Applications of DL_POLY and DL_MULTI to organic molecular crystals

Sarah Lois Price, Said Hamad, Antonio Torrisi, Panagiotis Karamertzanis, Maurice Leslie, Richard Catlow

▶ **To cite this version:**

Sarah Lois Price, Said Hamad, Antonio Torrisi, Panagiotis Karamertzanis, Maurice Leslie, et al.. Applications of DL_POLY and DL_MULTI to organic molecular crystals. Molecular Simulation, 2007, 32 (12-13), pp.985-997. 10.1080/08927020600880810 . hal-00514994

HAL Id: hal-00514994
https://hal.science/hal-00514994

Submitted on 4 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Molecular Simulation
Journal of
Experimental Nanoscience

Taylor & Francis
Taylor & Francis Group

Applications of DL_POLY and DL_MULTI to organic molecular crystals

Journal:	*Molecular Simulation/Journal of Experimental Nanoscience*
Manuscript ID:	GMOS-2006-0082
Journal:	Molecular Simulation
Date Submitted by the Author:	11-May-2006
Complete List of Authors:	Price, Sarah; University College London, Chemistry Hamad, Said; Royal Insitution of Great Britain Torrisi, Antonio; Royal Insitution of Great Britain Karamertzanis, Panagiotis; University College London, Chemistry Leslie, Maurice; CCLRC Daresbury Laboratory Catlow, Richard; University College London, Chemistry; Royal Institution of Great Britain; Royal Institution of Great Britain
Keywords:	organic solid state, polymorphism, distributed multipoles

For Review Only

SCHOLARONE™
Manuscripts

Title: **Applications of DL_POLY and DL_MULTI to organic molecular crystals**

Authors: Sarah L. Price[1], Said Hamad[2], Antonio Torrisi[2], Panagiotis G. Karamertzanis[1], Maurice Leslie[3] and C. Richard A. Catlow[1,2]

[1] Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, Tel: 020 7679 4622, Fax: 020 7679 7463, s.l.price@ucl.ac.uk; p.karamertzanis@ucl.ac.uk; c.r.a.catlow@ucl.ac.uk

[2] The Royal Institution of Great Britain, 21 Albemarle Street, London W1S 4BS, Tel: 020 7670 2901, Fax: 020 7629 3569, said@ri.ac.uk; antonio@ri.ac.uk; richard@ri.ac.uk

[3] CCLRC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, Tel: 01925 603507, Fax: 01925 603634; m.leslie@dl.ac.uk

Word Count: 9936

For Peer Review Only

Applications of DL_POLY and DL_MULTI to organic molecular crystals

S.L. Price, S. Hamad, A Torrisi, P.G. Karamertzanis, M. Leslie and C.R.A. Catlow

Abstract

Molecular Dynamics simulations are capable of giving considerable insight into the polymorphism of organic molecules, a problem of major concern to the pharmaceutical and other speciality chemicals industries. We illustrate some of the challenges involved in small organic systems which have complex solid state phase behaviour, including characterising rotationally disordered phases, modelling polymorphs with very different hydrogen bonding motifs and explaining the solvent dependence of a polymorphic system. Simulating the dynamics within the organic solid state can be very demanding of the model for the weak forces between the molecules. This has led to the development of DL_MULTI so that a distributed multipole electrostatic model can be used to describe the orientation dependence of hydrogen bonding and π–π stacking more realistically. Once a simulation is correctly reproducing the known crystal structures, there are also considerable system-specific challenges in extracting novel insights from the MD simulations.

Keywords: organic solid state, polymorphism, distributed multipoles,

Introduction

The ability of chemists to synthesize an almost infinite range of organic molecules has led to a huge diversity in the organic systems of industrial interest, such as

For Peer Review Only

pharmaceuticals and pigments. For some speciality chemicals, such as energetic materials, organic conductors and non-linear optics, the desired physical properties are so critically dependent on the crystal packing, that the ability to predict the crystal structure prior to synthesis would be a great aid to the design of new materials. For all organic materials, the possibility of polymorphism, the adoption of more than one crystalline form, is a major quality control problem, as different polymorphs can have very different physical properties, such as dissolution rates. Hence pharmaceuticals are only licensed in a specific polymorphic form, and trying to establish that all polymorphs are known and can be controllably produced is an important part of the drug development. Since new polymorphs can appear after decades of manufacture [1] and indeed, an anti-HIV pharmaceutical ritonavir had to be urgently reformulated after a new more stable polymorph appeared in the manufacturing process [2], a computational method of predicting polymorphism [3] would greatly aid the solid form development of pharmaceuticals.

Considerable effort has gone into the development of methods of organic crystal structure prediction, as evidenced by the international blind tests organized by the Cambridge Crystallographic Data Centre [4-6]. The majority of successful methods have been based on searches for the global minimum in the lattice energy of the molecular crystals. However, for many molecules, such searches produce many more energetically feasible structures than known polymorphs. Whilst one of the computed hypothetical structures predicted for both aspirin [7] and paracetamol [8] have later been found experimentally [9,10], in general, it is likely that the majority of the low energy unobserved structures will not correspond to polymorphs because they may be

For Peer Review Only

kinetically inaccessible, or are not minima in the free energy at experimental crystallization temperatures.

Thus, realistic Molecular Dynamics simulations at accessible temperatures have the potential to make a considerable contribution to our ability to understand the polymorphism of organic crystals. In this article, we outline some early attempts to use Molecular Dynamics studies to understand different aspects of the solid state phase behaviour of a few simple molecules (Figure 1). One key theme is how to build on the considerable previous work on the solid state simulation of small polyatomic and more symmetrical organic molecules [11] to extract the information that is relevant to the specific polymorphism problem for more complex molecules. An important requirement for polymorphism studies is that the force-field has to be equally realistic for all phases, as it is the relative stability that is important. Using a distributed multipole model for the molecular charge distribution does produce a significant improvement in the modelling of organic crystal structures and their relative energies [12,13]. Hence DL_POLY has been developed to use more realistic distributed multipole models for the electrostatic forces between rigid organic molecules, and some of the first studies using DL_MULTI [14] are outlined. This paper presents case studies - simulations of cyclopentane, glycine, imidazole, 5-azauracil and 5-fluorouracil – which illustrate the present status of the application of dynamical simulations in the study of organic crystals.

Simulating order-disorder phase changes in cyclopentane.

For Peer Review Only

Cyclopentane, C_5H_{10}, is one of the simplest organic molecules, with only two atomic types C and H, and weak intermolecular forces. The Molecular Dynamics study was undertaken following an experimental determination of the phase diagram by X-ray powder diffraction [15]. The low temperature ordered phase III was solved from the powder data simultaneously with a single crystal X-ray diffraction at low temperature [16]. The high temperature form I was clearly of hexagonal symmetry [17]. However, there was an intermediate phase II over the temperature range 120-132 K, whose unit cell could not be indexed, but was likely to be large, with the possibility of the structure being incommensurate.

Form I was clearly rotationally disordered, possibly a plastic phase, of the type simulated by MD for SF_6 [18], adamantane [19] and cubane [20]. The challenge was to perform simulations that could reproduce the ordered form III and form I sufficiently well to provide confidence that qualitative insights into form II could be obtained if this were seen in the simulations. The crystal structure prediction lattice energy minima search had form III at or close to the global minimum (depending on the electrostatic model used) but there were a huge number of alternative structures within a small energy range. (For example, the search with a distributed multipole electrostatic model had about 10 structures in the 0.24 kJ mol^{-1} between the known form III and the global minimum in lattice energy.) This certainly indicates that a large variety of ordered crystals is thermodynamically feasible, which may also indicate a very flat free energy surface for different types of disorder, either static or dynamic depending on the barriers for reorientation.

For Peer Review Only

Initial simulations used a rigid molecule, optimized at the MP2/6-31G** level of theory and the corresponding CHELPG [21] potential derived charges in conjunction with a simple atom-atom potential (*6-exp* form with the parameters fitted to organic crystal structures and heats of sublimation [22]). The phase change behaviour was simulated in the NST ensemble with DL_POLY [23] for a box of 180 molecules corresponding to a 3x5x3 supercell of the predicted lattice energy minima with the same potential. This gave a monoclinic cell, with unique angle of 70 ° and cell lengths of about 28-30 Å. Initially simulations were carried out for 1 ns (with 50 ps equilibration) at temperature intervals of 10 K in the range 100-170 K. It appeared (Figure 2) that form III was being simulated between 100 and 120 K, and hexagonal phase I between 140 K and 170 K, with signs of an intermediate phase or transitional behaviour between the two. The average cell parameters with temperature showed good agreement with the experimentally determined cell constants for forms I (after hexagonal to orthorhombic transformation) and III. (Table 1). The monoclinic to hexagonal transition takes place with the monoclinic angle β changing from 113 ° to 90 °, and by a simultaneous change of the *b* and *c* cell constants to have a ratio $c = \sqrt{3}\, a$.

