2/c</i> | 11.598 | 7.55 | 11.786 | 90 | 93.94 | 90 | 2.8 | sheet 1 |
| A2_4_a fa/78 | 34 | -160.8 | -163.0 | 19 | <i>P2₁/c</i> | 6.705 | 11.706 | 9.281 | 90 | 130.7 | 90 | 2.9 | not sheet |
| A3_1_e da/57 | 12 | -160.8 | -162.8 | 21 | <i>Cc</i> | 11.676 | 7.482 | 17.437 | 90 | 160.208 | 90 | 2.4 | sheet 1 |
| A2_2_c am/86 | 5 | -160.7 | -163.0 | 20 | <i>P2₁/c</i> | 6.935 | 12.468 | 6.966 | 90 | 120.13 | 90 | 1.4 | sheet 2 |
| A1_1_a ab/19 | 50 | -160.7 | -165.3 | 3 | <i>P2₁/m</i> | 6.584 | 5.775 | 7.065 | 90 | 101.8 | 90 | 0.4 | sheet expt |

^a Structures are denoted by the acceptor, the placement of the water, and the orientation of the water, MOLPAK packing type and number of one example of the cluster, with N_{find} being the number of searches (max 66) in which this crystal structure was found at least once. ^b An estimate of the Helmholtz free energy at 298 K derived from the second derivative properties at the lattice energy minimum, with free energy rank being the order of this stability. ^c The lowest eigenvalue of the shear submatrix of the elastic tensor. Bold type corresponds to the experimental structure (ExpMinOpt).

**Figure 4.** Hydrogen bonding in hypothetical sheets 1 and 2. Orange dashed lines show hydrogen bonds out of the plane of the sheet to molecules in the sheet below.

if fragments of the observed sheet formed during the nucleation process. However, the idea that water of hydration gets trapped into the hydrate structure is less compelling for this sheet than for a structure in which the water had multipoint hydrogen bond contacts to a single solute molecule.⁷³ Thus, we should assess the uncertainties in calculating the relative thermodynamic stabilities of the structures in Table 6. The application of the rigid body approximation is a significant limitation even for the hydrate molecules considered here, as demonstrated by the ExpMinOpt vs ExpMinExp energies in Tables 3 and 5. While methods of ab initio optimization of the molecular structure under the influence of the packing forces have been successfully applied for variations in single bond torsion angles to optimize similar strength intermolecular hydrogen bonds,⁷⁴ getting the inter- and intramolecular energy balance correct for these more rigid molecules would be challenging, particularly because of the limited confidence that can be

**Figure 5.** Overlay of the experimental ExpMinOpt sheet with the predicted sheet 1 motif. Carbon and hydrogen atoms are colored gray and white in both structures, with oxygen red and nitrogen blue in the experimental ExpMinOpt and green and purple in the lowest energy sheet 1 structure, A2_1_d aa/82.

placed in the intermolecular potential used. While the superiority of the water...water model to simple models developed for liquid simulation has been clearly demonstrated for lattice energy minimization modeling, the repulsion–dispersion potential has a poor basis compared with more modern water potentials with more complex potential forms. The close carbonyl...carbonyl head on contact found in the most stable predicted structures is unlikely to have been sampled during the initial empirical parameter fitting to oxohydrocarbons or during subsequent validation in predominantly heteroatomic environments within organic crystal structures. Since this oxygen repulsion potential

and all the water...organic molecule interactions are derived from the geometric combining rules, with their poor theoretical basis,⁶³ it is encouraging that this starting point for development of the model potential has worked so well. The predominant reason for this success is surely that the distributed multipoles electrostatic model reflects the charge distribution of the molecules realistically. Developing more accurate intermolecular potentials including the effect of polarization,⁷⁵ and appropriately modeling molecular flexibility will require considerably more research as well as computational resources for their implementation. However, it is plausible that improvements in the model potential and allowing the water and 5-azauracil geometries to relax under the crystal packing forces may well stabilize the observed structure over the alternatives.

