REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2006 Name of Author LEWIS T.C.

COPYRIGHT

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

COPYRIGHT DECLARATION

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

LOANS

Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION

University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).

B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.

C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.

D. 1989 onwards. Most theses may be copied.

This thesis comes within category D.

☑ This copy has been deposited in the Library of UCL

☐ This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.

Investigating polymorphic behaviour of organic molecules using theoretical and experimental techniques

Thomas Charles Lewis

A thesis submitted to the University of London in partial fulfilment of the requirements for the degree of Doctor of Philosophy

University College London
September 2005

Christopher Ingold Laboratories
Department of Chemistry
University College London
20 Gordon Street
London
WC1H 0AJ
United Kingdom

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

For my family

Family is not an important thing, it's everything.

"I am among those who think that science has a great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale"

Marie Curie (1867 – 1934)

Academic Acknowledgements

I would very much like to thank my supervisors, Professor Sally Price and Dr Derek Tocher for the continuous support and encouragement throughout my PhD studies. Their dedication to science has been an inspiration to me, and it has been a privilege to work with them.

There are many people who have helped me throughout my PhD studies. I would like to thank both Dr Graeme Day and Dr Carole Ouvrard who assisted me greatly in the early stages of my PhD, in addition to the countless people I have met at conferences and within the scientific field for which I had useful discussions with. My thanks also go to Dr Royston Copley and Miss Lucie Deprez for the collaboration forming the basis of chapter 8, in addition to the authors of the many programs that were used in this study, including Dr Maurice Leslie, Dr Graeme Day (DMAREL) and Professor Herman Ammon (MOLPAK).

Not forgetting everyone in the research group, in which there have been many changes over the last couple of years. Thanks for the snippets of helpful information and useful discussions which greatly aided my completed of this PhD.

Financially, I would also like to thank the following organisations:
The EPSRC for funding this research through a departmental research studentship, and for a bursary to attend the 9th BCA/CCG X-ray structure analysis teaching school in Durham, 2003. The graduate school research projects fund for providing me with the opportunity to attend the 35th EuroSummerSchool in Erice on the Diversity amidst Similarity crystallography course.

Personal Acknowledgements

I would like to thank the following people...apologies for those people I forgot to mention:

My family for their continuous support.

Sally and Derek for giving me the opportunity to contribute to crystal polymorphism research for this PhD. I would also like to thank Bob for giving me his expertise in this exciting scientific field, and also for the friendly banter.

Everyone from the research group over the years, from the early days with Graeme, Michael, Carole and Aileen; to my fellow colleagues who worked on the CPOSS project, Louise, Ashley, Sarah, Harriott, Panos and Ian. I would also like to thank the undergraduate project students who allow me to improve my teaching and motivation skills! It has been a privilege to work with everyone.

My friends to whom I tried to explain my PhD research with varied degrees of success! It has been great to wind down from my studies with you guys... whether it is on holidays or just from a trip down the pub for a swift half. Thanks!

Finally I would like to thank the people of London. The courage and friendliness of people since the tragic events of this summer have been an inspiration to me and have undoubtedly given me huge encouragement to complete this PhD. God bless you all.

Abstract

This thesis describes complementary theoretical and experimental investigations into polymorphism of several structurally similar small organic molecules.

Five biologically important molecules (DNA/RNA bases) and five barbiturate type compounds were subjected to a wide range of crystallisation conditions which resulted in the characterisation of new products, including new polymorphs of barbituric acid and 6-methyluracil. These results were correlated with computational polymorph predictions and in some cases rationalised to investigate why some structures have unused hydrogen bond acceptors.

The crystal structure prediction process was evaluated on hydantoin as part of the formal third crystal structure prediction blind test organised by the Cambridge Crystallographic Data Centre. This involved performing the calculations from just the known molecular structure. The compounds 3-oxauracil and 5-hydroxyuracil were also studied in an informal blind test scenario in collaboration with GlaxoSmithKline. In all cases the experimental crystal structure was found within the lowest energy structures.

This work demonstrates that structurally similar organic molecules can have different patterns of the relative energies of the hypothetical low energy crystal structures, along with differences in the experimental polymorphic behaviour. Molecular flexibility and the model applied for the intermolecular forces can reorder the relative stabilities of the low energy structures. These variations are often comparable to the differences in the energies of the different crystal structures which considerably decrease the confidence in the computational predictions. The kinetics of crystallisation and the limitations in the range of crystallisation techniques available are important in determining which polymorphs are seen experimentally. The range of experimental and computational polymorphic behaviour exhibited by the molecules described in this thesis highlights the challenges involved in polymorph prediction.

Table of contents

List of Figures

List of Tables

1. Introduction

"...a crystal is like a class of children arranged for drill, but standing at ease, so that while the class as a whole has regularity both in time and space, each individual child is a little fidgety."[1;2]

This statement by Dame Kathleen Lonsdale, the first female fellow of the Royal Society, describes the nature of crystalline material. She briefly mentions[1;3] the fact that some molecules have at least two different arrangements of the 'children', which is known as polymorphism[2]. Polymorphism is the ability of a compound to crystallise in more than one distinct form. There is considerable debate over how common it is to encounter several polymorphs of a single compound under normal laboratory conditions. Some state it is quite common[4], whilst others are inclined to disagree, with Bernstein stating in reference to just finding polymorphs of a single compound of interest for a PhD thesis – *'The author has yet to encounter an academic research advisor who would be prepared to take the responsibility of assigning such a research project to a PhD student'*[5]. Although crystal polymorphism is the different arrangements of the molecules, this has an important impact on the physical properties. This is highlighted by the use of the crystalline substances in the pharmaceutical industry. Crystalline polymorphs can differ in density, refractive index, melting point, and solubility[6]. It is therefore crucial to safely develop and market one specific polymorph of a particular compound, and comply with the current International Conference of Harmonisation (ICH) guidelines[7]. The importance and relevance of crystal polymorphism has not always been a priority within the industry, as in the past scientists may have carried out only a few dozen crystallisation experiments, and possibly proposed a handful of different salts of the compound[8]. However due to recent cases involving multi million pound drug compounds, for example Ritonavir[9;10] in which a more stable conformational polymorph[11] was found two years into bulk manufacture, it is now essential that polymorphism studies are performed prior to the initiation of clinical studies[12]. This has become more facile by the advent of high-throughput crystallisation techniques[8;10;13;14] in which many crystallisations can be performed in a short time frame. Another example of the importance of crystal polymorphism is in the explosives industry, where a number of polymorphs can have very different sensitivity to detonation[15-17].

The crystallisation of a compound is governed by thermodynamics and kinetics. The Gibbs free energy G = H – TS (where H is the enthalpy, and S the entropy of the system) shows the balance

between the tendency towards maximal disorder of the molecules in the system, and the opposing tendency for attractive intermolecular forces to bring the system into a state where the potential energy is minimised[18]. Thermodynamics does predict that any substance must crystallise providing it is pure and the temperature is low (or the pressure is high) enough[19]. Nevertheless crystallisation does not necessarily produce the most thermodynamically stable form, even at a given temperature and pressure. Indeed Ostwalds 'law' indicates that in a polymorphic system the crystallisation processes may be complex, starting with the appearance of the least stable form and finishing with the most stable[20;21]. There are many exceptions to this rule, with metastable structures found from a variety of solvents[22-24] that can usually be explained either on a structural basis or on the basis of irreversible thermodynamics[25]. Therefore a particular crystal form may not be at the global minimum in terms of free energy, but it may be the most common outcome because it is kinetically dictated by the reaction conditions[26]. Indeed the differences in polymorphic behaviour is highlighted by systems in which an observed polymorph can be metastable with respect to other known polymorphs over the whole temperature range. This monotropic behaviour is illustrated in Figure 1.1 and is contrasted with an enantiotropic relationship where the relative stability of two polymorphs changes at a transition temperature T_p.

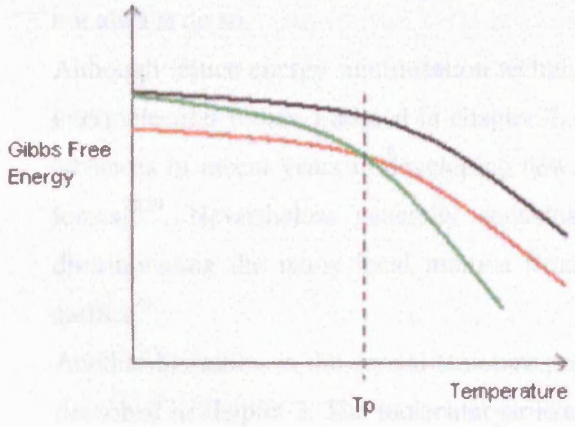


Figure 1.1 Gibbs free energy vs temperature plot of a trimorphic system, Form 1 (red), Form 2 (green) and Form 3 (black). Form 1 and Form 2 are enantiotropically related, with Form 1 being more stable below the transition temperature T_p, and Form 2 above T_p. Form 3 is monotropically related to both Form 1 and Form 2 and is metastable over the whole temperature range

With the continuing importance of polymorphism, can we predict polymorphs of a compound just by knowing the molecular structure? A well-known quote by John Maddox summed up the frustration of crystal structure prediction in the late 1980's – *"One of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids from the knowledge of their chemical composition"*[27]. In 1994 Gavezzotti stated that the concise answer is *"no"*'[19], with Dunitz nine years later concluding that

it could still be deemed a *"no"*, although at certain levels of discussion a *"maybe"*[18]. This is further demonstrated by the international blind tests on crystal structure prediction[28-30] in which some participants correctly predicted some crystal structures, with generally more success for rigid molecules rather than those with conformational flexibility. Flexible molecules do pose a significantly greater challenge as the number of variables needed to define the crystal structure is considerably increased. A molecule capable of strong intermolecular interactions may compete with intramolecular interactions to induce a strained conformer away from the stable gas phase conformation, an example being for 2,4-dibromoestradiol[31]. However a study on conformer distributions found that crystal packing effects in the majority of cases investigated do not have a strong systematic effect on molecular conformers[32]. Hence it is preferable to develop these crystal structure prediction techniques on molecules that can be assumed to be rigid.

Many methods of modern crystal structure prediction are based on finding the minimum in the lattice energy, as one of the main assumptions in this area is that the lower the lattice energy the more stable the structure. In crystals molecules typically pack together to optimise the intermolecular interactions, which tends to be comparable to the close packing principle[33]. Therefore the most stable polymorph lacking strongly directional intermolecular interactions at 0 K will be the one with the highest density[34]. Crystal packing dominated by van der Waals interactions usually follow this rationalisation, however crystals containing hydrogen bonds may not always do so.

Although lattice energy minimisation techniques are well developed, choosing a model for the intermolecular forces, outlined in chapter 2, still can prove a challenge. There have been some advances in recent years in developing new potentials[35-37] and ways to model the electrostatic forces[38;39]. Nevertheless generally speaking no existing force field is up to the task of discriminating the many local minima from the one global energy minimum on the energy surface[40].

Another limitation in the crystal structure prediction process is the nature of the search method described in chapter 3. The molecular structure in the computational predictions in this thesis is assumed to be rigid, with the crystal structures having one molecule in the asymmetric unit. The asymmetric unit can be defined as the smallest portion of the crystal structure (consisting of a molecule or molecules that are not related by symmetry) to which crystallographic symmetry can be applied to generate one unit cell. In addition many crystal structure prediction methods only search the most popular space groups, as this is more practical and less computer intensive. This can be partly justified as it has been found that the majority of crystal structures (with $Z' \leq 1$) can be found in only six space groups[41]. However there is a small risk on missing low energy

minima corresponding to other (higher symmetry) space groups. While attempts have been made to predict crystal structures with Z' = 2[42;43] (despite only approximately 10 % of known organic crystal structures have Z' > 1[44]), and with more conformationally flexible molecules[45-47], this remains one of the stumbling blocks in successful crystal structure prediction.

Despite there being considerable success at predicting crystal structures based on lattice energy alone[28;47-51], the role of free energies[52] is also important as it has been suggested that entropy differences up to 4.5 kJ mol^{-1} at room temperature between polymorphs are possible[53]. Some ideas on kinetics have been explored, including mechanical stability[54;55] and growth volumes[56] of the hypothetical structures, both outlined in chapter 3. Nevertheless understanding the kinetic factors in crystal structure prediction is in its very early stages, with major projects such as the Basic Technology project on the 'control and prediction of the organic solid state' (www.cposs.org.uk) attempting to address kinetic effects during crystallisation.

To evaluate the crystal structure prediction methodology it is essential that all polymorphs are known. Consequently some of the research in this thesis will involve both computational and experimental polymorph studies, chapter 3. As a result of the experimental polymorph screens reported in this thesis, twenty three crystal structures were characterised (of which six were published in Acta Crystallographica section E[57-62]). These structures include new polymorphs of barbituric acid[63] and 6-methyluracil.

A range of approximately rigid polar organic molecules will be studied in this thesis. In chapter 4 crystal structure prediction will be used to study polymorphism of several DNA bases, and will investigate the relative energies of the hypothetical low energy structures and the effects of slight molecular changes on the computational predictions. Chapter 5 will look at using crystal structure prediction to highlight trends in unused hydrogen bond acceptors in related heterocyclic compounds[63;64]. Chapters 6, 7 and 8 describe some computational aspects of crystal structure prediction. The effects of slight conformational changes in uric acid on the low energy crystal structures will be investigated in chapter 6. Chapter 7 will report the formal third international blind test predictions on hydantoin[28], whilst chapter 8 looks at the informal blind test predictions on both 3-oxauracil and 5-hydroxyuracil[49], in collaboration with GlaxoSmithKline. Therefore this thesis evaluates the strengths and weaknesses of a crystal structure prediction process based mainly on lattice energy minimisation.

2. Intermolecular forces within organic crystals

The intermolecular forces are an important part of determining the structure of the crystal and are dependent on the specific molecule being considered. These intermolecular forces can be classified into two types, long range and short range contributions to the intermolecular energy. In this context long range is defined when the energy of interaction behaves as an inverse power of R (atom-atom intermolecular distance) due to the lack of overlap of the molecular charge distributions, while short range is when there is some overlap and the energy decreases exponentially with distance. The sum of the energies of these intermolecular interactions (i.e. the intermolecular potentials) between all the molecules in the crystal defines the lattice energy of the system[65]. For these contributions to the intermolecular energy only neutral closed shell molecules will be considered, ignoring the possibility of magnetic and resonance interactions.

2.1 Long range contributions to the intermolecular energy

2.1.1 Electrostatics

The electrostatic energy is the first order term in the long range perturbation theory with the electrostatic energy arising from the rearrangement of the valence electron density on bonding. This major contribution to the intermolecular forces can be both repulsive and attractive depending on the orientation of the molecules, and is strictly pair wise additive as it is defined in terms of the undistorted charge densities. The Coulombic interaction between the undistorted molecular charge distributions is defined as (equation 2.1):

$$U_{electrostatic} = \int_{\text{all space}} \frac{\rho^A(r_1)\rho^B(r_2)}{|r_1 - r_2|} d^3r_1 d^3r_2 = < 0^A 0^B \mid H' \mid 0^A 0^B >$$

Equation 2.1

where ρ^A is the charge distribution corresponding to the ground state wave function 0^A of molecule A in isolation. The H' term is the perturbation operator which represents the sum of the charges of atoms A and B as $\sum \frac{e_i^A e_j^B}{4\pi\varepsilon r_{ij}}$. Integration of equation 1 is needed to model the Coulombic interaction, however this is fairly limited in its use and therefore some approximations are required. The traditional method of quantifying electrostatic forces is using a central multipole expansion, whereby the series of multipole moments determined directly from the definition in terms of ground state expectation values of the operators (equation 2.2):

$$\hat{Q}_{ik} = \rho r' C_{ik}(\theta,\phi)$$

Equation 2.2

where ρ is the charge density operator and the $C_{ik}(\theta,\phi)$ term is the modified spherical harmonic.

However the central multipole expansion is poorly convergent at shorter separations, so its use is fairly limited, mainly to model reasonably spherical molecules like N_2. This series is also only valid when there is no overlap of spheres around each molecule which contain the molecular charge distribution[66]. For example many terms in the central multipole expansion of cytosine (chapter 4) will be needed before it includes even the crudest representation of the charge density of the amine group, and the expansion is not valid for many orientations where the molecules are in van der Waals contact.

2.1.2 Theoretical modelling of the electrostatic interactions

A variety of computational methods have been used to model the electrostatic contribution to the lattice energy. One of the simplest is to use a point charge assigned to every nuclear position, which assumes that the charge density around each atom is spherical. These point charges can be derived in a number of ways from the molecular charge distribution[67;68], for example potential derived charges which are optimised to approximately reproduce the electrostatic potential outside the molecule. Nevertheless no atomic charge model is capable of accurately representing the electrostatic potential for a range of molecules[69]. Atomic point charge models are still widely used in a number of areas of scientific research, including calculating morphology attachment

energies and growth volumes[56;70]. The atomic point charge model can be improved by using additional sites, such as satellite point charges[39;71].

2.1.3 Distributed Multipole Analysis (DMA)

Charge density maps from X-ray and neutron diffraction data, for example for parabanic acid[72] and alloxan[73], shows that the electron density around the atoms in a molecule is not spherical. Therefore a more elaborate computational method is required to model the non-spherical nature of the electrostatic forces arising from lone pairs and π- π electron density[74]. The method used in this thesis involves using multipoles, derived from the distributed multipole analysis (DMA) technique of Stone[75;76]. This is based on the density matrix of the molecule, expressed in terms of Gaussian primitives η that comprise the atomic orbital basis set (equation 2.3):

$$\rho(r) = \sum_{ij} \rho_{ij} \eta_i \eta_j$$

Equation 2.3

Each contribution to the charge density can be exactly represented by a series of atomic multipoles at a point determined by the origin and the exponents of the two Gaussian primitives involved. If the contribution arises from orbitals on different atoms the point multipoles would be sited between the nuclei, so this charge density is represented by a multipole series on the nearer nucleus[74]. This method automatically generates significant atomic quadrupoles and dipoles to represent the anisotropic electrostatic potential arising from non-spherical features in the charge density. The accuracy of this method is very dependant on the relative separation and orientation of the molecules, and the order of atomic multipoles used. In this work the electrostatic contribution to the lattice energy was obtained by summing all the terms in the atom-atom multipole expansion[77] up to R^{-5}, using Ewald summation for the charge-charge, charge-dipole and dipole-dipole contributions and direct summation over entire molecules whose centres were separated by up to 15 Å for the higher multipole interactions.

The distributed multipole analysis is sensitive to the quality of the wave function used. In this thesis both the SCF (self-consistent field) and MP2 (second-order Møller-Plesset perturbation theory[78]) levels of theory will be used to calculate the wave function, as outlined in the various chapters. MP2 gives a more realistic charge density as it includes electron correlation. The

calculated molecular dipoles are overestimated when using the SCF wavefunction with MP2 giving a better approximation. This is shown in Table 2.1 for a variety of heterocyclic compounds, with a 7 – 30 % overestimation of the molecular dipoles when using a SCF quality wavefunction. This can have a significant effect on the accuracy of the calculation of free energies in some simulations[79].

Table 2.1. Calculated molecular dipoles for a variety of heterocyclic compounds (chapter 5) using both SCF and MP2 quality wave functions.

Molecule	Wave function	Dipole (D)	% difference
Barbituric Acid	SCF	0.90	29
	MP2	0.70	
Alloxan	SCF	2.90	4
	MP2	2.80	
Urazole	SCF	1.73	15
	MP2	1.51	
Parabanic Acid	SCF	2.31	7
	MP2	2.15	

2.1.4 Visualisation of the electrostatic potential

For discussions on the influence of the intrinsic electrostatic contribution to the hydrogen bonded energy for a variety of structurally similar molecules, the DMA representations of the molecular charge density were used in the program ORIENT[80]. This then calculates the electrostatic potential on a grid of points 1.4 Å from the van der Waals surface of each molecule, as defined by the Pauling van der Waals radii of 1.5 Å for nitrogen, 1.4 Å for oxygen, and 2.0 Å for carbon. The hydrogen atoms have no explicit radius. This corresponds to the water accessible surface which approximates to the position for the closest approach of a water molecule. Points are generated at intervals of 0.3 Å and viewed using the program ESTGEN[81;82]. The V_{max} and V_{min} are the values corresponding to the maximum and minimum on the electrostatic potential surface.

2.1.5 Induction energy

The induction energy is the attractive term arising from the distortions of the charge density of each molecule due to the field arising from the other (undistorted) molecule, and can be defined using second order perturbation theory[77;83] (Equation 2.4):

$$U^{A}_{induction} = -\sum_{n^A \neq 0^A} \frac{|<0^A 0^B|H'|n^A 0^B>|^2}{E^A_n - E^A_0}$$

Equation 2.4

This term describes the additional energy due to the changes in the charge density of A (a molecule whose excited state wave functions n^A are of energy E^A_n) caused by the presence of B, interacting with the undistorted ground state charge distribution of B. The energy is always attractive, because the distortions occur only to lower the energy of the pair and is not pairwise additive. It is very difficult to model this contribution to the intermolecular forces for small neutral molecules, hence it is ignored throughout this thesis.

2.1.6 Dispersion energy

In the case of argon there are no electrostatic or induction forces due to the spherical nature of the atoms. Nevertheless argon liquefies at low temperature, therefore there must still be a long range attractive force present. This is the dispersion energy which appears in second order perturbation theory as (equation 2.5):

$$U_{dispersion} = -\sum_{n^A \neq 0^A, n^B \neq 0^B} \frac{|<0^A 0^B|H'|n^A n^B>|^2}{E^A_n - E^A_0 + E^B_n - E^B_0}$$

Equation 2.5

The sum over all the excited states of both molecules A and B shows that the dispersion arises from correlated distortions in the two molecular charge densities. The dispersion energy for two argon atoms can be defined as (equation 2.6):

$$U_{dispersion} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} - \ldots$$

Equation 2.6

For polyatomic molecules the dispersion coefficients depend on the molecular orientation. The C_n coefficients can be obtained from the properties of the isolated molecules, for example from the polarisabilities of the two molecules calculated from imaginary frequencies[77]. In this thesis the $-\frac{C_6}{R^6}$ term will be used to model the dispersion forces within the intermolecular potential.

2.2 Short range contributions to the intermolecular energy

2.2.1 Exchange/repulsion

As two molecules approach each other the overlapping charge densities cannot occupy the same space, due to the Pauli Exclusion Principle, as it is not possible for closed shell molecules to accept other electrons in their doubly occupied molecular orbitals. This produces a repulsive force that dominates over the attractive exchange force. In this thesis this interaction is represented by the 6:exp or Buckingham's form[84] (equation 2.7) where a and b are to be determined.

$$E_{nb}(R) = a\exp(-bR)$$

Equation 2.7

2.2.2 Charge transfer and other short range terms

A further term that arises in second order perturbation theory is the transfer of charge from the occupied orbitals of one molecule to the unoccupied orbitals of the other. This type of interaction can be evaluated[85], and has been shown to decay approximately exponentially with separation, and is intrinsically non-additive. If the charge transfer is relatively large then it approximates to a covalent interaction, whilst if it is small then it can be absorbed into the intermolecular potential. One drawback of evaluating this interaction is that its magnitude can be grossly overestimated due to the *ab initio* basis set superposition error (BSSE)[86]. BSSE involves the use of orbitals based on one molecule to describe the charge distribution on the other, which can improve the description of the intramolecular interactions within one molecule. This type of problem can be overcome[86;87], but the results should be viewed with caution.

The other short range terms arise from the overlap of molecular wavefunctions, mainly penetration and damping effects, Table 2.2.

2.2.3 Summary of intermolecular forces

Table 2.2 summarises the short and long range intermolecular interactions between molecules.

Table 2.2 Summary of the contributions to the energy of interactions between molecules[77].

Contribution	Additive?	Sign	Comment
Long –range ($U \sim R^{-n}$)			
Electrostatic	Yes	+/-	Strong orientation dependence
Induction	No	-	
Dispersion	Approx.	-	Always present
Resonance	No	+/-	Degenerate states only
Magnetic	Yes	+/-	Very small
Short-range ($U \sim e^{-\alpha R}$)			
Exchange	No	-	
Repulsion	No	+	Dominates at very short range
Charge Transfer	No	-	Donor-acceptor interaction
Penetration	Yes	-	Can be repulsive at very short range
Damping	Approx.	+	Modification of dispersion and induction

2.3 Hydrogen bonding

Hydrogen bonds are the highest energy interactions in molecular crystals, and they greatly affect the way in which certain molecules pack in the crystalline environment. Hydrogen bonding consists of a donor and acceptor, D-H···A, with stronger hydrogen bonds associated with the most electronegative atoms, mainly N, O, F and Cl.

There has been considerable debate on the nature of hydrogen bonding. Early work suggested electrostatic[88] or covalent character[89;90] for rather weak, or very strong hydrogen bonds respectively. In the past the term 'hydrogen bond' was restricted to interactions like N-H···O, O-H···O and F-H···F[91], however today the concept of the hydrogen bond is extended to weaker interactions like C-H···O, C-H···N and O-H···π , which have hardly any covalent character and are only marginally electrostatic[92]. It is now recognised that weak hydrogen bonds are electrostatic in nature but become increasingly covalent with increasing strength[93;94]. The differences between the relative strengths of hydrogen bonds (including very strong and weak hydrogen bonds) are shown in Table 2.3. In this thesis the hydrogen bonding that will be studied will be of the types

N-H···O, O-H···O, and N-H···N interactions. Often hydrogen bonds are defined either by the H···A or the D···A distance. The D···A distance is preferred as this does not take into account the position of the hydrogen atom, which can be subjective if the hydrogen positions have been determined from X-ray data.

The DMA, section 2.1.3, used throughout this thesis will model the electrostatic contribution to the hydrogen bonding, and reproduces the directionality of the hydrogen bonding interactions fairly adequately. This is highlighted by the Buckingham-Fowler model in which using atomic point multipoles gives correctly predicted geometries of ten van der Waals complexes of the hydrogen bonded type[95].

Table 2.3 The differences between the relative strengths of hydrogen bonds[96]

	Very strong	**Strong**	**Weak**
Bond energy, kJ mol^{-1}	163-167	17-63	< 17
Examples	[F···H···F]$^-$, [N···H···N]$^+$	N-H···N, O-H···O, N-H···O	C-H···O, O-H···π
D···A, Å	2.2-2.5	2.5-3.2	3.0-4.0
H···A, Å	1.2-1.5	1.5-2.2	2.0-3.0
Angle, °	175-180	130-180	90-180
Effect on crystal packing	Dominant	Distinctive	Variable
Covalency	Pronounced	Weak	Vanishing
Electrostatics	Significant	Dominant	Moderate

2.4 Model potentials

In the computational studies the intermolecular forces need to be accurate enough for the calculation of the lattice energy for the relative stabilities of the predicted polymorphs. In this work the repulsion and dispersion forces within the crystalline environment will be modelled using a potential. The minimal requirement of a potential is that it should reproduce the observed crystal structures without excessive changes in geometry[43]. As the energy differences between the various predicted polymorphs are very small[97] it is vital that the potential used in the computational predictions is as accurate as possible for modelling the intermolecular forces.

The usual model intermolecular potential for organic molecules is based on the assumption that the interaction between the molecules is the sum of the interactions between their constituent atoms. In this work the electrostatic term is defined in terms of the multipoles calculated from the distributed multipole analysis (section 2.1.3), with the rest of the interaction energy between molecules M and N given by (equation 2.8):

$$U(R,\Omega)=\sum_{i\in M,k\in N}[A_{\iota\kappa}\exp(-B_{\iota\kappa}R_{ik})-\frac{C_{\iota\kappa}}{R_{ik}^{6}}]$$

Equation 2.8

where atom i in molecule M is of type ι, and atom k in molecule N is of type κ.

Even with this assumption there are still a considerable number of parameters that need to be considered, depending on the number of types of atoms in the molecule. Another common approach to reduce the number of parameters is to assume some form of combining rules (equation 2.9):

$$A_{\iota\kappa}=(A_{\iota\iota}A_{\kappa\kappa})^{1/2}\quad B_{\iota\kappa}=0.5(B_{\iota\iota}+B_{\kappa\kappa})\quad C_{\iota\kappa}=(C_{\iota\iota}C_{\kappa\kappa})^{1/2}$$

Equation 2.9

The $C_{\iota\kappa}$ dispersion combining rule has some physical relation to the R^{-6} dispersion coefficient, but the other rules, though still widely used, are poorly justified and have been found to be limited in their accuracy[98]. The two potentials considered for the computational studies in this thesis will be the W99[99] and FIT[100-102] potentials.

The parameters present in the FIT potential are for C, N, O, and polar (H_p) and non-polar (H_c) hydrogens, to keep the number of types to a minimum[100]. The empirical potential repulsion-dispersion C, H_c, and N parameters were derived by Williams and Cox[102], and the O parameters taken from the fitting to oxohydrocarbons[101]. The polar hydrogen parameters (H_p) were originally fitted to intermolecular perturbation theory calculations of the exchange-repulsion, penetration and dispersion interactions between formamide and formaldehyde[103], and then improved by fitting the parameters to a dataset of thirteen hydrogen bonded structures[100] in conjunction with a DMA electrostatic model. The FIT potential itself, or some of the empirical parameters present have been used in a wide range of studies including on three furazan derivatives[104], 5-fluorouracil[51], and aspirin[47]. The FIT potential with the carbon repulsion parameters decreased by 25 % (for reasons described in chapter 5) will be used for the computational research in this thesis, unless stated otherwise.

The W99 potential was derived from fitting the empirical parameters to observed hydrocarbon crystal structures (with 5-16 carbon atoms)[105], crystalline oxohydrocarbons which include

hydrogen bonding[106], and azahydrocarbon compounds[99], from which their molecular wave functions and molecular electric potential (MEP) were calculated from HF/6-31G** basis sets.

The W99 potential is an improvement on the FIT potential[100-102] as it has different structural classes defined for the H, C, N, and O atoms. For example there are four structure classes for hydrogen, to model the repulsion and dispersion forces for a hydrogen in a C-H, hydroxy, carboxyl and a N-H group. Another major difference compared to the FIT potential is that the X-H bond distances were set to standard values for the wave function calculation and then foreshortened by 0.1 Å for the force field fitting. Therefore when calculating the DMA used in conjunction with this model potential, standard neutron X-H bond lengths[107] (X = C, N) are used, with the hydrogen interaction sites foreshortened by 0.1 Å along the bond to represent the displacement of electron charge. This however causes considerable problems in implementation of this model.

The W99 potential has recently been used for crystal structure prediction using a Monte Carlo prediction method[108], and for the assessment of lattice energy minimisation using 50 small rigid molecules[50]. In these cases it gave some improvement over the other potentials tested, including the FIT[100-102], Dreiding[109], CVFF95[110;111] and COMPASS[112] potentials. In addition the W99 potential was used in a study where it was found that using multipoles compared to using atomic point charge electrostatics in the calculations gave some important improvements in the reliability of lattice energy minimisation for the prediction of crystal structures[113].

The modelling of the intermolecular forces in this thesis will involve a good electrostatic model derived from distributed multipole analysis. The dispersion and repulsion forces will be represented by a model potential, nevertheless it should be noted that there will still be some inadequacies present due to the empirically fitted nature of this technique.

3. Computational and experimental methods

In this chapter the computational and experimental techniques used in this thesis will be described. These involve computational crystal structure prediction, along with experimental polymorph screens to determine the polymorphic behaviour of a selection of small organic compounds.

3.1 Computational methods

Initially studies will be performed to test the computational model for the intermolecular forces on the known crystal structure(s), described in sections 3.1.1 to 3.1.3 and shown in Figure 3.1. This computational model will then be used in the crystal structure prediction process, described in sections 3.1.4 to 3.1.7. The comparison of the hydrogen bonding graph sets of the crystal structures is described in section 3.1.8, with the property calculations on the low energy and known crystal structures described in sections 3.1.9 to 3.1.14.

3.1.1 Initial computational studies

The initial studies will involve testing the computational model for the intermolecular forces (chapter 2) within the crystalline environment for the specific molecules. If this modelling is unsatisfactory, another model potential could be used and/or the potential or the quality of wavefunction for the DMA could be modified within the boundaries suggested by the theory of intermolecular forces and the derivation of the potential, chapter 2.

The solid state crystal structure (with the X-H bond lengths adjusted to standard neutron values[107] if determined using X-ray data), usually obtained from the Cambridge Structural Database[114], is lattice energy minimised using DMAREL[115] (section 3.1.6) and is denoted ExptMinExpt. This is to determine how well the crystal structure is reproduced to judge the modelling of the intermolecular forces. However if there is no solid state crystal structure available (as in the International Blind Tests[28-30], chapter 7), then crystal structures of closely chemically related molecules could be used.

The experimentally determined molecular structure is the fundamental input in this process and another conformation has to be used for genuine predictions, and to investigate the effect of non-rigidity of the molecule. This molecular structure is usually the *ab initio* 'gas phase' conformation of the isolated molecule, optimised using Gaussian98[116] using the MP2/6-31G** level of theory (unless otherwise stated) starting from the solid state molecular structure or one built using MOLDEN[117]. This *ab initio* molecular structure is not affected by packing forces which differ between polymorphs and this conformation is placed in the solid state crystal structure by generally optimising the RMS overlap of the molecules. This crystal structure is then lattice energy minimised and denoted ExptMinOpt. A comparison between ExptMinExpt and ExptMinOpt (by assessing the lattice energy minimisation, section 3.1.2) can highlight the influence of molecular conformation on the theoretical calculations. Other molecular conformations in the calculations could also be considered if necessary.

Once a satisfactory method of modelling the intermolecular forces has been found, this is then used in the computational polymorph predictions.

3.1.2 Assessing lattice energy minimisations

To assess how well a crystal structure has been reproduced after lattice energy minimisation, a number of parameters can be compared. This includes analysing the percentage and root-mean-square percentage errors in the unit cell lengths and angles, and the hydrogen bonding distances in the structure.

The F value 'figure of shame'[118], equation 3.1, gives a general indication of how well a lattice energy minimisation has reproduced the initial crystal structure, with the higher the value the less accurate the reproduction. An F value of order 50 or more indicates that the lattice energy minimised crystal structure differs significantly from the experimental structure and therefore there is a problem with the modelling.

$$F = (\Delta\theta/2)^2 + (10\Delta x)^2 + (100\Delta a/a)^2 + (100\Delta b/b)^2 + (100\Delta c/c)^2 + \Delta\alpha^2 + \Delta\beta^2 + \Delta\gamma^2$$

Equation 3.1

The $\Delta\theta$ represents the total rigid-body rotational displacement after minimisation (°), with the Δx representing the total rigid-body translational displacement (Å). The other six terms depend on the changes in cell parameters (Å and °). The 0.5, 10, and 100 bring the contributions from the different displacements to a comparable scale.

The errors in the lattice energy minimisation, and the sensitivity to experimental variations in the molecular structure of paracetamol[119] suggested that differences of around 3 % in the unit cell parameters between the minimised and solid state structures may arise from approximations such as the neglect of thermal effects rather than inadequacies in the model potential. A 'typical thermal expansion' argument is that a few % in the cell parameters is satisfactory agreement between the lattice energy minima and solid state structures for neutral organic molecules. For this reason, it was deemed in this thesis that any error above 5 % in the cell parameters would be unacceptable.

3.1.3 Comparison of lattice and sublimation energies

The enthalpy of sublimation, ΔH_{sub} of a solid is the experimental thermodynamic quantity describing the stability of the crystal structure, and is a measure of the strength of the intermolecular interactions[120]. ΔH_{sub} can be estimated as $\Delta H_{sub} = -E_{lattice} - 2RT$[121], where 2RT represents a correction factor for the difference between the gas phase enthalpy and the vibrational contribution to the crystal enthalpy[122]. This can be useful when comparing how accurate the lattice energy minimisation has approached the lattice energy of the crystal structure, as this energy should approximate the sublimation energy at 0 K within the errors present. It has been argued that differences between experimental sublimation energies and calculated lattice energies is generally within the 3 – 4 kcal mol^{-1} (12 – 17 kJ mol^{-1}) that 'should not cause any concern when judging the quality of the parameters of the potential'[123]. Nevertheless these experimental and theoretical errors need to be taken into account when gauging the confidence in this comparison, which is made in this thesis in the limited number of cases where ΔH_{sub} is available (Chapter 4).