Thus, to link to the experimental data, it was necessary to simulate the powder pattern corresponding to the MD simulations. To provide insight into form II, we needed to contrast the dynamic motions of the molecules in the different phases, quantifying the behaviour observed in the simulation movies generated by Materials Studio [24] from the history files generated on the HPC(x). Both forms of analysis become ill-defined in NST simulations because of the fluctuations in the units cell parameters. Hence, a snapshot of the simulation with as close to the average cell dimensions as possible,

For Peer Review Only

was used for the initial configuration in an NVE simulation for each temperature of interest. The powder patterns were simulated by first calculating the structure factor F_{hkl} for each snapshot by summing the product of the angle-dependent atomic scattering factors with the exponential function of the *hkl* and atomic fraction coordinates over all the atoms in the simulation cell. The intensity $I_{hkl}(\theta)$ is then calculated from the structure factor and Lorentz-polarization factor. Finally, these intensities were averaged over all the timesteps of the equilibrated simulation. These powder patterns gave good agreement with the experimental powder patterns for forms I and III. The simulations agreed with the experiment to the extent that there was a distinct and significantly more complex powder pattern in the small region that appeared to be form II, but the detailed agreement was poor. This was not unexpected, as it would be impossible to represent an incommensurate phase in such an MD simulation box, as well as the many other limitations of the simulations.

The simulation movies clearly show that, in the ordered phase III, the molecules were just vibrating around their crystallographic positions (Figure 2). In the high temperature form I, the molecules rotate to give this high symmetry structure, but the time average is far from spherical, and the motion is not continuous, with most molecules undergoing very large amplitude librations and frequent complete flips. In the intermediate transitional phase II, a smaller proportion of molecules were rotating and the phase seemed closer to being statically disordered.

Overall, these simulations give an atomistic level insight into the behaviour of cyclopentane, clearly showing the transition between the ordered form III and high symmetry phase I, is complicated, and that there are intermediate stages to the

For Peer Review Only

disorder that are apparent experimentally as form II. It is perhaps surprising that such a simple model should simulate the ordered and disordered phases (III and I) and the temperature range of the transition so well. Further details will be reported subsequently.

Simulating the polymorphs of glycine

Here we review recent simulation studies of glycine which emphasize the sensitivity of the results to the parameters and form of the potential employed. Glycine is at the other extreme of neutral organic molecules, in that it is a zwitterion in all its polymorphs and hence the intermolecular forces are dominated by the electrostatic term. The flexibility of amino acids around C_α is also apparent in the polymorphs of this smallest example, and so needs to be modelled by a combined inter and intra-molecular force-field. Fortunately, force-fields for polypeptide simulations have been extensively developed [25-27]. Three polymorphs of glycine have been widely studied and have very different hydrogen bonding motifs (Figure 3). The β form is metastable and undergoes a solvent mediated transformation to the more stable α form [28]. The α and γ forms are more stable, with the γ form believed to be the most stable at room temperature [29,30], though most crystallization conditions lead to the α form. The two structures differ significantly, in that the α structure is based on hydrogen bonded dimers, whereas the hydrogen bonded sheets in the γ form pack in a polar crystal [31,32]. A wide range of studies of the different experimental conditions that cause the formation of phases have been reported, with factors such as pH and additives as well as solvent influencing the polymorphic outcome [31-33]. Recently crystallizations under pressure [34, 35] have found further polymorphs. Calculations

For Peer Review Only

on the known and some hypothetical structures of glycine using Density Functional methods [36] reveal the difficulty in modelling the structures and their relative stabilities.

In order to investigate the factors which influence the polymorphic outcome of crystallization of glycine, we require a model force-field that can reproduce the three atmospheric pressure polymorphs equally well, before it can be combined with solvent models. The electrostatic term would be expected to be the most influential in modelling the crystal structures well, and so a variety of point charge models specifically derived for glycine were investigated in conjunction with the AMBER force-field for all other terms. An atomic point charge model for glycine cannot be simply constructed by fitting to the electrostatic potential around the *ab initio* charge density of glycine (e.g. CHELPG charges [37]) as would normally be done for organic molecules, because the zwitterionic charge density rearranges to the neutral form for the isolated molecule. Hence sets of charges were obtained for the zwitterions, either stabilized by a surrounding dielectric continuum to represent water, or in a cluster with a few water molecules. In the latter case, Natural Bond Orbital charges were investigated as well as CHELPG charges, since the fitting of the latter was liable to errors in portioning the potential between the zwitterions and the water molecules. Since most of these methods gave rise to inequivalent charges on the different hydrogen and oxygen atoms, which seemed inappropriate for NH_3^+ and CO_2^- groups where the conformation was changing between polymorphs, we derived models in which the average charges for these groups were used on the H and O atoms were also investigated.

For Peer Review Only

The various charge models were used in conjunction with the bond-stretching, angle-bending, torsion and non-bonded parameters of the AMBER force-field [25]. In this force-field, the charge-charge interaction between atoms separated by 3 bonds (1-4 interactions) are scaled by 1.2, and the corresponding non-bonded terms by 2, which required careful implementation in DL_POLY by entering additional parameters to the DIHEDRALS parameterization. This scaling makes a very significant difference to the conformations seen in the simulations. The choice of supercells for equivalent MD NPT simulations of different polymorphs whose unit cells differ significantly in dimensions is problematic as these monoclinic (α and β) and trigonal (γ) cells have different numbers of molecules per unit cell. In the end, for these simulations aimed at potential testing, we chose simulation boxes of 4x4x4 unit cells for β (128 molecules) and γ (192 molecules) and 6x3x6 unit cells for α (432 molecules) to ensure that the all simulation box edges were at least 20 Å, to allow a reasonable van der Waals cutoff of 10 Å for the direct summation. The electrostatic terms were summed by Ewald summation, which assumes that the net cell dipole of the polar β and γ forms is effectively neutralized. (This is probably the safest assumption, given the debate as to whether the surface dipole correction term which can destabilize polar crystals should be included in the energy [38-40]).

The results of the simulation were certainly sensitive to the choice of the charges, and the relative reproduction of the three phases varied with the choice of model. The best simulation results (Figure 3) with the *ab initio* glycine-derived charges were quite satisfactory for the α and γ phases, but in the β phase the molecules rotated to form additional hydrogen bonds, with a significant change in conformation. Finally, a series of simulations were attempted using the published AMBER charges for the

For Peer Review Only

NH_3^+ and CO_2^- terminal groups, with charges for the CH_2 group hydrogen atoms being taken from glycine amino-acid residue, and the charge on C_α being adjusted to give a neutral zwitterion. These simulations worked very satisfactorily (Table 2), even down to a reasonable prediction of the relative stability of the three phases. This superiority of the published AMBER potential really underlies the extent to which empirically derived force-fields include a considerable degree of cancellation of errors between the terms that nominally represent different effects and absorb the energy terms which are not explicitly included. In the case of flexible molecule force-fields, a theoretical improvement in the intermolecular electrostatic interactions may well alter the torsional potential and result in a conformationally distorted molecule within the crystal [41]. Hence, in the case of peptides and other systems where a carefully parameterized force-field is available, it will be very difficult, for a model based on a better theoretical description for the individual terms, to be more reliable [42], at least in the regions of the potential sampled in the empirical parameterization. In the case of glycine, we are now proceeding with simulations to assess various theories for why different solution crystallization conditions can produce different polymorphs, with confidence in at least the solute-solute interactions.

DL_MULTI for using more realistic intermolecular potentials for polyatomic molecules.

DL_POLY simulations are restricted to the use of isotropic atom-atom potentials, which essentially assume that molecules interact with each other as if they were a superposition of spherical atomic charge densities. Although this procedure is often adequate, as in the study of glycine in the previous section, it neglects the orientation

For Peer Review Only

dependence of the intermolecular forces that can arise from non-spherical features in the molecular charge density such as lone pair and π electron density. It has long been recognized that this factor is important in modelling the directionality of hydrogen bonding and π–π interactions, since the structures of hydrogen bonded van der Waals complexes were reproduced [46] using a distributed multipole model for the electrostatic interactions. The tendency of hydrogen bonds to form along the direction of lone pair density in the acceptor, or to π density, cannot generally be reproduced by atomic charge or central multipole electrostatic models. A distributed multipole model generally represents the molecular charge density by sets of atomic charges, dipoles, octapoles and hexadecapoles, which are derived by various methods of partitioning the *ab initio* molecular charge density between atoms. The most widely used method is the Distributed Multipole Analysis (DMA) method [47,48], which uses the program GDMA [49] to analyse *ab initio* charge densities calculated using GAUSSIAN [50]. Although some organic crystals can be satisfactorily modelled using atomic charges, the reproduction of the structures of polar and hydrogen bonded molecules [51], the relative lattice energies in crystal structure prediction searches [52], and harmonic mode lattice frequencies [53] are usually improved by using this more theoretically accurate model for the electrostatic forces [12,13]. Thus a general method of modelling the intermolecular forces between organic molecules, that is sufficiently realistic for simulating the organic solid state of a range of organic molecules, will require the use of anisotropic atom-atom multipolar electrostatic models.