There is no reason to expect that the intermolecular potentials developed in this study will not be suitable for similar studies to suggest possible hydrate structures for other small organic molecules, since so many experimental hydrate structures are adequately reproduced. The success of the search strategy employed for 5-azauracil suggests that it could be thoughtfully adapted to a range of other molecules with competing hydrogen bond donors and acceptors to investigate ordered hydrogen-bonded hydrate structures. Indeed, far fewer input clusters would be required to contrast different possible hydrate hydrogen-bonding motifs as a complement to an experimental solid form screen.⁷⁶ More mathematically complete search algorithms for two (or more) independent molecules in the asymmetric unit cell^{23,24} would avoid the need to carefully consider the hydrogen-bonding possibilities for the molecules' functional groups and be capable of finding a wider range of structures. Indeed the crystal structure of 5-azauracil monohydrate can be readily found with the search methods of van Eijck²⁵ and Karamertzanis,²³ although the relative lattice energies of the structures found with point charge electrostatic models are significantly reordered when reminimized with the distributed multipole model developed in this study.

Although this work represents progress in the prediction of the crystal structures of a specific type of hydrate, it clearly demonstrates that the accuracy of the lattice energy modeling is not capable of predicting whether a hydrate will form on energetic grounds. The comparison of the lattice energies of the hydrates compared with the extrapolation from the corresponding lattice energies for the anhydrate and ice (Table 5) does not suggest such a marked stabilization compared with the uncertainties in the thermodynamic modeling that the formation of the observed hydrate is clearly predicted. However, this analysis has necessarily been limited to molecules where both the hydrate and anhydrate crystallize sufficiently readily to produce single crystals suitable for X-ray determination.

5. Conclusion

A methodology currently used for organic crystal structure prediction has been adapted to monohydrates and shown to be successful in generating plausible 5-azauracil monohydrate structures with the known crystal structure found

within 5 kJ mol⁻¹ of the global minimum in lattice energy, 1 kJ mol⁻¹ in free energy. The approach is certainly suitable for determining the range of plausible hydrogen-bonding motifs for a hydrate of a rigid polar molecule. However, further developments in the thermodynamic modeling are required to predict the formation of a hydrate in preference to an anhydrate. Nevertheless, computational studies clearly have potential in understanding hydrate formation.

The computed crystal structures for 5-azauracil monohydrate are stored on CCLRC e-Science Centre data portal and are available from the authors.

Acknowledgment. ATH was funded by the Basic Technology program of the Research Councils UK through the project 'Control and Prediction of the Organic Solid State'. Dr. Panagiotis Karamertzanis is thanked for useful discussions and computational assistance, Dr. Bouke van Eijck for contrasting search methods, and Dr. Louise Price for assistance in preparing the manuscript.

Supporting Information Available: CSD analysis of hydrogen bonding of water to specific acceptors, detailed lattice energy minimizations for the ice and hydrate structures, an analysis of the occurrence of the low-energy structures from the different bimolecular cluster searches, and a list of the final potential parameters. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Threlfall, T. L. Analysis of Organic Polymorphs - A Review. *Analyst* **1995**, *120*, 2435–2460.
- (2) Morris, K. R. Structural aspects of hydrates and solvates. In *Polymorphism in Pharmaceutical Solids*; Brittain, H. G., Ed.; Marcel Dekker, Inc.: New York, 1999; Vol. 95, pp 125–181.
- (3) Infantes, L.; Chisholm, J.; Motherwell, S. Extended motifs from water and chemical functional groups in organic molecular crystals. *CrystEngComm* **2003**, *5*, 480–486.
- (4) Khankari, R. K.; Grant, D. J. W. Pharmaceutical Hydrates. *Thermochim. Acta* **1995**, *248*, 61–79.
- (5) Byrn, S. R.; Pfeiffer, R. R.; Stowell, J. G. *Solid-state Chemistry of Drugs*; 2nd ed.; SSCI Inc.: West Lafayette, IN, 1999.
- (6) Price, S. L. The computational prediction of pharmaceutical crystal structures and polymorphism. *Adv. Drug Delivery Rev.* **2004**, *56*, 301–319.
- (7) Infantes, L. Water Clusters in Organic Molecular Crystals. *CrystEngComm* **2002**, *4*, 454–461.
- (8) Etter, M. C.; MacDonald, J. C.; Bernstein, J. Graph-Set Analysis of Hydrogen-Bond Patterns in Organic Crystals. *Acta Crystallogr., Sect. B: Struct. Sci.* **1990**, *46*, 256–262.
- (9) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380–388.
- (10) Gillon, A. L.; Feeder, N.; Davey, R. J.; Storey, R. Hydration in molecular crystals - A Cambridge Structural Database analysis. *Cryst. Growth Des.* **2003**, *3*, 663–673.
- (11) Desiraju, G. R. Hydration in Organic Crystals: Prediction from Molecular Structure. *Chem. Commun.* **1991**, 426–428.