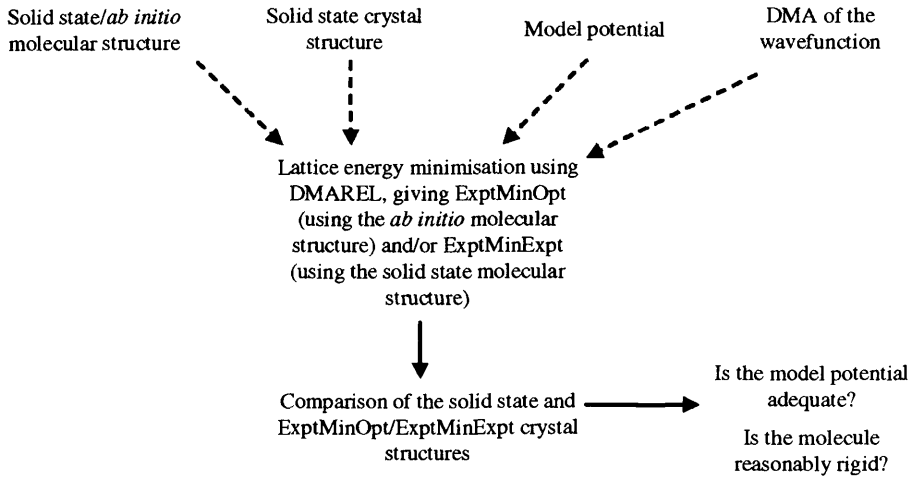


Figure 3.1 The process for testing the computational model for the intermolecular forces in the known crystal structures

3.1.4 Methods to search for initial trial crystal packings

For a computational crystal structure prediction search to be a success, many trial structures need to be considered which, after lattice energy minimisation, will correspond (if low enough in energy) to the energetically feasible crystal structures. There are a variety of current methods that can be used to produce these trial structures, a selection are outlined in Table 3.1.

Table 3.1 A selection of current methods of generating initial trial crystal structures

Program	Method	Selection of studies
ICE9[124]	Vectorised grid search, only input being molecular geometry, multipole moments and van der Waals volume	On a variety of rigid hydrocarbons, including benzene, naphthalene, tetracene, and pentacene[124]
MGAC[125]	Modified genetic algorithm	Benzene, naphthalene, and antracene[125]; benzene II at 25 Kbar[126]; silicon clusters[127]; L-alanine and DL-alanine[128]
MPA[129;130]	*Ab initio* molecular packing analysis, Monte Carlo	Benzene[131]; *m*-nitroaniline[132]
MSI Polymorph Predictor	Monte Carlo, simulated annealing	Primidone and progesterone[133]; aspirin[134]; diasteromeric salts[135]
Perlstein[136]	Monte Carlo	13 molecules randomly chosen[136] from the CSD[114]; molecular aggregates[137]; semi-flexible organic molecules[138]
PROMET[139;140]	Stepwise construction of dimers and layers	7-dimethylaminocyclopenta[c]comarin[141]; t-butyldiazopyruvate[142]; HNN radical[143]
UPACK[42;144]	Random/systematically generated starting structures	Small carbohydrate molecules[144]; acetic acid[145;146]; ethanol and benzene[147]

An extensive survey of the current crystal structure prediction methods and molecules studied can be found at www.cposs.org.uk (Control and Prediction of the Organic Solid State). The program used to generate the trial crystal structures in this thesis will be MOLPAK[148], described in section 3.1.5, with DMAREL[115] as the lattice energy minimisation algorithm, outlined in section 3.1.6. This gives a comparatively efficient method of generating good initial crystal structures which is necessary to balance the relatively expensive cost of using atomic multipole models in the lattice energy minimisation.

3.1.5 MOLPAK (MOLecular PAcKing) program

MOLPAK[148] is a program which uses a rigid body packing probe to build trial structures for lattice energy minimisation, which has been previously used in a variety of studies including on paracetamol[149], uracil[150], and 5-azauracil[151]. It performs a systematic grid search on orientations of the central molecule in 29 common co-ordination geometries of organic molecules in 13 space groups, outlined in Table 3.2. During the course of this PhD in early 2004, it was expanded to include 11 new packing types in a total of 20 space groups. It is currently limited to structures in which there is only one molecule in the asymmetric unit.

MOLPAK was originally developed in the early 1990's for the prediction of crystal structures of energetic materials where density is a key factor. This involved an analysis of 242 C, H, N, O and F containing compounds (in the primitive, triclinic, monoclinic and orthorhombic space groups with $Z \leq 4$) to determine the common co-ordination sphere patterns. This co-ordination sphere of a molecule consists of molecules that are in contact or close to the van der Waals contact with the central molecule. MOLPAK uses a pseudo hard-sphere repulsion potential and predetermined docking thresholds when packing molecules within the various co-ordination types. This type of potential is used because it is considerably faster than one that minimises the energy between molecules, due to the need to examine all atom-to-atom interactions between molecules at each step. The search consists of 10 ° steps for the orientation of the molecule to seek the densest structures, thus using the close packing principle[33] in the search method. For every initial orientation the possible packing is constructed by the approach of two molecules to form a structure line, then the approach of two structure lines to form a two-dimensional grid. A three-dimensional grid is then generated by the approach of two-dimensional grids of molecules to each other. The packing type/space group symmetry is imposed during the process. Many trial

structures are generated and the 50 to 200 densest structures are then taken forward to be used as input for the lattice energy minimisations.

Table 3.2 The space groups and co-ordination types[a] considered in the MOLPAK process, with the ones used in the extended MOLPAK shown in bold

Space group	Co-ordination types	Space group	Co-ordination types
P1	AA	$Pna2_1$	**AU**, AV, BD, BF
$P\bar{1}$	AB, CA	Pba2	**AW, BG**
Pc	**AD**	$Pca2_1$	AY, BH
$P2_1$	AF, AH	$Pmn2_1$	**BJ**
$P2_1/c$	AI, AK, AM, FA, FC	Pma2	**BK**
P2/c	**AJ, AL**	Pbca	CB, CC
$P2_1/m$	**FB**	Pbcn	CD, CE
$P2_12_12$	AP, BA, BB	Cc	DA
$P2_12_12_1$	AQ, AZ	C2	DB
Pnn2	**AR, BE**	C2/c	DC, DD, DE

[a]Co-ordination types labelled as in the original MOLPAK paper[148]

3.1.6 DMAREL lattice energy minimisation algorithm

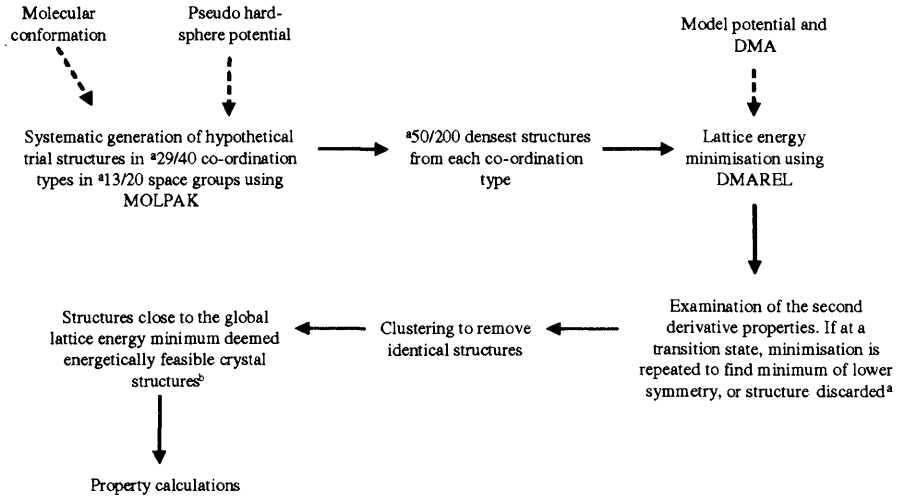
DMAREL[115] is a program in which the lattice energy is calculated and then minimised for a particular crystal structure, with the assumption that the molecules within the unit cell are rigid. This program lattice energy minimises while maintaining space group symmetry[152]. One of the key features is the ability of DMAREL to use atomic multipoles (from the distributed multipole analysis of the wave function[75;76], section 2.1.3) to model the electrostatic forces within the crystalline environment, in addition to using the model potential. The electrostatic sums for charges and dipoles are evaluated using the Ewald approach[153;154], higher multipolar interactions use a molecule based cut-off for a direct summation, with the short range potentials summed using an atom based cut-off.

DMAREL uses a modified Newton-Rhapson method to optimise the cell parameters and the positions and orientations of the rigid molecules, which involves calculating the forces and the second derivatives of the lattice energy. The lattice energy is minimised with respect to the changes in the shape of the unit cell and the orientation and centre of mass vectors of each molecule. If the second derivatives show that the crystal structure is at a transition state within

the space group, then symmetry reduction can be performed. This involves removing the symmetry operation corresponding to the negative eigenvalue, and restarting the minimisation procedure. This process often leads to crystal structures with Z' = 2.

3.1.7 Clustering of the low energy structures

The initial generation of trial structures and lattice energy minimisations result in many crystal structures at the same minima on the potential energy surface. To cluster these equivalent crystal structures the conversion of the arbitrary primitive cell into the reduced cell is performed. Any crystal lattice can be represented by a positive ternary quadratic form[155], the reduced cell. Such a cell provides a unique description of the lattice and is defined independently of lattice symmetry[156]. Originally Niggli[155] derived geometrically the reduced forms for all the Bravais lattices, with a unified algorithm for determining the reduced cell developed in the late 1970's[157]. This is implemented in PLATON[158] and in the MOLPAK/DMAREL procedure. In addition clustering can be aided by comparing simulated powder patterns[159;160] and contrasting the co-ordination spheres using COMPACK[161;162]. The low energy crystal structures were visualised using Mercury[163] or Cerius2[164]. The whole crystal structure prediction process is shown in Figure 3.2.



[a]MOLPAK parameters/process that are specified in the individual studies in the thesis, since MOLPAK, DMAREL and the other available computing facilities were developed during this thesis work. [b]The approximate energy range between real polymorphs[5] is often 0 – 10 kJ mol^{-1}. In this thesis the energy range considered for polymorphism in the computational polymorph searches will be within 7 kJ mol^{-1} of the global lattice energy minimum. This can be adjusted by a few kJ mol^{-1} depending on the specific molecule.

Figure 3.2 The process of computational crystal structure prediction

3.1.8 Graph Sets

The hydrogen bonding is an important characteristic of the organic crystal structures studied in this thesis. One of the problems with defining hydrogen bonds is the distance criteria used. Different strength hydrogen bonds can usually be classified according to the D···A or H···A distance (as shown in section 2.3), however the H···A distance can be prone to error if the crystal structure has been determined from X-ray data. Therefore the definition of a hydrogen bonding within crystal structures is problematic, and based on a length cutoff. The method used for the comparison in this thesis is the classification of hydrogen bonded structures based on the representation of these structures as graphs. This concept was first introduced in the early 1980's[165] and developed further in the early 1990's[166;167]. This method is based on graph theory for categorising hydrogen bond motifs in such a way that complex hydrogen bond patterns can be disentangled, or decoded, systematically and consistently. This also involves viewing these patterns topographically as if they were intertwined nets with molecules as the nodes and hydrogen bonds as the lines[167]. The graph set analysis consists of different levels that are associated with the different hydrogen bonded nets present. Therefore the combination of the different levels defines the hydrogen bonding in the crystal.

The graph set analysis in this thesis was performed by RPluto[168], with the default program maximum H···A distance for the hydrogen bonding being 2.52 Å for the O-H···O and N-H···O interactions, and 2.55 Å for the O-H···N and N-H···N interactions. The minimum ∠D-H···A angle considered is 90 °. The graph set notation implemented is given by:

G a,d(n)

Where **G** is the pattern descriptor:

R *ring (intermolecular);* **C** *infinite (chain);* **D** *discrete (finite);***S** *self (intramolecular ring)*

a *is the number of acceptors;* **d** *is the number of donors;* and **n** *is the degree of the pattern (path length).*

An example of this graph set notation is R2,2(8), as shown in a hypothetical crystal structure of barbituric acid, Figure 3.3. This shows a ring (intermolecular) pattern with two hydrogen bond donors and acceptors.

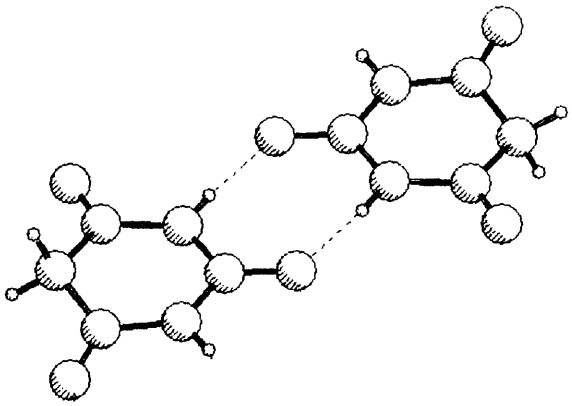


Figure 3.3 A graphical representation of the R2,2(8) graph set as shown in a hypothetical crystal structure of barbituric acid

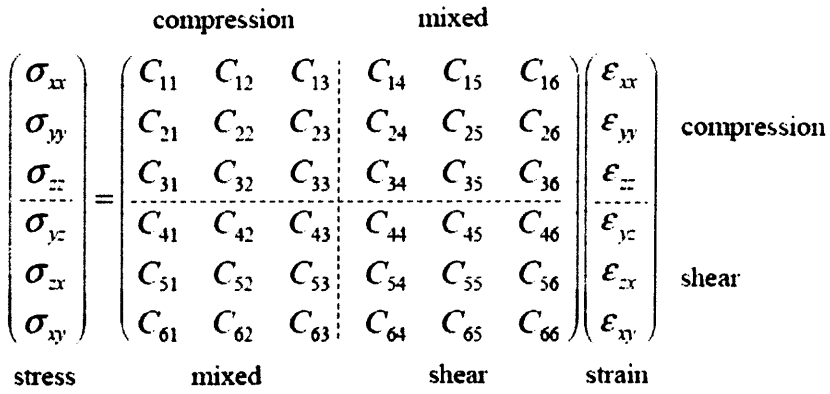
3.1.9 Property calculations

Kinetic effects play an important part in whether we see certain crystals experimentally. Nevertheless the current methods to model some of these aspects of kinetics (the properties) are in the very early stages of development. The low energy structures generated in this thesis will be used to establish whether some of these properties can distinguish those low energy crystal structures that could be deemed more likely to be seen experimentally.

3.1.10 Elastic constants

The elastic (mechanical) properties of solids are determined by the interatomic forces acting on the atoms when they are displaced from the equilibrium positions. At small deformations these forces can be assumed proportional to the displacement of atoms (the harmonic approximation). The elasticity of the crystal, being represented by a fourth-rank tensor, relates the second-rank stress and strain tensors. The stress tensor is defined as the series of external forces acting on the crystal, with the resulting crystal deformation represented in the strain tensor as the change in dimensions of a body as a result of subjecting the body to this system of forces which are in equilibrium. This may result in a crystal extension, compression, or a shear. The elastic constant matrix C, which has 36 elastic constants in total, relate the strain, σ_{ij}, and stress, ε_{ij}, tensors in a linear fashion by $\sigma_{ij} = \sum_{kl} C_{ijkl}\varepsilon_{kl}$. These constants are denoted by C_{mn}, where m and n are

defined as 1 = *xx*, 2 = *yy*, 3 = *zz* for the compression components and as 4 = *yz*, 5 = *zx*, and 6 = *xy* for the shear components. Therefore the general form of the matrix C is given by:



The elastic constants of molecular organic crystals are fundamental to the tableting properties of pharmaceuticals[169;170], however the number of systems in which there are experimentally determined elastic constants available are small[55].

In DMAREL[115] the elastic constants are calculated directly from the potential energy surface, therefore they are only applicable at 0 K for the perfect crystal. For an easier comparison of the mechanical stability for a number of crystal structures the lowest resistance to shear in any direction, calculated as the smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants (C_{ij}, i,j = 4, 5, 6), are presented. This reveals whether the crystal packing has resulted in any weak shear planes. This elastic constant information can give an approximation to whether new polymorphs have better mechanical properties than other solid state crystal structures, as the elastic tensors vary between polymorphs. For example the room temperature experimental elastic constants for urea range from 0.5 to 51.0 GPa for C_{12} and C_{33} respectively[171], showing the considerable anisotropy often seen in organic crystal structures.

3.1.11 Phonons and free energies

The lattice energy minimisations using DMAREL[115] calculate the lattice energy from the potential energy surface, and are strictly relevant to 0 K due to the neglect of thermal effects. An estimation of the free energy at room temperature (298 K) is found by using the second derivative matrix at the lattice energy minimum of the low energy structure to calculate the corresponding 0 K elastic constants[54;55] (outlined in section 3.1.10) and the $k = 0$ intermolecular

phonons[52]. The intermolecular $k = 0$ frequencies correspond approximately to the low frequency IR and Raman spectroscopy range. These phonons are used to estimate the vibrational zero point energy and thermal contributions at 298 K, by using the Debye-Einstein phonon dispersion model[172]. The elastic constants are used to estimate the Debye frequency for the acoustic mode vibrations. It would be hoped that the errors in these approximations might be roughly the same when comparing the different lattice energy minimised structures of the same rigid molecule, particularly when Z is constant[52]. The contributions to the free energy are summarised as:

Free Energy (298 K) = Zero point vibrational energy + Total thermal energy (298 K) + Lattice energy

The more elaborate methods of modelling the electrostatic contribution to the energy (for example using multipoles instead of point charges) give good estimates of the zero point energies[52]. Changes to the model potential also affect each lattice vibration independently, as the vibrational properties of molecular crystals are sensitive to the shape of the repulsive wall[54]. This separate consideration of rigid body motions will be unreliable for flexible molecules, as the different crystal structures will couple the soft intramolecular and intermolecular modes to give very different atomic motions in the crystal[173]. For all the unique crystal structures generated within the energy range of polymorphism in the computational polymorph searches in this thesis, the room temperature free energies will be calculated. This is to see whether there is any significant change to the thermodynamic stabilities of these structures at room temperature rather than 0 K, even within the perfect harmonic lattice approximation for rigid molecules.

3.1.12 Morphology calculations

The prediction and investigation of crystal morphologies is a field of active research[174;175]. Previous computational polymorph prediction studies, one example being on paracetamol[149], have used attachment energy information to identify crystals with at least one face with a low attachment energy, suggesting difficulties in crystal growth of that face (at least by the mechanisms for which the attachment energy model is appropriate[56]). Therefore if a crystal has a very slow growing morphologically important face it is unlikely to be observed. The minimum attachment energies (section 3.1.13) are calculated for a selection of small organic compounds in chapters 5 and 7. The recent ability to estimate the relative growth volumes[56] of the crystal structures has meant that post computational search analysis has made a step towards modelling

some aspects of the kinetics in the predictions. These relative growth volumes (section 3.1.14) are calculated in chapters 5 and 7.

3.1.13 Attachment energy calculations for the morphology prediction

Initial work on predicting crystal morphologies was based on the interplanar spacings, d_{hkl}, of the different crystal faces, with the lowest growth rates occurring at the faces with the largest interplanar spacing[176;177]. Bravais-Friedel-Donnay-Harker (BFDH) theory was then developed which states that the relative growth rate of faces[178] on crystals is inversely proportional to the interplanar distance d_{hkl} and thus their morphological importance (i.e. those faces with the slowest growth rates and the largest surface area are the most morphologically important). However the BFDH method is commonly now only used to select the faces for consideration, as it takes no account of the detailed molecular structure or intermolecular forces[70]. The role of intermolecular forces in the crystallisation process was considered by Hartman and Perdok[179], who developed a growth attachment energy model. In this model the attachment energy $E_{att} = \sum_{i=1}^{\infty} E_i(hkl)$, where $E_i(hkl)$ is the interaction energy per molecule between a slice of thickness d_{hkl} and the ith underlying slice, is proportional to the growth rate and inversely proportional to its morphological importance.

The attachment energy model does not include effects of solvent[175;180] or additives[181], hence the calculations are strictly relevant to a vapour grown crystal. In addition the model cannot predict the differential growth rate of the (h,k,l) and (-h,-k-l) faces in polar crystals such as urea[182], and these is less confidence in the results using the attachment energy model for $Z' \neq 1$ structures.

The attachment energy information can also be used to predict the vapour grown habit of a particular crystal structure[183], as shown in chapter 5.

3.1.14 Growth volumes

A recent significant development in the modelling of crystal morphologies is the ability to use attachment energy information to give an estimate and comparison of the relative growth volumes of the crystals, assuming that the proportionality constant is the same for all faces of

crystals composed of the same molecules. While this is strictly applied to vapour grown crystals it nevertheless gives an indication as to which crystals have a relatively fast growth rate, and hence we can investigate whether such crystals are likely to be seen experimentally.

These growth volume calculations[56] have been performed in collaboration with Dr. David Coombes at the Royal Institution, using the program GDIS[184] to generate the surfaces for study, via a BFDH analysis[185]. The attachment energy for the surfaces were calculated using GULP 1.4[186], using the same model potential, and CHelpG charges[187] (calculated using the same Gaussian98[116] with a MP2/6-31G** wave function) instead of a DMA. It has been found that only minor variations occur in the predicted morphologies when using different model potentials in the calculations[188]. The growth volumes were estimated by calculating the volume within the Wulff shape (in which the distance from the origin to the (h,k,l) face, R_{hkl}, is proportional to the magnitude of the attachment energy) by means of a numerical integration technique[56] using the in-house program CALCVOL[189].

3.2 Experimental methods

In addition to carrying out computational polymorph screens, some experimental studies were performed concurrently to try and establish the experimental polymorphic behaviour of the compound. This involves an initial solubility screen, described in section 3.2.1, followed by an experimental polymorph screen, described in sections 3.2.2 to 3.2.5.

3.2.1 The solubility of a compound

The solubility of a compound depends on the polarity of the solvent, which is related to the dielectric constant (a measure of the ability of a solvent to insulate opposite charges from one another). Generally 'like dissolves like', for example polar compounds will dissolve in polar solvents. Molecules with large dipole moments and high dielectric constants are considered polar, and those with low dipole moments and small dielectric constants are classified as non-polar. Table 3.3 shows the variations in the dipole moments and dielectric constants of a variety of solvents used in the experimental screens in this thesis. These variations show differences in the degree of interaction between the solvent and solute which could affect the way the compound crystallises from solution. This is emphasised in studies in which crystallisation

conditions were manipulated via solvent selection in an attempt to produce polymorphs containing specific packing modes[190-193].

Table 3.3 Typical dipole moments and dielectric constants[194] for a selection of solvents used in the experimental screens, with the full list shown in Table 3.1 SI

Name	Molecular formula	[a]Dipole moment /D	[b]Dielectric constant	[c]T/K
Water	H_2O	1.84	80.1	
Hydrogen bromide	HBr	0.8	8.23	186.8
Hydrogen chloride	HCl	1.1	4.6	300.9
Dichloromethane	CH_2Cl_2	1.6	17.26	298
Formaldehyde	CH_2O	2.3	-	
Nitromethane	CH_3NO_2	3.46	37.27	
Methanol	CH_4O	1.71	33	
Ethanol	CH_6O	1.89	25.3	
Chloroform	$CHCl_3$	1.04	4.8069	
Tetrachloroethylene	C_2Cl_4	0	2.268	303.2
2,2,2-trifluoroethanol	$C_2H_3F_3O$	-	27.68	
Acetonitrile	C_2H_3N	3.92	36.64	
1,2-dichloroethane	$C_2H_4Cl_2$	[d]1.8	10.42	

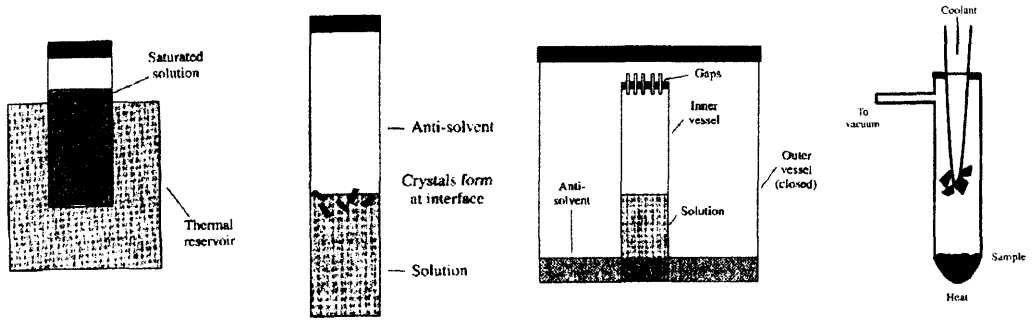
[a]Electric dipole moment in debye units
[b]Static relative permittivity measured in static fields or at low frequencies where no relaxation effects occur
[c]Temperature at 298.2 K for the determination of the dielectric constant, unless otherwise stated
[d]Measurement performed on the pure liquid or in solution, these are less reliable than the other values, which were obtained in the gas phase

Solubility screens were performed on each compound (chapters 4 and 5) so that appropriate solvents can be selected for the crystallisation experiments. The screen involved attempting to dissolve 0.03 g of the compound in 10 cm^3 of the solvent at room temperature. The qualitative solubility of the sample was performed by optical observation and classed into three categories, soluble, partially soluble and insoluble. Soluble is defined when the entire compound has dissolved, partially soluble is when some or most of the compound has dissolved, while insoluble is when it appears that none of the compound has dissolved in the solvent.

3.2.2 Crystallisation techniques

For the experimental polymorph screens a number of techniques will be adopted in an attempt to induce crystallisation. Generally if the solute is not very soluble, the result is often crystals of very small size. Mechanical disturbance of the crystal growing vessel, for example vibration, could also result in smaller crystals. In addition the number of nucleation sites at which the crystals begin to grow is also important, since fewer sites will result in fewer crystals, each of

larger size. The size is important as the crystal has to be of adequate size and quality for the single crystal X-ray studies.



(a) (b) (c) (d)

Figure 3.4 A selection of crystallisation methods[195] showing (a) cooling, (b) solvent diffusion, (c) vapour diffusion, and (d) sublimation.

There are a variety of crystallisation techniques which can be employed in these types of experimental screens[195]. The typical approaches used in the research described in this thesis are:

- *Slow evaporation* – This is the simplest way to grow crystals and the method that works best for compounds that are not sensitive to ambient conditions in the laboratory. The solvent is evaporated off, which causes the solution to become supersaturated which induces crystallisation.
- *Slow cooling* – This method of crystallisation is usually based on the assumption that solubility decreases with temperature. The cooling rate can be adjusted by using different surface areas of crystallisation vessel, or by using a water bath/eurotherm controller device, which allows gradual heating and cooling of the crystallisation vessel. This method is good for solute-solvent systems that are less than moderately soluble, Figure 3.4(a).
- *Solvent diffusion* – This involves placing a solution containing a compound into a narrow tube, where an anti-solvent is injected carefully into the solution. The layers diffuse into each other slowly and crystals grow at the interface, Figure 3.4(b). One example of this procedure is the acid-base layering attempted in the experimental polymorph screen on guanine, Chapter 4.
- *Vapour diffusion* – This method involves vapour diffusion of a second solvent (the anti solvent) into a solution of the compound thereby reducing the solubility and hence

promoting crystallisation. This has the advantage of a relatively slow rate of diffusion, and is good for dealing with milligrams amounts of material, Figure 3.4(c).

- *Sublimation* – This technique, Figure 3.4(d), involves using a closed vessel in which the solid is heated and vapourized. The vapours diffuse towards the cold finger, where they are trapped. In favourable circumstances crystals then grow on the cold finger. To gain better quality crystals lower sublimation temperatures should be used. One limitation of this technique is the possibility of decomposition of the sample, an example of which being the sublimation of cytosine in chapter 4, in which ammonium hydrogen carbonate was obtained as a decomposition product.

Other techniques have been developed to induce crystallisation, including the use of solvent-mediated polymorphic transformation[196] (slurrying), solvent drop grinding[197], and crystallisation from the melt[192;193].

3.2.3 Characterisation of crystallisation samples

The samples obtained from the crystallisations were characterised by powder and single crystal X-ray diffraction. Scattering of X-rays by the electrons of the atoms in the lattice is governed by Bragg's law, which gives the conditions under which a diffracted beam can be observed. Bragg's law is expressed as $2d\sin\theta = n\lambda$, where θ is the Bragg angle, λ is the wavelength of the X-rays, and d is the spacing between adjacent planes in the parallel set. This is shown in Figure 3.5.

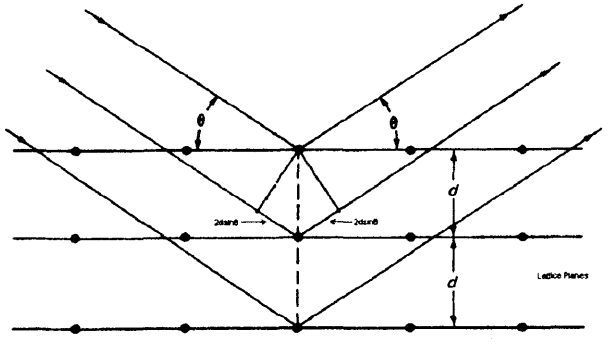


Figure 3.5 Diffraction of X-rays from crystal lattice planes illustrating Bragg's law

To get constructive interference the pathlength difference must be a whole number of wavelengths. The spacings between scattering waves give information regarding the unit cell dimensions and the intensities of the diffracted X-rays are related to identity and position of the

unit cell contents. Powder X-ray diffraction was used when the crystallisation does not give adequate quality crystals for single crystal X-ray diffraction. Single crystal X-ray diffraction was used for unit cell checking when there are adequate single crystals available, and for a full data collection when necessary (usually at 150 K) so that the crystal structure can be fully solved and refined.

3.2.4 X-ray powder diffraction

Powder diffraction is a quick way of comparing crystalline compounds, and hence identifying different polymorphs present[198], as each crystal structure has a unique X-ray powder pattern that may be used as a 'fingerprint' for its identification. If a new crystal structure is identified, then further crystallisations are attempted to gain suitable crystals for single crystal X-ray analysis. For the X-ray diffraction scan, a 2θ range of 10 to 60 degrees will be performed to give adequate comparisons with other diffractograms. This will involve using a Siemens D5000 diffractometer with monochromated $CuK_{\alpha 1}$ radiation (λ_1 = 1.5406 Å) utilising a Position Sensitive Detector (PSD). The quality of the powder X-ray diffractogram can be affected by a number of factors, including how microcrystalline the sample is and any preferred orientation effects[199;200]. In addition the simulated powder diffractograms of the hypothetical and solid state crystal structures can be calculated using Cerius2[164] or Mercury[201] for comparisons to be made.

3.2.5 Single crystal X-ray diffraction

For the single crystal analysis, unless stated otherwise, X-ray diffraction was performed on a Bruker SMART APEX diffractometer, equipped with graphite-monochromated Mo-Kα radiation (λ = 0.71073 Å) and a nominal crystal-to-area detector distance of 60 mm. If a data collection is performed the intensities were integrated using SAINT+[202] and the absorption correction was applied used SADABS[203]. The structure was solved with direct methods (SHELXS97) and refined against F^2 (SHELXL97)[204]. Details of the structure solutions and refinements for each crystal structure reported in this thesis can be found in the SI. In general all non-hydrogen atoms were refined anisotropically and the hydrogen atoms refined freely with an isotropic model.

3.3 Discussion

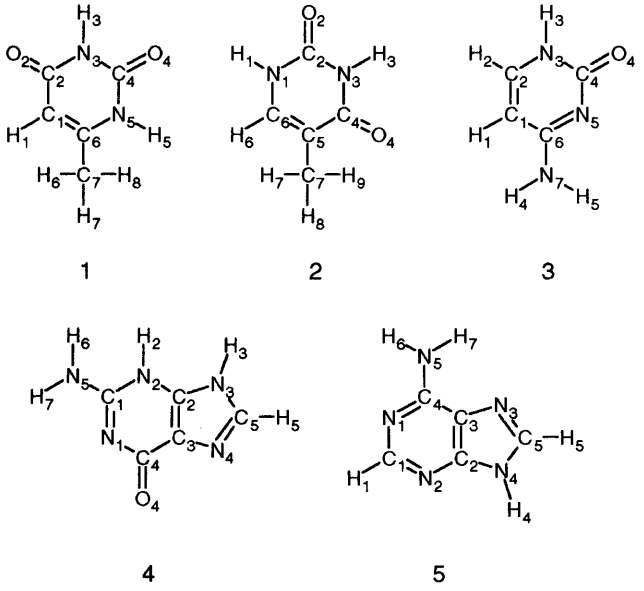
All the computational methods discussed in this chapter will be used throughout this thesis, with the exception being the morphology calculations (sections 3.1.12 to 3.1.14) which are only used in chapters 5 and 7 which were performed in collaboration with Dr. David Coombes. The experimental methods described here are used in chapters 4 and 5 on a large selection of small, organic hydrogen bonded molecules.

4. Computational and experimental polymorphism of DNA/RNA bases

4.1 Introduction

The key discovery in molecular biology in the last half century was the discovery of the two-stranded helical structure of DNA[205] (deoxyribonucleic acid), with the hydrogen bonding between the purines and pyrimidines present in RNA and DNA widely distributed in biological systems[206]. These hydrogen bonding interactions play a crucial part in the structure of DNA, with hydrogen bonding base pairings between adenine/thymine, and guanine/cytosine. The importance of this hydrogen bonding is exemplified by the fact that if the bases were to pair in other ways, the resulting structures would not be a double helix[206].

The molecules that will be studied in this chapter will be the DNA/RNA bases 6-methyluracil, thymine, cytosine, guanine and adenine, Scheme 4.1, in which there is limited molecular flexibility present, mainly in the methyl and amine groups. This chapter will investigate the differences in the solid state polymorphic behaviour and hydrogen bonding patterns of the hypothetical stable homo crystal packings, along with the effect of the limited molecular flexibility on the computational polymorph predictions.



Scheme 4.1 The molecular structures of (1) 6-methyluracil, (2) thymine, (3) cytosine, (4) guanine and (5) adenine

4.2 Outline

There have been no previously reported separate or combined experimental and computational polymorph studies on these five molecules. Initially *ab initio* studies and modelling of the known crystal structures will be performed to determine how sensitive the computational methods are to the limited molecular flexibility present, and to establish how well the method of modelling the intermolecular forces is at reproducing the known crystal structures after lattice energy minimisation. The experimental and computational polymorph searches will then be discussed for each molecule.

For thymine[207] and cytosine[208] there is one anhydrous crystal structure of each, with no anhydrous crystal structures known for guanine and adenine. There are two known polymorphs[209;210] of 6-methyluracil. In this thesis low temperature determinations have been performed on the crystal structures of anhydrous thymine, cytosine and 6-methyluracil Form *i*, as described in the relevant sections in this chapter. The computational polymorph searches used the extended version of MOLPAK[148], section 3.1.5, using the FIT potential[100-102] with the carbon repulsion parameters decreased by 25 % (the reasons given in chapter 5) and MP2 DMA. The other methodologies were the same as used throughout this thesis, unless stated otherwise. In the experimental polymorph screens all the solvents in the solubility studies were used in an attempt to induce crystallisation.