Hence, we have extended the general purpose MD simulation package DL_POLY [23] to simulate rigid organic models whose intermolecular interactions are described by a distributed multipole electrostatic model. The resulting program DL_MULTI

For Peer Review Only

[14] allows Molecular Dynamics calculations to be performed with the same models that are used in simulating molecular clusters and surfaces in the programme ORIENT [54] and organic crystal structures and properties by static lattice energy minimization in the programme DMAREL [55,56]. The extension of DL_POLY to use anisotropic atom-atom interactions of the form dictated by the multipole expansion of the electrostatic energy, with their associated non-central forces and torques [56-58], essentially follows ORIENT and DMAREL, but with a few modifications outlined below that are necessary for the Molecular Dynamics method. One major difference necessary for MD simulations of the organic solid state, proved to be the need to develop the methodology for sufficiently accurate summation of the electrostatic energy contribution arising from the atomic dipoles, quadrupoles, octapoles and hexadecapoles. Normally, convergent long range forces, which decay as the inverse fourth or greater power of the interatomic separation would be evaluated by direct summation to a given cutoff distance. The tiny error introduced by two atoms moving so that their separation goes over the cutoff distance (U(cutoff)) cancels when the atoms move to being within the cutoff distance again if the potential only depends on the atomic separation. However, when the atoms interact by an anisotropic potential, the changes in the relative orientation whilst the atoms are separated by more than the cutoff distance, prevent this cancellation being exact. However, in the case of simulating crystal structures, the approximate translational symmetry within the simulation cell considerably exacerbates the problem. An efficient solution of this problem is to extend the Ewald summation techniques for all the terms in the multipolar expansion of the electrostatic energy, up to R^{-5}, which has been derived and programmed in DL_MULTI. The zero wavevector reciprocal lattice term in the Ewald sum is finite for the dipole-dipole term and needs to be included for fluids,

For Peer Review Only

although it should be omitted for polar solids as they can have a macroscopic electric field across the crystal. The effects of the higher multipole moments have been shown to have a significant effect in simulating the properties of liquid hydrogen fluoride [59] and water [59,60] and anyway allow for the use of more accurate potentials. Applications to the structural lattice dynamical properties of organic crystals are discussed below.

Simulations of dynamical motions in imidazole and 5-azauracil

Imidazole was chosen as a simple rigid molecule whose structure [61] and lattice modes (Far-Infrared [62] and Raman frequencies [63]) at 103 K were known. Furthermore, a previous study [64] had reported difficulties in simulating the crystal structures and phonon modes by Molecular Dynamics, despite using an intermolecular potential that had been fitted to the experimental lattice frequencies using lattice dynamics calculations. To provide a contrast to imidazole, whose monoclinic crystal structure is composed of hydrogen bonded chains, the crinkled hydrogen-bonded sheet crystal structure of 5-azauracil [65] was also simulated, but at room temperature. Both molecules were simulated as rigid molecules using the experimental molecular structure, with their electrostatic interactions modelled by the DMA of this MP2/6-31G(d,p) charge density, and all other terms represented by an isotropic atom-atom *exp-6* potential with parameters [51]. In both cases, this model predicts the known crystal structure as the global minimum in the lattice energy [65,66], with the hypothetical structures which are close in lattice energy comprised of similar hydrogen bonding motifs. The DL_MUTLI simulation cells were chosen as 5x5x4 unit cells for imidazole (400 molecules) and a 4x2x3 unit cells for 5-azauracil

For Peer Review Only

(192 molecules) to give a roughly isodimensional supercell, and started from the lattice energy minima (0 K structure) found using DMAREL [67] with the same molecular structure and intermolecular potential.

The first stage to test the adequacy of the simulation model [68] was to run an NST simulation at atmospheric pressure and 100 K for imidazole and 310 K for 5-azauracil. As shown (Table 3), the averaged Molecular Dynamics simulation gave reasonable reproductions of both crystal structures, though the cell volume was overestimated by 5% for imidazole and 7% for 5-azauracil. Since even the 0 K lattice energy minimum gave a larger cell than the experimental structures, this could be attributed to the use of a repulsion-dispersion potential whose parameters had been fitted to room temperature crystal structures. The hydrogen bonding motifs were well reproduced in the simulations, and even the thermal expansion (estimated by comparing the DL_MUTLI simulation cell and the lattice energy minimum) was very reasonable, being smaller in the cell directions with a large hydrogen bonding component. Thus the use of an anisotropic DMA based potential model was clearly justified in terms of providing a realistic simulation of both crystals. We now turn our attention to the modelling of lattice dynamical properties of these structures.

In order to contrast the Molecular Dynamics description of the molecular motions with harmonic lattice modes estimated by lattice dynamics, it was necessary to analyse the motions in the NVE ensemble to remove the fluctuations in the simulation cell. The initial structure for these simulations was an NST configuration, which was chosen to have nearly identical cell parameters to the average cell (Table 3), and so could be used to provide undistorted coordinates for the rigid molecules. In these

For Peer Review Only

NVE simulations the molecules move as dictated by the intermolecular potential, not its second derivatives at the minimum, though the motions are restricted by the periodic boundary conditions and can only be those whose *k* vectors fit into the supercell. This allowed 100 *k* vectors for imidazole but only 24 for 5-azauracil which has twice as many molecules in the crystallographic unit cell. In order to quantify the motions of the molecules about their crystallographic positions in the Molecular Dynamics simulations to compare with the lattice dynamics calculations and experimental results, we needed to extract the frequencies of the 45 optical k=0 modes for 5-azauracil and 21 for imidazole from the simulation, and characterize the symmetry of each mode. This was done by collecting the following data for each molecule every 5 timesteps for over 12000 timesteps of 0.003 ps of the equilibrated NVE simulations: the Cartesian positions of the centres of mass and the component of the translational velocity parallel to each Cartesian axis, the quaternions determining the orientation of the principal axes of each molecule and the angular velocities of each molecule about its principal axes of inertia. Symmetry analysis allowed the definition [68] of the normalized N-dimensional symmetry adapted normal mode coordinate for each specific symmetry representation and specific degree of freedom in terms of the masses, moments of inertia, velocities and angular velocities of the N molecules in the simulation cell. The calculation of the power spectrum of the Fourier transform of the symmetry adapted velocity autocorrelation functions provided the normalized phonon densities of states for each symmetry representation and its contributions from the different rotations and translations of the molecules. As shown in Figure 4, the normalized phonon densities of states for each symmetry representation gave peaks which were quite sharp and close to the frequencies calculated by harmonic lattice dynamics, though both the width of the peaks and the

For Peer Review Only

shifts in the frequencies were larger for the room temperature simulation of 5-azauracil. For imidazole, most of the modes were shifted by less than 5 cm^{-1} to lower frequencies by the Molecular Dynamics allowing for the anharmonicity in the intermolecular interactions, though some of the hydrogen bond bending modes were surprisingly shifted to slightly higher frequencies. Overall the rms error in comparison with experiment improved from 19.9 cm^{-1} for the harmonic modes to 18.8 cm^{-1} for the Molecular Dynamics modes, and it seems possible that the rigid body approximation may be a more significant contribution to the errors than the harmonic approximation. In contrast, the higher temperature simulation for 5-azauracil had most of the assigned peaks within 20 cm^{-1} of the harmonic values and were generally shifted to lower frequencies. Overall, these results [68] were encouraging for the use of the lattice dynamics to estimate the lattice mode contribution to the relative free energies of crystals of rigid organic molecules in crystal structure prediction studies. This is fortunate, as the lattice dynamics calculations with DMAREL take minutes, as opposed to the days required for the Molecular Dynamics simulations and weeks for the analysis of the motions.

Understanding the polymorphism of 5-fluorouracil.

The anti-cancer agent, 5-fluorouracil (Figure 1) provides an example of a rigid molecule capable of a range of different types of intermolecular interactions. Indeed, the crystal structure [70] determined in 1973 was unusual, in that it featured close contacts between the fluorine atoms and two single NH···O=C hydrogen bonds to other neighbouring molecules and only one double NH···O=C hydrogen bond motif of the type that dominates amide crystal structures (Figure 5). A lattice energy minimization search [71] predicted that there were many more conventional

For Peer Review Only

structures, with the amide groups forming doubly hydrogen bonded ribbons, up to 6 kJ mol^{-1} more stable than the known form. A manual polymorph screen found a new polymorph with this motif, which corresponded to the most stable crystal structure found in the polymorph search [71]. This success immediately raised the question as to whether the crystal structure prediction work had led to the discovery of a thermodynamically more stable polymorph, and why the new form had been so difficult to crystallize. These are important questions for pharmaceutical development, as once a thermodynamically more stable form has been discovered, by whatever means, then seeds of the more stable form can nucleate the new form under more conventional crystallization conditions. This can lead to "disappearing polymorphs" [72] or severe problems in reliably manufacturing the previously known form, as required the rapid reformulation of the pharmaceutical ritonavir [2].

In the case of 5-fluorouracil, thermal analysis did not show any transformations between the two polymorphs, consistent with the radically different hydrogen bonding motifs. However, the relative melting points and enthalpies of melting indicated that the original form I was more thermodynamically stable at room temperature, and probably at all temperatures. Hence, the DMA based model potential for 5-fluorouracil was incorrect in the relative lattice energies, although it reproduced both structures well. The replacement of the DMA by CHELPG charges of the same MP2/6-31G(d,p) charge density also gave the wrong stability order and worse reproduction of the unit cells, particularly the cell angles (Table 4).

The effects of temperature on the crystal structures were studied by means of a steepest-descent dynamic free energy minimization with respect to the simulation box

For Peer Review Only

vectors, using the recently reported metadynamics algorithm [73], using DL_MULTI. For 5-fluorouracil, approximately isodimensional supercells were created, (4x4x3 for form I containing 384 molecules and 7x3x6 for form II with 504 molecules), starting from the lattice energy minimum geometry with the corresponding electrostatic model. The convergence of cell lengths and angles required approximately 100 metadynamics steps each one of which involved a 2500 steps NVT simulation with a 3 fs timestep. The pressure tensor used to calculate the free energy derivative with respect to the cell geometry was averaged over the last 1500 steps.