- (12) Infantes, L.; Fabian, L.; Motherwell, W. D. S. Organic crystal hydrates: what are the important factors for formation. *CrystEngComm* **2007**, *9*, 65–71.
- (13) Thanki, N.; Thornton, J. M.; Goodfellow, J. M. Distributions of Water Around Amino-Acid Residues in Proteins. *J. Mol. Biol.* **1988**, *202*, 637–657.
- (14) Gorbitz, C. H.; Hersleth, H. P. On the inclusion of solvent molecules in the crystal structures of organic compounds. *Acta Crystallogr., Sect. B: Struct. Sci.* **2000**, *56*, 526–534.
- (15) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *J. Phys. Chem.* **1996**, *100*, 7352–7360.
- (16) Guillot, B. A reappraisal of what we have learnt during three decades of computer simulations on water. *J. Mol. Liq.* **2002**, *101*, 219–260.
- (17) Abascal, J. L. F.; Sanz, E.; Fernandez, R. G.; Vega, C. A potential for the study of ices and amorphous water: TIP4P/Ice. *J. Chem. Phys.* **2005**, *122*, 234511.
- (18) Price, S. L. Quantifying intermolecular interactions and their use in computational crystal structure prediction. *CrystEngComm* **2004**, *6*, 344–353.
- (19) Stone, A. J.; Alderton, M. Distributed Multipole Analysis - Methods and Applications. *Mol. Phys.* **1985**, *56*, 1047–1064.
- (20) Millot, C.; Soetens, J. C.; Costa, M. T. C. M.; Hodges, M. P.; Stone, A. J. Revised anisotropic site potentials for the water dimer and calculated properties. *J. Phys. Chem. A* **1998**, *102*, 754–770.
- (21) Bukowski, R.; Szalewicz, K.; Groenenboom, G.; van der Avoird, A. Interaction potential for water dimer from symmetry-adapted perturbation theory based on density functional description of monomers. *J. Chem. Phys.* **2006**, *125*, 044301.
- (22) Karamertzanis, P. G.; Price, S. L. Challenges of crystal structure prediction of diastereomeric salt pairs. *J. Phys. Chem. B* **2005**, *109*, 17134–17150.
- (23) Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction - I. Rigid molecules. *J. Comput. Chem.* **2005**, *26*, 304–324.
- (24) Bazterra, V. E.; Thorley, M.; Ferraro, M. B.; Facelli, J. C. A Distributed Computing Method for Crystal Structure Prediction of Flexible Molecules: An Application to N-(2-Dimethyl-4,5-dinitrophenyl) Acetamide. *J. Chem. Theory Comput.* **2007**, *3*, 201–209.
- (25) van Eijck, B. P.; Kroon, J. Structure predictions allowing more than one molecule in the asymmetric unit. *Acta Crystallogr., Sect. B: Struct. Sci.* **2000**, *56*, 535–542.
- (26) Kamb, B. Ice II: A Proton-Ordered Form of Ice. *Acta Crystallogr.* **1964**, *17*, 1437–1449.
- (27) Kamb, B.; Hamilton, W. C.; LaPlaca, S. J.; Prakash, A. Ordered Proton Configuration in ice II, from Single-Crystal Neutron Diffraction. *J. Chem. Phys.* **1971**, *55*, 1934–1945.
- (28) Kuhs, W. F.; Finney, J. L.; Vettier, C.; Bliss, D. V. Structure and hydrogen ordering in Ices VI, VII, VIII by neutron powder diffraction. *J. Chem. Phys.* **1984**, *81*, 3612–3623.
- (29) La Placa, S. J.; Hamilton, W. C.; Kamb, B.; Prakash, A. On a nearly proton-ordered structure for ice IX. *J. Chem. Phys.* **1972**, *58*, 567–580.