4.3 Initial computational modelling

4.3.1 *Ab initio* conformational analysis

A study was performed to see whether there are any low energy minima on the potential energy surface associated with conformational changes in the flexible functional groups present in these molecules. This will allow any other *ab initio* molecular arrangements to be identified for the computational polymorph studies. The CH_3 and NH_2 groups were rotated through 360 ° in 10 ° steps, being the C1C6C7H6, C4C5C7H7, C1C6N7H4, N2C1N5H6 and C3C4N5H7 torsion angles for 6-methyluracil, thymine, cytosine, adenine and guanine respectively, and relaxing the rest of the molecule. The starting molecular structures are either the solid state forms, or for adenine and guanine a molecular structure built using MOLDEN[117]. These conformational scans were performed using a SCF/6-31G** wavefunction using Gaussian98[116], as it is less

computationally expensive than using a MP2 wavefunction. However it should be noted that using the SCF wavefunction does not typically give pyramidisation character to the NH_2 group. Therefore the *ab initio* conformations used in the computational studies were obtained from the conformations associated with the minima on the potential energy surface which were then optimized using a MP2/6-31G** wave function, in which the NH_2 groups show some pyramidisation character. A comparison of the molecular parameters between the low temperature solid state and *ab initio* molecular structures of 6-methyluracil, thymine and cytosine are shown in Table 4.1 SI. The results for the conformational scans are shown in Figures 4.1 to 4.5, with the 0 ° conformation associated with one of the methyl or amine hydrogens being coplanar with the ring. The direction of the scan is from left to right.

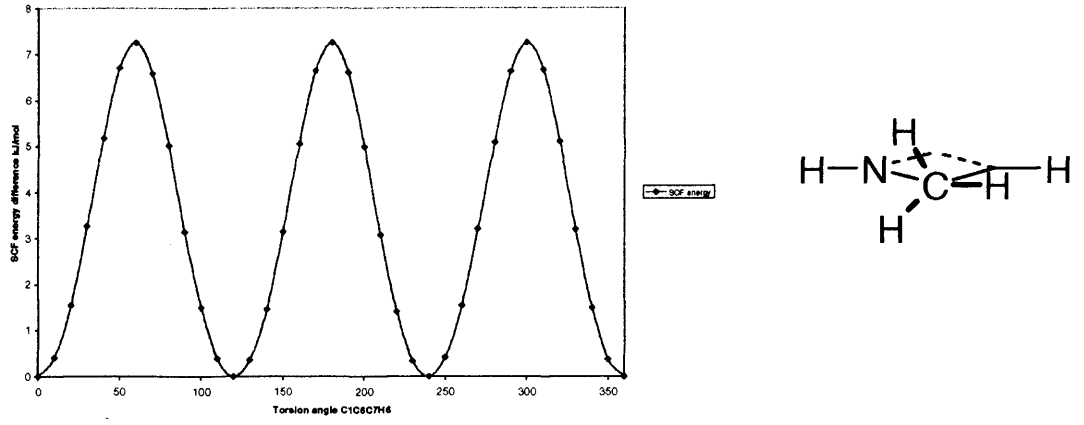


Figure 4.1 The SCF conformational energy scan for the torsion angle C1C6C7H6 for 6-methyluracil, with the molecular conformation corresponding to the low energy minima also shown

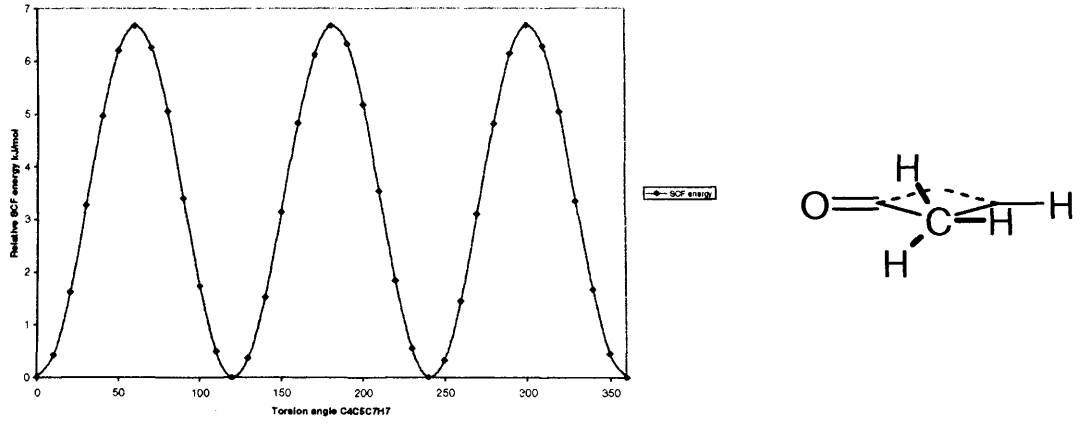


Figure 4.2 The SCF conformational energy scan for the torsion angle C4C5C7H7 for thymine, with the molecular conformation corresponding to the low energy minima also shown

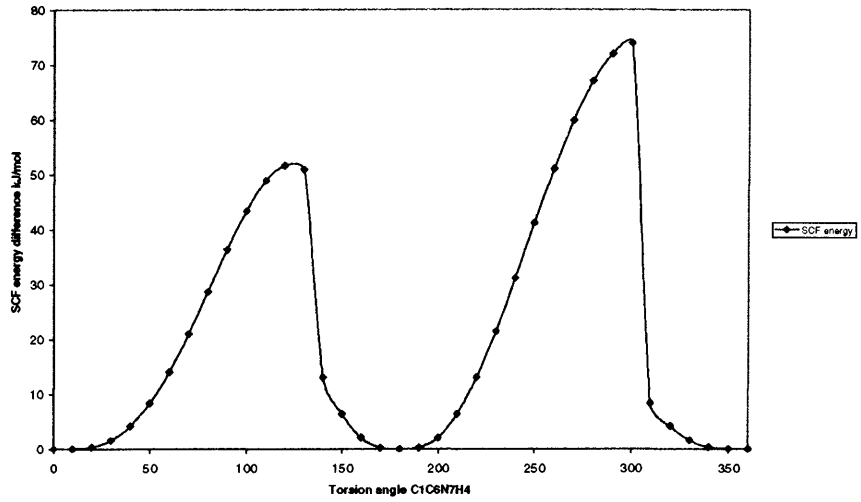


Figure 4.3 The SCF conformational energy scan for the torsion angle C1C6N7H4 for cytosine

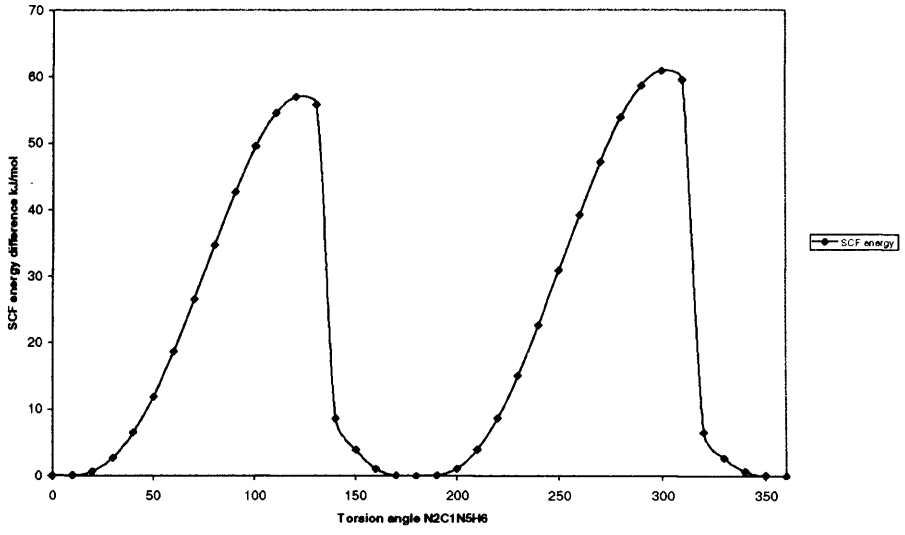


Figure 4.4 The SCF conformational energy scan for the torsion angle N2C1N5H6 for adenine

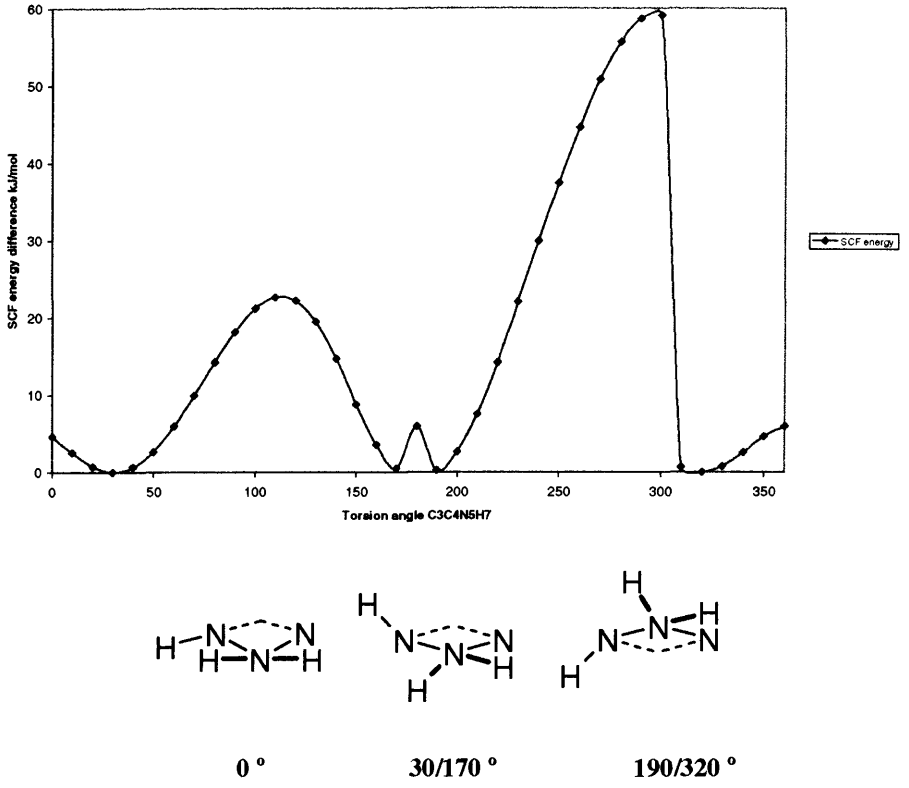


Figure 4.5 The SCF conformational energy scan for the torsion angle C3C4N5H7 for guanine. The schematic exaggerated conformations of the NH_2 group associated with the low energy minima on the potential energy surface are also shown, with the N-H ring hydrogen allowed to relax during the optimisation

It is clear from the *ab initio* energy scans that the most stable methyl conformation is when one of the hydrogens is co-planar with the ring. However slight rotation of this functional group is still energetically feasible. This is the same for the amine group in cytosine and adenine, however for guanine there are two conformations that are equally stable. The energy differences between the different conformations are relatively large which is an artefact of the quality of the basis set used, and how sensitive these intramolecular energies are to slight changes in bond lengths, for example for the carbonyl groups. Optimizing the DNA bases using HF and MP2 level of theory (using a 6-31G** basis set) found that variations in the bond lengths are significant, from as small as 0.001 Å to as large as 0.045 Å (with the largest deviation associated with the C5N4/C5N3 double bonds in guanine and adenine respectively)[211]. It has been shown however that the geometrical parameters, other than the amino group geometry for cytosine, are quite insensitive to the choice of basis set[212].

4.3.2 Modelling of the known crystal structures

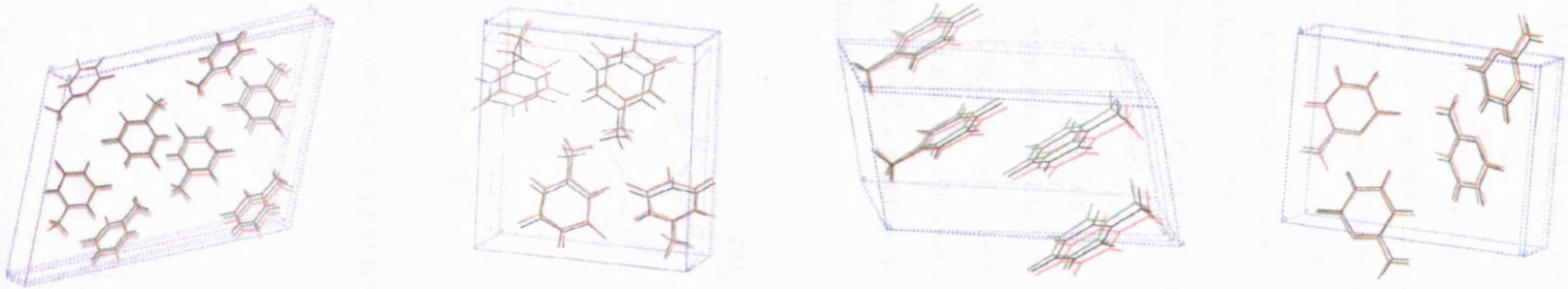
To ascertain whether the method of modelling the intermolecular forces reproduces the low temperature solid state crystal structures of 6-methyluracil Form *i*, thymine and cytosine satisfactory, lattice energy minimisations were performed using both the solid state and *ab initio* molecular conformations. The calculations were also carried out on 6-methyluracil Form *ii*, which was determined from X-ray powder diffraction[210;213]. The hydrogen positions in this powder X-ray diffraction study were not determined, therefore the hydrogens were added to the molecules at standard bond lengths and angles using SHELXP[204]. This will undoubtedly affect the results of the lattice energy minimisation, as shown by the high errors in the cell parameters and is only included as a comparison to the other structures. The results are shown in Table 4.1 and Figure 4.6.

The results show that the method of modelling the intermolecular forces reproduces the crystal structures of 6-methyluracil Form *i*, thymine and cytosine satisfactorily and therefore is adequate for use in the computational studies.

Table 4.1 Results of the lattice energy minimisations on the solid state crystal structures of 6-methyluracil Form *i*/Form *ii*, thymine and cytosine using the solid state and *ab initio* molecular structures. All the minimisations use the low temperature molecular structures, with the exception of 6-methyluracil Form *ii* which was determined at room temperature elsewhere[210;213]. The relative % errors compared to the solid state crystal structures are shown in brackets

	6-methyluracil		6-methyluracil		Thymine		Cytosine	
	Form *i*	Form *i*	Form *ii*	Form *ii*				
Molecular structure	Expt	*Ab initio*	*Expt	*Ab initio*	Expt	*Ab initio*	Expt	*Ab initio*
Space Group	C2/c	C2/c	$P2_1/c$	$P2_1/c$	$P2_1/c$	$P2_1/c$	$P2_12_12_1$	$P2_12_12_1$
Lattice energy kJ mol^{-1}	-117.029	-109.568	-107.715	-108.733	-116.783	-108.294	-143.463	-126.61
a/Å	20.769 *(1.32)*	20.912 *(2.02)*	4.047 *(-10.34)*	4.030 *(-10.71)*	12.373 *(-2.81)*	13.170 *(3.45)*	3.752 *(0.04)*	3.702 *(-1.30)*
b/Å	3.853 *(0.15)*	3.819 *(-0.73)*	11.938 *(8.64)*	10.989 *(8.14)*	7.003 *(2.38)*	7.006 *(2.42)*	9.541 *(0.67)*	9.557 *(0.83)*
c/Å	14.818 *(0.85)*	14.998 *(2.08)*	11.637 *(-0.73)*	11.722 *(-0.57)*	6.594 *(-0.40)*	6.401 *(-3.32)*	12.751 *(-1.94)*	13.009 *(0.05)*
β/°	110.462 *(-0.36)*	109.692 *(-1.05)*	89.023 *(-8.74)*	87.430 *(-10.37)*	103.071 *(-1.21)*	105.175 *(0.81)*		
Volume/$Å^3$	1111.011 *(2.60)*	1127.902 *(4.16)*	562.176 *(-2.47)*	557.705 *(-3.24)*	556.551 *(-0.36)*	570.026 *(2.05)*	456.517 *(-1.24)*	460.259 *(-0.44)*
F	**5**	**19**	**325**	**372**	**21**	**39**	**13**	**11**

*The hydrogens were added to the molecule at standard geometries and bond lengths using SHELXP[204]



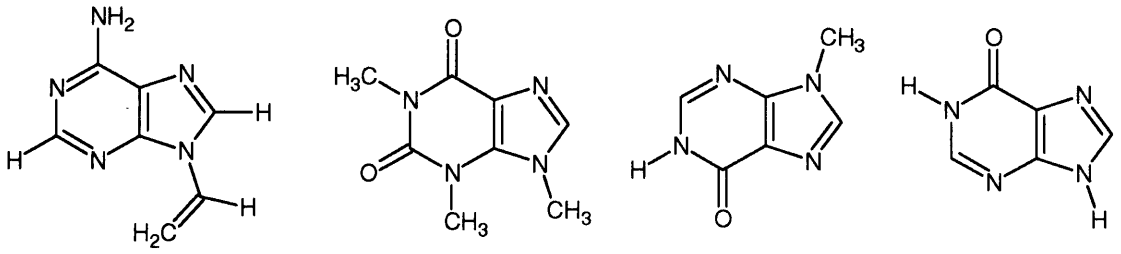
(a) (b) (c) (d)

Figure 4.6 Superimposed unit cells of (a) 6-methyluracil Form *i*, (b) 6-methyluracil Form *ii*, (c) thymine, and (d) cytosine showing the solid state crystal structure (black), and the lattice energy minimised structure using the solid state molecular structure (green), and the *ab initio* molecular structure (red).

4.3.3 Guanine and Adenine related crystal structures

To give an indication of how well the method of modelling the intermolecular forces might be expected to reproduce the crystal structures of guanine and adenine, a search in the Cambridge Structural Database[114] was performed to find crystal structures of structurally similar molecules, by allowing substitution of ring atoms and on the nitrogens. Out of the 30 or so crystal structures found most were either solvates or hydrates, or too flexible/large to realistically carry out lattice energy minimisation. Nevertheless, 9-vinyladenine[214] (VAVTOB), 1,3,9-trimethyl-2,6-dioxopurine[215] (ISCOFF), 9-methylhypoxanthine[216] (VEXMEQ) and hypoxanthine[217] (GEBTUC) were chosen, scheme 4.2, for further study.

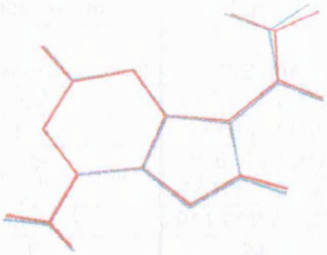
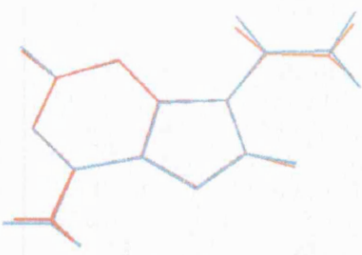
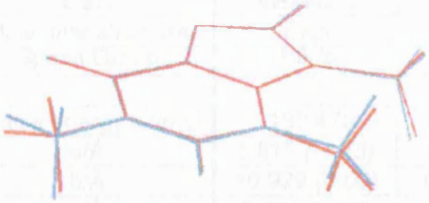
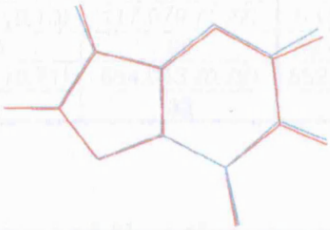
Both VAVTOB and GEBTUV have two molecules in the asymmetric unit, with VAVTOB having two different molecules that differ in the orientation of the vinyl group relative to the adenine frame. The solid state molecular structures were *ab initio* optimised using a MP2/6-31G** wave function using Gaussian98[116], shown in Table 4.2. The two different conformations in the asymmetric unit of VAVTOB were optimised separately to gain two *ab initio* conformations for use in the lattice energy minimisation.



6 7 8 9

Scheme 4.2 The molecular structures of 6, VAVTOB; 7, ISCOFF; 8, VEXMEQ; and 9, GEBTUC.

Table 4.2 A comparison between the solid state conformation (red) and *ab initio* conformation (blue) for VAVTOB (molecules 1 and 2), ISCOFF, VEXMEQ, and GEBTUC.

VAVTOB molecule 1	VAVTOB molecule 2
	
ISCOFF	**VEXMEQ**
	
GEBTUC	
	

The majority of the crystal structures are reproduced satisfactory, Table 4.3, despite the 20 ° rotation of one of the methyl groups from the *ab initio* conformation in ISCOFF. The VAVTOB crystal structure was not reproduced satisfactorily using the *ab initio* molecular structure, possibly due to the flexibility present in the vinyl groups. Also the crystal structure reproduction was not satisfactory for GEBTUC, using both molecular conformations, since the hydrogen bonded sheets present slip over each other giving unacceptable errors in the cell dimensions.

These results do show that the modelling of the intermolecular forces seems sufficient for use in the computational predictions for guanine and adenine in addition to the other DNA bases, despite some issues related to molecular flexibility.

Table 4.3 Results on the lattice energy minimisations on the adenine and guanine related compounds obtained from the Cambridge Structural Database[114], with the % errors compared to the solid state crystal structure shown in brackets

CSD	VAVTOB	VAVTOB	ISCOFF	ISCOFF
Molecular structure	Expt	*Ab initio*	Expt	*Ab initio*
Space Group	C2/c	C2/c	$P2_1/n$	$P2_1/n$
Lattice energy kJ/mol	-122.244	-116.02	-136.252	-125.652
a/Å	13.311 *(-2.72)*	13.515 *(-1.22)*	7.610 *(-1.39)*	7.767 *(0.65)*
b/Å	9.958 *(-2.85)*	10.250 *(-0.002)*	8.149 *(2.96)*	8.060 *(1.84)*
c/Å	21.676 *(0.01)*	21.429 *(-1.13)*	13.807 *(1.18)*	14.126 *(3.52)*
β/°	97.105 *(-0.63)*	94.205 *(-3.60)*	92.346 *(-0.55)*	92.109 *(-0.81)*
Volume/Å^3	2851.002 *(-5.35)*	2960.518 *(-1.71)*	855.510 *(2.77)*	883.753 *(6.62)*
F	**24**	**63**	**15**	**23**
CSD	**VEXMEQ**	**VEXMEQ**	**GEBTUC**	**GEBTUC**
Molecular structure	Expt	*Ab initio*	Expt	*Ab initio*
Space Group	$P2_1/c$	$P2_1/c$	P-1	P-1
Lattice energy kJ/mol	-133.478	-129.332	-145.741	-137.878
a/Å	6.815 *(-1.13)*	6.782 *(-1.62)*	6.764 *(-4.76)*	6.776 *(-4.59)*
b/Å	10.929 *(0.08)*	10.908 *(-0.12)*	9.808 *(0.50)*	9.814 *(0.56)*
c/Å	9.821 *(1.88)*	10.012 *(3.86)*	10.822 *(4.18)*	11.029 *(6.18)*
α/°	90	90	60.898 *(3.48)*	61.128 *(3.87)*
β/°	116.614 *(0.10)*	117.979 *(1.27)*	63.242 *(-6.51)*	63.052 *(-6.78)*
γ/°	90	90	69.469 *(-3.52)*	70.579 *(-1.97)*
Volume/Å^3	653.987 *(0.71)*	654.043 *(0.72)*	552.057 *(-2.12)*	566.458 *(0.44)*
F	**11**	**32**	**93**	**128**

4.3.4 Comparison of known sublimation energies

A second comparison was made which involved comparing the calculated lattice energy and the experimental sublimation energy of the crystal structures, Table 4.4, which is possible as there is substantial sublimation energy data available.

Table 4.4 Comparison of the known sublimation energies and the calculated lattice energies for the DNA/RNA bases. There are no ExptMinOpt and ExptMinExpt structures for guanine and adenine, so a comparison is made with the crystal structure at the global lattice energy minimum in the computational polymorph searches, sections 4.8.3 and 4.9.3

Compound	Sublimation energy kJ mol^{-1}	Lattice energy kJ mol^{-1}	% difference[h] between ΔH_{sub} and E_{latt}
6-methyluracil	131 (406 – 503 K)[218]	*ExptMinExpt1 (Form i) = -117.03*	*-12*
		[a]*ExptMinExpt2 (Form ii) = -107.73*	*-22*
		ExptMinOpt1 (Form *i*) = -109.57	-16
		ExptMinOpt2 (Form *ii*) = -108.73	-20
Thymine	131.3 ± 4[219]	*ExptMinExpt = -114.60*	*-13*
	[d]138 ± 10[220]	ExptMinOpt = -108.29	-20
	[e]134.1 ± 4.2[221]		
	[f]124.3[178]		
	[c]125.7 ± 3.6 (383 – 438 K)[219]		
	[g]124.4 (378 – 428 K)[222]		
Cytosine	[d]167 ± 10[220]	*ExptMinExpt = -143.46*	*-12*
	[e]176 ± 10[223]	ExptMinOpt = -126.61	- 27
	[c]150.6[178]		
	147.2 ± 2.6 (423 – 483 K)[219]		
Guanine	[g]186.2[178]	*Ab initio* molecular structure: -179.32	-3.8
		Planar molecular structure: -183.79	-1.3
Adenine	[f]126.3[178]	*Ab initio* molecular structure: -143.06	26
	[b]110[224]	Planar molecular structure: -142.87	26
	[c]109 ± 8 [224]		
	109.2 (448 – 473 K)[225]		

[a]From Powder X-ray structure, with no hydrogens determined[210]. Hydrogens added at standard bond lengths and angles. [b]by vapour pressure; [c]by mass effusion-knudsen effusion/torsion effusion; [d]by torsion effusion, [e]by calorimetry; [f]by langmuir evaporation; [g]by quartz resonator. [h]When there are multiple sublimation energies available, an average was taken ignoring the experimental errors

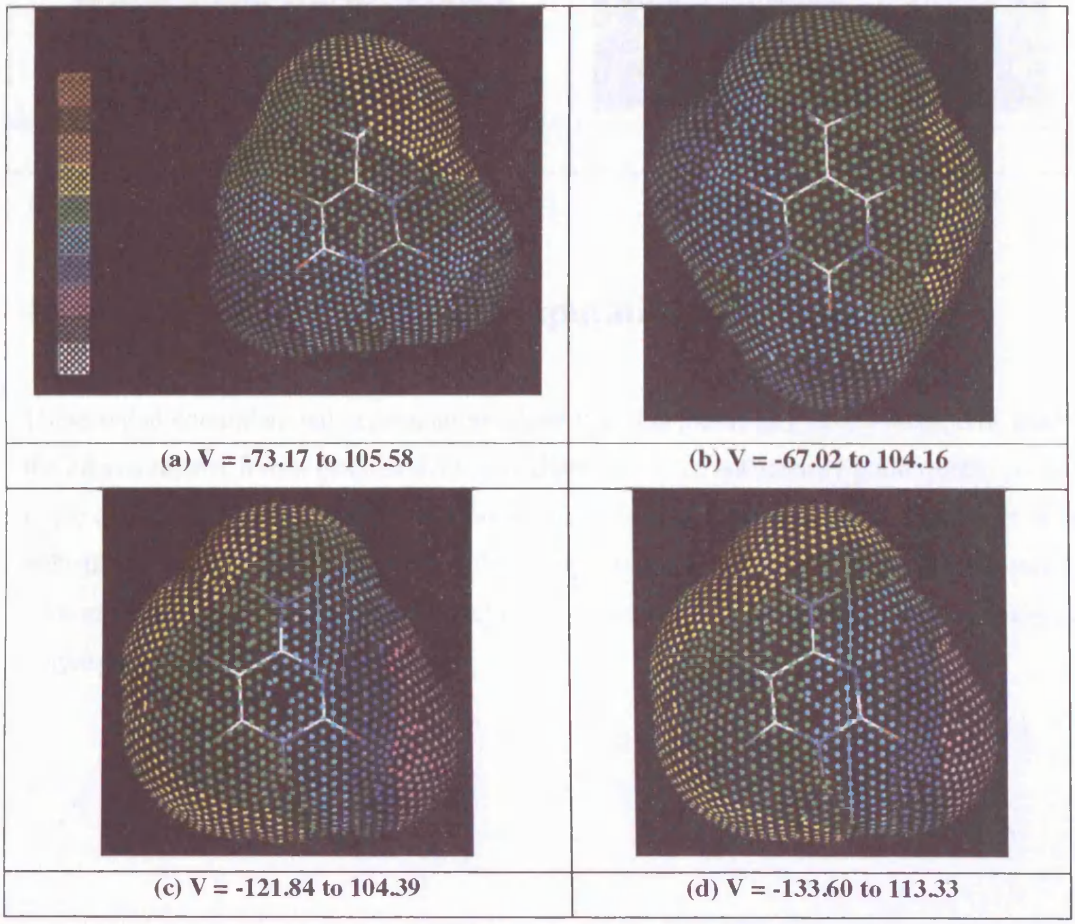
The results show that the lattice energies and sublimation energies compare reasonably well, within the errors present. However the results do show a systematic underestimation of the calculated lattice energies, with the exception of the results for adenine. This does highlight the fact that despite the majority of the crystal structures being satisfactorily reproduced after lattice energy minimisation, there are still inadequacies in the potential which in turn is reflected in the differences between the experimental sublimation and calculated lattice energies.

4.3.5 Electrostatic potentials

To determine how much intramolecular conformational changes affect the intrinsic electrostatic contribution to the hydrogen bond energy, the electrostatic potentials on the water accessible surfaces were calculated for each molecule. These electrostatic potentials, Figure 4.7, show that the molecular environment of the hydrogen bond donors and acceptors has a significant effect on the electrostatic potential around the molecule, as exemplified by 6-methyluracil and thymine.

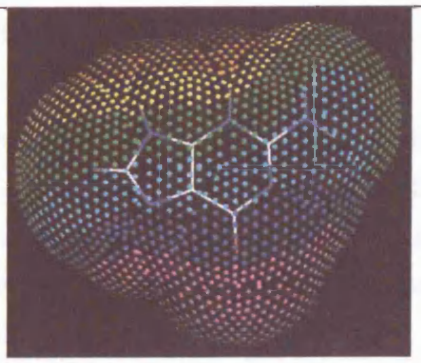
The conformational changes in the NH_2 group in cytosine and guanine have a significant effect on the V_{min} and V_{max}, however it should be noted that in the case of guanine the conformational deviations are not just restricted to the NH_2 group but also include the puckering of the ring to which it is attached, and also the position of the N-H bond adjacent to this amine group. Adenine has a much weaker electrostatic potential around it than cytosine and guanine indicating a weaker contribution to the hydrogen bonding, mainly due to the absence of the carbonyl group. The conformational change of the amine group in this case does not have a significant affect on V_{min} and V_{max}. These results suggest that slight intramolecular changes can have a significant affect on the electrostatic contribution to the lattice energy in the crystalline environment.

Figure 4.7 The electrostatic potential V (kJ mol^{-1}) on the water accessible surface of (a) 6-methyluracil, (b) thymine, (c) cytosine, (d) cytosine solid state, (e) guanine, (f) guanine planar, (g) adenine, and (h) adenine planar molecular structures as calculated from DMA derived from the MP2/6-31G wave function, colour coded: -200 < white < -160 < grey < -120 < magenta < -80 < blue < -40 < cyan < 0 < green < 40 < yellow < 80 < orange < 120 < brown < 160 < red < 200. The *ab initio* molecular structures were used, unless otherwise stated**

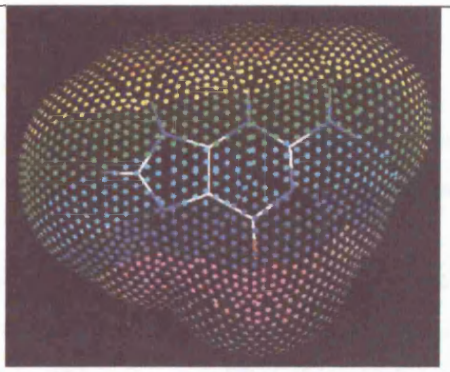


(a) V = -73.17 to 105.58 | **(b) V = -67.02 to 104.16**

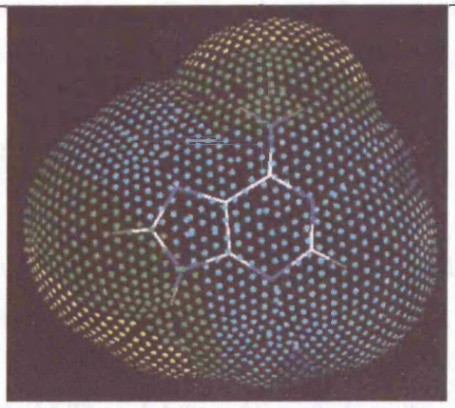
(c) V = -121.84 to 104.39 | **(d) V = -133.60 to 113.33**



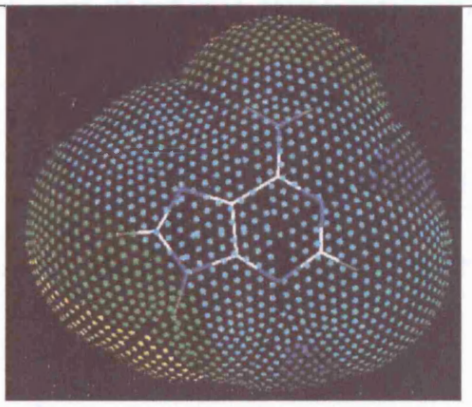
(e) V = -150.64 to 160.75



(f) V = -161.01 to 185.05



(g) V = -56.51 to 106.14



(h) -55.19 to 103.67

4.4 Discussion on the initial computational investigations

These initial computational investigations show that this previously tested method of modelling the intermolecular forces (section 2.14, and chapter 5) is of sufficiently good quality to be used in the computational polymorph predictions with these DNA bases. Nevertheless there are some indications that the limited molecular flexibility, which includes the slight differences in the conformations of the amine groups, could have an effect on the accuracy of the computational polymorph predictions.

4.5 6-methyluracil

6-methyluracil is a simple analogue of the natural nucleic acid pyrimidine bases, and is in the World Drug Index[226] as it possesses anabolic and anticatabolic activity, including inhibition of human spleen dihydroorotate dehydrogenase[227].

The crystal structure of Form *i* (also named betamecil) was originally determined in 1984[228], and was redetermined in 1993 to give a C2/c, Z = 8 structure[209]. A second polymorph, deemed Form *ii*, was originally determined from X-ray powder diffraction data[210], being a $P2_1/c$ Z = 4 structure. A new third polymorph, deemed Form *iii*, was found in an experimental screen run concurrently with this computational study.

4.5.1 Solubility

The results of the solubility study, carried out using the method outlined in section 3.2.1, are shown in Table 4.5.

Table 4.5 The solubility of 6-methyluracil in various solvents.

Observation	Solvent
Soluble	H_2O, methanol, DMF, formaldehyde, DMSO, 2,2,2-trifluoroethanol, 1,methyl-2,pyrolidinone
Partially Soluble	Ethanol, dichloromethane, butan-1-ol, propan-2-ol, acetone, diethyl ether, propan-1-ol, acetonitrile, THF, butan-2-ol, tetrachloroethylene, ethylene glycole, hydrochloric acid, hydrobromic acid, dimethylamine in H_2O
Insoluble	Chloroform, 1,4-dioxane, nitromethane, 1,2-dichloroethane, ethyl acetate, toluene, cyclohexane, aniline, hexane, methyl benzoate, xylene, isopropyl ether, n-octane, diethyl oxalate, ethyl methyl ketone, tert-butylmethyl ether, di-n-butyl ether

4.5.2 Experimental results

The results of the experimental polymorph screen are shown in Table 4.2 SI. The majority of crystallisations gave Form *i*, section 4.5.3, as a crystalline or a microcrystalline solid. A new polymorph was found, designated Form *iii*, described in section 4.5.5. Oxalic acid dihydrate, described in section 4.15 SI, was obtained from crystallisation in diethyl oxalate solvent.

4.5.3 Crystallisation of Form *i*

Form *i* was crystallised from the majority of solvents in the experimental polymorph screen, (Table 4.2 SI) either as colourless, block/needle-like crystals or as a microcrystalline solid. The low temperature refinement of this crystal structure has a greater precision in the metric parameters compared to the previous refinement[209], as outlined in section 4.3 SI. The packing in Form *i* consists of centrosymmetric N-H···O dimer units, with these dimer units hydrogen bonded to other units through N-H···O bonds forming a chain motif.