Static lattice energy minimizations demonstrated that the multipole model yields better agreement with experiment than atomic point charges, as shown by the reproduction of cell angles of the triclinic polymorph I and cell length *a* in the monoclinic polymorph II (Table 4), though form II is still predicted to be more stable than form I at 0 K. The modelling of thermal effects at 310 K also does not reverse the relative enthalpic stability. The predicted thermal expansion between 0 and 310 K of 3-4% is plausibly larger that the experimentally observed change of 3.4% and 2.4% for forms I and II respectively between 150 K and room temperature. It is encouraging that the anisotropic thermal expansion reflects the directionality of the hydrogen bond network. The largest expansion in form II is along the *c*-direction which corresponds to the slip planes of the hydrogen bonded layers. In form I, the smallest expansion is along *c*, which cannot be altered without significantly distorting the hydrogen bonded sheet. Since the density was already underestimated by lattice energy minimization, the results in Table 4 strongly suggest that further improvements in the ability to simulate 5-fluorouracil must go beyond the empirically fitted isotropic atom-atom potential [51]. Model potentials which have been fitted to

For Peer Review Only

organic crystal structures by lattice energy minimization have already absorbed some thermal effects, and so ideally Molecular Dynamics studies should use non-empirical model intermolecular potentials, though methods of deriving these for organic molecules are in their early stages [74,75].

The lack of observed solid state transformations between form I and form II of fluorouracil in the thermal analysis make it clear that kinetic rather than thermodynamic factors are responsible for the formation of form II. A key clue came from the experimental observation that form II could only be crystallized from nitromethane, and even then, it had to be dry nitromethane as form I crystallized from samples where this solvent had been allowed to absorb water from the atmosphere. The first problem in trying to devise a series of MD simulations [76] to account for the solvent dependent polymorphism of 5-fluorouracil was its low solubility. Eventually a NPT simulation of 16 5-fluorouracil and 1550 water molecules which gave an equilibrated cubic cell of side of 36.9 Å was used, despite this corresponding to approximately 6 times the saturated concentration at the simulated room temperature and pressure. Thus the simulations were clearly not going to be able to simulate nucleation, as even if we could afford to simulate for long enough for all 16 5-fluorouracil molecules to form a cluster (which would take orders of magnitude longer than the 4 ns simulation time), a 16 molecule cluster is unlikely to represent a critical nucleus. However, although the concentration of 5-fluorouracil molecules was unphysically supersaturated, this was unlikely to affect the initial associations of the molecules which were the focus of interest. 5-fluorouracil is even less soluble in the hydroscopic solvent nitromethane. Thus the 16 fluorouracil molecules were simulated with 480 nitromethane molecules for this solution, and with 64 water and 496

For Peer Review Only

nitromethane molecules to represent wet nitromethane, though this is on the low side of the proportion of water molecules likely to be in nitromethane that had not been carefully dried.

These simulations [76] were seeking to contrast the initial association of 5-fluorouracil molecules in different solvents, to provide an understanding why such different hydrogen bonding motifs resulted from the crystallization experiments. Like the previous pioneering work on simulating 2-pyridone [77] and tetrolic acid [78] in carbon tetrachloride, it was essential that the solvent-solute interactions were realistically balanced with the solvent-solvent and solute-solute interactions, for a meaningful assessment of the extent to which singly versus doubly hydrogen bonded dimers existed in solution. (These simulations accounted for crystal structures containing a hydrogen bond chain rather than doubly hydrogen bonded dimer motif crystallising from non-polar solvents in which latter motif dominated). In addition, the 5-fluorouracil problem required appropriate relative strengths of interaction between water and nitromethane. For such a study, there was no experimental data, apart from the crystallization results, for validating and or developing the potential. Hence, we had to rely on literature potentials for 5-fluorouracil–5-fluorouracil, water–water and nitromethane–nitromethane and the standard Lorentz-Berthelot combinations rules being sufficiently realistic for the qualitative purposes of the simulation.

This fortunately proved the case, with the results being consistent with chemical intuition and the experimental outcome. Contrasting the radial distribution functions between 5-fluorouracil molecules confirmed the impression from direct visualization, that pairs of 5-fluorouracil molecules tend to form close F···F interactions and single

For Peer Review Only

N-H···O=C hydrogen bonds in water solution, but double N-H···O=C hydrogen bonds in nitromethane. However, confirming the observation that only single hydrogen bonds were ever seen in water, whereas doubly hydrogen bonded N-H···O=C formed readily in nitromethane, required a careful analysis of the secondary maxima in the radial distribution functions (RDFs), in contrast with the distances involved in the different doubly hydrogen bonded motifs that are possible for 5-fluorouracil. Thus, a very detailed, system specific analysis of a range of atom-atom RDFs involving both solute-solute and solute-solvent interactions was needed to justify the qualitative conclusion, that the hydration of the amide groups of 5-fluorouracil molecules by water is sufficiently strong to favour initial association of the hydrophobic F ends of the molecule, and in the cases where a N-H···O=C hydrogen bond is formed, the hydrating water was not displaced (over the timescale of the simulations) to form the doubly hydrogen bonded motif. In contrast, in nitromethane, once one N-H···O=C hydrogen bond is formed, the molecules can quickly reorient themselves to form the second hydrogen bond, giving rise to chains of motif of form II. Any water present in the simulation tended to aggregate quickly either in clusters, or to hydrate the 5-fluorouracil molecules where is effectively prevented the formation of any doubly hydrogen bonding motifs.

Conclusions and future outlook

There are many problems in understanding the organic solid state that need to be solved before we can hope to predict polymorphism. One is understanding solid state transformations, though the number of martensitic, second order phase changes that are known for organic molecules, and so can hope to be simulated within the

For Peer Review Only

Molecular Dynamics simulation cell is extremely limited [79] and restricted to systems where the differences between the two polymorphs are rather subtle and so highly demanding of the modelling. Indeed, it has been argued that all phase transformations between organic polymorphs occur by a nucleation and growth [80] and indeed the very subtle polymorphic phase change between the polymorphs of tetrachlorobenzene shows significant hysteresis and is first order [81]. However, even if Molecular Dynamics does not simulate real phase changes, it is likely to be a very useful tool in showing which lattice energy minima remain as stable structures once the dynamic motions of the molecules are taken into account. It is likely that many lattice energy minima are related by low energy barriers, and so would not be minima on the free energy surface. The method of metadynamics has shown considerable promise in predicting the phase diagram of benzene [73] and offers an exciting prospect for polymorph prediction. What is certain is that such methods will need to use the most accurate potentials and hence need DL_MULTI. It is less clear whether a metadynamics led exploration of the free energy surface would find known polymorphs, such as 5-fluorouracil form II, that are probably metastable over the entire temperature range.

The intriguing question as to how the nucleation process can lead to different polymorphs is an area where Molecular Dynamics should give valuable insights, even when the polymorphic outcome is not determined by the initial associations as in 5-fluorouracil. However, simulating nucleation and crystallization from solution is a major challenge, even for the simplest systems [82], let alone for polymorphic organic molecules. Turning insights into to how nucleation controls which energetically feasible structures are observed polymorphs into a computationally implementable

For Peer Review Only

theory of polymorph prediction will be even more challenging. For example, an experimental crystallization screen on the anti-epileptic carbamazepine [83] established that form IV of carbamazepine cannot be crystallized from solvent, and that the formation of the other metastable polymorphs was more determined by supersaturation and cooling rate than solvent. This illustrates how solving the industrial important problem of polymorph prediction will provide challenges to Molecular Dynamics simulation for decades to come. The development of DL_MULTI is an important step towards making the intermolecular potentials used in the simulations more realistic. Unfortunately, as Figure 1 suggests, many applications to pharmaceutical molecules will also require the realistic modelling of molecular flexibility, with the inherent challenge of balancing the energy penalties for intramolecular distortion [84,85] with the improvement in crystal packing.

Acknowledgements

The Basic Technology Program of the Research Councils UK is thanked for funding Control and Prediction of the Organic Solid State (CPOSS, www.cposs.org.uk), including the HPC(x) facilities which were used for much of this work.

References

[1] G. M. Day, A. V. Trask, W. D. S. Motherwell, W. Jones. Investigating the Latent Polymorphism of Maleic Acid. *Chem. Commun.*, 54-56 (2006)

[2] S. R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubauer, B. A. Narayanan, M. Soldani, D. Riley, K. McFarland. Dealing With the Impact

For Peer Review Only

of Ritonavir Polymorphs on the Late Stages of Bulk Drug Process Development. *Organic Process Res. Dev.*, 4, 413-417 (2000)

[3] S. L. Price. The Computational Prediction of Pharmaceutical Crystal Structures and Polymorphism. *Adv. Drug Deliver. Rev.*, 56, 301-319 (2004)

[4] J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, W. T. M. Mooij, S. L. Price, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, D. E. Williams. A Test of Crystal Structure Prediction of Small Organic Molecules. *Acta Crystallogr. B*, 56, 697-714 (2000)

[5] W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, J. P. M. Lommerse, W. T. M. Mooij, S. L. Price, H. Scheraga, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, D. E. Williams. Crystal Structure Prediction of Small Organic Molecules: a Second Blind Test. *Acta Crystallogr. B*, 58, 647-661 (2002)

[6] G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. Scheraga, Y. A. Arnautova, M. U. Schmidt, P. Verwer. A Third Blind Test of Crystal Structure Prediction. *Acta Crystallogr. B*, 61, 511-527 (2005)

For Peer Review Only

[7] C. Ouvrard, S. L. Price. Towards Crystal Structure Prediction for Conformationally Flexible Molecules: the Headaches Illustrated by Aspirin. *Cryst. Growth Des.*, 4, 1119-1127 (2004)

[8] T. Beyer, G. M. Day, S. L. Price. The Prediction, Morphology, and Mechanical Properties of the Polymorphs of Paracetamol. *J. Am. Chem. Soc.*, 123, 5086-5094 (2001)