- (30) Jackson, S. M.; Nield, V. M.; Whitworth, R. W.; Oguro, M.; Wilson, C. C. Single-Crystal Neutron Diffraction Studies of the Structure of Ice XI. *J. Phys. Chem. B* **1997**, *101*, 6142–6145.
- (31) Leadbetter, A. J.; Ward, R. C.; Clark, J. W.; Tucker, P. A.; Matsuo, T.; Suga, H. The equilibrium low-temperature structure of ice. *J. Chem. Phys.* **1985**, *82*, 424–428.
- (32) Salzmann, C. G.; Radaelli, P. G.; Hallbrucker, A.; Mayer, E.; Finney, J. L. The Preparation and Structures of Hydrogen Ordered Phases of Ice. *Science* **2006**, *311*, 1758–1761.
- (33) Finney, J. L. What's so special about water? *Philos. Trans. R. Soc. London, Ser. B* **2004**, *359*, 1145–1165.
- (34) Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. The Relaxation of Molecular Crystal Structures Using a Distributed Multipole Electrostatic Model. *J. Comput. Chem.* **1995**, *16*, 628–647.
- (35) Spek, A. L. *PLATON - a multipurpose crystallographic tool*; Utrecht University: Utrecht, 2001.
- (36) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Dordrecht, 1981; p 331.
- (37) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (38) Boulougouris, G. C.; Economou, I. G.; Theodorou, D. N. Engineering a Molecular Potential for Water Phase Equilibrium over a Wide Temperature Range. *J. Phys. Chem. B* **1998**, *102*, 1029–1035.
- (39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (40) Errington, J. R.; Panagiotopoulos, A. Z. A Fixed Point Charge Model for Water Optimised to the Vapour-Liquid Coexistence Properties. *J. Phys. Chem. B* **1998**, *102*, 7470–7475.
- (41) Cox, S. R.; Hsu, L. Y.; Williams, D. E. Nonbonded Potential Function Models for Crystalline Oxohydrocarbons. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1981**, *37*, 293–301.
- (42) Beyer, T.; Price, S. L. Dimer or catemer? Low-energy crystal packings for small carboxylic acids. *J. Phys. Chem. B* **2000**, *104*, 2647–2655.
- (43) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles From Molecular Electrostatic Potentials - the Need For High Sampling Density in Formamide Conformational-Analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.
- (44) Filippini, G.; Gavezzotti, A. Empirical Intermolecular Potentials For Organic-Crystals: the '6-Exp' Approximation Revisited. *Acta Crystallogr., Sect. B: Struct. Sci.* **1993**, *49*, 868–880.
- (45) Williams, D. E.; Cox, S. R. Nonbonded Potentials For Azahydrocarbons: the Importance of the Coulombic Interaction. *Acta Crystallogr., Sect. B: Struct. Sci.* **1984**, *40*, 404–417.
- (46) Williams, D. E.; Houpt, D. J. Fluorine Nonbonded Potential Parameters Derived From Crystalline Perfluorocarbons. *Acta Crystallogr., Sect. B: Struct. Sci.* **1986**, *42*, 286–295.
- (47) Price, S. L.; Price, L. S. Modelling Intermolecular Forces for Organic Crystal Structure Prediction. In *Intermolecular Forces and Clusters I*; Wales, D. J., Ed.; Springer-Verlag: Berlin, Heidelberg, Germany, 2005; pp 81–123.