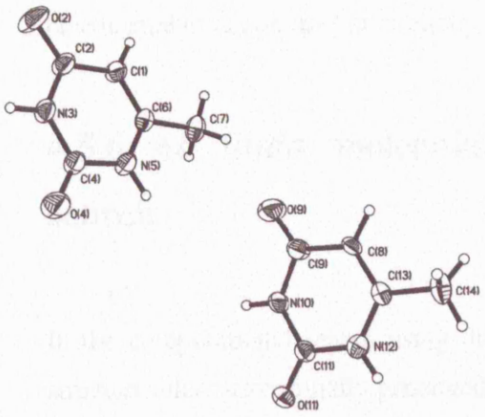
4.5.4 Form *ii*

Form *i* was obtained as a stock sample from Aldrich, whilst Form *ii* was obtained as a stock sample from ACROS, as confirmed by powder diffraction. This suggests that it might be the different syntheses of this compound that could give rise to different polymorphs. Using Form *ii* as the starting material in a range of solvent crystallisations only gave Form *i*. The crystal packing present in the X-ray powder crystal structure solution[210] is described in section 4.4 SI.

4.5.5 Crystallisation of Form *iii*

This new polymorph was crystallised by sublimation of a sample at 220 ° C and at low pressure. Small plate-like crystals was formed after about a day. Data collection was performed at low temperature, however it was found that the crystal was twinned, with two components in an approximate ratio of 75:25. The structure was solved satisfactorily on one component. The thermal ellipsoid plot is shown in Figure 4.8, along with the crystallographic data shown in Table 4.6 SI.

Form *iii* crystallises in the space group $P2_1/c$, with two molecules in the asymmetric unit. The metric parameters and the hydrogen bonding are shown in Table 4.5 SI. The molecular conformations in the asymmetric unit differ in the orientations of the methyl group. In molecule A one of the methyl hydrogens is close to planar with the ring, similar to the *ab initio* molecular structure, section 4.3.1. In molecule B the methyl group is rotated through 40 °. This agrees with the *ab initio* conformational analysis, section 4.3.1, which demonstrates that slight rotations of the methyl group are energetically feasible.



A B

Figure 4.8 The thermal ellipsoid plot (50 % probability level) of Form *iii*, showing the two molecules in the asymmetric unit, A and B

The packing in Form *iii*, Figure 4.9, is very similar to that of Form *ii*, section 4.4 SI. Each molecule in the asymmetric unit only uses one hydrogen bond acceptor, with these two molecules forming a hydrogen bonded dimer unit. This dimer unit is hydrogen bonded to adjacent dimer units through N-H···O bonds, at an angle approximately 74 ° from the plane of the dimer. This is different to Form *i* that uses both hydrogen bond acceptors in the crystalline environment.

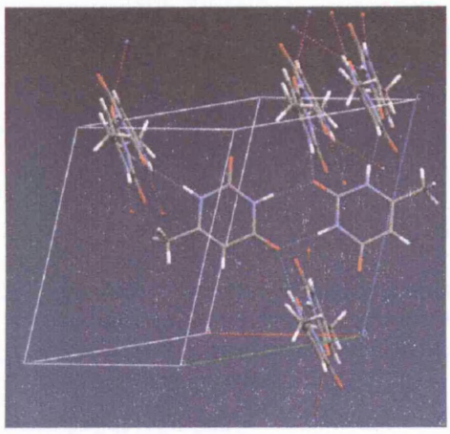


Figure 4.9 The hydrogen bonded dimer present in Form *iii* of 6-methyluracil, forming a three-dimensional hydrogen bond motif

Lattice energy minimisation using the solid state molecular structures of the two molecules in the asymmetric unit gave a satisfactory reproduction of this new solid state crystal structure. However when the *ab initio* molecular structure is used the reproduction becomes unacceptable, with the highest error in the cell parameters being 8 %. These results suggest that the flexibility

present in the methyl groups is important, with slight conformational changes having a detrimental effect on the lattice energy minimisation.

4.5.6 *Ab initio* molecular structure computational polymorph search

In the computational search using the *ab initio* molecular structure, of the 3300 hypothetical structures that were initially generated, there were 33 unique crystal structures within 7 kJ mol^{-1} of the global lattice energy minimum. ExptMinOpt1, ExptMinOpt2 and ExptMinOpt3 are the associated lattice energy minima corresponding to Form *i*, Form *ii*, and Form *iii* respectively. The results are shown in Figure 4.10 and Table 4.6.

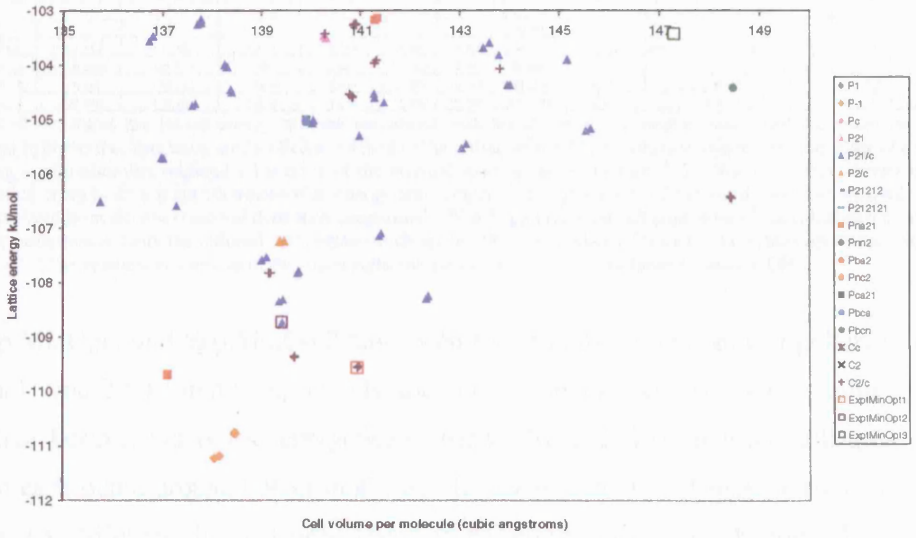


Figure 4.10 The results of the computational polymorph search on 6-methyluracil, using the *ab initio* molecular structure. Only the hypothetical crystal structures within 9 kJ mol^{-1} of the global lattice energy minimum are shown, with ExptMinOpt1, ExptMinOpt2 and ExptMinOpt3 also shown for comparison.

Table 4.6 The low energy crystal structures[a] found within 5 kJ mol^{-1} of the global lattice energy minimum for 6-methyluracil, using the *ab initio* molecular structure. The full list of structures within 7 kJ mol^{-1} is shown in Table 4.7 SI. ExptMinOpt1 and ExptMinOpt2 are also shown for comparison, with ExptMinOpt3 found approximately 8 kJ mol^{-1} above the global lattice energy minimum.

Structure	Space	Lattice Energy	[b]Free energy		[c]Reduced Cell				Hydrogen bonding acceptors		[d]Graph set			[e]Elastic
	group	kJ mol^{-1}	at 298 K kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles/°	used and motif		Level 1	Level 2	Level 3	constant
AB1	P-1	-111.231	-124.316	1.517	3.864	7.312	10.433	α 108.830	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	1.30
								β 96.167						
								γ 92.994						
CA39	P-1	-110.783	-122.454	1.512	4.630	7.756	8.686	α 65.865	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	1.74
								β 83.048						
								γ 76.807						
AV62	Pna2$_1$	-109.702	-122.492	1.527	4.538	8.656	13.963		O2,O4	3D	C1,1(6)	C1,1(4)	C2,2(8)	2.33
ExptMinOpt1	**C2/c**	**-109.589**	**-123.573**	**1.486**	**3.819**	**10.629**	**14.996**	**α 70.64**	**O2,O4**	**Chains**	**R2,2(8)**	**R2,2(8)**	**C2,2(10)**	**0.79**
								β 90.0						
								γ 79.65						
DE56	C2/c	-109.57	-123.343	1.486	3.819	10.629	14.997	α 70.640	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	0.79
								β 90.00						
								γ 79.650						
DD109	C2/c	-109.383	-121.531	1.500	6.196	10.764	17.102	α 101.551	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	0.50
CA82	P-1	-109.366	-121.426	1.499	6.196	6.209	8.551	α 79.977	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	0.50
								β 89.945						
								γ 60.084						
ExptMinOpt2	**P2$_1$/c**	**-108.733**	**-122.31**	**1.502**	**4.030**	**11.656**	**11.884**	**γ 92.57**	**O2**	**Jagged Sheet**	**C1,1(6)**	**R2,2(8)**	**C2,2(10)**	**2.48**
AM107	P2$_1$/c	-108.744	-122.32	1.502	4.030	11.655	11.884	γ 92.568	O2	Jagged Sheet	C1,1(6)	R2,2(8)	C2,2(10)	2.48
FC57	P2$_1$/c	-108.353	-123.954	1.503	3.805	7.375	19.868	α 90.068	O2,O4	Jagged Sheet	R2,2(8)	C1,1(4)	C2,2(10)	1.59
AI33	P2$_1$/c	-108.297	-123.165	1.471	3.877	9.991	15.256	α 105.504	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	2.03
DE75	C2/c	-107.829	-123.085	1.504	3.784	14.465	21.443	α 108.407	O2,O4	Jagged Sheet	R2,2(8)	C1,1(4)	C2,2(10)	2.17
AI22	P2$_1$/c	-107.823	-121.955	1.499	4.464	8.828	14.217	α 93.752	O2,O4	Jagged Sheet	R2,2(8)	C1,1(4)	C2,2(10)	4.94
AJ70[sg]	P2$_1$/c	-107.758	-123.746	1.466	3.803	10.510	14.830	α 105.400	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	1.55
AL97[sg]	P2$_1$/c	-107.595	-122.441	1.507	4.053	9.647	14.749	α 105.380	O2,O4	Jagged Sheet	R2,2(8)	C1,1(4)	C2,2(10)	3.05
AJ25[sg]	P-1	-107.171	-120.039	1.501	6.974	8.354	9.895	α 85.630	O2,O4	Chains	R2,2(8)	D1,1(2)	R2,2(8)	4.28
								β 76.120						
								γ 88.070						
AM89	P2$_1$/c	-107.152	-121.724	1.513	4.114	9.991	13.799	γ 94.346	O2,O4	ninite ribbon of dimer	C1,1(6)	R2,2(8)	R4,4(16)	1.97
AM128	P2$_1$/c	-106.589	-121.871	1.525	3.849	10.368	13.760	γ 90.805	O2,O4	Sheets	C1,1(6)	R2,2(8)	C2,2(8)	1.52
FA1	P2$_1$/c	-106.526	-120.712	1.543	3.924	7.238	19.148	γ 93.297	O2,O4	Jagged Sheet	C1,1(4)	R2,2(8)	C2,2(10)	3.66
DE119	C2/c	-106.448	-119.387	1.410	4.571	15.624	18.182	α 113.861	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	2.34

[a]All calculated structures are lattice energy minima calculated with the *ab initio* molecular model and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK coordination geometry and order of density, with 'sg' denoting a minimum that required a lowering of the original space group symmetry. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated by PLATON[158] are given for comparison. Only the reduced cell angles which are not 90° are tabulated. [d]Only the first three levels shown, calculated using RPluto[168]. [e]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.

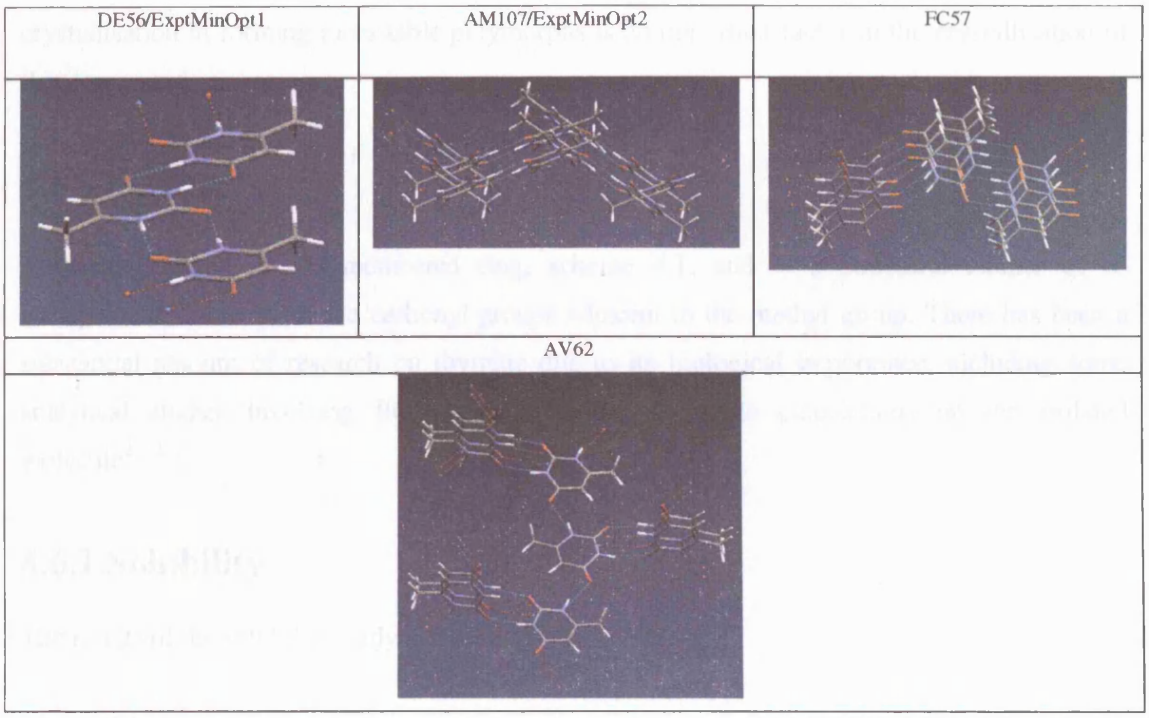
Both ExptMinOpt1 and ExptMinOpt2 have been found in the computational polymorph search, 1.6 kJ mol^{-1} and 2.5 kJ mol^{-1} respectively above the global lattice energy minimum. Form *ii* is denser than Form *i*, but is less energetically stable. Nevertheless they are still quite close in energy to each other, around 0.9 kJ mol^{-1}, which increases to 1.3 kJ mol^{-1} at room temperature estimates. ExptMinOpt3 has not been found in the computational predictions, due to the two different molecular conformations in the asymmetric unit, however it is predicted to be around 8 kJ mol^{-1} above the global lattice energy minimum and is significantly less dense than both ExptMinOpt1 and ExptMinOpt2.

There are three hypothetical crystal structures energetically more stable and denser than Form *i*, with the majority of the low energy structures consisting of molecules that use both hydrogen bond acceptors, forming mainly sheets and chains. The two structures at the global lattice energy minimum consist of the same chain structure as Form *i*. AV62, ranked 3rd in energy, consists of a

three-dimensional hydrogen bond network, suggesting that other hydrogen bond motifs can be competitive in energy. These structures are shown in Figure 4.11.

Within the low energy crystal structures only one other structure, AM141 (Table 4.7 SI, -104.606 kJ mol^{-1}) consists of molecules that use only the O2 hydrogen bond acceptor, in a similar way to Form *ii* and Form *iii* and two others are found that consist of molecules that use only the O4 hydrogen bond acceptor. This suggests that using only one acceptor is not that energetically favourable. Indeed, it is unusual that any crystal structures with unused hydrogen bond acceptors are competitive, despite the fact that only Form *i* uses all of the acceptors in the crystal structure.

Figure 4.11 The hydrogen bonded chains, three-dimensional and jagged sheet structures present in the low energy crystal structures of 6-methyluracil.



4.5.7 Conclusions

The experimental screen yielded a new polymorph, deemed Form *iii*. The importance of this new crystal structure is highlighted by the fact that there is a significant difference in the conformation of the methyl groups in the two molecules in the asymmetric unit, showing that 6-methyluracil does possess some limited flexibility within its molecular structure. There is the

possibility that these conformational changes allow this molecule to pack more favourably in the crystal. Nevertheless these differences have a significant effect on the lattice energy minimisations.

The computational study suggests that Form *i* is the thermodynamically stable crystal structure, with Form *ii* predicted to be quite close in energy. Form *iii* is significantly less stable than both Form *i* and Form *ii*. However the relative stabilities of these polymorphs are only within the limitations of the accuracy of the calculations.

The computational study has predicted a variety of energetically favourable polymorphs, which exhibit a variety of hydrogen bond motifs. Nevertheless despite three known polymorphs, only Form *i* has been crystallised from an experimental polymorph solvent screen, suggesting that crystallising other polymorphs via solvents is kinetically disfavoured. Although it has been shown that other polymorphs can be obtained by different methods, it is clear the kinetics of crystallisation in forming metastable polymorphs is an important factor in the crystallisation of this compound.

4.6 Thymine

Thymine consists of a 6-membered ring, scheme 4.1, and is a structural isomer of 6-methyluracil, with one of the carbonyl groups adjacent to the methyl group. There has been a substantial amount of research on thymine due to its biological importance, including some analytical studies involving IR spectra[229;230], and *ab initio* calculations on the isolated molecule[211;212].

4.6.1 Solubility

The results of the solubility study are shown in Table 4.7.

Table 4.7 The solubility of thymine in various solvents.

Observation	Solvent
Soluble	DMF, formaldehyde, DMSO, 1,methyl-2,pyrolidinone,
Partially Soluble	H_2O, acetone, methanol, ethanol, diethyl ether, aniline, 2,2,2-trifluoroethanol, tetrachloroethylene, diethyl oxalate, methyl benzoate, isopropyl ether, tert-butyl methyl ether, ethylene glycol, hydrochloric acid, hydrobromic acid, dimethylamine in H_2O
Insoluble	Chloroform, propan-2-ol, butan-1-ol, dichloromethane, 1,4-dioxane, nitromethane, propan-1-ol, 1,2-dichloroethane, toluene, acetonitrile, THF, cyclohexane, hexane, butan-2-ol, xylene, di-n-butyl ether, n-octane, ethyl methyl ketone, disopropylether

4.6.2 Experimental results

The results of the experimental polymorph screen are shown in Table 4.8 SI. The majority of crystallisations gave the anhydrous crystal form, with the only other crystal structure found being thymine monohydrate, described in section 4.10 SI.

4.6.3 Crystallisation of anhydrous thymine

Anhydrous thymine crystallises in a variety of morphologies including block, plate and needle-like crystals from the majority of solvents, as outlined in Table 4.8 SI. A suitable single crystal of anhydrous thymine was found from evaporation of an acetone solution. The refinement of the structure from the low temperature data was not significantly better than previously[207], section 4.9 SI.

4.6.4 *Ab initio* molecular structure computational polymorph search

In the computational search using the *ab initio* molecular structure, of the 3600 structures generated, there are 59 unique crystal structures within 7 kJ mol^{-1} of the global lattice energy minimum. The results of the search are shown in Figure 4.12 and Table 4.8.

ExptMinOpt was found around 0.2 kJ mol^{-1} above the global lattice energy minimum, with three structures denser and lower in energy. Within 2 kJ mol^{-1} of the global lattice energy minimum there are 17 other energetically competitive crystal structures, indicating a shallow global minimum on the energy surface. Taking into account the room temperature energy estimates, there is a significant reordering of the stabilities of the hypothetical structures, in which there are now eight structures more thermodynamically stable than ExptMinOpt.

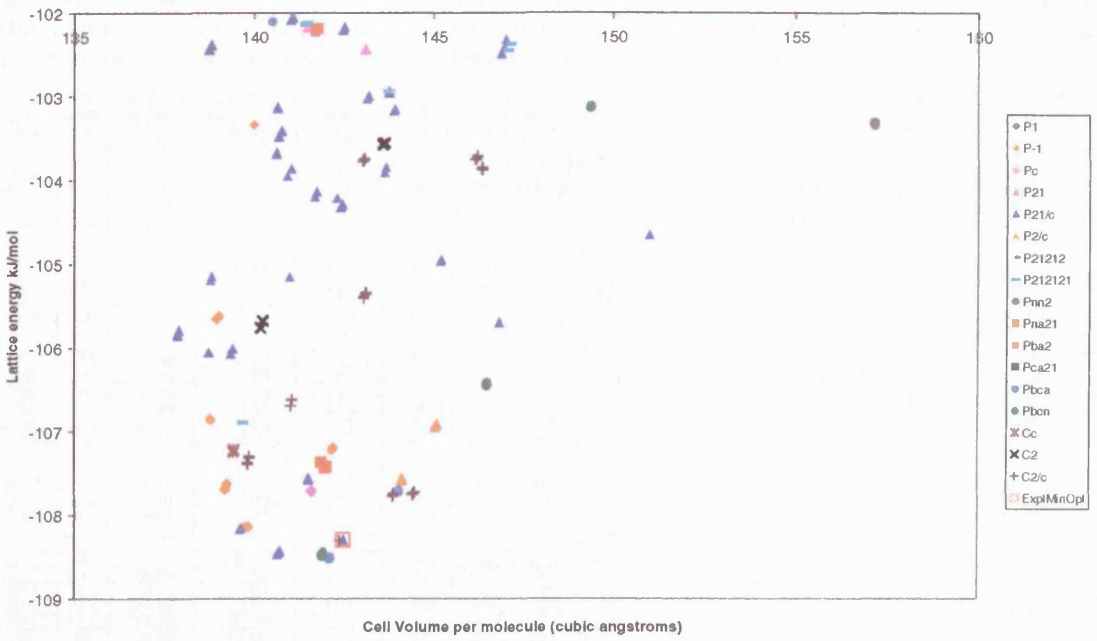


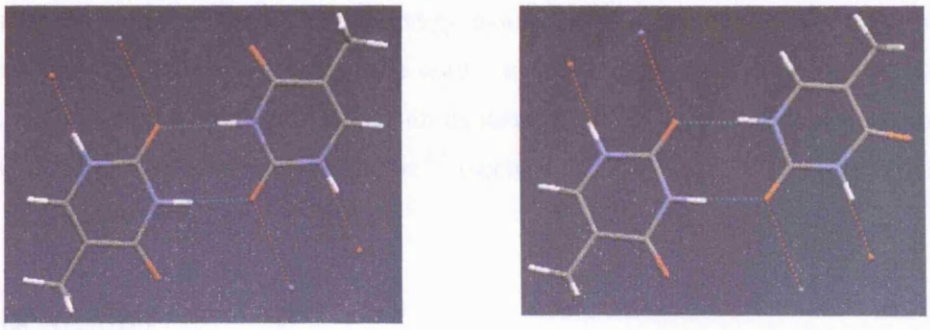
Figure 4.12 The results of the computational polymorph search on thymine, using the *ab initio* molecular structure. Only the hypothetical crystal structures within 7 kJ mol^{-1} of the global lattice energy minimum are shown, with the ExptMinOpt structure shown for comparison

All the crystal structures within 2 kJ mol^{-1} of the global lattice energy minimum only use one hydrogen bond acceptor per molecule in the crystal, however this gives rise to two different types of hydrogen bonded chain structures, one is centrosymmetric and the other non-centrosymmetric, Figure 4.13. The solid state crystal structure consists of chains of the non-centrosymmetric dimers, which is the same motif as the structure at the global lattice energy minimum, CC14. However, the 2nd and 3rd ranked structures contain chains of the centrosymmetric dimers, suggesting energetically competitive packings for both types of dimers. These dimers form hydrogen bonded chains, with the structures close to the global lattice energy minimum showing slight variations in the relative orientations of these chains, shown in Figure 4.14. Some of these crystals are denser and energetically more favourable than the solid state structure. However this could be due to inadequacies with the modelling of the molecular stacking due to the errors present in the potential. Around 2.5 kJ mol^{-1} above the global lattice energy minimum there are hypothetical structures that consist of molecules that use both hydrogen bond acceptors, leading to three-dimensional/sheet hydrogen bond motifs.

Table 4.8 The low energy crystal structures[a] found within 3 kJ mol^{-1} of the global lattice energy minimum for thymine, using the *ab initio* molecular structure. The full list of structures within 7 kJ mol^{-1} is shown in Table 4.11 SI. ExptMinOpt is also shown for comparison.

Structure	Space	Lattice Energy	[b]Free energy		[c]Reduced Cell				Hydrogen bonding acceptors		[d]Graph set			[e]Elastic
	group	/kJ mol^{-1}	at 298 K kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles/°	used and motif		Level 1	Level 2	Level 3	constant
CC14	Pbca	-108.509	-122.969	1.474	6.155	7.007	25.1238		O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	1.52
FA84	$P2_1/c$	-108.468	-124.682	1.489	3.7964	5.8766	25.2231	γ 90.859	O2	Chains	R2,2(8)	R2,2(8)	C2,2(8)	2.17
DE102	C2/c	-108.304	-120.913	1.471	6.3805	7.0174	25.457	β 91.802	O2	Chains	R2,2(8)	R2,2(8)	C2,2(8)	1.07
AI61	$P2_1/c$	-108.302	-122.278	1.470	6.3984	7.0061	13.0522	β 103.053	O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	1.41
ExptMinOpt	**$P2_1/c$**	**-108.294**	**-122.27**	**1.470**	**6.401**	**7.006**	**13.049**	**β 103.060**	**O2**	**Chains**	**C1,1(4)**	**C1,1(4)**	**R2,2(8)**	**1.40**
AK30	$P2_1/c$	-108.167	-121.292	1.500	7.1345	8.2673	9.4738	α 92.030	O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	2.25
CA121	P-1	-108.147	-121.865	1.498	3.7918	5.9595	12.4988	α 87.325	O2	Chains	R2,2(8)	R2,2(8)	C2,2(8)	4.51
								β 83.719						
								γ 90.237						
DE1	C2/c	-107.768	-120.734	1.455	6.5898	6.8899	24.993	β 90.823	O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	0.91
DE12	C2/c	-107.758	-119.817	1.450	7.022	8.1184	20.67	α 101.308	O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	1.80
AD34	Pc	-107.722	-122.673	1.479	3.782	5.7768	12.9385	γ 92.060	O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	0.77
CA129	P-1	-107.696	-120.19	1.504	4.6308	5.2182	11.7476	α 99.699	O2	Chains	R2,2(8)	R2,2(8)	C2,2(8)	5.63
								β 95.555						
								γ 90.397						
FA28	$P2_1/c$	-107.581	-124.837	1.480	3.8004	5.763	25.8652	γ 92.065	O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	0.67
AU89	$Pna2_1$	-107.446	-124.448	1.475	3.7849	5.8763	25.5331		O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	1.45
DC19	C2/c	-107.404	-121.406	1.498	7.035	7.247	22.181	α 98.470	O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	0.76
DD6	C2/c	-107.389	-121.104	1.498	7.0052	7.2478	12.1805	α 98.470	O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	1.12
DA61	Cc	-107.241	-117.615	1.502	6.9417	7.0752	12.5109	γ 114.817	O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	0.75
CA96	P-1	-107.22	-118.761	1.473	6.3079	6.7849	7.0021	α 77.570	O2	Chains	R2,2(8)	R2,2(8)	C2,2(8)	3.02
								β 81.770						
								γ 77.616						
BK25[sg]	Pc	-107.169	-123.827	1.471	3.849	5.834	25.482	α 95.85	O2	Chains	D1,1(2)	D1,1(2)	D1,1(2)	1.25
AQ7	$P2_12_12_1$	-106.895	-120.585	1.499	6.3685	7.0569	12.4349		O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	0.49
CA10	P-1	-106.86	-119.295	1.508	3.9039	8.1345	9.2924	α 72.306	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	1.37
								β 82.808						
								γ 82.505						
DD41	C2/c	-106.681	-120.207	1.484	6.9925	7.0367	22.9282	α 90.117	O2	Chains	R2,2(8)	R2,2(8)	C2,2(8)	0.74
CC68[sg]	$Pca2_1$	-106.388	-120.543	1.518	5.193	11.853	17.831		O2	Chains	D1,1(2)	D1,1(2)	D1,1(2)	5.83
AI115	$P2_1/c$	-106.068	-120.216	1.503	5.0891	10.2947	11.8029	α 115.651	O2	3D	C1,1(4)	R2,2(8)	C2,2(8)	4.90
AM13	$P2_1/c$	-106.058	-119.961	1.508	6.9426	7.1128	12.4501	γ 115.488	O2	Chains	R2,2(8)	R2,2(8)	C2,2(8)	0.80
AI120	$P2_1/c$	-105.848	-120.334	1.519	3.8234	11.6324	12.6604	α 101.635	O2	3D	C1,1(4)	R2,2(8)	C2,2(8)	3.89
DB8	C2	-105.767	-117.043	1.494	7.0765	7.2854	10.9645	α 97.230	O2	Chains	C1,1(4)	C1,1(4)	R2,2(8)	0.60
FA1	$P2_1/c$	-105.707	-121.688	1.426	5.1225	5.871	19.9142	γ 101.265	O2	Chains	R2,2(8)	R2,2(8)	C2,2(8)	0.34
CA16	P-1	-105.652	-117.873	1.507	6.2227	6.8246	7.7109	α 111.925	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	3.59
								β 95.325						
								γ 109.249						
DD62	C2/c	-105.406	-121.005	1.464	3.8526	13.3341	22.5776	α 99.313	O2	Chains	R2,2(8)	R2,2(8)	C2,2(8)	1.52

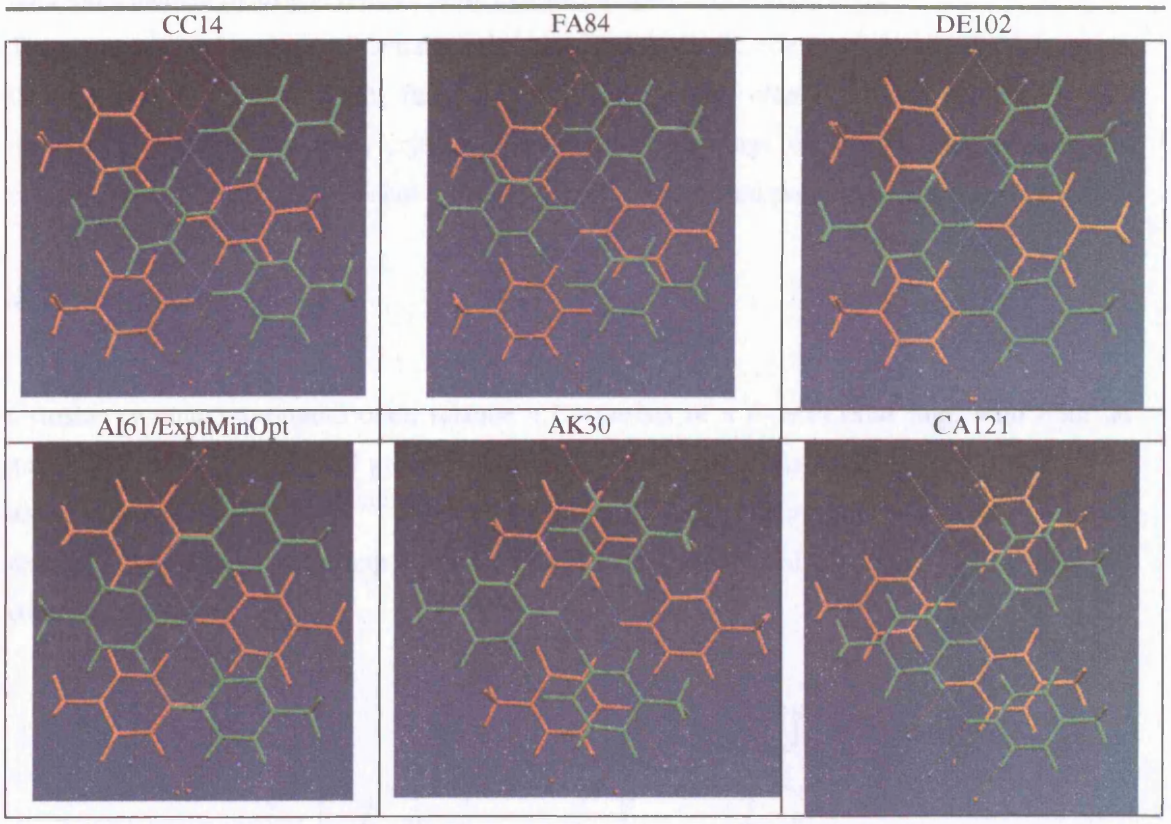
[a]All calculated structures are lattice energy minima calculated with the *ab initio* molecular model and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK coordination geometry and order of density, with 'sg' denoting a minimum that required a lowering of the original space group symmetry. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated by PLATON[158] are given for comparison. Only the reduced cell angles which are not 90° are tabulated. [d]Only the first three levels shown, calculated using RPluto[168]. [e]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.



(a) (b)

Figure 4.13 The centrosymmetric dimer present in (a) AI61 and the non-centrosymmetric dimer present in the (b) solid state and CC14 crystal structures

Figure 4.14 The variations in the hydrogen bond chain motifs in the low energy hypothetical structures of thymine (red and green representing the different stackings of these chains)



With the search producing a shallow global energy minimum with very similar packings of the molecules, the question does arise whether the search is predicting possible disorder within the crystal structure of thymine. However the anhydrous form shows no indication of disorder in the diffraction studies, and in thymine monohydrate[231] (section 4.10 SI) the only disorder present is in the water of crystallisation.

4.6.5 Conclusions

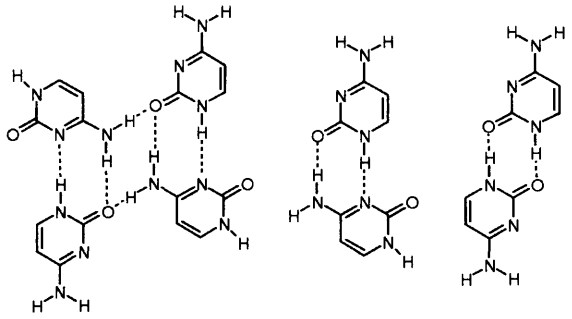
The computational predictions show a range of energetically competitive crystal structures, consisting of comparable stackings of hydrogen bonded dimers. This suggests that any intrinsic differences in the van der Waals contacts between these hydrogen bonded chains does not have a significant effect on the lattice energy of the crystal. Even though there are two distinct hydrogen bonded dimer motifs present within the low energy structures, this shallow minimum in the energy surface could be an artefact of inadequacies in the model potential at accurately

modelling this molecular stacking. Nevertheless crystal structures consisting of both sheet and three-dimensional hydrogen bonded networks are energetically competitive.

The computational and experimental results suggest that kinetic effects play an important part in the crystallisation of thymine, further highlighted by the effects of the solvent on the morphology of the anhydrous crystals. Nevertheless the range of energetically competitive crystal structures predicted does not rule out finding experimental polymorphs of this compound.

4.7 Cytosine

Cytosine (6-amino-pyrimid-2-one), scheme 4.1, consists of a 6-membered ring, with both an amine and carbonyl functional group. There is one anhydrous crystal structure of cytosine,[208;232] along with a monohydrate[208;233;234] and a hydrochloride salt[235]. There has been a wide range of research on cytosine, including *ab initio* studies on the isolated molecule[211;212;236], and on complexes/solvates[237-239].



(a) (b) (c)

Scheme 4.3 The hydrogen bonding found between cytosine residues[206] in (a) cytosine, (b) cytosine monohydrate[208;233;234] and (c) the complex cytosine/5-fluorouracil monohydrate[240]

Cytosine, with its arrangement of hydrogen bond donors and acceptors, has the ability to form a range of hydrogen bonding motifs in the crystalline state, as shown in Scheme 4.3, including hydrogen bonded tetramer and dimer units.

4.7.1 Solubility

The results of the solubility study are shown in Table 4.9.

Table 4.9 The solubility of cytosine in various solvents.