[9] P. Vishweshwar, J. A. McMahon, M. Oliveira, M. L. Peterson, M. J. Zaworotko. The Predictably Elusive Form II of Aspirin. *J. Am. Chem. Soc.*, 127, 16802-16803 (2005)

[10] M. L. Peterson, S. L. Morissette, C. McNulty, A. Goldsweig, P. Shaw, M. LeQuesne, J. Monagle, N. Encina, J. Marchionna, A. Johnson, J. Gonzalez-Zugasti, A. V. Lemmo, S. J. Ellis, M. J. Cima, O. Almarsson. Iterative High-Throughput Polymorphism Studies on Acetaminophen and an Experimentally Derived Structure for Form III. *J. Am. Chem. Soc.*, 124, 10958-10959 (2002)

[11] M. L. Klein. Computer Simulation Studies of Solids. *Annu. Rev. Phys. Chem.*, 36, 525-548 (1985)

[12] S. L. Price. Quantifying Intermolecular Interactions and Their Use in Computational Crystal Structure Prediction. *CrystEngComm*, 6, 344-353 (2004)

[13] S. L. Price. Modelling Intermolecular Forces for Organic Crystal Structure Prediction. In *Intermolecular Forces and Clusters I*, D. J. Wales, Ed., Vol. 115, Chapter 3, pp 81-124, Springer-Verlag, Berlin, (2005).

For Peer Review Only

[14] M. Leslie. DL_MULTI - A Program to Use Distributed Multipole Electrostatic Models to Simulate the Dynamics of Organic Crystals. *Mol. Phys.*, in press (2005)

[15] C. K. Broder, K. Shankland, W. I. F. David, R. Ibberson. Neutron Study of the Phase Behaviour of Solid Cyclopentane. *in preparation*, (2006)

[16] R. Boese, M. T. Kirchner, D. Blaeser. *in preparation*, (2006)

[17] B. Post, R. S. Schwartz, I. Fankuchen. X-Ray Investigation of Crystalline Cyclopentane and Neohexane. *J. Am. Chem. Soc.*, 73, 5113-5114 (1951)

[18] M. T. Dove, G. S. Pawley. A Molecular-Dynamics Simulation Study of the Orientationally Disordered Phase of Sulfur-Hexafluoride. *J. Phys. C Solid State*, 17, 6581-6599 (1984)

[19] C. G. Windsor, D. H. Saunderson, J. N. Sherwood, D. Taylor, G. S. Pawley. Lattice-Dynamics of Adamantane in Disordered Phase. *J. Phys. C Solid State*, 11, 1741-1759 (1978)

[20] N. A. Murugan. Orientational Melting and Reorientational Motion in a Cubane Molecular Crystal: A Molecular Simulation Study. *J. Phys. Chem. B*, 109, 23955-23962 (2005)

[21] C. M. Breneman, K. B. Wiberg. Determining Atom-Centered Monopoles From Molecular Electrostatic Potentials - the Need For High Sampling Density in Formamide Conformational-Analysis. *J. Comput. Chem.*, 11, 361-373 (1990)

For Peer Review Only

[22] D. E. Williams, S. R. Cox. Nonbonded Potentials For Azahydrocarbons: the Importance of the Coulombic Interaction. *Acta Crystallogr. B*, 40, 404-417 (1984)

[23] W. Smith, T. R. Forester. DL_Poly 2.0 - A General Purpose Parallel Molecular Dynamics Simulation Package. *J. Mol. Graphics*, 14, 136-141 (1996)

[24] Accelrys Inc. Materials Studio.(2.2). (2002).

[25] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. A 2nd Generation Force-Field For the Simulation of Proteins, Nucleic- Acids, and Organic-Molecules. *J. Am. Chem. Soc.*, 117, 5179-5197 (1995)

[26] S. Lifson, A. T. Hagler, P. Dauber. Consistent Force Field Studies of Intermolecular Forces in Hydrogen Bonded Crystals. 1. Carboxylic Acids, Amides and the C=O···H- Hydrogen Bonds. *J. Am. Chem. Soc.*, 101, 5111-5121 (1979)

[27] W. R. P. Scott, P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, W. F. van Gunsteren. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A*, 103, 3596-3607 (1999)

[28] E. S. Ferrari, R. J. Davey, W. I. Cross, A. L. Gillon, C. S. Towler. Crystallization in Polymorphic Systems: The Solution-Mediated Transformation Beta to Alpha Glycine. *Cryst. Growth Des.*, 3, 53-60 (2003)

For Peer Review Only

[29] G. L. Perlovich, L. K. Hansen, A. Bauer-Brandl. The Polymorphism of Glycine - Thermochemical and Structural Aspects. *J. Therm. Anal. Calorim.*, 66, 699-715 (2001)

[30] E. V. Boldyreva, V. A. Drebushchak, T. N. Drebushchak, I. E. Paukov, Y. A. Kovalevskaya, E. S. Shutova. Polymorphism of Glycine - Thermodynamic Aspects. Part I. Relative Stability of the Polymorphs. *J. Therm. Anal. Calorim.*, 73, 409-418 (2003)

[31] C. S. Towler, R. J. Davey, R. W. Lancaster, C. J. Price. Impact of Molecular Speciation on Crystal Nucleation in Polymorphic Systems: The Conundrum of Gamma Glycine and Molecular 'Self Poisoning'. *J. Am. Chem. Soc.*, 126, 13347-13353 (2004)

[32] I. Weissbuch, V. Y. Torbeev, L. Leiserowitz, M. Lahav. Solvent Effect on Crystal Polymorphism: Why Addition of Methanol or Ethanol to Aqueous Solutions Induces the Precipitation of the Least Stable Beta Form of Glycine. *Angew. Chem. Int. Edit.*, 44, 3226-3229 (2005)

[33] Y. Lian, N. G. Kingman. Glycine Crystallization During Spray Drying: The PH Effect on Salt and Polymorphic Forms. *J. Pharm. Sci.*, 91, 2367-2375 (2002)

[34] A. Dawson, D. R. Allan, S. A. Belmonte, S. J. Clark, W. I. F. David, P. A. McGregor, S. Parsons, C. R. Pulham, L. Sawyer. Effect of High Pressure on the Crystal Structures of Polymorphs of Glycine. *Cryst. Growth Des.*, 5, 1415-1427 (2005)

For Peer Review Only

[35] S. V. Goryainov, E. V. Boldyreva, E. N. Kolesnik. Raman Observation of a New Polymorph of Glycine? *Chem. Phys. Lett.*, 419, 496-500 (2006)

[36] J. A. Chisholm, S. Motherwell, P. R. Tulip, S. Parsons, S. J. Clark. An *Ab initio* Study of Observed and Hypothetical Polymorphs of Glycine. *Cryst. Growth Des.*, 5, 1437-1442 (2005)

[37] C. M. Breneman, K. B. Wiberg. Determining Atom-Centered Monopoles From Molecular Electrostatic Potentials - the Need For High Sampling Density in Formamide Conformational-Analysis. *J. Comput. Chem.*, 11, 361-373 (1990)

[38] B. P. van Eijck, J. Kroon. Coulomb Energy of Polar Crystals. *J. Phys. Chem. B*, 101, 1096-1100 (1997)

[39] B. P. van Eijck, J. Kroon. Comment on "Crystal Structure Prediction by Global Optimization As a Tool for Evaluating Potentials: Role of the Dipole Moment Correction Term in Successful Predictions". *J. Phys. Chem. B*, 104, 8089 (2000)

[40] W. J. Wedemeyer, Y. A. Arnautova, J. Pillardy, R. J. Wawak, C. Czaplewski, H. A. Scheraga. Reply to "Comment on 'Crystal Structure Prediction by Global Optimization As a Tool for Evaluating Potentials: Role of the Dipole Moment Correction Term in Successful Predictions'" by B. P. Van Eijck and J. Kroon. *J. Phys. Chem. B*, 104, 8090-8092 (2000)

[41] S. Brodersen, S. Wilke, F. J. J. Leusen, G. Engel. A Study of Different Approaches to the Electrostatic Interaction in Force Field Methods for Organic Crystals. *Phys. Chem. Chem. Phys.*, 5, 4923-4931 (2003)

For Peer Review Only

[42] J. B. O. Mitchell, S. L. Price. A Systematic Nonempirical Method of Deriving Model Intermolecular Potentials for Organic Molecules: Application to Amides. *J. Phys. Chem. A*, 104, 10958-10971 (2000)

[43] P. Langan, S. A. Mason, D. Myles, B. P. Schoenborn. Structural Characterization of Crystals of Alpha-Glycine During Anomalous Electrical Behaviour. *Acta Crystallogr. B*, 58, 728-733 (2002)

[44] A. Kvick, W. M. Canning, T. F. Koetzle, G. J. B. Williams. Experimental-Study of the Influence of Temperature on A Hydrogen-Bonded System - Crystal-Structure of Gamma-Glycine at 83-K and 298-K by Neutron-Diffraction. *Acta Crystallogr. B*, 36, 115-120 (1980)

[45] L. Yu, J. Huang, K. J. Jones. Measuring Free-Energy Difference Between Crystal Polymorphs Through Eutectic Melting. *J. Phys. Chem. B*, 109, 19915-19922 (2005)

[46] A. D. Buckingham, P. W. Fowler. A Model for the Geometies of Van Der Waals Complexes. *Can. J. Chemistry*, 63, 2018-2025 (1985)

[47] A. J. Stone, M. Alderton. Distributed Multipole Analysis - Methods and Applications. *Mol. Phys.*, 56, 1047-1064 (1985)

[48] A. J. Stone. Distributed Multipole Analysis: Stability for Large Basis Sets. *Journal of Chemical Theory and Computation*, 1, 1128-1132 (2005)

[49] A. J. Stone. GDMA: a program for performing Distributed Multipole Analysis of wavefunctions calculated using the Gaussian program system.(1.0). (1999). University of Cambridge (U. K.).