- (48) Hulme, A. T.; Price, S. L.; Tocher, D. A. A New Polymorph of 5-Fluorouracil Found Following Computational Crystal Structure Predictions. *J. Am. Chem. Soc.* **2005**, *127*, 1116–1117.
- (49) Hulme, A. T.; Tocher, D. A. The Discovery of New Crystal Forms of 5-Fluorocytosine Consistent with the Results of Computational Crystal Structure Prediction. *Cryst. Growth Des.* **2006**, *6*, 481–487.
- (50) Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R. Tables of Bond Lengths Determined By X-Ray and Neutron-Diffraction .1. Bond Lengths in Organic-Compounds. *J. Chem. Soc., Perkin Trans. 2* **1987**, S1–S19.
- (51) Frisch, M. J. et al. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2003.
- (52) Stone, A. J. *GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System, version 1.0*; University of Cambridge: Cambridge, United Kingdom, 1999.
- (53) Potter, B. S.; Palmer, R. A.; Withnall, R.; Chowdhry, B. Z.; Price, S. L. Aza analogues of nucleic acid bases: experimental determination and computational prediction of the crystal structure of anhydrous 5-azauracil. *J. Mol. Struct.* **1999**, *486*, 349–361.
- (54) Gray, A. E.; Day, G. M.; Leslie, M.; Price, S. L. Dynamics in crystals of rigid organic molecules: contrasting the phonon frequencies calculated by molecular dynamics with harmonic lattice dynamics for finidazole and 5-azauracil. *Mol. Phys.* **2004**, *102*, 1067–1083.
- (55) Holden, J. R.; Du, Z. Y.; Ammon, H. L. Prediction of Possible Crystal-Structures For C-, H-, N-, O- and F-Containing Organic Compounds. *J. Comput. Chem.* **1993**, *14*, 422–437.
- (56) Chisholm, J. A.; Motherwell, S. COMPACK: a program for identifying crystal structure similarity using distances. *J. Appl. Crystallogr.* **2005**, *38*, 228–231.
- (57) de Gelder, R.; Wehrens, R.; Hageman, J. A. A generalized expression for the similarity of spectra: Application to powder diffraction pattern classification. *J. Comput. Chem.* **2001**, *22*, 273–289.
- (58) van de Streek, J.; Motherwell, S. Searching the Cambridge Structural Database for polymorphs. *Acta Crystallogr., Sect. B: Struct. Sci.* **2005**, *61*, 504–510.
- (59) Day, G. M.; Price, S. L.; Leslie, M. Elastic constant calculations for molecular organic crystals. *Cryst. Growth Des.* **2001**, *1*, 13–26.
- (60) Day, G. M.; Price, S. L.; Leslie, M. Atomistic calculations of phonon frequencies and thermodynamic quantities for crystals of rigid organic molecules. *J. Phys. Chem. B* **2003**, *107*, 10919–10933.
- (61) Anghel, A. T.; Day, G. M.; Price, S. L. A study of the known and hypothetical crystal structures of pyridine: why are there four molecules in the asymmetric unit cell? *CrystEngComm* **2002**, *4*, 348–355.
- (62) Buckingham, A. D.; Fowler, P. W.; Stone, A. J. Electrostatic Predictions of Shapes and Properties of Vanderwaals Molecules. *Int. Rev. Phys. Chem.* **1986**, *5*, 107–114.
- (63) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, U.K., 1996.
- (64) Gavezzotti, A. *Molecular Aggregation: Structure Analysis and Molecular Simulation of Crystals and Liquids*; Oxford University Press: Oxford, U.K., 2007; Vol. 19.
- (65) Whalley, E. The difference in the intermolecular forces on H₂O and D₂O. *Trans. Faraday Soc.* **1957**, *53*, 1578–1585.
- (66) Motherwell, W. D. S.; Shields, G. P.; Allen, F. H. Visualization and characterization of non-covalent networks in molecular crystals: automated assignment of graph-set descriptors for asymmetric molecules. *Acta Crystallogr., Sect. B: Struct. Sci.* **1999**, *55*, 1044–1056.
- (67) Hulme, A. T. Experimental and Computational Studies of Polymorphism of Small Organic Molecules, Ph.D. Thesis, University College London, 2007.
- (68) Rowland, R. S.; Taylor, R. Intermolecular Nonbonded Contact Distances in Organic-Crystal Structures - Comparison With Distances Expected From Van-Der-Waals Radii. *J. Phys. Chem.* **1996**, *100*, 7384–7391.
- (69) Allen, F. H.; Baalham, C. A.; Lommerse, J. P. M.; Raithby, P. R. Carbonyl-carbonyl interactions can be competitive with hydrogen bonds. *Acta Crystallogr., Sect. B: Struct. Sci.* **1998**, *54*, 320–329.
- (70) Coombes, D. S.; Nagi, G. K.; Price, S. L. On the lack of hydrogen bonds in the crystal structure of alloxan. *Chem. Phys. Lett.* **1997**, *265*, 532–537.
- (71) Day, G. M.; Motherwell, W. D. S. An experiment in crystal structure prediction by popular vote. *Cryst. Growth Des.* **2006**, *6*, 1985–1990.
- (72) Day, G. M. et al. A third blind test of crystal structure prediction. *Acta Crystallogr., Sect. B: Struct. Sci.* **2005**, *61*, 511–527.
- (73) Nangia, A.; Desiraju, G. R. Pseudopolymorphism: occurrences of hydrogen bonding organic solvents in molecular crystals. *Chem. Commun.* **1999**, 605–606.
- (74) Karamertzanis, P. G.; Price, S. L. Energy Minimization of Crystal Structures Containing Flexible Molecules. *J. Chem. Theory Comput.* **2006**, *2*, 1184–1199.
- (75) Stone, A. J.; Misquitta, A. J. Atom-atom potentials from ab initio calculations. *Int. Rev. Phys. Chem.* **2007**, *26*, 193–222.
- (76) Johnston, A.; Florence, A. J.; Shankland, N.; Kennedy, A. R.; Shankland, K.; Price, S. L. Crystallization and crystal energy landscape of hydrochlorothiazide. *Cryst. Growth Des.* **2007**, *7*, 705–712.

CT700045R