Observation	Solvent
Soluble	H_2O, DMSO, 2,2,2-triflouroethanol, formaldehyde, hydrochloric acid
Partially Soluble	Ethanol, butan-1-ol, methanol, propan-2-ol, acetone, 1,4-dioxane, diethyl ether, DMF, ethyl acetate, acetonitrile, aniline, THF, 1,methyl-2,pyrilidinone, disopropylether, diethyloxalate, hydrobromic acid, dimethylamine in H_2O
Insoluble	Dichloromethane, chloroform, nitromethane, propan-1-ol, 1,2-dichloroethane, toluene, butan-2-ol, hexane, cyclohexane, methyl benzoate, xylene, tetrachloroethylene, n-octane, di-n-butyl ether, ethyl methyl ketone, tert-butylmethyl ether, ethylene glycol

4.7.2 Experimental results

The results for the experimental polymorph screen are shown in Table 4.12 SI. The majority of crystallisations gave the monohydrate form, as cytosine is known to be sensitive to moisture[232]. The anhydrous crystal structure was also obtained and redetermined at low temperature, outlined in section 4.7.3. Sublimation of cytosine at 220 ° C yielded plate like crystals of a decomposition product ammonium hydrogen carbonate[241;242]. The two other crystal structures found were cytosine hydrochloride, described in section 4.14 SI, and oxalic acid dihydrate, described in section 4.15 SI.

4.7.3 Crystallisation of anhydrous cytosine

The anhydrous crystals of cytosine were obtained, mostly in a plate-like habit, from methanol/acetone, methanol/chloroform and methanol/disopropylether mixed solvent systems; and from vapour diffusion of chloroform and toluene into ethanolic solutions. The vapour diffusion techniques produced very small crystals of sufficient quality for single crystal X-ray analysis. However the refinement of the low temperature data set was not significantly better than previously reported[208], section 4.13 SI.

In the low temperature solid state molecular structure, the N(3)-H(3) bond is out of the plane of the ring by around 10 °, shown in Figure 4.15. The NH_2 group is in a slight pyramidal conformation, rotated around 8 ° from the mean ring plane. This agrees with the *ab initio*

conformational analysis, section 4.3.1, which demonstrates that slight rotations of the amine group are energetically feasible.

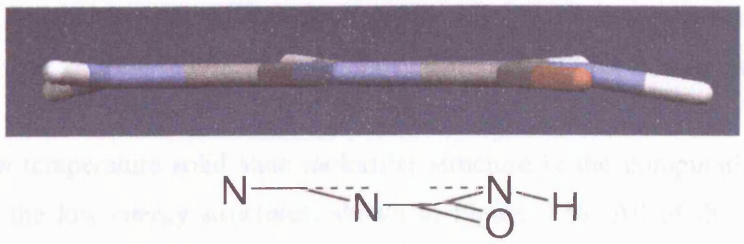


Figure 4.15 The cytosine molecule in the anhydrous low temperature crystal structure, showing the non-planarity of the ring

4.7.4 Computational polymorph searches

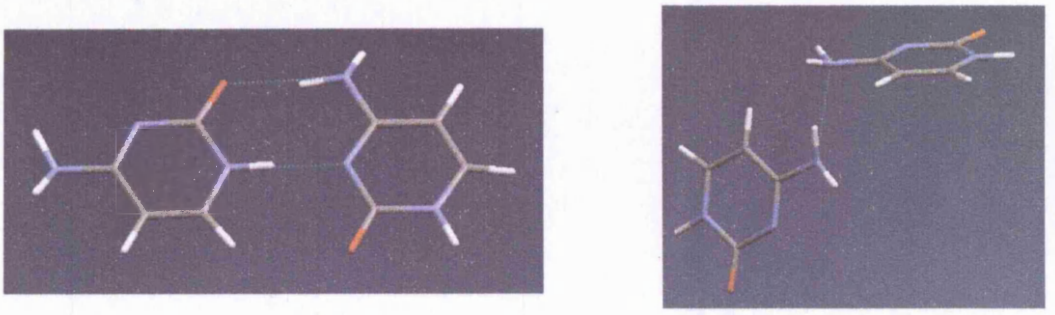
In the *ab initio* molecular structure computational search, of the 3200 structures that were generated, there were 12 unique crystal structures within 10 kJ mol^{-1} of the global lattice energy minimum. To determine whether the search is sensitive to the assumed molecular structure, another search was performed using the low temperature solid state molecular structure. In this search there are 18 unique crystal structures within the same energy range. The results are shown in Figure 4.17 and Table 4.10.

Both ExptMinOpt and ExptMinExpt have been found in the computational searches, however there is a large difference in the relative energies between the two. ExptMinOpt was found 9 kJ mol^{-1} above the global lattice energy minimum (6 kJ mol^{-1} at room temperature estimates), and at the global lattice energy minimum for ExptMinExpt. This shows that the search is very sensitive to the differences in the amine conformation in cytosine. This is further highlighted by (U_{latt}ExptMinExpt – U_{latt}ExptMinOpt) being around 17 kJ mol^{-1}, which is much larger than the estimated MP2 intramolecular energy difference between the solid state (with full molecular optimisation with the NH_2 constrained to the solid state conformation) and *ab initio* molecular structures, *ca.* 0.005 kJ mol^{-1}.

In the *ab initio* molecular structure computational search, the structure at the global lattice energy minimum, AQ68, is around 4 kJ mol^{-1} more stable than the structure ranked 2nd in energy, which decreases to 3 kJ mol^{-1} at room temperature estimates. There are five hypothetical structures that are energetically more favourable than ExptMinOpt, all using O4 and N5 as hydrogen bond acceptors, with the majority also using N7 as an acceptor, Figure 4.16. All these

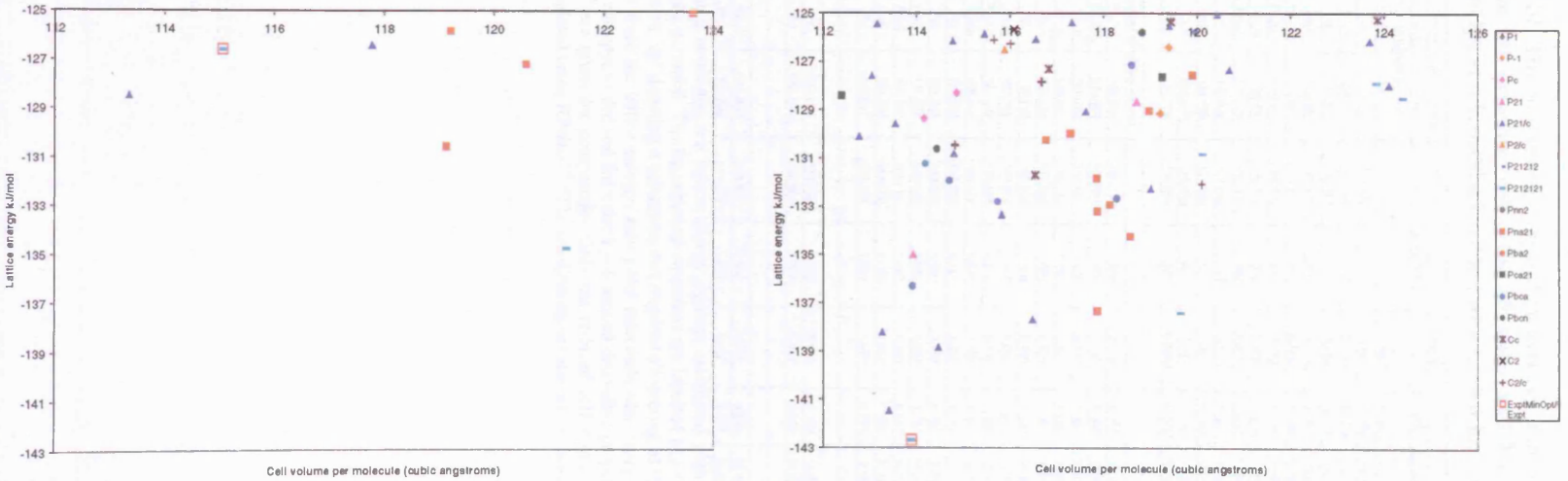
structures form either three-dimensional or sheet hydrogen bonded networks, which consist of the characteristic hydrogen bonded dimer, shown in Figure 4.16(a). The hydrogen bonded N-H···N (amine-amine) interaction is common in crystals, with 1,5-diamino-1*H*-1,2,3,4-tetrazole[243], aniline[244] and 2,5-dichloroaniline[245] as examples. In the solid state crystal structure, the adjacent chains are linked by N-H···O hydrogen bonds, with the N7 not acting as a hydrogen bond acceptor.

Using the low temperature solid state molecular structure in the computational search gave a reordering of the low energy structures, shown in Figure 4.18. All of the low energy crystal structures use the same hydrogen bonded dimer (Figure 4.16) as in the *ab initio* molecular structure computational search, with a greater variety of orientations of these chains relative to each other. The majority of these hypothetical structures use N7 as a hydrogen bond acceptor, strengthening the idea that this type of hydrogen bonding arrangement is energetically favourable.



(a) (b)

Figure 4.16 (a) The characteristic dimer structure present within the hydrogen bonding networks in the low energy crystal structures of cytosine, using the *ab initio* molecular structure, and (b) the N-H···N amine-amine hydrogen bond in AQ68, showing N7 used as an acceptor



(a) (b)

Figure 4.17 The results of the computational polymorph search on cytosine, using the (a) *ab initio*, and the (b) low temperature solid state molecular structure. The ExptMinOpt and ExptMinExpt structures are shown for comparison

Table 4.10 The low energy crystal structures[a] found within 10 kJ mol^{-1} of the global lattice energy minimum for cytosine, using the *ab initio* and low temperature solid state molecular structures. ExptMinOpt and ExptMinExpt are also shown for comparison

Structure	Space	Lattice Energy	[b]Free energy		[c]Reduced Cell				Hydrogen bonding acceptors		[d]Graph set			[e]Elastic
	group	/kJ mol^{-1}	at 298 kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles/°	used and motif		Level 1	Level 2	Level 3	constant
Ab initio molecular structure search														
AQ68	$P2_12_12_1$	-134.778	-145.144	1.521	3.8255	9.3841	13.1745		O4,N5,N7	3D	C1,1(4)	C1,1(2)	C2,2(12)	2.19
AR5[sg]	$P2_1$	-130.79	-142.424	1.561	3.565	8.868	15.004	α 94.44	O4,N5,N7	3D	D1,1(2)	D1,1(2)	D1,1(2)	2.09
AM1	$P2_1/c$	-128.509	-140.066	1.628	3.7364	9.5039	12.777	β 92.06	O4,N5	Sheet	C1,1(6)	C1,1(6)	C1,1(4)	2.78
AR59[sg]	$P2_1$	-127.318	-138.951	1.537	3.638	8.779	15.072	α 94.30	O4,N5,N7	3D	D1,1(2)	D1,1(2)	D1,1(2)	1.81
AW50[sg]	$P2_1/c$	-126.937	-139.113	1.557	3.596	9.083	14.519	γ 87.65	O4,N5,N7	Jagged sheet	C1,1(6)	C1,1(4)	R2,2(4)	0.95
AQ6	$P2_12_12_1$	-126.613	-138.506	1.603	3.7022	9.5566	13.0091		O4,N5	3D	C1,1(6)	C1,1(6)	C1,1(4)	4.40
ExptMinOpt	**$P2_12_12_1$**	**-126.61**	**-138.511**	**1.603**	**3.702**	**9.557**	**13.009**		**O4,N5**	**3D**	**C1,1(6)**	**C1,1(6)**	**C1,1(4)**	**4.40**
AM27	$P2_1/c$	-126.43	-138.212	1.566	3.5822	8.8534	14.8624	γ 91.76	O4,N5,N7	Jagged sheet	C1,1(6)	C1,1(4)	R2,2(4)	1.05
AV54	$Pna2_1$	-125.073	-139.921	1.492	3.5321	9.4553	14.8087		O4,N5,N7	3D	C1,1(6)	C1,1(4)	C1,1(2)	0.38
CD43[sg]	Pc	-124.898	-135.892	1.531	6.814	9.585	14.851	β 96.24	O4,N5,N7	Jagged sheet	D1,1(2)	D1,1(2)	D1,1(2)	1.06
AU62	$Pna2_1$	-124.363	-137.399	1.538	3.6477	9.4354	13.9419		O4,N5	Chains	C1,1(6)	C1,1(4)	R2,2(8)	1.99
CD68[sg]	$P2_1/c$	-124.305	-135.506	1.534	7.141	9.595	14.847	γ 108.98	O4,N5,N7	3D	C1,1(6)	C1,1(6)	C1,1(4)	1.49
CB53	Pbca	-124.096	-134.851	1.598	6.7486	9.7362	14.0581		O4,N5	Jagged sheet	C1,1(6)	C1,1(6)	C1,1(4)	1.11
Solid state molecular structure search														
AQ35	$P2_12_12_1$	-142.689	-153.108	1.620	3.792	9.460	12.702		O4,N5	3D	C1,1(6)	C1,1(6)	C1,1(4)	4.52
ExptMinExpt	**$P2_12_12_1$**	**-143.463**	**-154.085**	**1.616**	**3.792**	**9.460**	**12.701**		**O4,N5**	**3D**	**C1,1(6)**	**C1,1(6)**	**C1,1(4)**	**4.11**
FC42	$P2_1/c$	-141.469	-152.007	1.627	3.762	9.438	12.777	β 90.40	O4,N5,N7	Sheets	C1,1(6)	C1,1(6)	C1,1(4)	2.95
AM120	$P2_1/c$	-138.847	-149.101	1.612	3.650	8.906	14.087	γ 90.63	O4,N5,N7	3D	C1,1(6)	C1,1(6)	C1,1(4)	1.92
AM44	$P2_1/c$	-138.226	-148.056	1.629	6.932	7.557	9.585	γ 115.54	O4,N5,N7	Sheets	C1,1(6)	C1,1(6)	C1,1(4)	0.86
AR58[sg]	$P2_1$	-137.758	-148.905	1.576	3.648	8.822	14.639	α 83.81	O4,N5,N7	3D	D1,1(2)	D1,1(2)	D1,1(2)	2.20
AM35	$P2_1/c$	-137.685	-147.884	1.584	6.892	7.217	9.388	γ 93.94	O4,N5,N7	Sheets	C1,1(6)	C1,1(4)	R2,2(4)	0.11
AQ66	$P2_12_12_1$	-137.422	-148.324	1.542	3.782	9.388	13.477		O4,N5,N7	3D	C1,1(6)	C1,1(4)	C1,1(2)	1.19
CB21	Pbca	-136.312	-146.612	1.620	6.754	9.674	13.947		O4,N5,N7	Sheets	C1,1(6)	C1,1(6)	C1,1(4)	0.88
AR14[sg]	$P2_1$	-135.806	-146.307	1.581	3.685	8.772	14.510	α 84.44	O4,N5,N7	3D	D1,1(2)	D1,1(2)	D1,1(2)	3.10
AH6	$P2_1$	-135.008	-146.857	1.619	3.698	6.787	9.250	γ 101.05	O4,N5	Chains	C1,1(6)	C1,1(4)	R2,2(8)	1.81
CD28[sg]	$Pca2_1$	-134.921	-146.004	1.581	6.931	9.457	14.242		O4,N5,N7	Jagged Sheet	D1,1(2)	D1,1(2)	D1,1(2)	0.74
CD130[sg]	$Pca2_1$	-134.402	-145.44	1.549	7.384	9.523	13.547		O4,N5,N7	Jagged Sheet	D1,1(2)	C1,1(6)	D1,1(2)	1.20
CD46[sg]	Pca21	-133.762	-144.782	1.550	6.869	9.526	14.623	β 95.61	O4,N5,N7	Jagged Sheet	D1,1(2)	D1,1(2)	D1,1(2)	0.78
AJ78[sg]	P-1	-133.607	-144.191	1.594	6.984	7.813	9.580	α 83.38	O4,N5,N7	Ribbons	D1,1(2)	D1,1(2)	D1,1(2)	0.57
								β 79.63						
								γ 64.32						
BJ53[sg]	Pc	-133.590	-145.168	1.555	3.929	9.370	12.996	α 97.27	O4,N5,N7	3D	D1,1(2)	C1,1(6)	D1,1(2)	3.60
CD69[sg]	P-1	133.392	-143.982	1.581	6.817	9.582	14.663	α 93.58	O4,N5,N7	3D	D1,1(2)	D1,1(2)	D1,1(2)	1.07
								β 100.70						
								γ 95.42						
AM82	$P2_1/c$	-133.350	-144.527	1.593	3.678	8.248	15.464	γ 99.03	O4,N5,N7	Sheets	C1,1(6)	C1,1(4)	R2,2(4)	1.07
CB106	Pbca	-132.782	-144.122	1.584	6.412	9.399	15.364		O4,N5,N7	Sheets	C1,1(6)	C1,1(4)	R2,2(4)	0.31

[a]All calculated structures are lattice energy minima calculated with the *ab initio* or solid state molecular model and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK coordination geometry and order of density, with 'sg' denoting a minimum that required a lowering of the original space group symmetry. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated by PLATON[158] are given for comparison. Only the reduced cell angles which are not 90° are tabulated. [d]Only the first three levels shown, calculated using RPluto[168]. [e]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.

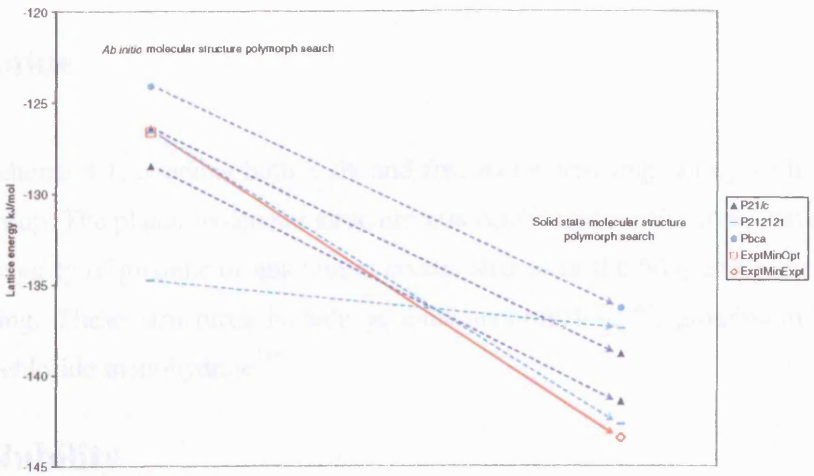


Figure 4.18 A comparison of the relative energies within a select number of equivalent low energy structures in both computational polymorph searches on cytosine, within 10 kJ mol^{-1} of the global lattice energy minimum. The ExptMinOpt and ExptMinExpt are also shown for comparison.

4.7.5 Conclusions

It has been shown, Figure 4.7, that the electrostatic potential around the molecule changes significantly due to slight conformational changes of the amine group. It is clear that this limited flexibility has a significant effect on the computational predictions, which is highlighted by the large relative energy gap between the ExptMinOpt and ExptMinExpt crystal structures. However despite the questionable relative stabilities, there are a wide range of low energy crystal structures predicted which comprise of both three-dimensional or sheet hydrogen bonding networks. The majority of structures use N7 as a hydrogen bond acceptor in N-H···N interactions, which imply that this type of hydrogen bond arrangement is energetically favourable, despite the experimental solid state structure not using this acceptor.

It is most probable that the anhydrous crystal structure is the most energetically stable, however the sensitivity to the limited molecular flexibility, combined with the errors in the model potential, makes it difficult to gauge the relative energies of the energetically competitive hypothetical structures. These computational results, along with the fact that cytosine is sensitive to moisture in the laboratory, suggests that finding polymorphs is unlikely.

4.8 Guanine

Guanine, scheme 4.1, contains both a six and five membered ring, along with an amine and a carbonyl group. The planar molecular structure was considered for the computational predictions as in the majority of guanine or guaninium crystal structures the NH_2 group is nearly co-planar with the ring. These structures include guanine monohydrate[246], guaninium chloride[247] and guaninium chloride monohydrate[248].

4.8.1 Solubility

The results of the solubility study are shown in Table 4.11.

Table 4.11 The solubility of guanine in various solvents.

Observation	Solvent
Soluble	Hydrochloric acid, hydrobromic acid
Partially Soluble	H_2O, DMF, 1,2-dichloroethane, 2,2,2-trifluoroethanol, diethyl oxalate
Insoluble	Acetone, choloform, propan-2-ol, methanol, butan-1-ol, ethanol, 1,4-dioxane, diethyl ether, nitromethane, propan-1-ol, ethyl acetate, toluene, acetonitrile, THF, cyclohexane, aniline, formaldehyde, DMSO, hexane, butan-2-ol, 1,methyl-2-pyrolidinone, tetrachloroethylene, methyl benzoate, xylene, diisopropylether, octane, di-n-butyl ether, ethyl methyl ketone, ethylene glycol

4.8.2 Experimental results

A summary of the results of the experimental screen are shown in Table 4.12. A variety of other methods to attempt crystallisation were performed, shown in Table 4.16 SI. Attempts to crystallise the monohydrate form[246] were not successful. The vapour diffusion experiments did not yield any sample for characterisation.

Table 4.12 A summary of the results of the crystallisation experiments on guanine. The other crystallisation experiments did not give any sample for analysis

Method	Solvents	Result
Slow evaporation	Dichloromethane, DMF, propan-1-ol, 1,2-dichloromethane, butan-1-ol	Small particles of powder, not adequate for analysis
Slow evaporation	Hydrochloric acid	Guaninium chloride monohydrate was obtained by slow evaporation, section 4.17 SI. Guaninium chloride dihydrate[59] was obtained by virtue of a faster evaporation rate, section 4.18 SI
Slow evaporation	Diethyl oxalate solution	Crystals of oxalic acid dihydrate
Slow evaporation	Aqueous butyamide solution of adenine	Crystals of butyramide[61], section 4.19 SI
Slow evaporation	Hydrochloric acid solution of guanine, with triethylenediamine added	Crystals of triethylenediaminium chloride, and triethylenediaminium dichloride dihydrate[62]
Sublimation	-	Small particles of powder, not adequate for characterisation analysis

4.8.3 Computational polymorph searches

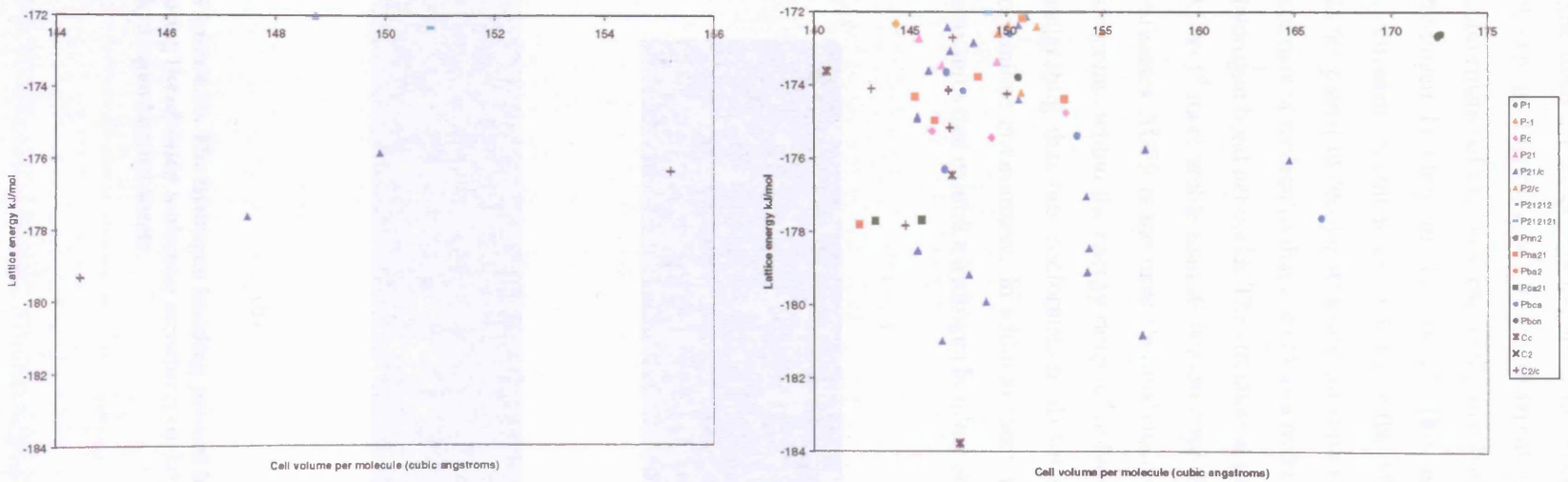
The two stable *ab initio* conformations (section 4.3.1) gave qualitatively the same results in the computational polymorph search. Of the 1600 structures generated, 10 unique structures were found within 7 kJ mol^{-1} of the global lattice energy minimum. Using the planar molecular structure in the computational search, there are 22 unique structures found within this energy range. The results of both searches are shown in Figure 4.19 and Table 4.13.

The two computational polymorph searches show significantly different patterns of the low energy structures. In the *ab initio* molecular structure polymorph search, all the structures consist of three-dimensional hydrogen bonded motifs, with the structure at the global lattice energy minimum, DE43, around 2 kJ mol^{-1} energetically more stable than its nearest rival. This is the only structure within the energy range of polymorphism that consists of molecules that do not use O4 as a hydrogen bond acceptor, although it has a similar density to the others. There are 5 unique crystal structures within 4 kJ mol^{-1} of the global lattice energy minimum with molecules that use a variety of different hydrogen bond acceptors and donors. Nevertheless when comparing the room temperature energy estimates there is a significant reordering of the stabilities, with the 3rd ranked structure at 0 K now at the global energy minimum.

Table 4.13 The low energy crystal structures[a] found within 7 kJ mol^{-1} of the global lattice energy minimum for the computational polymorph searches on guanine, using the *ab initio* and planar molecular structures

Structure	Space	Lattice Energy	[b]Free energy		[c]Reduced Cell				Hydrogen bonding acceptors		[d]Graph set			[e]Elastic
	group	/kJ mol^{-1}	at 298 kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles/°	used and motif		Level 1	Level 2	Level 3	constant
Gas phase molecular structure search														
DE43	C2/c	-179.324	-188.43	1.738	7.4653	12.427	12.6949	α 101.161	N1,N4,N5	3D sheets	C1,1(6)	R2,2(8)	C1,1(7)	1.97
AI79	$P2_1/c$	-177.64	-188.243	1.701	7.0613	7.3905	11.3765	β 96.435	O4,N1,N4,N5	3D sheets	C1,1(6)	R2,2(8)	C1,1(7)	2.32
CC19[sg]	$Pca2_1$	-177.164	-191.556	1.741	3.742	12.745	24.182		O4, N1,N5	3D sheets	C1,1(6)	C1,1(6)	C1,1(6)	2.87
DE87	C2/c	-176.361	-186.643	1.616	6.7858	13.2987	14.1622	α 103.69	O4,N1,N4,N5	3D	C1,1(6)	R2,2(8)	C1,1(7)	1.89
AM18	$P2_1/c$	-175.863	-189.541	1.674	3.6927	11.271	14.4174	β 92.166	O4,N1,N4	Sheets	C1,1(6)	C1,1(5)	C1,1(7)	1.73
CD74[sg]	$Pca2_1$	-175.221	-190.2	1.741	3.439	12.11	27.691		O4,N1,N5	3D sheets	C1,1(6)	C1,1(6)	C1,1(6)	3.13
CE55[sg]	$Pna2_1$	-173.764	-189.102	1.747	3.515	12.53	26.089		O4,N1,N5	3D sheets	C1,1(6)	C1,1(6)	C1,1(6)	1.94
DC14[sg]	Cc	-173.636	-183.861	1.680	7.011	12.495	13.742	β 97.088	O4,N1,N4,N5	3D	C1,1(6)	D1,1(2)	D1,1(2)	3.01
FA54[sg]	$P2_1$	-173.235	-187.009	1.727	3.994	6.589	22.597	γ 112.18	O4,N1,N5	3D sheets	C1,1(6)	C1,1(6)	C1,1(6)	4.85
AQ91	$P2_12_12_1$	-172.347	-185.477	1.664	3.7833	11.4592	13.9159		O4,N4,N5	3D	C1,1(6)	C1,1(5)	C1,1(7)	5.11
Planar molecular structure search														
DA10	Cc	-183.788	-192.283	1.7	3.7583	10.8485	14.6136	γ 97.838	O4,N1,N4	3D	C1,1(6)	C1,1(6)	C1,1(7)	2.87
CD89[sg]	$Pca2_1$	-181.454	-193.194	1.705	6.691	13.055	13.486		O4,N1,N4	3D	C1,1(6)	C1,1(6)	C1,1(7)	1.13
AM15	$P2_1/c$	-180.829	-193.322	1.598	4.6886	11.4216	11.7649	β 94.297	O4,N1,N4	3D	C1,1(6)	C1,1(5)	C1,1(7)	2.67
CE12[sg]	$P2_1$	-180.512	-193.507	1.733	7.477	11.948	13.021	γ 95.220	O4,N1,N4,N5	3D sheets	D1,1(2)	D1,1(2)	D1,1(2)	1.47
AU79[sg]	P-1	-180.39	-193.054	1.765	6.772	7.765	11.549	α 85.46	O4,N1,N4,N5	3D sheets	D1,1(2)	D1,1(2)	D1,1(2)	1.76
								β 74.34						
								γ 76.60						
AM75	$P2_1/c$	-179.912	-192.004	1.685	5.117	7.8962	14.8335	γ 96.248	O4,N1,N4	3D	C1,1(6)	C1,1(6)	C1,1(5)	0.62
AM85	$P2_1/c$	-179.177	-191.405	1.695	6.8829	7.1024	12.2051	α 96.969	O4,N1,N4	3D	C1,1(6)	C1,1(6)	C1,1(7)	0.93
AM23	$P2_1/c$	-179.111	-192.807	1.627	7.2408	7.3514	11.7645	γ 99.954	O4,N1,N4	Sheets	C1,1(6)	C1,1(5)	R2,2(8)	0.01
AM31[sg]	Pc	-179.918	-192.748	1.641	5.03	7.568	16.135	γ 95.12	O4,N1,N4	Sheets	C1,1(6)	C1,1(6)	C1,1(6)	0.28
AM28	$P2_1/c$	-178.517	-191.95	1.726	4.0656	9.6311	14.8852	γ 93.779	O4,N1,N4,N5	3D	C1,1(6)	C1,1(6)	C1,1(5)	2.80
AM45	$P2_1/c$	-178.454	-192.233	1.626	6.4378	6.8316	15.0464	γ 111.137	O4,N1,N4	Sheets	C1,1(6)	C1,1(6)	C1,1(5)	0.80
DE138	C2/c	-177.893	-190.044	1.734	8.0298	11.5917	12.9271	α 105.75	O4,N4,N5	3D sheets	C1,1(6)	C1,1(6)	C1,1(5)	0.76
AV6	$Pna2_1$	-177.82	-189.794	1.763	4.6547	8.6576	14.1325		O4,N1,N4	3D	C1,1(6)	C1,1(7)	C1,1(4)	1.52
AY31	$Pca2_1$	-177.71	-191.365	1.752	6.8917	7.0794	11.7409		O4,N4,N5	3D	C1,1(6)	C1,1(6)	C1,1(6)	2.89
AY30	$Pca2_1$	-177.694	-192.584	1.723	3.577	11.3337	14.3679		O4,N4	3D	C1,1(6)	C1,1(6)	C1,1(6)	5.67
CB56	Pbca	-177.629	-193.009	1.508	6.4842	11.8563	17.3132		O4,N1,N4	Sheets	C1,1(6)	C1,1(5)	R2,2(8)	0.07
FC90	$P2_1/c$	-177.035	-189.832	1.628	7.278	7.686	11.0759	α 95.518	O4,N1,N4	3D	C1,1(6)	C1,1(6)	C1,1(5)	1.55
BG86[sg]	Cc	-176.469	-187.629	1.705	6.779	6.949	6.649	α 64.86	N1,N4	Sheets	C1,1(6)	C1,1(7)	C1,1(8)	4.22
								β 84.46						
								γ 84.46						
DA104	Cc	-176.463	-185.56	1.706	6.7793	7.4525	11.7297	β 96.567	N1,N4	Sheets	C1,1(6)	C1,1(7)	C2,2(8)	0.36
CB67[sg]	$P2_1/c$	-176.068	-189.095	1.675	6.366	12.815	14.697	α 90.03	N1,N4,N5	3D	D1,1(2)	C1,1(7)	D1,1(2)	0.02
CB39	Pbca	-176.32	-188.485	1.709	6.953	11.9469	14.1387		O4,N1,N4	Sheets	C1,1(6)	C1,1(6)	C1,1(6)	3.18
AM2	$P2_1/c$	-176.056	-190.967	1.524	3.788	9.9252	17.658	γ 97.072	O4,N1,N4	Sheets	C1,1(6)	C1,1(5)	R2,2(8)	0.15

[a]All calculated structures are lattice energy minima calculated with the the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK coordination geometry and order of density, with 'sg' denoting a minimum that required a lowering of the original space group symmetry. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated by PLATON[158] are given for comparison. Only the reduced cell angles which are not 90° are tabulated [d]Only the first three levels shown, calculated using RPluto[168]. [e]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.



(a) (b)

Figure 4.19 The results of the computational polymorph searches on guanine, using the (a) *ab initio*, and (b) planar molecular structure. The low energy structures are shown.

The three lowest energy hypothetical structures in the *ab initio* molecular structure search are shown in Figure 4.20. These different packings of guanine molecules shows that slight conformational changes, especially involving the amine group, have a significant effect on the hydrogen bonding in the crystal. This is consistent with the differences observed in the electrostatic potential associated with the different conformations, Figure 4.7.

In the planar molecular structure computational search the lowest energy hypothetical structures consists of molecules that use O4 as a hydrogen bond acceptor, and consist of three-dimensional hydrogen bond networks. The structure at the global lattice energy minimum, DA10, is around 2 kJ mol^{-1} more stable than its nearest rival. However when comparing room temperature energy estimates AU79 is the most thermodynamically stable crystal structure. There are many more structures within the energy range of polymorphism when using the planar molecular structure, suggesting that this conformation allows more energetically favourable packings within the crystalline environment. In addition there are an increased number of energetically favourable structures that exhibit a hydrogen bonding sheet motif.

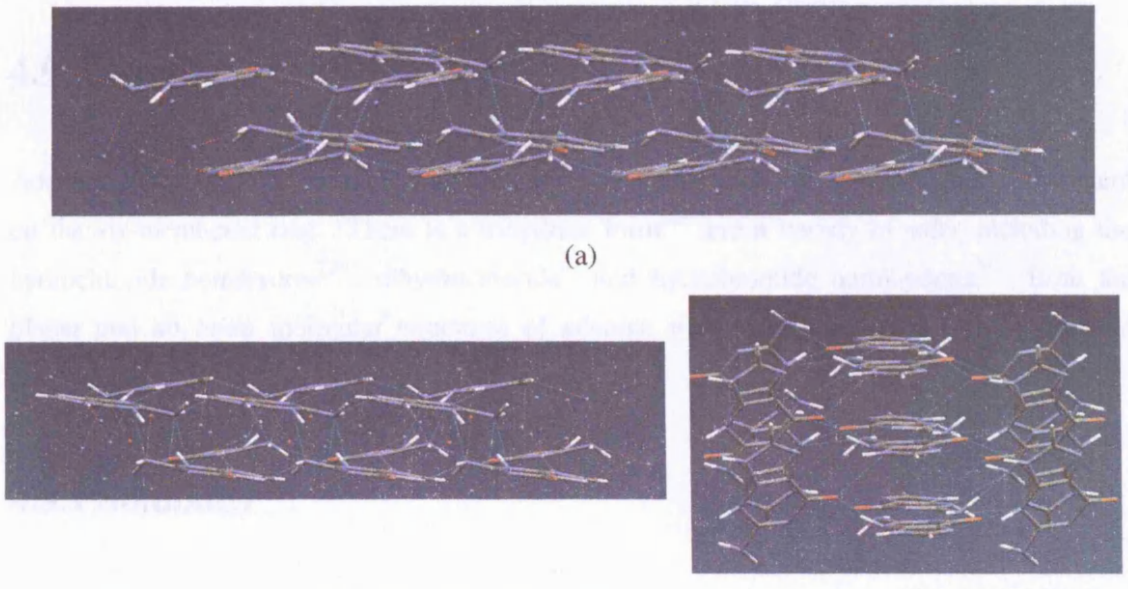


Figure 4.20. The hydrogen bonding present in the three lowest energy crystal structures of guanine using the *ab initio* molecular structure; (a) DE43, (b) AI79, and (c) CC19 showing three-dimensional hydrogen bonded sheets

4.8.4 Conclusions

This study shows that the computational predictions are sensitive to the assumed molecular structure of guanine, highlighting the problems of limited flexibility for crystal structure prediction. The majority of low energy structures are unique in both computational searches, suggesting that further conformational changes could generate further packings of molecules. This could be a consequence of the significant differences in the electrostatic potential between the two conformations, shown in Figure 4.7, showing different strengths of the contribution to the hydrogen bonding.