For Peer Review Only

[50] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Chesseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayalla, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, J. A. Pople. GAUSSIAN 98.(A6). (1998). Pittsburgh, Gaussian Inc.

[51] D. S. Coombes, S. L. Price, D. J. Willock, M. Leslie. Role of Electrostatic Interactions in Determining the Crystal- Structures of Polar Organic Molecules. A Distributed Multipole Study. *J. Phys. Chem.*, 100, 7352-7360 (1996)

[52] G. M. Day, W. D. S. Motherwell, W. Jones. Beyond the Isotropic Atom Model in Crystal Structure Prediction of Rigid Molecules: Atomic Multipoles Versus Point Charges. *Cryst. Growth Des.*, 5, 1023-1033 (2005)

[53] G. M. Day, S. L. Price, M. Leslie. Atomistic Calculations of Phonon Frequencies and Thermodynamic Quantities for Crystals of Rigid Organic Molecules. *J. Phys. Chem. B*, 107, 10919-10933 (2003)

[54] A. J. Stone, P. L. A. Popelier, D. J. Wales. ORIENT: a program for calculating electrostatic interactions.(3.2). (1997). University of Cambridge.

For Peer Review Only

[55] S.L.Price, D.J.Willock, M.Leslie, G. M. Day. DMAREL 3.02. (2001). http://www.ucl.ac.uk/~ucca17p/dmarelmanual/dmarel.html.

[56] D. J. Willock, S. L. Price, M. Leslie, C. R. A. Catlow. The Relaxation of Molecular Crystal Structures Using a Distributed Multipole Electrostatic Model. *J. Comput. Chem.*, 16, 628-647 (1995)

[57] A. J. Stone. *The Theory of Intermolecular Forces*, Clarendon Press, Oxford (1996).

[58] P. L. A. Popelier, A. J. Stone. Formulas For the 1st and 2nd Derivatives of Anisotropic Potentials With Respect to Geometrical Parameters. *Mol. Phys.*, 82, 411-425 (1994)

[59] S. Y. Liem, P. L. A. Popelier. High-Rank Quantum Topological Electrostatic Potential: Molecular Dynamics Simulation of Liquid Hydrogen Fluoride. *J. Chem. Phys.*, 119, 4560-4566 (2003)

[60] S. Y. Liem, P. L. A. Popelier, M. Leslie. Simulation of Liquid Water Using a High-Rank Quantum Topological Electrostatic Potential. *Int. J. Quantum Chem.*, 99, 685-694 (2004)

[61] R. K. McMullan, J. Epstein, J. R. Ruble, B. M. Craven. Neutron Study on Imidazole. *Acta Crystallogr. B*, 35, 688-691 (1979)

[62] C. Perchard, A. Novak. Far-Infrared Spectra and Hydrogen-Bond Frequencies of Imidazole. *J. Chem. Phys.*, 48, 3079-3084 (1968)

[63] M. Majoube, G. Vergoten. Lattice-Vibrations of Crystalline Imidazole and N-15 and D Substituted Analogs. *J. Chem. Phys.*, 76, 2838-2847 (1982)

For Peer Review Only

[64] S. Besainou, G. Cardini, D. A. Does. Molecular Dynamics of Crystalline Imidazole. *Chem. Phys.*, 156, 71-77 (1991)

[65] B. S. Potter, R. A. Palmer, R. Withnall, B. Z. Chowdhry, S. L. Price. Aza Analogues of Nucleic Acid Bases: Experimental Determination and Computational Prediction of the Crystal Structure of Anhydrous 5-Azauracil. *J. Mol. Struct.*, 486, 349-361 (1999)

[66] S. L. Price, B. Patel, P. Pridhanani-Jethani, A. Torrisi. Crystal Structure Prediction and Polymorphism - Some Mutual Insights. *Trans. ACA*, 39, 2-13 (2004)

[67] S. L. Price, D. J. Willock, M. Leslie, G. M. Day. DMAREL 3.02. (2001). http://www.ucl.ac.uk/~ucca17p/dmarelmanual/dmarel.html.

[68] A. E. Gray, G. M. Day, M. Leslie, S. L. Price. Dynamics in Crystals of Rigid Organic Molecules: Contrasting the Phonon Frequencies Calculated by Molecular Dynamics With Harmonic Lattice Dynamics for Imidazole and 5-Azauracil. *Mol. Phys.*, 102, 1067-1083 (2004)

[69] P. Jiminez, M. V. Roux, C. Turrion. Thermochemical Properties of N-Heterocyclic Compounds .1. Enthalpies of Combustion, Vapor-Pressures and Enthalpies of Sublimation, and Enthalpies of Formation of Pyrazole, Imidazole, Indazole, and Benzimidazole. *J. Chem. Thermodyn.*, 19, 985-992 (1987)

[70] L. Fallon. The Crystal and Molecular Structure of 5-Flurouracil. *Acta Crystallogr. B*, 29, 2549-2556 (1973)

For Peer Review Only

[71] A. T. Hulme, S. L. Price, D. A. Tocher. A New Polymorph of 5-Fluorouracil Found Following Computational Crystal Structure Predictions. *J. Am. Chem. Soc.*, 127, 1116-1117 (2005)

[72] J. D. Dunitz, J. Bernstein. Disappearing Polymorphs. *Accounts Chem. Res.*, 28, 193-200 (1995)

[73] P. Raiteri, R. Martonak, M. Parrinello. Exploring Polymorphism: The Case of Benzene. *Angew. Chem. Int. Edit.*, 44, 3769-3773 (2005)

[74] W. T. M. Mooij, F. B. van Duijneveldt, J. van Duijneveldt-van de Rijdt, B. P. van Eijck. Transferable *Ab initio* Intermolecular Potentials. 1. Derivation From Methanol Dimer and Trimer Calculations. *J. Phys. Chem. A*, 103, 9872-9882 (1999)

[75] G. M. Day, S. L. Price. A Nonempirical Anisotropic Atom-Atom Model Potential for Chlorobenzene Crystals. *J. Am. Chem. Soc.*, 125, 16434-16443 (2003)

[76] S. Hamad, C. Moon, C. R. A. Catlow, A. T. Hulme, S. L. Price. Kinetic Insights into the Role of the Solvent in the Polymorphism of 5-Fluorouracil. *J. Phys. Chem. B*, 110, 3323-3329 (2006)

[77] A. Gavezzotti. Computer Simulations of Organic Solids and Their Liquid-State Precursors. *Faraday Discuss.*, 106, 63-78 (1997)

[78] A. Gavezzotti, G. Filippini, J. Kroon, B. P. vanEijck, P. Klewinghaus. The Crystal Polymorphism of Tetrolic Acid (CH3C CCOOH): A Molecular

For Peer Review Only

Dynamics Study of Precursors in Solution, and a Crystal Structure Generation. *Chem. -Eur. J.*, 3, 893-899 (1997)

[79] S. C. Tuble, J. Anwar, J. D. Gale. An Approach to Developing a Force Field for Molecular Simulation of Martensitic Phase Transitions Between Phases With Subtle Differences in Energy and Structure. *J. Am. Chem. Soc.*, 126, 396-405 (2004)

[80] Y. Mnyukh. *Fundamentals of Solid-State Phase Transitions, Ferromagnetism and Ferroelectricity.*, Bloomington: 1st Books Library, (2001).

[81] S. A. Barnett, C. K. Broder, K. Shankland, W. I. F. David, R. Ibberson, D. A. Tocher. Single Crystal X-Ray and Neutron Powder Diffraction Investigation of the Phase Transition in Tetrachlorobenzene. *Acta Crystallogr. B*, 62, 287-295 (2006)

[82] J. Anwar, P. K. Boateng. Computer Simulation of Crystallization From Solution. *J. Am. Chem. Soc.*, 120, 9600-9604 (1998)

[83] A. J. Florence, A. Johnston, S. L. Price, H. Nowell, A. R. Kennedy, N. Shankland. An Automated Parallel Crystallization Search for Predicted Crystal Structures and Packing Motifs of Carbamazepine. *J. Pharm. Sci.*, in press, (2006)

[84] T. van Mourik, P. G. Karamertzanis, S. L. Price. Molecular Conformations and Relative Stabilities Can Be As Demanding of the Electronic Structure Method As Intermolecular Calculations. *J. Phys. Chem. A*, 110, 8-12 (2006)

For Peer Review Only

[85] P. G. Karamertzanis, S. L. Price. Energy Minimization of Crystal Structures Containing Flexible Molecules. *Journal of Chemical Theory and Computation*, in press, (2006)

For Peer Review Only

Table 1: Lattice parameters of the experimental and MD simulated phases III and I of cyclopentane.

	a/Å	***b***/Å	***c***/Å	***α***/°	***β*** /°	***γ*** /°
Form III 100K expt	9.578	5.329	10.004	90.00	113.20	90.00
100K calc	9.134	5.762	10.307	90.00	104.78	90.00
Form I expt	9.330	5.830	10.100	90.00	90.00	90.00
170K calc	9.577	5.962	10.326	89.99	90.03	90.06

Form I is the orthorhombic equivalent to the hexagonal unit cell a=9.330 Å, b=c=5.830 Å α=120 °, β=γ=90 °.