Despite the sensitivity to the assumed molecular structure, there are a variety of energetically feasible crystal structures consisting of complex three-dimensional or sheet hydrogen bond networks. Therefore it could be plausible that the anhydrous crystal structure, if found, could comprise of similar molecular packings. However since guanine is sparsely soluble in the majority of organic solvents, the experimental limitations will be the deciding factor in whether any polymorphs can be readily found.

4.9 Adenine

Adenine, Scheme 4.1, consists of both a six and five-membered ring, with an amine substituent on the six-membered ring. There is a trihydrate form[249] and a variety of salts, including the hydrochloride hemihydrate[250], dihydrochloride[60] and hydrobromide hemihydrate[251]. Both the planar and *ab initio* molecular structures of adenine were considered in the computational predictions.

4.9.1 Solubility

The results of the solubility study are shown in Table 4.14.

Table 4.14 The solubility of adenine in various solvents.

Observation	Solvent
Soluble	Hydrochloric acid, dimethylamine + H_2O, 2-methoxyethylamine
Partially Soluble	Nitromethane, benzonitrile, 2,2,2-trifluoroethanol, 2-chloroethanol, DMSO, acetic acid, hydrobromic acid, 2-methoxyethylcyanoacetate, triethylamine
Insoluble	H_2O, methanol, acetone, 1,4-dioxane, ethanol, acetonitrile, propan-1-ol, ethyl acetate, tert-butylmethyl ether, THF, issopropylether, di-n-butyl ether, toluene, ethylmethyl ketone, 1,methyl-2-pyrrolidinone, n-octane, o-xylene, butan-1-ol, 1,2-dichloromethane, butan-2-ol, propan-1-ol, n-hexane, diethyl ether, cyclohexane, methylbenzoate, DMF, dichloromethane, chloroform, tetachloroethylene, dimethylamine

4.9.2 Experimental results

The results of the crystallisation experiments are described in Table 4.15. The vapour diffusion experiments gave no solid products for analysis.

Table 4.15 The results of the crystallisation experiments on adenine. The other crystallisation experiments did not give any sample for analysis

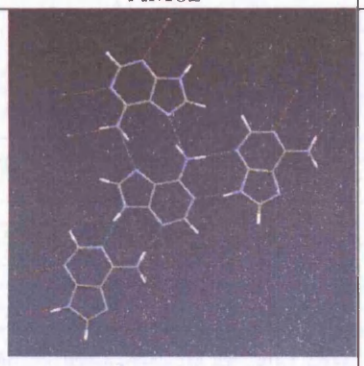
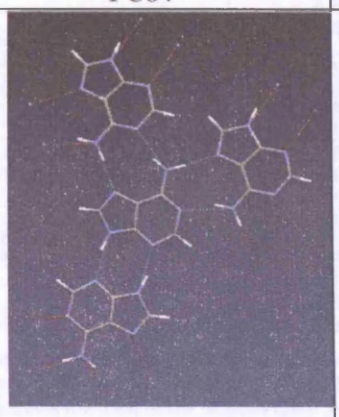
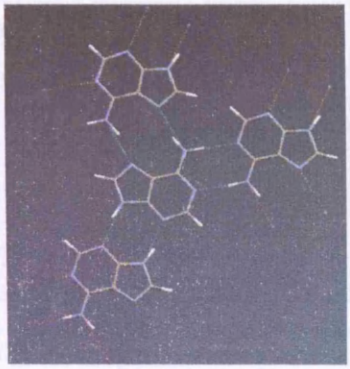
Method	Solvents	Result
Slow evaporation	Methanol, propan-2-ol, tert-butylmethylether, THF, butan-1-ol, propan-1-ol	Small particles of solid, not enough for characterisation analysis
Slow evaporation	2,2,2-trifluoroethanol, acetic acid, dimethylamine in H_2O, 2-methyoxyethylamine	Microcrystalline solid, powder pattern matched stock sample
Slow evaporation	2-chloroethanol	Adeninium chloride hemihydrate[250;252;253]
Slow evaporation	Equimolecular amounts of thymine/adenine and cytosine/adenine in dilute hydrochloric acid, dilute hydrochloric acid, concentrated hydrochloric acid	Adeninium dichloride[60], described in section 4.20 SI
Slow evaporation	H_2O	Microcrystalline solid, poor powder pattern which could not be indexed, did not match stock sample or known trihydrate structure[249]
Sublimation	-	Microcrystalline solid, powder pattern matched stock sample

4.9.3 Computational polymorph searches

In the computational polymorph search based on the *ab initio* molecular structure, of the 2900 structures generated in the search, there are 11 unique structures within 7 kJ mol^{-1} of the global lattice energy minimum. When using the planar molecular structure, there are 14 unique structures in the same energy range. The results are shown in Figure 4.21 and Table 4.17.

In the *ab initio* molecular structure computational search the structure at the global lattice energy minimum, AM62, is around 2 kJ mol^{-1} more stable than the closest rival, with about the same estimated energy gap at room temperature. There are three hypothetical crystal structures within 4.5 kJ mol^{-1} of the global lattice energy minimum, which decreases to 2.5 kJ mol^{-1} at room temperature. Using the planar molecular structure in the computational search there are 5 unique hypothetical structures within 1 kJ mol^{-1} of the global lattice energy minimum, showing a range of densities. In both computational polymorph searches all but one of the crystal structures consists of hydrogen bonding sheets. There are three main types of sheets present differing in the combinations of acceptors and donors used, resulting in some subtle rearrangements of molecules in the sheets, Table 4.16. There are also slight changes in the orientations of the sheets relative to each other. All the structures consisting of molecules that use N1, N2 and N3 as hydrogen bond acceptors.

Table 4.16 The three different hydrogen bonded sheet structures present in the low energy structures in both computational polymorph searches on adenine

AM62	FC84	AB102
		

The majority of elastic constants of the low energy structures, Table 4.17, are relatively low, < 1 GPa. This could indicate a tendency for the hydrogen bonded sheets to slip over each other, decreasing the mechanical stability. This would suggest that the energetically favourable crystal packings of adenine give relatively soft crystals, and hence could indicate problematic crystal growth.

Table 4.17 The low energy crystal structures found within 7 kJ mol^{-1} of the global lattice energy minimum for adenine, using the *ab initio* and planar molecular structures. All the structures consist of molecules that use N1, N2, and N3 as hydrogen bond acceptors, and all contain hydrogen bonded sheets, except for DE56 which contains a three-dimensional hydrogen bonding network

Structure	Space group	Lattice Energy /kJ mol^{-1}	[b]Free energy at 298 kJ mol^{-1}	Density/g cm^{-3}	[c]Reduced Cell a/Å	b/Å	c/Å	Angles/°	[d]Graph set Level 1	Level 2	Level 3	[e]Elastic constant
Gas phase molecular structure												
AM62	$P2_1/c$	-143.063	-152.263	1.627	6.637	6.8249	12.1814	γ 90.121	C1,1(6)	C1,1(6)	R2,2(10)	0.96
FC84	$P2_1/c$	-141.431	-150.427	1.602	4.696	8.1672	14.7514	β 97.984	R2,2(8)	C1,1(5)	C1,1(4)	1.86
CB24	Pbca	-138.54	-149.743	1.592	6.4401	12.1864	14.3705		C1,1(6)	C1,1(6)	R2,2(10)	0.79
FC4	$P2_1/c$	-138.408	-148.383	1.550	4.7183	8.1635	15.1273	β 96.315	R2,2(8)	C1,1(5)	C1,1(4)	0.49
AB102	P-1	-138.032	-147.059	1.625	5.3147	6.9824	7.97	α 101.390	R2,2(8)	R2,2(8)	R2,2(10)	7.09
								β 106.251				
								γ 93.485				
DC46	C2/c	-137.912	-146.967	1.596	8.3516	10.0284	13.7504	α 102.380	R2,2(8)	C1,1(5)	C1,1(4)	0.29
AB16	P-1	-137.428	-147.022	1.564	4.5964	8.1356	8.5411	α 65.623	R2,2(8)	R2,2(8)	R2,2(10)	7.56
								β 89.734				
								γ 81.346				
DD101	C2/c	-136.97	-145.472	1.546	6.8836	12.2225	13.8558	β 95.076	C1,1(6)	C1,1(6)	R2,2(10)	0.17
AM21	$P2_1/c$	-136.677	-147.112	1.540	5.2338	7.4804	14.9694	γ 96.036	C1,1(6)	C1,1(6)	R2,2(10)	0.51
CD110[sg]	$Pna2_1$	-136.489	-148.028	1.547	6.398	12.161	14.913		D1,1(2)	D1,1(2)	D1,1(2)	0.10
CB89[sg]	$P2_1/c$	-136.377	-147.142	1.588	7.027	12.161	13.23	β 89.750	C1,1(6)	C1,1(6)	D1,1(2)	0.72
CE49[sg]	P1	-135.477	-146.068	1.554	6.679	7.967	21.715	α 90.76	D1,1(2)	D1,1(2)	D1,1(2)	0.28
Planar molecular structure												
FC33	$P2_1/c$	-142.874	-151.506	1.597	4.8509	8.0923	14.4496	β 97.855	R2,2(8)	C1,1(5)	C1,1(4)	1.68
AM9	$P2_1/c$	-142.456	-151.945	1.605	6.4576	7.1205	12.1696	γ 91.858	C1,1(6)	C1,1(6)	R2,2(10)	0.77
AK71	$P2_1/c$	-141.902	-150.933	1.562	6.65	7.8624	11.0287	β 94.825	R2,2(8)	C1,1(5)	C1,1(4)	0.78
CC71[sg]	$P2_1/c$	-141.816	-152.796	1.585	7.485	7.496	21.831	γ 112.4	D1,1(2)	D1,1(2)	D1,1(2)	0.52
FC1	$P2_1/c$	-141.685	-150.906	1.552	4.8788	7.8707	15.126	β 95.304	R2,2(8)	C1,1(5)	C1,1(4)	0.81
AJ80[sg]	P-1	-140.85	-150.037	1.560	4.725	8.089	15.226	α 86.54	R2,2(8)	D1,1(2)	D1,1(2)	0.72
								β 83.41				
								γ 85.08				
CE73[sg]	$Pca2_1$	-140.125	-150.426	1.559	6.555	8.007	21.958		D1,1(2)	D1,1(2)	D1,1(2)	0.33
AM43	$P2_1/c$	-139.919	-150.472	1.547	5.3012	7.3655	14.9515	γ 96.478	C1,1(6)	C1,1(6)	R2,2(10)	0.09
CB108[sg]	$P2_1/c$	-139.836	-151.121	1.541	6.736	11.912	14.943	γ 103	C1,1(6)	C1,1(6)	D1,1(2)	0.06
AB72	P-1	-139.082	-148.088	1.617	4.9257	7.0008	8.426	α 99.783	R2,2(8)	R2,2(8)	R2,2(10)	10.21
								β 103.833				
								γ 90.967				
BF37[sg]	P-1	-138.534	-148.595	1.587	5.429	6.882	15.403	α 81.67	R2,2(8)	D1,1(2)	D1,1(2)	10.11
								β 85.29				
								γ 84.61				
AI116	$P2_1/c$	-138.491	-147.801	1.583	8.1118	8.3821	8.8323	α 109.213	R2,2(8)	C1,1(5)	C1,1(4)	0.65
CA73	P-1	-138.089	-147.515	1.552	4.77	7.9944	8.5013	α 65.271	R2,2(8)	R2,2(8)	R2,2(10)	8.52
								β 89.140				
								γ 79.870				
DE56	C2/c	-136.984	-147.913	1.572	3.7788	15.7094	19.452	α 98.530	R2,2(8)	C1,1(4)	R2,2(10)	1.26

[a]All calculated structures are lattice energy minima calculated with the *ab initio* or the planar molecular model and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK coordination geometry and order of density, with 'sg' denoting a minimum that required a lowering of the original space group symmetry. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated by PLATON[158] are given for comparison. Only the reduced cell angles which are not 90° are tabulated. [d]Only the first three levels shown, calculated using RPluto[168]. [e]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.

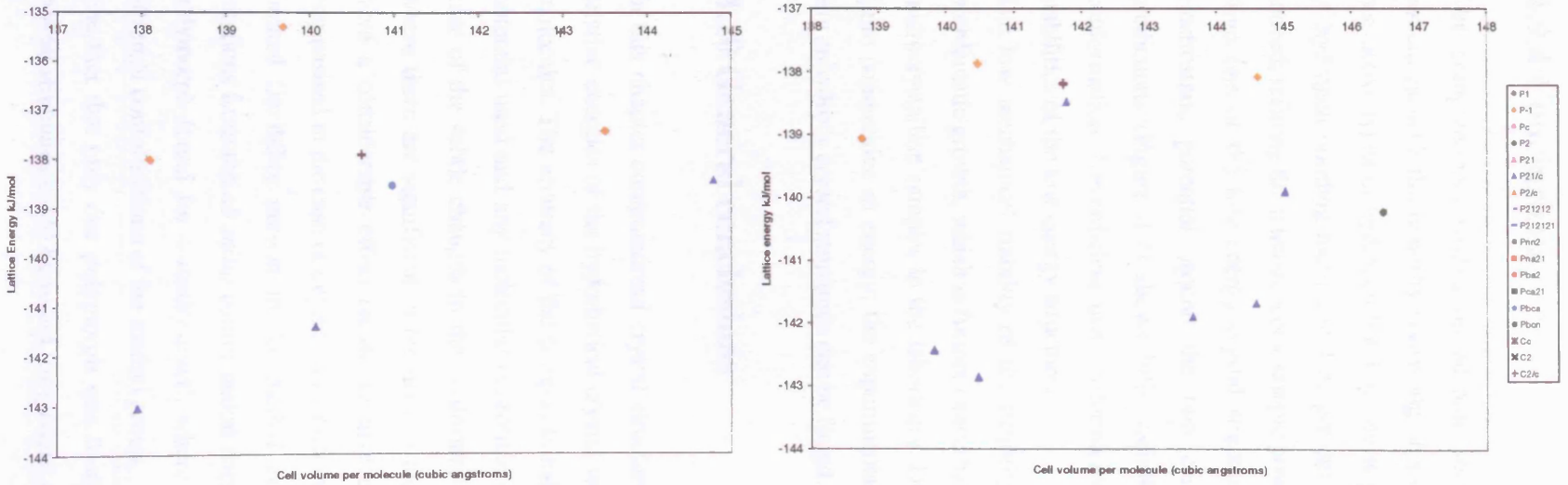


Figure 4.21 The results of the computational polymorph searches on adenine, using (a) the *ab initio*, and (b) planar molecular structure, showing the low energy structures.

4.9.4 Conclusions

The computational studies on adenine show a range of energetically competitive crystal structures, with the majority consisting of a hydrogen bonding sheet structure. There are three distinctive types of hydrogen bonding sheets present. Therefore it seems plausible that this type of hydrogen bonding motif could be present if an anhydrous crystal structure could be found. Indeed, refining the adenine stock sample powder X-ray diffraction data against a model starting from one of the low energy crystal structures (CC71, Table 4.17) seems promising[254]. The electrostatic potential around the two conformations considered in these computational predictions (Figure 4.7) shows little variation, despite the differences in the amine group conformation. Nevertheless this conformation difference does have an affect on the relative stabilities of the low energy structures.

The low mechanical stability of the majority of hypothetical crystal structures does indicate problematic growth, which is further exacerbated by the poor solubility and the tendency to form microcrystalline samples in the laboratory. Despite a variety of crystal structures predicted, all quite competitive in energy, the experimental limitations will undoubtedly determine whether any anhydrous crystal structures can be found.

4.10 General conclusions

In this chapter computational crystal structure prediction has shown the range of patterns and relative energies of the hypothetical crystal structures for a number of structurally similar small molecules. The accuracy of the computational predictions is sensitive to the errors in the model potential used and any molecular conformational flexibility present. This is highlighted in the case of the subtle changes in the conformation of the amine group for cytosine and guanine, where there are significant differences to the electrostatic potential. These structural changes have a considerable effect on the relative energies of the low energy molecular structures, emphasised in the case of cytosine in which the relative stabilities are problematic. There is also limited flexibility present in the methyl group in 6-methyluracil and thymine, with small rotations from the *ab initio* conformation energetically plausible. This is highlighted in the new polymorph found for 6-methyluracil, where the two molecules in the asymmetric unit have different conformations of the methyl group.

The fact that only one polymorph was found in the experimental screens, and no anhydrous crystal structures could be found for guanine and adenine is disappointing despite predictions of

a range of polymorphs. Just taking into account the lattice energy of the system gives many crystal structures within the energy range of polymorphism, and being able to accurately model some aspects of kinetics will increase the success of the predictions. This is emphasised for thymine in which there is a shallow global lattice energy minimum with many crystal structures of very similar energy. In the case of adenine and guanine one might make educated guesses regarding the molecular packing and hydrogen bonding present in the anhydrous crystal structure of these purines from the computational predictions, however it is still extremely difficult to identify one most probable structure.

This type of joint computational and experimental study helps in clarifying the challenges faced in developing crystal structure prediction, most critically in establishing whether there are readily obtainable polymorphs. The sensitivity of the computational predictions to slight changes in molecular conformation and to the errors in the intermolecular potential are key challenges which need to be addressed before aspects of kinetics can be incorporated into the predictive model.

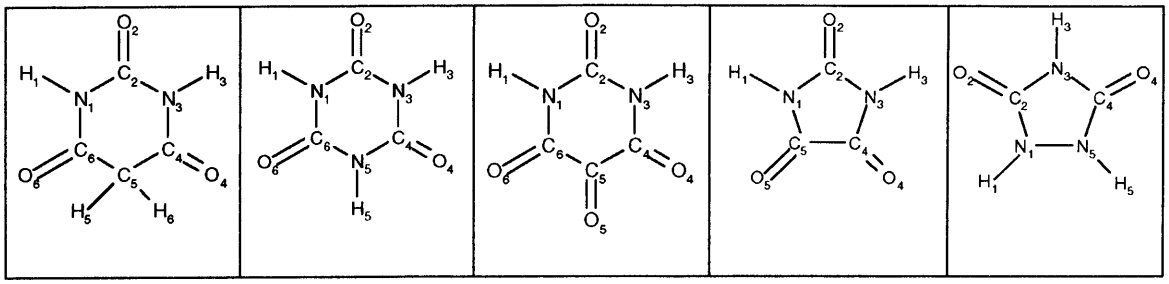
5. Investigating Unused Hydrogen Bond Acceptors Using Known and Hypothetical Crystal Polymorphism

5.1 Introduction

Hydrogen bonding is an important feature in many crystal structures and in biological systems, for example showing importance in the crystal structure of ice[255] and in the structure of DNA[256;257]. As this type of intermolecular interaction is present in a vast number of crystal structures, it is important to understand how the molecular configuration, including positions of hydrogen bond donors and acceptors, can be used to pack the molecules effectively and efficiently within the crystalline environment. It has been commented that - *'It is nearly axiomatic that a molecule with good hydrogen bond functionalities will use them when it packs in crystals'*[258]. Molecules with N-H and C=O functional groups form strong hydrogen bonds in crystals, with the classification strong implying that the hydrogen bond is dominated by the electrostatic interaction and the N···O separation is less than the sum of the van der Waals radii. In an attempt to understand hydrogen bonds as design elements in organic chemistry, general rules were published[166], in which it is clear that C=O and N-H functional groups are expected to participate. The first rule, that 'all good proton donors and acceptors are used in hydrogen bonding', is almost always reliable even though some exceptions are known[48;259]. The third rule that 'the best proton donors and acceptors remaining after intermolecular hydrogen bond formation form intermolecular hydrogen bonds to each other' implies the need to rank relative hydrogen bond strengths[260], and seems to mean that the lack of hydrogen bonds to a given donor or acceptor implies it is a poorer donor/acceptor. The hydrogen bond length and hence whether some acceptors are unused is defined by the length criteria, outlined in section 3.1.8. This study explores how computational modelling, including computational crystal structure prediction[261], can complement understanding of crystal structures with unused hydrogen bond acceptors. Computing the low energy crystal structures of a molecule can reveal whether there are alternative packings using different hydrogen bond donor and acceptor combinations that are

competitive in energy, in addition to establishing which polymorphs can be readily found in a simple experimental manual screen.

The molecules that will be studied are structurally similar to barbituric acid, the parent barbiturate. These molecules contain N-H and C=O ring systems, scheme 5.1, with multiple hydrogen bond donors and acceptors. This will provide an insight into a wide range of N-H···O hydrogen bonded systems.



Barbituric acid cyanuric acid alloxan parabanic acid urazole

Scheme 5.1 The five heterocyclic compounds studied in this chapter

Urazole is unique amongst the molecules in scheme 5.1, being the only one not to have examples of unused acceptors according to the depositions in the Cambridge Structural Database[114]. Cyanuric acid uses all its hydrogen bond acceptors in the anhydrous crystal structure, nevertheless in certain solvates and adducts (Table 5.20 SI) some acceptors are unused. Prior to this research there was only one anhydrous crystal structure known for each on these compounds. A low temperature redetermination of barbituric acid Form *i* was performed, described in section 5.5.3, and used in the computational modelling in this chapter. A new polymorph of barbituric acid was found in this research, denoted Form *ii*, described in section 5.5.4.

5.2 Solid state vs *ab initio* molecular structure

It is important to understand whether the *ab initio* conformation of the isolated molecule is significantly different from the solid state molecular structure for these five compounds. Any intramolecular differences could have an effect on the molecular packing, and could give an indication of any limited molecular flexibility present.

The solid state molecular structures of barbituric acid (Form *i*)[63], cyanuric acid[262], alloxan[73], urazole[263] and parabanic acid[72] were optimised using a MP2/6-31G** wave function using Gaussian98[116], with the comparisons shown in Figure 5.1, and the molecular parameters shown in Table 5.1 SI.

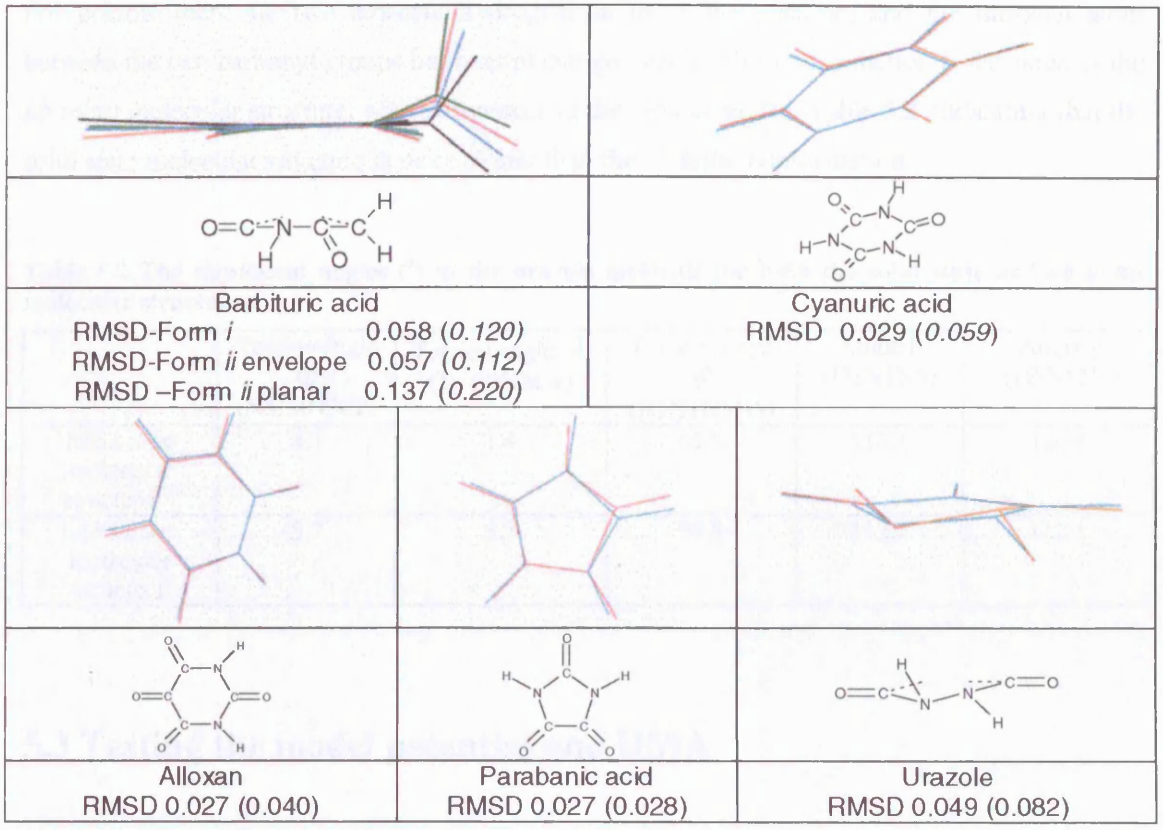


Figure 5.1 The *ab initio* (red) and solid state (blue) molecular structures of the five heterocyclic compounds qualitatively superimposed.

For barbituric acid the two conformations found in Form *ii* (section 5.5.4) are shown in green and black, in addition to the Form *i* conformation in blue. These qualitative overlays are supplemented by the smallest root mean square difference in the experimental and *ab initio* atomic co-ordinates, evaluated for the non-hydrogenic distance in Å, and in parenthesis, for all atoms.

These results show that for cyanuric acid, alloxan and parabanic acid, the solid state molecular structures compare well with the *ab initio* conformation, with parabanic acid becoming symmetric after optimization. However for the molecular structure corresponding to barbituric acid Form *i* there are significant differences in the deviations of the planarity of the ring, with the significant angles shown in Table 5.1. There is a 7 ° difference between the solid state and *ab initio* envelope conformation, which could suggest some aspect of limited molecular flexibility.

Table 5.1 The significant angles (°) in the barbituric acid molecule for both the solid state Form *i* and *ab initio* molecular structures.

	Angle between C4C5C6 and the mean C4N3C2N1C6 planes[a]	Angle α (H5C5H6)
Solid State molecular structure[a]	14.1	107.0
Ab initio molecular structure	21.1	107.1

[a]Low temperature single crystal data of Form *i*, section 5.5.3

For urazole there are two adjacent hydrogens in the *trans* position, and the nitrogen atom between the two carbonyl groups has near planar geometry. This conformation is the same as the *ab initio* molecular structure, with differences in the torsion angles, Table 5.2, indicating that the solid state molecular structure is more planar than the *ab initio* conformation.

Table 5.2 The significant angles (°) in the urazole molecule for both the solid state and *ab initio* molecular structures.

	Torsion angle Φ (N5C4N3C2)	Torsion angle λ (N1C2N3C4)	Torsion angle φ (H1N1N5H5)	Angle β (H1N1N5)	Angle γ (H5N5N1)
Solid state molecular structure[263]	4.7	1.4	65.5	116.1	120.4
Ab initio molecular structure	5.7	5.5	94.8	112.1	112.1

5.3 Testing the model potential and DMA

Two model potentials will be considered for the computational studies on these five heterocyclic compounds, the W99[99] (with non-nuclear hydrogen interaction sites, section 2.4) and FIT[100-102] potentials, along with two qualities of wave function for the DMA, at SCF and MP2 level of theory. It is therefore essential to determine how well these different combinations reproduce the solid state crystal structures after lattice energy minimisation. This will enable the best combination to be chosen for the computational studies for comparability between searches.

The *ab initio* molecular structures, section 5.2, will be used in the lattice energy minimisations. It was found in the previous parabanic acid study[264] that the W99 potential overestimated the carbon repulsion interactions in the crystalline environment and gave an unsatisfactory reproduction of the solid state crystal structure after lattice energy minimisation. This reproduction was improved when the carbon repulsion parameters were decreased by 25%. To

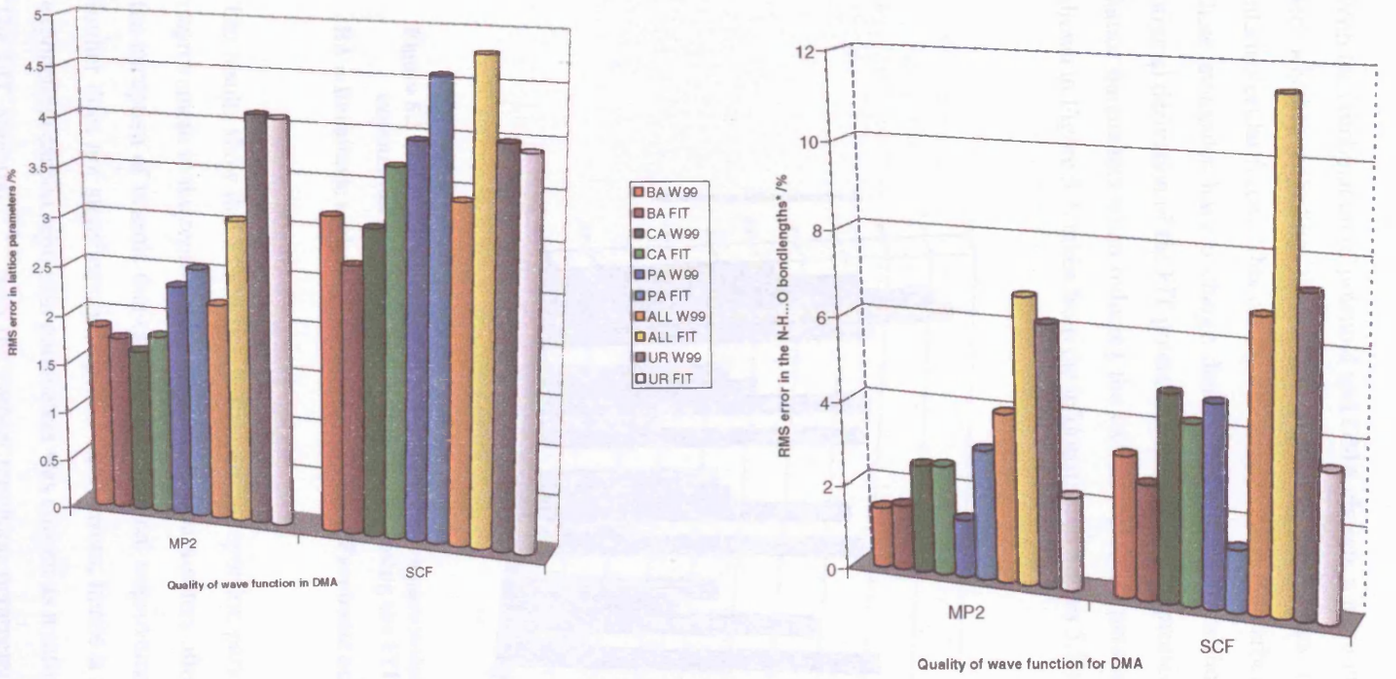
study this range of molecules, the carbon repulsion parameters were also decreased by 25%, 30% and 40% in this section of research. The results are shown for each molecule in Tables 5.2 to 5.6 SI.

5.3.1 Choosing the best model potential and DMA

To compare the results of using the different models for the intermolecular forces, the root mean square (RMS) percentage errors for the lattice parameters and hydrogen bond lengths were calculated, shown in Figure 5.2 (taken from the information in Tables 5.2 to 5.6 SI).

When comparing the lattice parameters, in general the FIT potential gives an inferior reproduction compared to the W99 potential, but in the majority of cases this is only marginal. Using SCF DMA gives an inferior reproduction of the lattice parameters for the majority of the crystal structures than using MP2 DMA, with the worst reproduction for both potentials for urazole, with approximately 4 % RMS errors in the lattice parameters when using MP2 DMA. However when using the SCF DMA, the results are more inconsistent with now the worst reproduction being for alloxan, when using the FIT potential. When comparing the hydrogen bond lengths (or (C=O)···(C=O) interactions for alloxan) in general the FIT potential gives slightly better reproduction of these interactions than the W99 potential, however it does depend on the quality of the wave function for the DMA and the molecule studied. Using SCF DMA gives an inferior reproduction of these interactions than using MP2, with the worst reproduction found for alloxan when using both qualities of wavefunction for the DMA and using the FIT potential. Nevertheless it should also be noted that using this potential with both MP2 and SCF DMA for urazole gave a much better reproduction of these interactions than using the W99 potential.

Both potentials reproduce the lattice parameters for the majority of these crystal structures satisfactory, with the FIT potential giving a slightly better modelling of the hydrogen bonding/carbonyl-carbonyl contacts. Therefore the FIT potential was chosen as the basis for deriving the model potential for the computational studies. In addition the MP2 DMA was chosen as this gives a better reproduction of the electrostatic contribution to the hydrogen bonding for this range of compounds. However it should be noted that both potentials struggled to satisfactory model the intermolecular forces in the crystalline environment for all the compounds.



(a) (b)

Figure 5.2 The RMS percentage errors in the (a) lattice parameters, and the (b) N···O interaction lengths when using both a MP2 and SCF quality wave function for the DMA, for both the FIT and W99 potentials. *For alloxan, the (C=O)···(C=O) interactions are compared instead of the unusually long hydrogen bonds.

BA = Barbituric acid, CA = Cyanuric acid, PA = Parabanic acid, ALL = Alloxan, UR = Urazole

5.3.2 Decreasing the carbon repulsion

With the combination of potential and DMA chosen, a more in depth study can now be made to see whether adjusting the carbon repulsion parameters improves the modelling of the intermolecular forces. This can be rationalised since the carbon atoms in the carbonyl groups in these molecules have a charge density that differs from the oxohydrocarbons[101] used in the original derivation of the FIT potential. These results, expressed as RMS percentage errors in the lattice parameters when reducing the carbon repulsion parameters by 0, 25, 30 and 40 %, are shown in Figure 5.3, taken from the information in Tables 5.2 to 5.6 SI.

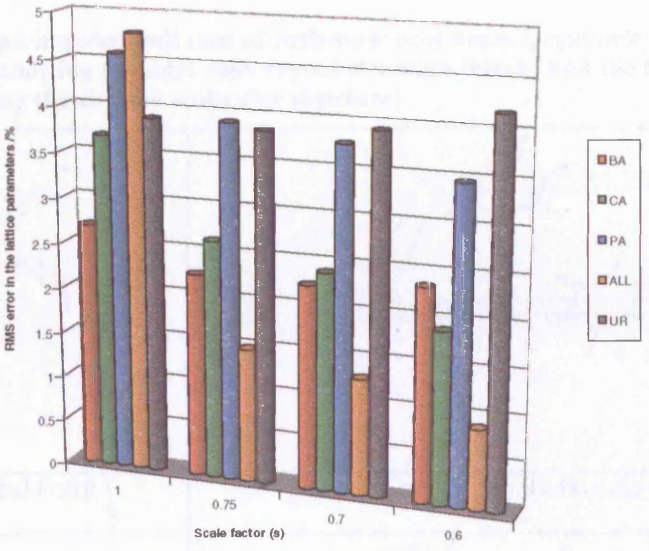


Figure 5.3 The RMS percentage errors in the lattice parameters shown as a function of the pre-exponential repulsion carbon scale factor (using the FIT potential and MP2 DMA)

BA = Barbituric acid, CA = Cyanuric acid, PA = Parabanic acid, ALL = Alloxan, UR = Urazole

The results show that a reduction in the carbon repulsion parameters by 25 % gives a significant improvement in the reproduction of the lattice parameters after lattice energy minimisation, with the exception of urazole that only shows marginal improvement. Reducing the carbon repulsion further does not significantly improve these errors, hence a scaling factor of 0.75 for the pre-exponential carbon repulsion parameters was chosen as a satisfactory compromise.

The FIT potential[100-102] with the carbon repulsion parameters decreased by 25 %, and a MP2 DMA will be used in the computational polymorph predictions. The modified potential

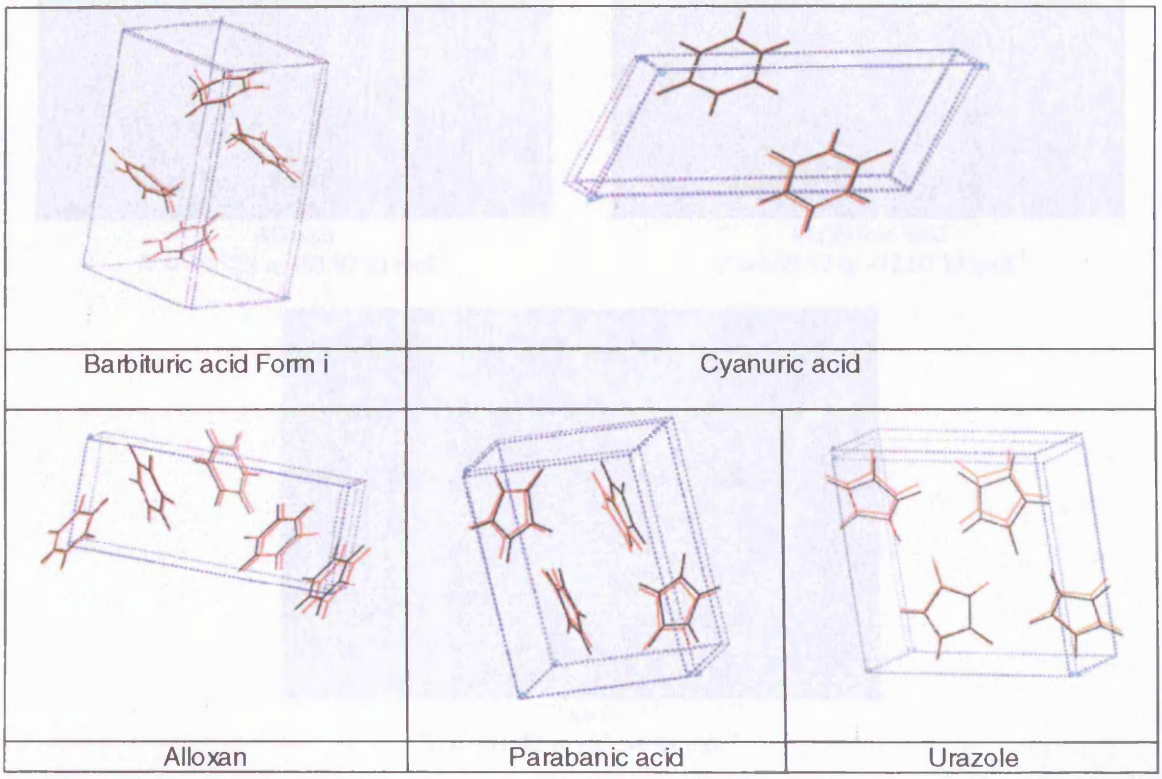
parameters are shown in Table 5.3, with the other parameters as derived in the original FIT potential[100-102].

Table 5.3 The repulsion-disperson carbon parameters of the modified FIT potential

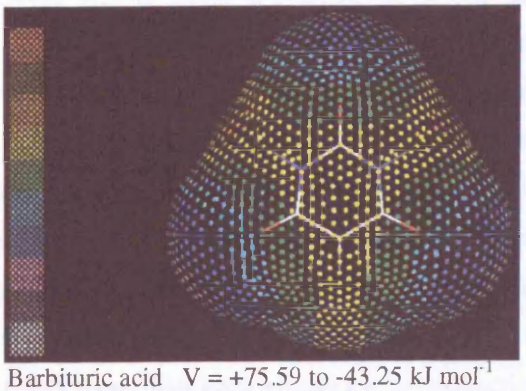
Interaction	A/ kJ mol^{-1}	B/ Å^{-1}	C/ kJ mol^{-1} Å^{6}
C-C	277180	3.60	2439
C-H_c	57590	3.67	577
C-H_p	37329	4.13	229
C-N	265553	3.69	1833
C-O	252468	3.78	1655

The solid state and lattice energy minimised unit cells of each molecule are shown in Figure 5.4.

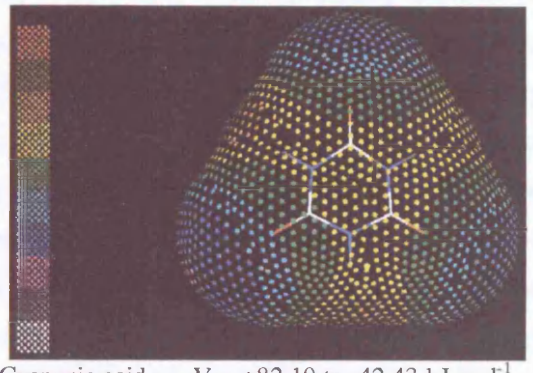
Figure 5.4 The superimposed unit cells of barbituric acid Form *i*, cyanuric acid, alloxan, parabanic acid, and urazole showing the solid state crystal structure (black) and the lattice energy minimised structure (red, using the *ab initio* molecular structure)



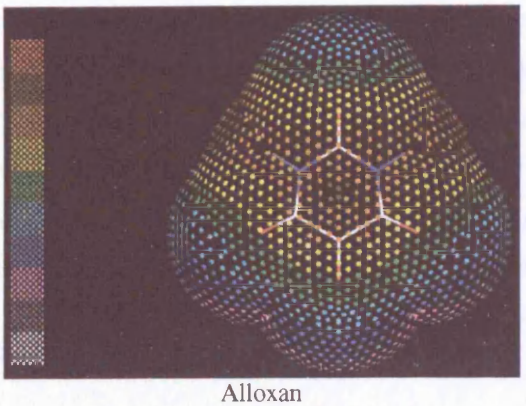
5.4 Electrostatic potentials



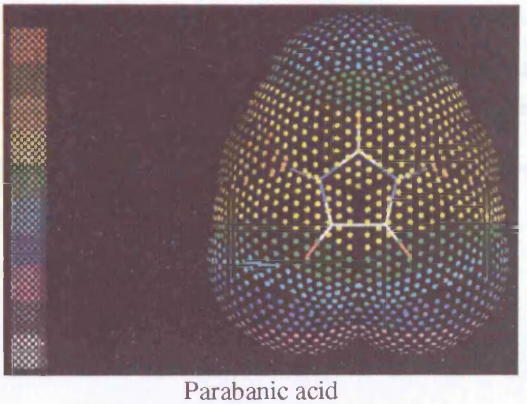
Barbituric acid V = +75.59 to -43.25 kJ mol^{-1}



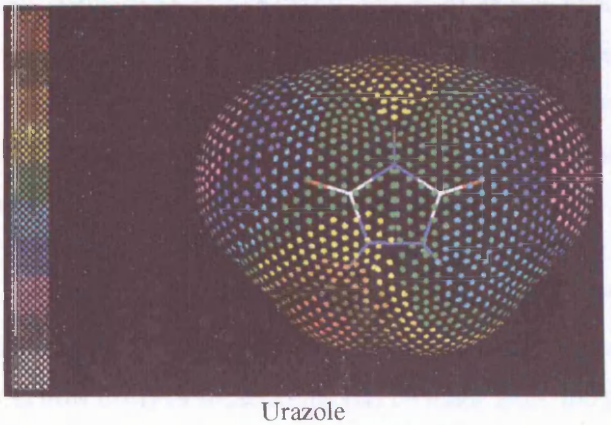
Cyanuric acid V = +82.19 to -42.43 kJ mol^{-1}



Alloxan
V = +96.23 to -63.87 kJ mol^{-1}



Parabanic acid
V = +89.92 to -72.67 kJ mol^{-1}



Urazole
V = +91.42 to -55.34 kJ mol^{-1}

Figure 5.5 The electrostatic potential V (kJ mol^{-1}), as calculated on the water accessible surface from DMAs derived from the MP2/6-31G wave functions for the *ab initio* molecular structures of barbituric acid, cyanuric acid, alloxan, parabanic acid and urazole. Colour coded: white < -80 < grey < -60 < magenta < -40 < blue < -20 < cyan < 0 < green < 20 < yellow < 40 < orange < 60 < brown < 80 < red.**

The electrostatic potentials on the water accessible surface were calculated for each molecule, Figure 5.5, to allow a comparison of the intrinsic contribution to the dominant electrostatic contribution to the hydrogen bonding energy. The electrostatic potential around the hydrogen bond donors and acceptors in this set of heterocyclic molecules show some variations. The long range nature of the electrostatic potential ensures that when two donors are adjacent, as in urazole, the potential maxima are larger and the potential is high between the two N-H bonds. Similarly adjacent carbonyl groups reinforce the potential giving the bonded carbonyls in parabanic acid a more negative potential than the unique carbonyl, and the most negative potential around the central carbonyl (O5) in alloxan. These results suggest that the long range effects more than compensate for any reduction in the C=O or N-H bond polarity caused by the competition for electron density when the functional groups are adjacent. Nevertheless, the close proximity of hydrogen bonding groups in these molecules probably does weaken their hydrogen bonds somewhat relative to more typical amide N-H and C=O groups. The corresponding[265] potential maxima for formamide and N-acetyl alanine N'methylamide (with α helix torsion angles) are 102 and 120 kJ mol^{-1}, and the minima -86 and -142 kJ mol^{-1} respectively, indicating a stronger electrostatic contribution to their hydrogen bonds.

5.5 Barbituric acid

Barbituric acid was first synthesised in 1864[266;267] and is itself not pharmacologically active, however by substituting groups attached to C5, scheme 5.1, the molecule becomes active. The barbiturates can be used as hyponotics, sedatives, anticonvulsants and anaesthetics, with examples being barbital[268] and phenobarbital[269;270]. Table 5.7 SI shows a selection of derivatives of barbituric acid which have either one or two carbonyl groups that do not participate in hydrogen bonding in the crystalline environment. It has been found that these molecules are *"notorious for their polymorphism"*[271], with polymorphs known for barbital[272], amobarbital[273] and thiopental[274]. Phenobarbital[275] is reported to have thirteen polymorphs[276], but there is considerable doubt about how many of these forms can be made from the pure compound[277].
In the known crystal structure[278] prior to these investigations (denoted Form *i*) the molecular structure uses both distinct hydrogen bond acceptors (i.e. one of C4O4 or C6O6 is unused), therefore the Form *i* structure only two carbonyl groups participate in hydrogen bonding. There is also a dihydrate[279;280] and a (1,4)-dioxane solvate[281] crystal structure.

5.5.1 Solubility

A solubility screen was performed using the method outlined in section 3.2.1, with the results shown in Table 5.4.

Table 5.4 The solubility of barbituric acid in various solvents

Observation	Solvents
Soluble	H_2O, methanol, ethanol, acetone, acetonitrile
Partially Soluble	Ethyl acetate
Insoluble	Chloroform, dichloromethane, nitromethane, toluene, diethyl ether

The solvents for which barbituric acid was deemed to be soluble and partially soluble were selected as being suitable for the crystallisation studies, described in Table 5.9 SI.

5.5.2 Experimental results

The results of the experimental polymorph screen are shown in Table 5.9 SI. The experimental screen gave Form *i* described in section 5.5.3, and also the dihydrate[279;280] crystal structure. Two new crystal structures were also found, a new polymorph denoted Form *ii*, described in section 5.5.4, and a reaction product, 5-isopropylidene-barbituric acid, described in section 5.13 SI.

5.5.3 Crystallisation of Form *i*

The crystal structure of Form *i* was originally determined in 1963[278], which at the time was refined to a high R value of 10.2 %. The single crystals used for the X-ray studies in this thesis were grown from a saturated solution of ethanol at 0 ° C in a sealed environment for several days. The low temperature determination gave superior crystallographic data for the computational modelling studies. The thermal ellipsoid plot is shown in Figure 5.6. Comparisons of the unit cell and metric parameters are shown in Tables 5.10 and 5.11 SI, along with the crystallographic data shown in Table 5.15 SI.

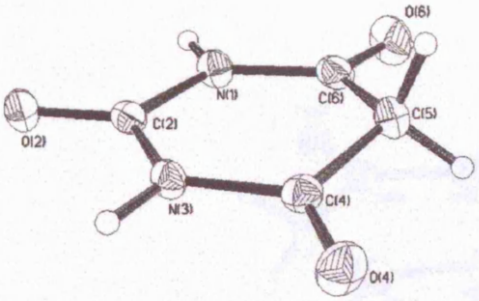


Figure 5.6 The thermal ellipsoid plot of barbituric acid Form *i*, showing 50 % probability level

The molecular structure of barbituric acid in Form *i* is an envelope conformation as previously reported[278], with the angle between the mean C(4)C(5)C(6) and C(4)N(3)C(2)N(1)C(6) planes, Figure 5.6, being 14.1 °. This is smaller than in the *ab initio* molecular structure, section 5.2, which could suggest some limited molecular flexibility present. The packing consists of infinite ribbons of dimers with both distinct hydrogen bond acceptors used, Figure 5.7.

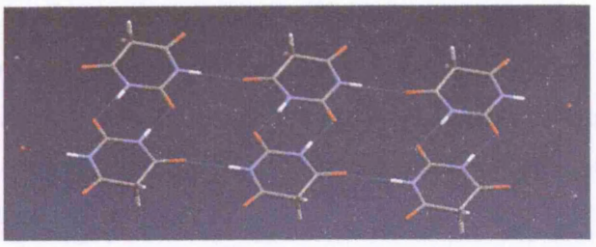


Figure 5.7 The hydrogen bonded infinite ribbons two molecules wide, present in Form *i* of barbituric acid.

In the low temperature determination of Form *i* the perpendicular distance between parallel ribbons is 2.97 Å, compared to 3.03 Å reported previously at room temperature[278]. Data derived from this new refinement will be used in the computational polymorph search, section 5.5.7.

5.5.4 Crystallisation of Form *ii*

A new polymorph, denoted Form *ii*, was crystallised from slow evaporation of both methanol and acetonitrile solutions forming plate like crystals. A superior single crystal for the X-ray diffraction studies was obtained from an acetonitrile solution. There was also concomitant polymorphism observed from this solution, with rapid evaporation giving Form *i* crystals in addition to Form *ii* from slower evaporation. This suggests that Form *i* is the kinetic structure whilst Form *ii* is the thermodynamic form. The thermal ellipsoid plot of Form *ii* is shown Figure 5.8, with the crystallographic data shown in Table 5.16 SI.

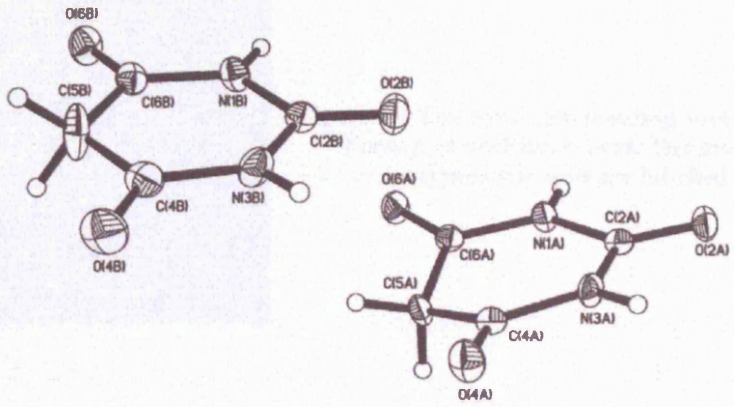


Figure 5.8 The thermal ellipsoid plot of the molecules in the asymmetric unit of Form *ii* of barbituric acid, showing the planar (A) and the envelope (B) conformations of the molecules.

The new polymorph crystallises in the space group $P2_1/c$. There are two crystallographically independent molecules in the asymmetric unit, in which one adopts an envelope conformation, like Form *i*, and the other refines as planar (Figure 5.8 molecule A). The metric parameters are shown in Table 5.12 SI.

The comparison between the molecular envelope conformation in Form *ii* and the molecule in Form *i* shows that the Form *ii* conformation has a greater deviation from planarity, with the angle between C(4)C(5)C(6) and the mean C(4)N(3)C(2)N(1)C(6) planes being 22.4 °, which is closer to the *ab initio* conformation, section 5.2. The thermal ellipsoid for C(5B), associated with the planar conformation, is more elongated above and below the plane than the others in the ring structure. This could be due to a static or dynamic disorder. However the C(5B)-C(6B) and C(4B)-C(5B) bond lengths (1.485(3) Å and 1.489(3) Å respectively) for molecule A do not show any signs of foreshortening, indicating that a static disorder is present.

Form *ii* consists of hydrogen bonded sheets. These sheets consist of a chain of A molecules and a chain of B molecules parallel to each other, with the two chains hydrogen bonded together, Figure 5.9. One of the distinct hydrogen bond acceptors does not participate in hydrogen bonding, which is a significant difference compared to Form *i*.

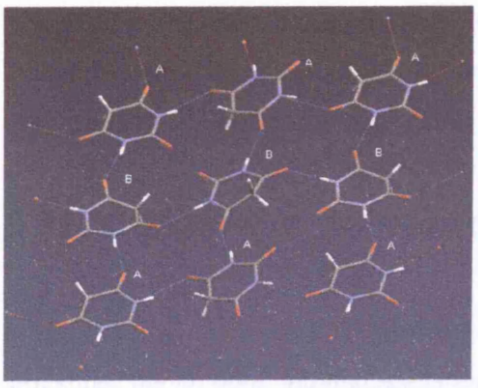


Figure 5.9 The hydrogen bonding sheet motif present in Form *ii* of barbituric acid. The molecules in the asymmetric unit are labelled A or B.

5.5.5 Conformational *ab initio* analysis

With the molecules in Form *ii* exhibiting some apparent limited flexibility, section 5.5.4, it is important to clarify whether the molecule is less rigid than first anticipated, as this will undoubtedly affect the validity of the computational polymorph search. An *ab initio* conformational analysis was performed on the barbituric acid molecule using Gaussian98[116] with a MP2/6-31G** wave function, which involved keeping the torsion angle C2N3C4C5 constant in each optimisation (from 0 – 17 °, in 1 ° steps) and allowing full relaxation of the rest of the molecule. This was to ascertain whether there could be any local minima on the potential energy surface which could indicate other energetically plausible conformations. The MP2 energy profile is shown in Figure 5.10.

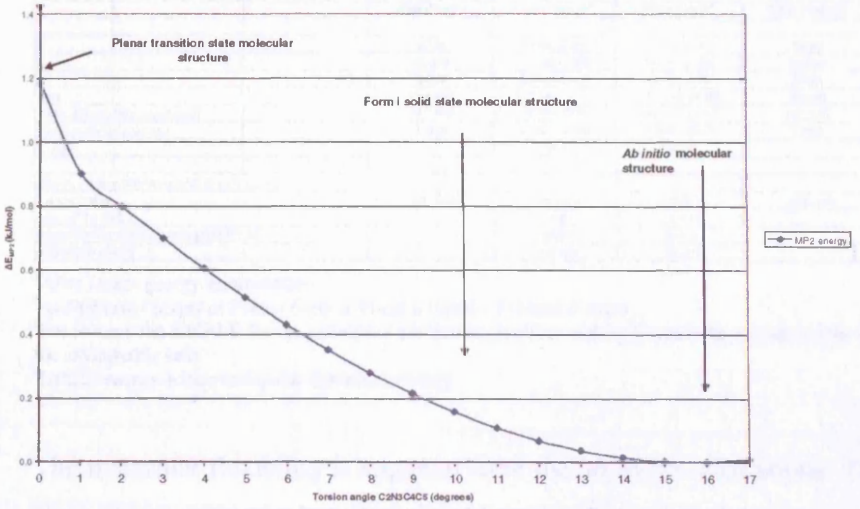


Figure 5.10 The MP2 energy torsion profile for barbituric acid, showing ΔE_{MP2} compared to the *ab initio* molecular structure

The MP2 energy scan found no local minima, with the planar conformation at a low energy transition state. There is an energy difference of around 1.1 kJ mol^{-1} between the planar and *ab initio* conformations, which suggests that the planar molecule is energetically plausible.

5.5.6 Modelling of the known crystal structures

Since there are a variety of envelope conformations found in the two anhydrous crystal structures, the *ab initio* intramolecular energies of these different conformations were compared and contrasted, Table 5.5. Both Form *i* and Form *ii*, using the solid state and *ab initio* conformations, were lattice energy minimised to determine how sensitive the computational methods are to any apparent molecular flexibility. These results are shown in Table 5.6.

Table 5.5 Intramolecular modelling of the known conformations of barbituric acid in Form *i* and Form *ii*

	Form *i* expt (env)	Form *ii* expt env	Form *ii* expt plan	Form *ii* opt plan	Opt Env
[a]Angle between mean planes	14.094	22.627	1.244	0.044	21.107
C6C5C4 and C6N1C2N3C4					
[a]Dihedral angle C2N1C6C5	10.791	10.423	0.53	0.00	15.62
[a]Dihedral angle C2N3C4C5	7.535	9.346	2.53	0.00	15.75
Molecular structure (intramolecular forces)		kJ/mol			
ab initio estimation of conformational energy difference (MP2/6-31G**)					
MP2 E	-1283154.099	-1283153.117	-1283142.864	-1283162.603	-1283163.788
MP2 ΔE^{b}	9.689	10.670	20.924	1.185	0

[a]Atomic labelling shown in scheme 1
[b]Relative *ab initio* energies calculated at MP2/6-31G** level using Gaussian98[116]. The experimental structures have been adjusted for the systematic foreshortening of bonds to hydrogen in X-ray structures.

Table 5.6 Comparison of the lattice energy minimisations of Form *i* and Form *ii* of barbituric acid

	Expt Form *i*	Form *i (expt)*[a]	Form *i (opt)*[a]	Expt Form *ii*	Form *ii (expt)*[a]	Form *ii* (opt and planar)[a]
a	**6.731**	6.935 (*3.03*)	7.024 (*4.35*)	**8.083**	8.019 (*-0.79*)	8.142 (*0.73*)
b	**14.029**	14.223 (*1.38*)	14.148 (*0.85*)	**12.583**	12.479 (*-0.83*)	12.481 (*-0.81*)
c	**6.231**	6.019 (*-3.40*)	6.204 (*-0.43*)	**9.764**	9.998 (*2.40*)	10.080 (*3.24*)
beta	**116.368**	116.446 (*0.07*)	117.125 (*0.65*)	**96.150**	95.64 (*-0.53*)	95.302 (*-0.88*)
Cell Volume (cubic angstroms)	**527.125**	537.919 (*2.05*)	548.757 (*4.10*)	**987.356**	985.770 (*0.85*)	1019.970 (*3.30*)
Cell Density (g/cubic cm)	**1.614**	1.582 (*-1.98*)	1.550 (*-3.97*)	**1.723**	1.709 (*-0.82*)	1.668 (*3.19*)
F value		20.1	26.7		14.9	30.0
Crystal packing (intermolecular forces) kJ/mol						
Lattice Energy	**-101.725**	-103.672	-98.815	**-109.148**	-110.704	-102.905
$\Delta E_{lattice}$[b] (at 0 K)		7.032	11.889		0.000	7.799
Intramolecular distortion energy (ΔMP2 E)[c]		9.689	0.000		15.797	0.593
0K relative energy[d]		-93.983	-98.815		-94.907	-102.312

[a]After lattice energy minimisation
[b]Δ=E[(Form *i* (expt) or Form *i* (opt) or Form *ii* (opt)] – E(Form *ii* expt)
[c]For Form *i*, the ΔMP2 E for the associated molecular structure and for Form *ii* the average of the two ΔMP2 E for both molecules in the asymmetric unit
[d]Lattice energy + intermolecular distortion energy

The molecular flexibility is apparent from the *ab initio* calculations, Table 5.5, with a transition state planar molecule less than 1.2 kJ mol^{-1} above the *ab initio* envelope conformation. This

small energy barrier suggests that the various experimentally observed conformations in the crystalline environment have virtually the same internal energy, to within a few kJ mol^{-1}. Therefore it seems that barbituric acid has the ability to change its envelope conformation considerably due to packing forces in the crystalline environment. This is evident in the dihydrate crystal structure, whereby a recent variable-temperature study found that at 100 K the envelope angle in the barbituric acid molecule is around 6 °, whilst at 270 K it is planar[282].

All lattice energy minimisations satisfactory reproduce the Form *i* and Form *ii* crystal structures, Table 5.6. In Form *ii* there are two molecules in the asymmetric unit with two different conformations, hence this satisfactory minimisation does suggest that molecule A, Figure 5.8, could be planar and flexing. The calculated lattice energies show that Form *ii* is more stable than Form *i*, by a margin of between 4 and 10 kJ mol^{-1} depending on the molecular model. Since the intramolecular energy loss for Form *ii* for having half its molecules planar appears to be less than 1 kJ mol^{-1} from the *ab initio* estimates, the total 0 K energy still favours Form *ii*. Since the planar conformation of half the molecules in Form *ii* is producing a more stable crystal lattice, the molecule that refines as planar could actually be planar.

5.5.7 Computational polymorph searches

In the computational search using the *ab initio* molecular structure, of the 1500 structures that were lattice energy minimised, 38 unique crystal structures were found within 7 kJ mol^{-1} of the global lattice energy minimum. To see whether the search is sensitive to the assumed molecular structure, a further search was performed using the low temperature solid state Form *i* molecular structure, in which there were 49 unique crystal structures within this energy range. The results are shown in Figure 5.12 and Table 5.7.

In the *ab initio* molecular structure computational search, ExptMinOpt was not found in the predictions, with many other crystal structures present. In the computational search using the Form *i* solid state molecular structure, ExptMinExpt was again not found in the predictions. The energy gap between the observed Form *i* and the global lattice energy minimum reduced to 2.8 kJ mol^{-1}, with a variety of more energetically stable crystal structures present. By comparing the room temperature estimates the energy gap between Form *i* and the global lattice energy minimum is reduced to about 4 kJ mol^{-1} and 1 kJ mol^{-1} for the *ab initio* and solid state molecular structures respectively. Thus the 7 ° difference in the envelope angle has a significant effect on the predictions.

One reason why the solid state crystal structure was not found in the predictions is that Form *i* is significantly less dense than the other hypothetical structures (1.61 g cm^{-3} compared to 1.71 g cm^{-3} for the structure at the global lattice energy minimum). It is plausible that the search itself had not been able to find this structure simply due to the denser structure criterion used by MOLPAK (section 3.1.5).

Despite these differences in relative energies, there are still considerable similarities between the two computational searches. The global lattice energy minima in the two searches (AM14 and AM16, Table 5.7) are essentially the same crystal structure, differing primarily in the molecular envelope angle. Both searches find low energy structures in which all carbonyl groups are used as hydrogen bond acceptors, structures that use one unique and one non-unique hydrogen bond acceptor (like Form *i*), and structures that use both non-unique acceptors (like Form *ii*). The three lowest energy crystal structures in the *ab initio* molecular structure computational search are shown in Figure 5.11.

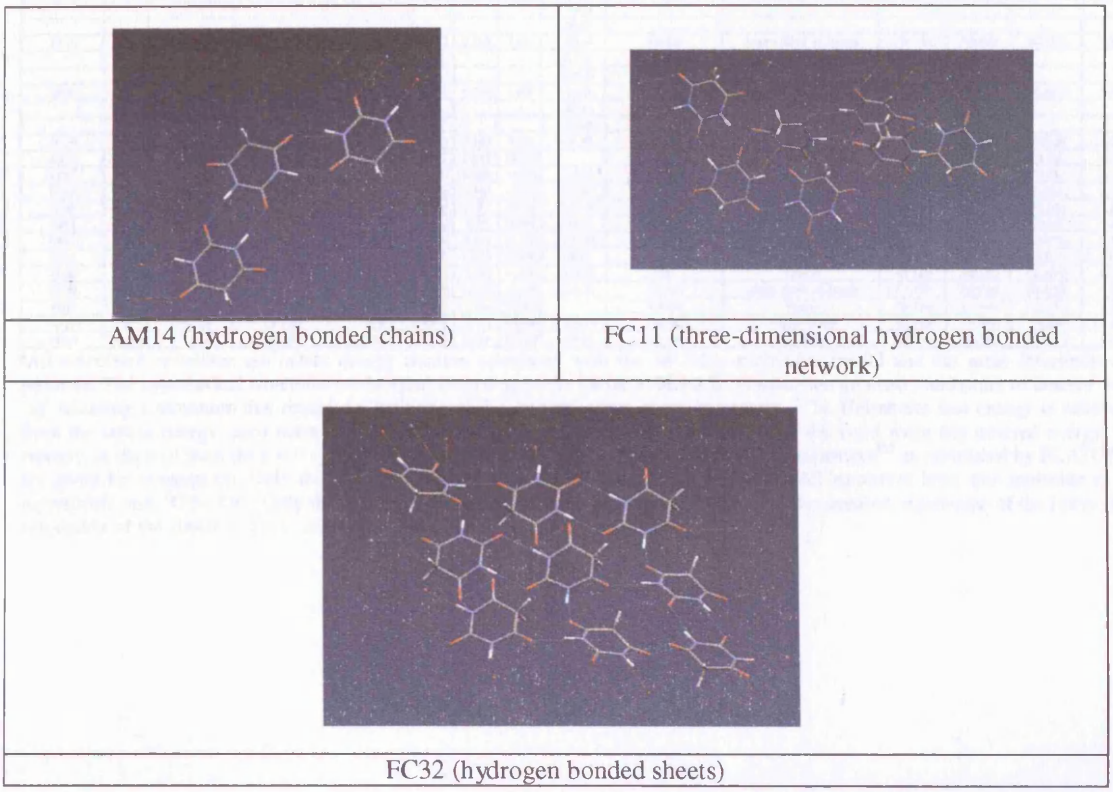
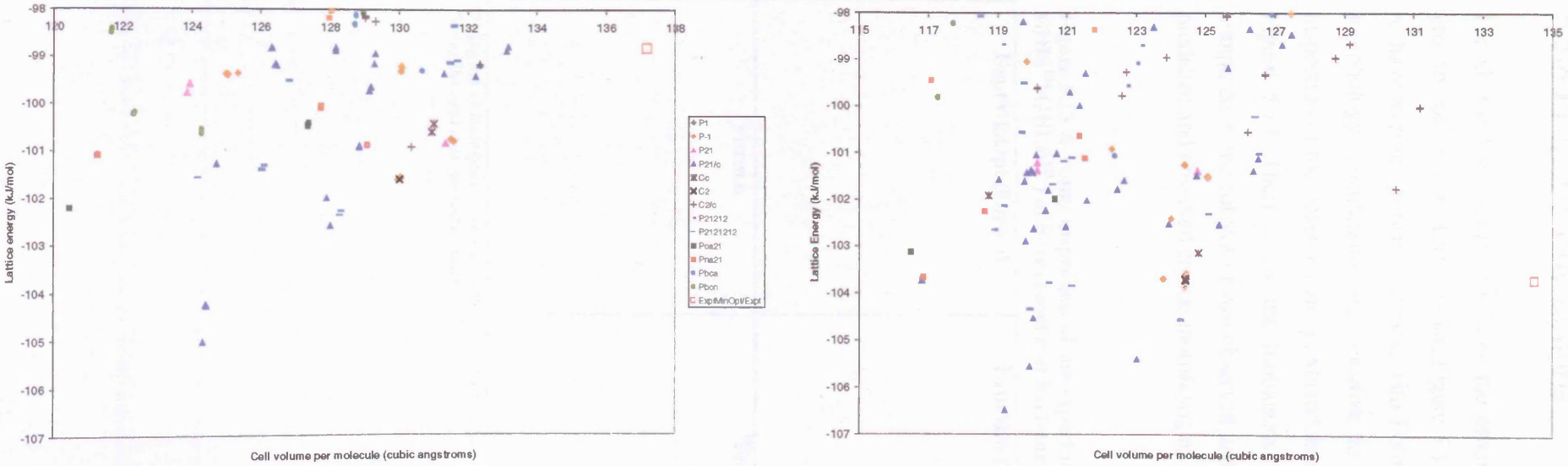


Figure 5.11 The three lowest energy hypothetical crystal structures found in the computational polymorph search using the *ab initio* molecular structure of barbituric acid

Table 5.7 The hypothetical structures for barbituric acid[a] using both the *ab initio* and Form *i* solid state molecular structures within 5 kJ mol^{-1} of the global lattice energy minimum. The full structure lists within an energy range of 7 kJ mol^{-1} are shown in Tables 5.18 and 5.19 SI