For Peer Review Only

Table 2 The cell parameters of the three polymorphs of glycine as simulated at 300 K by MD. In the NPT method employed the cell angles were fixed at their experimental values.

	a/Å	***b***/Å	***c***/Å	Relative energies / kJ mol^{-1}
Expt **α** [43]	5.099	11.942	5.461	+1.9[a]
MD	5.107	11.970	5.465	+1.5
Expt **β** [28]	5.078	6.192	5.387	>1.5[a]
MD	5.092	6.212	5.402	+3.1
Expt **γ** [44]	6.975	6.975	5.473	Most stable
MD	6.965	6.965	5.465	0.0

[a] Value of the enthalpy difference for T = -4 to 177 ºC, obtained from DSC experiments [45]. No value has been found in the literature for the enthalpy difference between γ and β, although it is known that the latter is the less stable polymorph.

For Peer Review Only

Table 3. The unit cell parameters of imidazole and 5-azauracil, contrasting simulations by lattice energy minimization (LE) and NST Molecular Dynamics, using distributed multipole electrostatic models.

	E*/kJ mol^{-1}	***a***/Å	***b***/Å	***c***/Å	***α***/°	***β***/°	***γ***/°	Vol. /Å^3
Imidazole								
Expt (103 K)	-88.1 ±0.2	7.569	5.366	9.779	90.0	119.1	90.0	347.32
LE (0 K)	-78.38	7.72 Δ=0.15	5.46 Δ=0.10	9.81 Δ=0.03	90.0	120.6 Δ=1.5	90.0	356.0 Δ=8.7
NST MD (100 K)	-76.12	7.783	5.509	9.810	89.998	119.92	90.003	364.51
Expansion (%)	-	0.843	0.875	-0.051	-	-	-	2.38
NVE MD (100 K)	-76.16	7.783	5.511	9.810	90.038	119.956	89.9806	364.53
5-azauracil								
Expt	-	6.5135	13.5217	9.5824	90.000	90.000	90.000	843.956
LE (0 K)	-117.4	6.72 Δ=0.20	13.89 Δ=0.35	9.31 Δ=-0.27	90.0	90.0	90.0	867.4 Δ=23.4
NST MD (310 K)	-110.45	6.916	13.979	9.361	89.976	89.986	89.9862	904.97
Expansion (%)	-	2.970	0.796	0.525	-	-	-	4.34
NVE MD (310 K)	-110.37	6.922	13.980	9.365	89.912	89.958	89.905	906.17

The energy quoted is -(ΔH_{sub} + 2RT) for the experimental value (only known for imidazole [69]), the lattice energy for the DMAREL calculations, and the average potential energy for the Molecular Dynamics calculations.

The NVE cell parameters are those used in the MD simulation of the phonon modes.

For Peer Review Only

Table 4. Free energy and lattice energy minima of the two polymorphs of 5-flurouracil, contrasting a point charge and distributed multipole model of the *ab initio* charge density for evaluating the electrostatic energy.

	enthalpy[1] (kJ mol^{-1})	*a*/Å	*b*/Å	*c*/Å	*α*/°	*β*/°	*γ*/°	density (g cm^{-3})
static lattice energy minimization[2]								
form I								
experimental, 150K		8.633	9.156	12.580	80.88	79.98	89.98	1.788
experimental, RT		8.786	9.220	12.660	81.40	80.53	89.41	1.728
Atomic charges	-92.40	8.769 +1.58%	9.274 +1.29%	13.083 +4.00%	87.48 +6.60	87.37 +7.39	92.72 +2.74	1.629 -8.89%
atomic multipoles	-96.58	8.839 +2.39%	9.279 +1.34%	13.048 +3.72%	83.39 +2.51	82.86 +2.88	91.69 +1.71	1.640 -8.28%
form II								
experimental, 150K		5.043	14.935	6.605	90.00	108.88	90.00	1.835
experimental, RT		5.154	15.001	6.654	90.00	110.34	90.00	1.791
Atomic charges	-96.58	5.767 +14.36%	14.942 +0.05%	6.444 -2.44%	90.00 -	114.78 +5.90	90.00 -	1.714 -6.59%
atomic multipoles	-102.48	5.350 +6.09%	15.262 +2.19%	6.508 -1.47%	90.00 -	110.29 +1.41	90.00 -	1.733 -5.56%
dynamic free energy minimization, 310K								
form I								
Atomic charges	-84.70	8.954	9.394	13.113	88.83	90.14	91.24	1.567
thermal expansion[3]		+2.10%	+1.29%	+0.23%	+1.35	+2.77	-1.48	-3.81%
atomic multipoles	-88.82	8.987	9.335	13.062	87.73	87.79	91.12	1.580
thermal expansion[3]		+1.67%	+0.60%	+0.11%	+4.34	+4.93	-0.57	-3.66%
form II								
Atomic charges	-89.19	5.814	15.094	6.603	89.99	115.35	90.00	1.650
thermal expansion[3]		+0.81%	+1.02%	+2.47%	-0.01	+0.50	0.00	-3.73%
atomic multipoles	-94.68	5.404	15.368	6.629	89.99	110.70	90.01	1.678
thermal expansion[3]		+1.01%	+0.69%	+1.86%	-0.01	+0.41	0.00	-3.17%

[1] for the dynamic free energy minimization results, enthalpy is taken equal to the time averaged potential energy over a 4.5 ps period following 3 ps equilibration time for the last metadynamics step

[2] for the lattice energy minimizations percentage differences in cell lengths and density and cell angle differences are with respect to the 150 K experimental determination

[3] thermal expansion with respect to the lattice energy minimization cell geometry

For Peer Review Only

Figure Captions

Figure 1 Chemical Diagrams for (top row) cyclopentane, glycine, imidazole, 5-azauracil, 5-fluorouracil, as molecules studied by Molecular Dynamics in this paper. (Middle row) aspirin, paracetamol, carbamazepine, (bottom row) ritonavir as molecules mentioned in text.

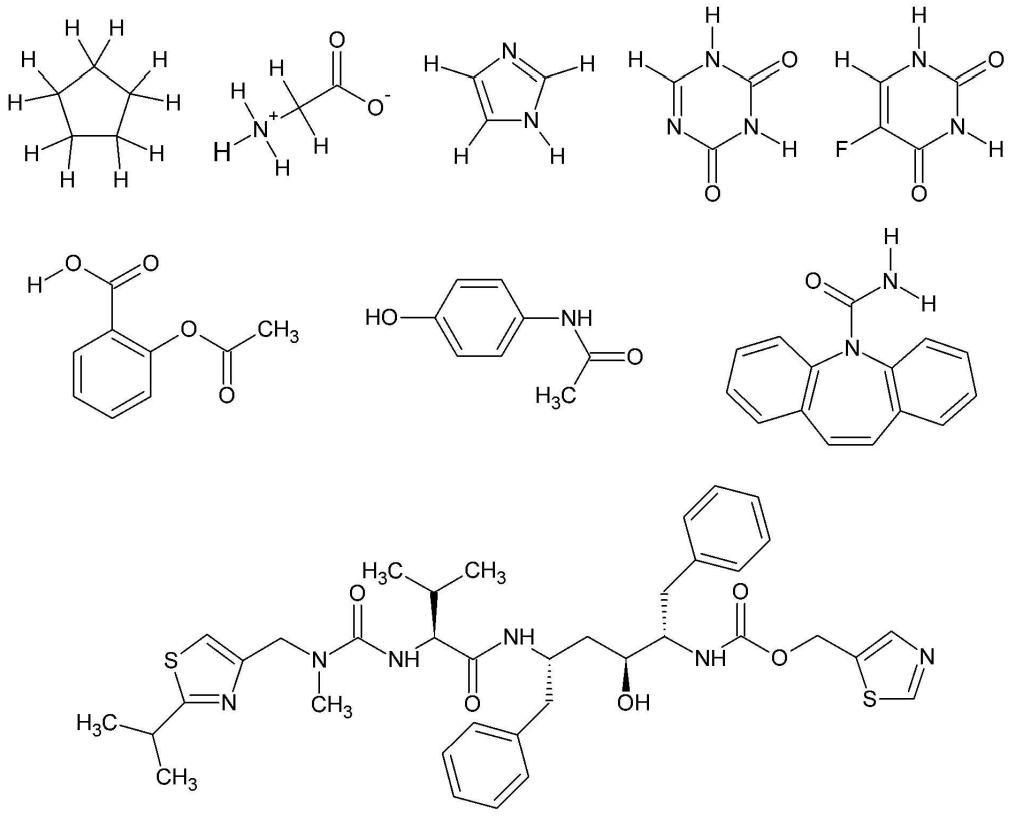
Figure 2 Snapshots of Molecular Dynamics simulations of cyclopentane. (a) monoclinic ordered phase III at 100 K viewed along *b* (b) hexagonal disordered phase I at 170 K, viewed along *b*, and (c) a plot of the variation in cell volume with temperature.

Figure 3. Snapshots of the Molecular Dynamics simulations of the (a) alpha, (b) beta and (c) gamma polymorphs of glycine. The dashed lines represent hydrogen bonds

Figure 4 Contrasting the k=0 modes of specific symmetry derived from the Molecular Dynamics simulation with the lattice dynamics harmonic frequencies (vertical lines), computed using the same distributed multipole model potential [68] for (a) the A_g modes of imidazole at 100 K and (b) the A_u modes for 5-azauracil at 310 K.

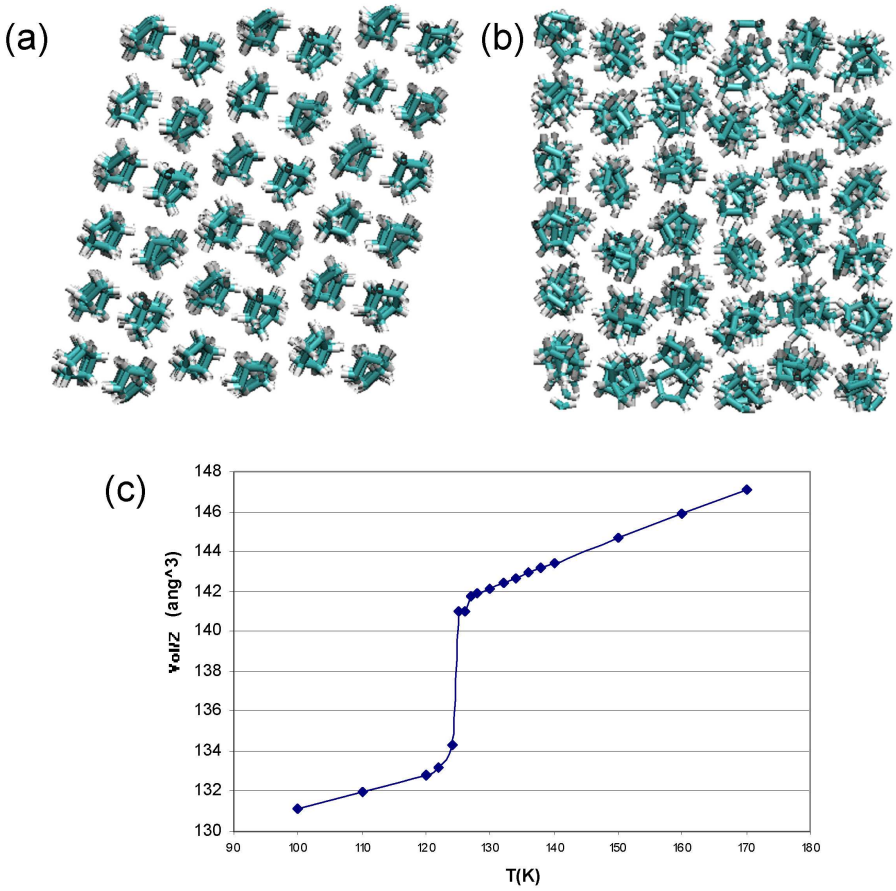
Figure 5. The crystal structures of 5-fluorouracil in (a) form I and (b) form II. (c) Snapshot of the simulations of 5-fluorouracil in nitromethane solution in which some water molecules are added, to take into account its hydroscopicity. Water molecules interact strongly with oxygen atoms in 5-fluorouracil, preventing the formation of double hydrogen bonded dimers.

For Peer Review Only

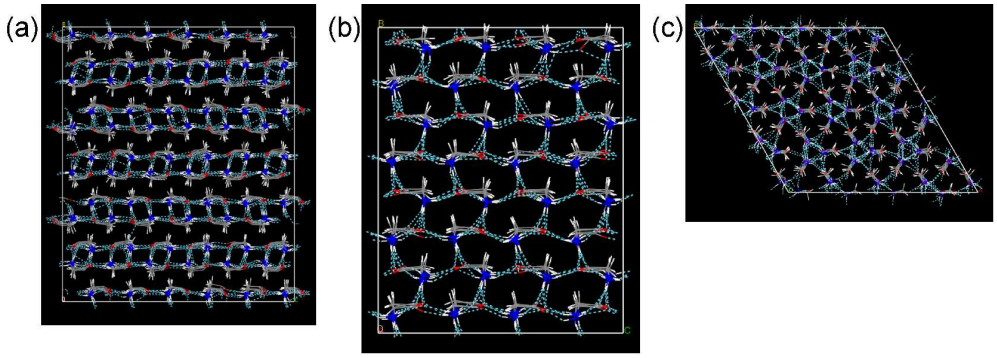


Chemical Diagrams for (top row) cyclopentane, glycine, imidazole, 5-azauracil, 5-fluorouracil, as molecules studied by Molecular Dynamics in this paper. (Middle row) aspirin, paracetamol, carbamazepine, (bottom row) ritonavir as molecules mentioned in text.
158x129mm (600 x 600 DPI)

For Peer Review Only

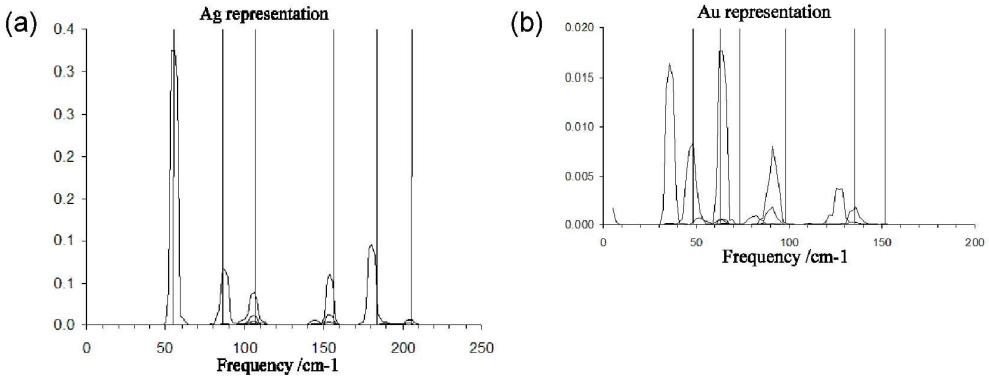


Snapshots of Molecular Dynamics simulations of cyclopentane. (a) monoclinic ordered phase III at 100 K viewed along *b* (b) hexagonal disordered phase I at 170 K, viewed along *b*, and (c) a plot of the variation in cell volume with temperature.
903x867mm (72 x 72 DPI)



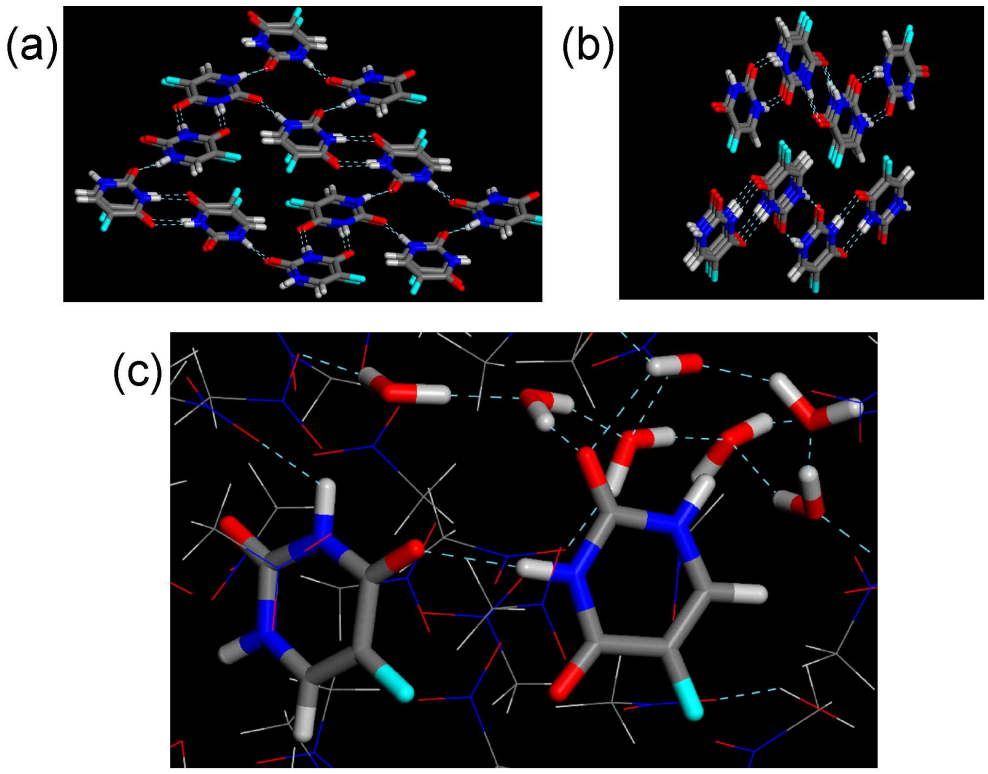
Snapshots of the Molecular Dynamics simulations of the (a) alpha, (b) beta and (c) gamma polymorphs of glycine. The dashed lines represent hydrogen bonds.
1183x414mm (72 x 72 DPI)

er Review Only



Contrasting the k=0 modes of specific symmetry derived from the Molecular Dynamics simulation with the lattice dynamics harmonic frequencies (vertical lines), computed using the same distributed multipole model potential [68] for (a) the A_g modes of imidazole at 100 K and (b) the A_u modes for 5-azauracil at 310 K.
1183x454mm (72 x 72 DPI)

For Review Only



The crystal structures of 5-fluorouracil in (a) form I and (b) form II.
(c) Snapshot of the simulations of 5-fluorouracil in nitromethane solution in which some water molecules are added, to take into account its hydroscopicity. Water molecules interact strongly with oxygen atoms in 5-fluorouracil, preventing the formation of double hydrogen bonded dimers.
726x555mm (72 x 72 DPI)

v Only