Structure	Space	Lattice Energy	[b]Free energy		[c]Reduced Cell parameters				[d]Hydrogen Bond		[e]Graph set			[f]Elastic
	group	/kJ mol^{-1}	at 298 K/kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles/°	acceptors and motif		level 1	Level 2	Level 2	constant
Expt 298K	$P2_1/c$			1.589	6.248	6.691	14.310	γ 116.52	O2,O4	Infinite ribbon of dimers	C1,1(6)	R2,2(8)	R4,4(16)	
Expt 150K	$P2_1/c$			1.614	6.231	6.731	14.029	γ 116.37	O2,O4	Infinite ribbon of dimers	C1,1(6)	R2,2(8)	R4,4(16)	
***Ab initio* molecular structure search**														
ExptMinOpt	$P2_1/c$	-98.815	-113.885	1.550	6.204	6.935	14.148	γ 115.64	O2,O4	Infinite ribbon of dimers	C1,1(6)	R2,2(8)	R4,4(16)	3.30
AM14	$P2_1/c$	-105.027	-117.757	1.711	4.470	9.779	11.517	β 98.97	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(10)	7.75
FC11	$P2_1/c$	-104.277	-116.460	1.709	6.482	8.000	10.141	α 108.83	O2,O4,O6	3D network	R2,2(8)	C1,1(4)	R2,2(8)	6.75
FC32	$P2_1/c$	-102.610	-115.230	1.662	6.485	7.825	10.542	α 106.87	O2,O4,O6	Sheets	R2,2(8)	R2,2(8)	C1,1(4)	1.05
AQ30	$P2_12_12_1$	-102.302	-116.077	1.657	6.371	7.543	10.681		O2,O6	3D network	C1,1(4)	C1,1(4)	C1,1(6)	10.81
AM42	$P2_1/c$	-102.010	-116.092	1.663	6.048	6.955	12.685	γ 106.50	O4,O6	Sheets	C1,1(6)	C1,1(6)	C2,2(8)	2.34
CA25	P-1	-101.635	-114.622	1.636	5.265	5.430	9.276	α 90.11	O2,O6	Infinite ribbon of dimers	R2,2(8)	C1,1(6)	R4,4(16)	3.52
								β 96.57						
								γ 99.23						
AQ28	$P2_12_12_1$	-101.586	-113.769	1.713	5.903	7.000	12.021		O2,O4,O6	3D network	C1,1(6)	C1,1(6)	C1,1(4)	5.11
AQ13	$P2_12_12_1$	-101.433	-115.073	1.687	4.507	10.310	10.852		O2,O6	Jagged Sheet	C1,1(4)	C1,1(4)	C2,2(10)	5.42
FC38	$P2_1/c$	-101.300	-114.049	1.705	4.742	6.234	16.884	β 91.778	O4,O6	Sheets	R2,2(8)	R2,2(8)	C1,1(4)	2.50
BB39[g]	Pnma	-101.244	-115.605	1.754	5.090	7.627	12.494		O2	Chains				1.37
DA39	Cc	-100.605	-111.290	1.624	6.858	7.950	9.610	α 91.50	O4,O6	Sheets	C1,1(6)	C1,1(6)	C2,2(8)	3.52
Form i solid state molecular structure search														
ExptMinExpt	$P2_1/c$	-103.672	-116.030	1.582	6.092	6.897	14.223	γ 115.81	O2,O4	Infinite ribbon of dimers	C1,1(6)	R2,2(8)	R4,4(16)	3.30
AM16	$P2_1/c$	-106.473	-119.120	1.784	4.494	9.629	11.158	β 99.12	O2,O4	Chains	R2,2(8)	C1,1(6)	C2,2(10)	7.22
FC23	$P2_1/c$	-105.549	-117.830	1.774	6.249	7.958	10.277	α 110.21	O2,O4,O6	3D network	R2,2(8)	C1,1(4)	R2,2(8)	6.25
AZ42	$P2_12_12_1$	-104.555	-118.020	1.771	6.733	7.003	10.542		O2,O4	3D network	C1,1(4)	C1,1(4)	C1,1(4)	7.34
AK38	$P2_1/c$	-104.510	-117.080	1.772	6.224	7.913	10.360	α 109.79	O2,O4,O6	3D network	C1,1(4)	R2,2(8)	R2,2(8)	6.26
AQ22	$P2_12_12_1$	-104.343	-116.550	1.774	5.808	6.925	11.927		O2,O4,O6	3D network	C1,1(6)	C1,1(4)	C1,1(6)	5.13
AB17	P-1	-103.839	-116.390	1.710	5.235	5.349	9.070	α 91.26	O2,O4	Infinite ribbon of dimers	C1,1(6)	R2,2(8)	R4,4(16)	2.97
								β 96.00						
								γ 99.68						
AQ32	$P2_12_12_1$	-103.837	-117.170	1.756	4.870	9.850	10.100		O2,O4	Jagged Sheet	C1,1(4)	C1,1(4)	C2,2(10)	4.55
AQ27	$P2_12_12_1$	-103.775	-117.140	1.765	4.588	10.215	10.284		O2,O6	Jagged Sheet	C1,1(4)	C1,1(4)	C2,2(10)	4.87
AK31	$P2_1/c$	-103.735	-117.410	1.821	5.053	7.566	12.223	α 90.01	O2	Chains	R2,2(8)	R2,2(8)	C2,2(8)	1.34
								β 105.3						
								γ 90.37						
AB10	P-1	-103.687	-117.080	1.718	5.277	6.323	7.914	α 93.44	O2,O4	Infinite ribbon of dimers	C1,1(6)	R2,2(8)	R4,4(16)	1.88
								β 97.02						
								γ 108.27						
AB40	P-1	-103.574	-116.041	1.710	5.291	5.325	8.998	α 94.61	O2,O6	Infinite ribbon of dimers	R2,2(8)	C1,1(6)	R4,4(16)	3.01
								β 90.19						
								γ 99.98						
DA29	Cc	-103.126	-113.658	1.704	6.773	7.580	9.356	α 91.30	O4,O6	Sheets	C1,1(6)	C1,1(6)	C2,2(8)	2.31
AZ47	$P2_12_12_1$	-102.652	-116.302	1.789	4.907	5.014	19.332		O2,O4	Infinite ribbon of dimers	C1,1(6)	C1,1(4)	C2,2(8)	11.43
FC40	$P2_1/c$	-102.635	-116.429	1.772	4.793	5.029	19.935	α 92.39	O2,O4	Jagged Sheet	C1,1(6)	C1,1(4)	C2,2(8)	7.51
FC18	$P2_1/c$	-102.588	-115.262	1.759	4.884	5.982	16.561	β 90.97	O4,O6	Jagged Sheet	R2,2(8)	C1,1(4)	C2,2(12)	2.84
FC32	$P2_1/c$	-102.528	-115.936	1.69	6.362	7.731	10.618	α 106.18	O2,O4,O6	Sheets	C1,1(4)	R2,2(8)	R2,2(8)	1.47
AK21	$P2_1/c$	-102.512	-116.048	1.716	6.799	8.23	8.868	α 92.49	O4,O6	Sheets	C1,1(6)	C1,1(6)	C2,2(8)	2.01
FA43	$P2_1/c$	-102.23	-115.341	1.767	4.862	6.174	16.706	γ 106.24	O2,O6	Infinite ribbon of dimers	R2,2(8)	C1,1(6)	R4,4(16)	8.57
FA33	$P2_1/c$	-102.015	-114.824	1.75	4.636	6.413	16.385	γ 93.27	O4,O6	Jagged Sheet	C1,1(4)	R2,2(8)	C2,2(12)	3.43
AM40	$P2_1/c$	-101.78	-116.100	1.737	6.209	7.095	12.333	γ 115.65	O4	Chains	C1,1(6)	R2,2(8)	C2,2(10)	3.29
DC32	C2/c	-101.745	-115.780	1.63	6.049	12.294	14.376	α 102.51	O2,O4	Infinite ribbon of dimers	C1,1(6)	R2,2(8)	R4,4(16)	7.34
AQ17	$P2_12_12_1$	-101.681	-116.370	1.764	6.157	7.397	10.592		O4,O6	Sheets	C1,1(4)	C1,1(4)	C2,2(12)	0.98
AK24	$P2_1/c$	-101.618	-114.820	1.776	6.661	7.164	10.255	β 101.74	O2,O6	Jagged Sheet	R2,2(8)	C1,1(6)	R2,2(8)	4.31
AM19	$P2_1/c$	-101.57	-115.020	1.787	4.118	9.810	11.822	γ 94.53	O2,O6	Chains	R2,2(8)	R2,2(8)	C2,2(10)	5.83

[a]All calculated structures are lattice energy minima calculated with the *ab initio* molecular model and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK coordination geometry and order of density, with 'sg' denoting a minimum that required a lowering of the original space group symmetry. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated by PLATON[158] are given for comparison. Only the reduced cell angles which are not 90° are tabulated. All structures have one molecule in the asymmetric unit. [d]O4 = O6. [e]Only the first three levels shown, calculated using RPluto[168]. [f]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.



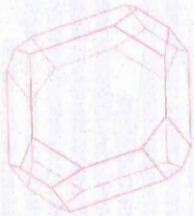
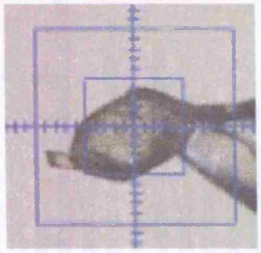
(a) (b)

Figure 5.12 Graphs showing the lattice energy vs cell volume per molecule for the low energy minima found using the (a) *ab initio*, and (b) solid state Form *i* molecular structures of barbituric acid. The ExptMinOpt and ExptMinExpt minima are also shown for comparison

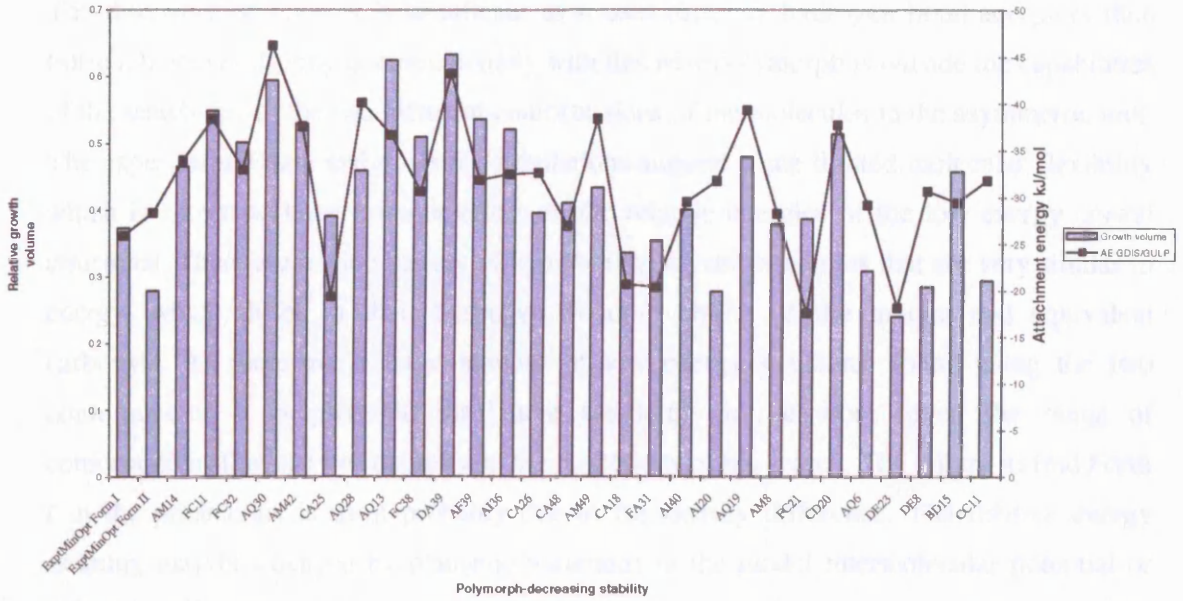
5.5.8 Property calculations

For all the low energy structures the minimum attachment energies and the majority of growth volumes are very similar, Figure 5.14. ExptMinOpt and ExptMinExpt are predicted to have average growth volumes, with Form *ii* having a slightly slower growth rate. Both morphology calculations are sensitive to the assumed molecular structure. A visual inspection of the observed and predicted morphologies for Form *i* and Form *ii* are shown in Figure 5.13. These show the limitations in predicting the vapour grown morphology compared to the solvent grown observed habit, despite some visual similarities between the predicted and observed crystal morphologies of Form *ii*.

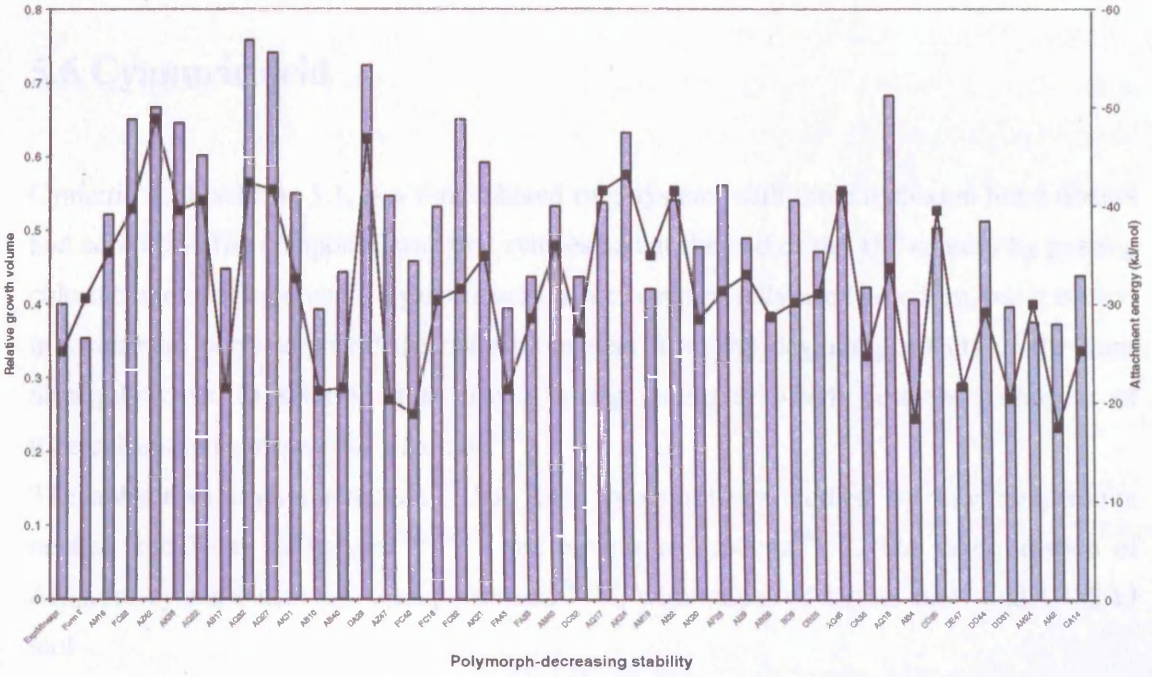
Figure 5.13 A visual inspection of the experimental and calculated crystal morphologies (using GDIS[184]/GULP[186]) of Form *i* and *ii* of barbituric acid

ExptMinOpt (Form *i*)	ExptMinExpt (Form *i*)	Experimental Form *i*, grown from an ethanol solution
		
[a]Form *ii*	[b]Form *ii*	Experimental Form *ii* grown from an acetonitrile solution
		

[a]Using the *ab initio*/planar transition state molecular structures
[b]Using the solid state molecular structures[63]



(a)



(b)

Figure 5.14 The relative growth volumes and minimum attachment energies calculated for the low energy crystal structures found in the (a) *ab initio*, and (b) solid state molecular structure computational search on barbituric acid. Only the crystal structures that have not been symmetry reduced are shown. ExptMinOpt, ExptMinExpt and the Form *ii*[63] (with the associated molecular structures) crystal structures are also shown for comparison.

5.5.9 Conclusions

The discovery of Form *ii* is significant as it uses different hydrogen bond acceptors than Form *i*, however dealing computationally with this new polymorph is outside the capabilities of the search due to the two different conformations of the molecules in the asymmetric unit. The experimental data and *ab initio* calculations suggest some limited molecular flexibility which is shown to have a major effect on the relative energies of the low energy crystal structures. There are a wide variety of hypothetical crystal structures that are very similar in energy, which differ in their hydrogen bonding ability of the unique and equivalent carbonyls. As there are a large number of low energy structures found using the two conformations it is plausible that there are likely to be more given the range of conformations that can be adopted and the limitations of the search. The failure to find Form *i* in the predictions is most probably due to the density difference. The relative energy ordering may be changed by plausible variations in the model intermolecular potential or better modelling of both the intra- and intermolecular contributions to the room temperature energy. Nevertheless experimentally establishing and understanding the polymorphism of barbituric acid will provide a major challenge to computational crystal structure prediction.

5.6 Cyanuric acid

Cyanuric acid, scheme 5.1, is a 6-membered ring system, with three hydrogen bond donors and acceptors. The compound was first synthesised at the end of the 18th century by passing chlorine into melting urea[283]. Cyanuric acid is not commercially used as a drug, but it is used in swimming pools to shield the chlorine present from the degrading effects of the Suns ultraviolet rays. In solution it is also a strong antiseptic which destroys pathogens of tuberculosis and various skin infections[284].

The anhydrous crystal structure[285] has been progressively studied by low temperature neutron and X-ray diffraction[262;286-288] and by infrared spectra[289;290]. The determination of ΔH_{sub} for cyanuric acid has been performed[291;292], with values of 131 kJ mol^{-1} and 139.5 kJ mol^{-1}.

There have been extensive studies on cyanuric acid adducts, shown in Table 5.20 SI. Even though the normal hydrogen bonded layer structure of cyanuric acid seems the most stable[262], other solvates and adduct structures of cyanuric acid are found with molecular tapes and chains in which the majority have unused hydrogen bond acceptors[293].

5.6.1 Solubility

A solubility study was performed with the results shown in Table 5.8.

Table 5.8 The solubility of cyanuric acid in various solvents.

Observation	Solvent
Soluble	DMF, DMSO
Partially Soluble	H_2O, methanol, ethanol, acetone, chloroform, acetonitrile, diethyl ether, ethyl acetate, formaldehyde
Insoluble	Propan-1-ol, dichloromethane, nitromethane, toluene, butan-2-ol

The solvents selected for the crystallisation studies were H_2O, methanol, ethanol, acetone, acetonitrile, propan-1-ol, DMF and DMSO. In addition, experiments were also set up using mixed solvent systems and vapour diffusion, outlined in Table 5.21 SI.

5.6.2 Experimental results

The results of the experimental polymorph screen are shown in Table 5.21 SI. Crystallisation from the majority of solvents gave the known anhydrous crystal structure[262], either as microcrystalline or crystalline material, section 5.6.3. Large block-like crystals were obtained from both acetone and acetonitrile solutions and were found to be cyanuric acid dihydrate, described in section 5.22 SI. In addition the DMF solvate was crystallised from a DMF solution, described in section 5.23 SI.

5.6.3 Crystallisation of anhydrous cyanuric acid

Crystallisation of anhydrous cyanuric acid from ethanol solution yielded crystals of a small block habit, whilst the crystallisation from the other solvents gave long, needle-like crystals. The packing in the crystal consists of hydrogen bonded sheets[286].

5.6.4 *Ab initio* molecular structure computational polymorph search

In the computational search using the *ab initio* molecular structure, of the 1500 structures that were lattice energy minimised, 17 unique structures were found within 6 kJ mol^{-1} of the global lattice energy minimum. This energy cut-off was chosen as increasing this energy range by 1 kJ mol^{-1} gave a substantial increase in the unique crystal structures found in the search. The results are shown in Figure 5.15 and Table 5.9.

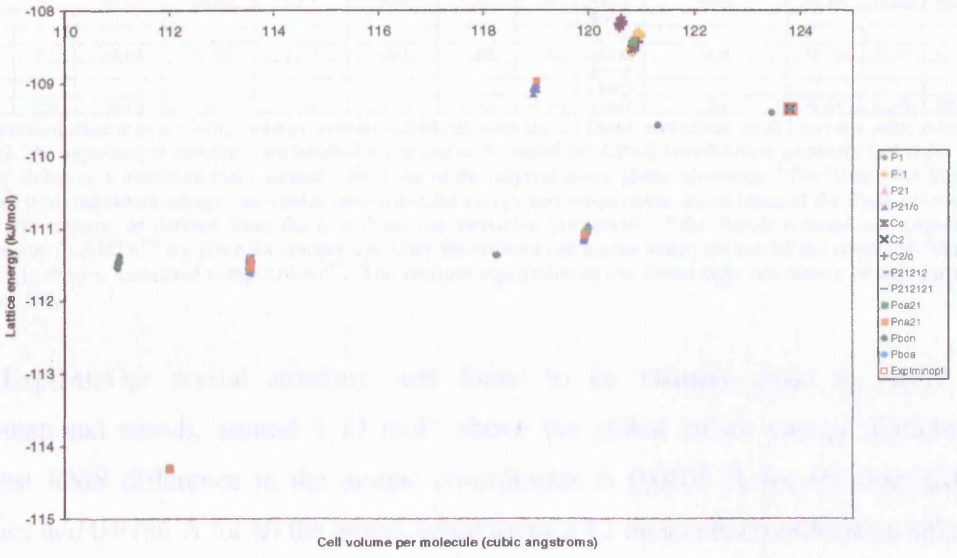


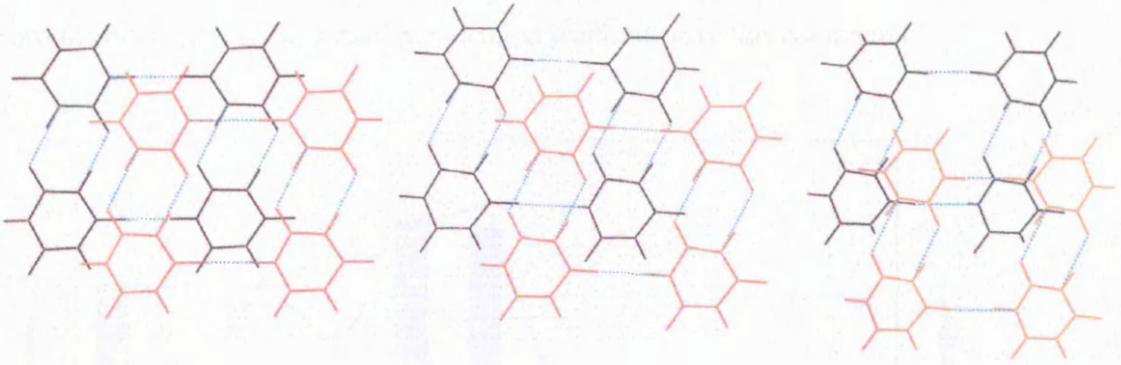
Figure 5.15 Graph showing the lattice energy vs cell volume per molecule for the lattice energy minima found for cyanuric acid in the energy range of potential polymorphs (6 kJ mol^{-1}). ExptMinOpt is also shown for comparison

Table 5.9 The low energy hypothetical crystal structures within 6 kJ mol^{-1} of the global lattice energy minimum of cyanuric acid[a]. All the molecules in the low energy crystal structures use all the hydrogen bond acceptors, with ExptMinOpt shown for comparison

Structure	Space	Lattice Energy	[b]Free energy		[c]Reduced Cell					[d]Graph set			[e]Elastic
	group	/kJ mol^{-1}	at 298 kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles/°	Hydrogen bonding motif	Level 1	Level 2	Level 3	constant
Expt 298K	C2/n			1.778	5.190	5.190	9.065	α 89.46	Sheet	C1,1(6)	R2,2(8)	C1,1(6)	
								β 89.46					
								γ 80.87					
Expt 100K	C2/n			1.819	5.134	5.134	9.032	α 89.92	Sheet	C1,1(6)	R2,2(8)	C1,1(6)	
								β 89.92					
								γ 82.00					
ExptMinOpt	**C2/n**	**-109.360**	**-121.796**	**1.731**	**5.133**	**5.133**	**9.032**	**α 89.91**	**Sheet**	**C1,1(6)**	**R2,2(8)**	**C1,1(6)**	1.12
								β 89.91					
								γ 81.99					
CA38	P-1	-114.317	-125.478	1.913	6.522	6.525	6.526	α 108.06	3D	R2,2(8)	R2,2(8)	R2,2(8)	1.17
								β 106.13					
								γ 108.1					
AZ9	$P2_12_12_1$	-111.642	-123.71	1.888	4.982	4.985	18.288		3D	C1,1(4)	C1,1(4)	C1,1(6)	13.99
CD11	Pbcn	-111.468	-123.565	1.930	7.435	10.499	11.378		3D	C1,1(6)	R2,2(8)	R2,2(8)	1.12
CD25[sg]	$P2_1/c$	-111.316	-124.442	1.812	6.145	6.827	12.187	γ 112.27	3D	C1,1(6)	R2,2(8)	C1,1(6)	1.52
AI38	$P2_1/c$	-111.178	-123.011	1.787	6.543	7.446	10.356	α 108.07	3D	C1,1(4)	C1,1(4)	C1,1(6)	3.96
CD24[sg]	$P2_1/c$	-109.575	-123.872	1.767	6.032	12.160	13.232	α 90.91	3D	C1,1(6)	D1,1(2)	D1,1(2)	0.23
AB47	P-1	-109.387	-121.689	1.731	5.265	5.322	8.943	α 89.28	Sheet	C1,1(6)	R2,2(8)	R2,2(8)	1.71
								β 87.9					
								γ 81.52					
FC23	$P2_1/c$	-109.152	-121.884	1.802	4.880	5.041	19.375	β 93.13	3D	R2,2(8)	C1,1(4)	C1,1(6)	4.36
AB18	P-1	-108.602	-120.869	1.774	5.267	6.596	7.666	α 98.31	Sheet	C1,1(6)	R2,2(8)	R2,2(8)	2.89
								β 98.37					
								γ 110.29					
AB48	P-1	-108.405	-120.969	1.774	5.204	6.569	7.752	α 97.91	Sheet	R2,2(8)	R2,2(8)	C1,1(6)	0.77
								β 98.41					
								γ 109.74					
DE7	C2/c	-108.219	-119.681	1.778	5.018	12.082	16.016	α 96.57	3D	R2,2(8)	C1,1(6)	R2,2(8)	8.42

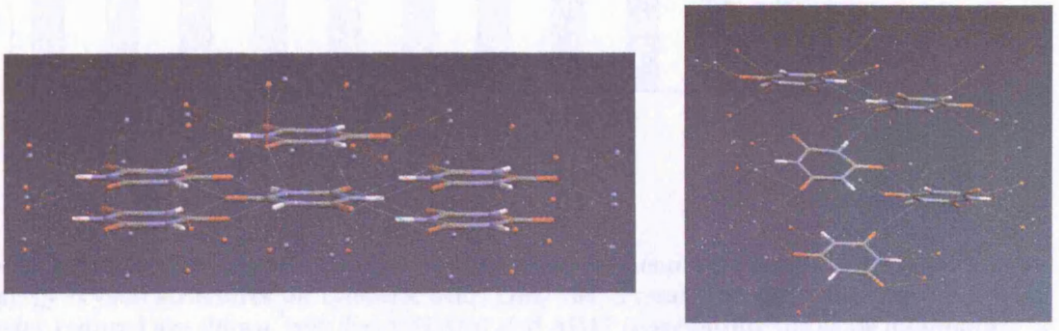
[a]All calculated structures are lattice energy minima calculated with the *ab initio* molecular model and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK coordination geometry and order of density, with 'sg' denoting a minimum that required a lowering of the original space group symmetry. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated by PLATON[158] are given for comparison. Only the reduced cell angles which are not 90° are tabulated. [d]Only the first three levels shown, calculated using RPluto[168]. [e]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa

The ExptMinOpt crystal structure was found to be visually close to AB47 in the computational search, around 5 kJ mol^{-1} above the global lattice energy minimum. The smallest RMS difference in the atomic co-ordinates is 0.0805 Å for the non-hydrogenic distance and 0.0786 Å for all the atoms, when using a 12 molecule co-ordination sphere. The hypothetical crystal structures predicted up to 5 kJ mol^{-1} (4 kJ mol^{-1} at room temperature estimates) more energetically stable are all based on a three-dimensional hydrogen bonded network using all donors and acceptors, in contrast to the sheet structure observed in the solid state. The majority of the crystal structures with the three-dimensional hydrogen bond network are denser than other higher energy structures. All these low energy structures seem plausible, none have extremely low mechanical stability, Table 5.9, and therefore these alternate structures seem thermodynamically and structurally plausible. Within the low energy crystal structures, with the hydrogen bonded sheet motif, there are subtle changes in the orientations of the sheets relative to each other, Figure 5.16. Two examples of the three-dimensional hydrogen bond network present in the low energy structures are shown in Figure 5.17.



(a) (b) (c)

Figure 5.16 The differences in the relative orientations of the hydrogen bonding sheet motif, as shown for (a) ExptMinOpt, (b) AB47, and (c) AB18



(a) (b)

Figure 5.17 Two low energy hypothetical structures in the computational search, showing variations in their three-dimensional hydrogen bond motif, (a) CA38 (molecules tilted towards the *ab* plane), and (b) AI38 (molecules tilted towards the *bc* plane)

5.6.5 Property calculations

The morphology calculations, Figure 5.18, show that the majority of the low energy crystal structures have similar minimum attachment energies, except for CA38, CD11 and ExptMinOpt which have a relatively fast growth of the dominant face. In ExptMinOpt this face is approximately parallel to the hydrogen bonded sheets. ExptMinOpt has the highest growth volume, with CA38 also having a comparable growth rate and comprising a different hydrogen bond motif.

A visual inspection of the predicted morphology of ExptMinOpt and the observed morphologies of cyanuric acid are shown in Figure 5.19. There is some agreement between the predicted morphology with the small block like morphology crystallised from the ethanol solution, even though the calculations are strictly of a vapour grown crystal. Nevertheless

cyanuric acid also crystallises as a needle-like habit from methanol solution, suggesting that solvent effects play an important part in the crystallisation of this compound.

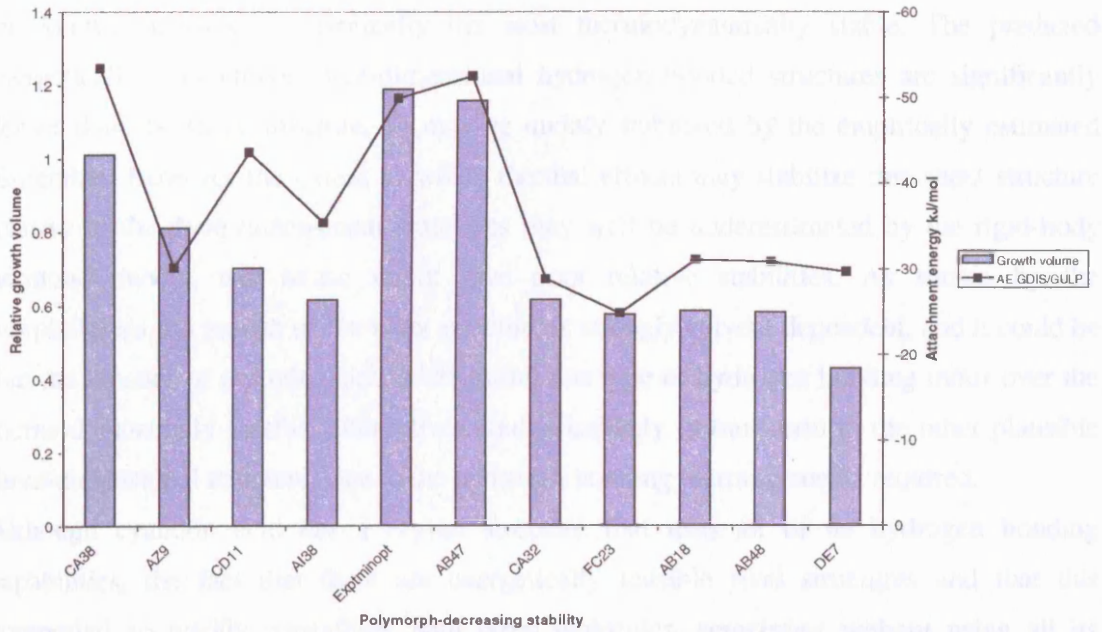
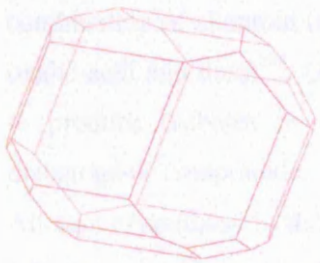


Figure 5.18 The relative growth volumes and minimum attachment energies calculated for the low energy crystal structures on cyanuric acid. Only the crystal structures that have not been symmetry reduced are shown, with ExptMinOpt and AB47 representing the same minimum

Figure 5.19 A visual inspection between the predicted ExptMinOpt and the two observed morphologies of cyanuric acid (not to scale).

[a]ExptMinOpt	Observed morphology from methanol solution (needle-like habit)	Observed morphology from ethanol solution (block-like habit)
		

[a]Calculated using GDIS[184]/GULP[186], using CHelpG charges[187] calculated using Gaussian98[116]

5.6.6 Conclusions

The accuracy of the computational studies does not confidently eliminate the possibility that the known structure is marginally the most thermodynamically stable. The predicted energetically competitive three-dimensional hydrogen bonded structures are significantly denser than the sheet structure, so may be unduly stabilized by the empirically estimated dispersion. However the extent to which thermal effects may stabilize this sheet structure relative to the three-dimensional structures may well be underestimated by the rigid-body harmonic model, and hence might give poor relative stabilities. As shown by the morphologies the growth of the sheet structure is strongly solvent dependent, and it could be that the kinetics of crystallisation is favouring this type of hydrogen bonding motif over the thermodynamically feasible alternatives, and is unlikely to transform to the other plausible three-dimensional structures due to the hydrogen bonding rearrangements required.

Although cyanuric acid has a crystal structure that uses all of its hydrogen bonding capabilities, the fact that there are energetically feasible rival structures and that this compound so readily crystallises with other molecules, sometimes without using all its hydrogen bonding capabilities, demonstrates that the known structure is not particularly kinetically or thermodynamically favourable.

5.7 Alloxan

Alloxan is a 6-membered heterocycle, with four hydrogen bond acceptors and two donors, with three of the acceptors in close proximity, scheme 5.1. Alloxan was first derived from a combination of allantoin (a product of uric acid) and 'oxalsuare' (oxaluric acid derived from oxalic acid and urea)[294]. One remarkable property of alloxan is the ability of the compound to produce diabetes in laboratory animals[295], making it a part of a small group of diabetogenic compounds.

Alloxan crystallises in the space group $P4_12_12$ with four molecules in the unit cell[296], $Z' = 0.5$, Figure 5.20. It has been found that this crystal structure contains no conventional hydrogen bonds[48;296], with the shortest contacts being (C=O)···(C=O) interactions. This is the unusual exception to the general rule that good proton donors and acceptors are used in hydrogen bonding in molecular crystal structures[166]. It is possible that the usual attractive forces between $C^{\delta+} = O^{\delta-}$ dipoles is sufficiently powerful to influence the arrangement of molecules to bring carbonyl atoms closer together than dispersion forces alone would allow[296]. The importance of dipolar interactions between carbonyl groups in stabilising the packing modes of small organic molecules is significant as the contribution of these

interactions is comparable to that of medium strength hydrogen bonds[297]. It was previously found that energetically competitive dimer structures of alloxan were found to contain (C=O)···(C=O) interactions, comparable in energy with hydrogen bonded dimers[48]. In the same study a computational polymorph search was also performed on alloxan[48].

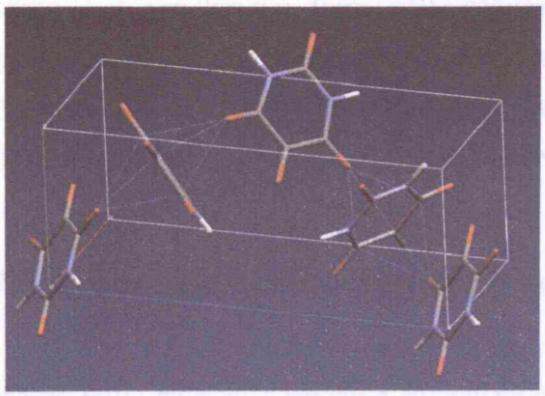


Figure 5.20 The anhydrous crystal structure of alloxan[296], showing the strong (C=O)···(C=O) contacts

5.7.1 Solubility

As anhydrous alloxan is not commercially available, alloxan monohydrate was obtained from Aldrich and heated to approx 200 ° C to acquire the dark pink alloxan, confirmed by powder diffraction, for use in the crystallisation experiments.

A solubility screen was performed with the results shown in Table 5.10.

Table 5.10 The solubility of alloxan in various solvents

Observation	Solvents
Soluble	H_2O, methanol, DMSO, DMF
Partially Soluble	Ethanol, acetone, acetonitrile, propan-1-ol, propan-2-ol, THF, 1,4-dioxane, formaldehyde
Insoluble	Chloroform, dichloromethane, butan-1-ol, nitromethane, toluene, diethyl ether, ethyl acetate, dichloroethane, butan-2-ol, hexane, aniline, cyclohexane

The solvents selected for the experimental polymorph screen were methanol, THF, acetone, diethyl ether, acetonitrile, ethyl acetate and 1,4-dioxane, outlined in Table 5.24 SI.

5.7.2 Experimental Results

The results of the experimental screen are shown in Table 5.24 SI. The majority of crystallisations yielded 5,5-dihydroxybarbituric acid (a reaction hydrate of alloxan) described in section 5.25 SI, confirming the supposition that alloxan is extremely sensitive to moisture[298]. Crystallisation from THF solution gave a new crystal structure of 5,5-dihydroxybarbituric acid monohydrate, described in section 5.26 SI. A low temperature redetermination of 5,5-dihydroxybarbituric acid trihydrate[58] was also performed to complement the results of this experimental screen, outlined in section 5.27 SI.

5.7.3 Computational polymorph search

In the *ab initio* molecular structure computational search, of the 1500 structures that were lattice energy minimised, 18 unique crystal structures were found within 7 kJ mol^{-1} of the global lattice energy minimum. The results of the search are shown in Figure 5.21 and Table 5.11.

This polymorph search gave qualitatively the same results as the previous study[48], which used a more limited search and a different, although DMA based, model potential. ExptMinOpt was found to be visually close (the smallest RMS difference in the atomic co-ordinates being 0.1402 Å for the non-hydrogenic distance, and 0.1421 Å for all atoms, when using a 12 molecule co-ordination sphere) to the structure at the global lattice energy minimum (which was found via symmetry reduction to $P2_12_12_1$ Z' = 2 structure), despite not having conventional hydrogen bonds. This structure is 1.2 kJ mol^{-1} more energetically stable, 1 kJ mol^{-1} at room temperature estimates, than the second lowest structure AQ9. AQ9 contains both hydrogen bonds to O2 and O6 and (C=O)···(C=O) contacts, Table 5.11. This decrease is consistent with the known structure being denser than the hypothetical structures with hydrogen bonds.

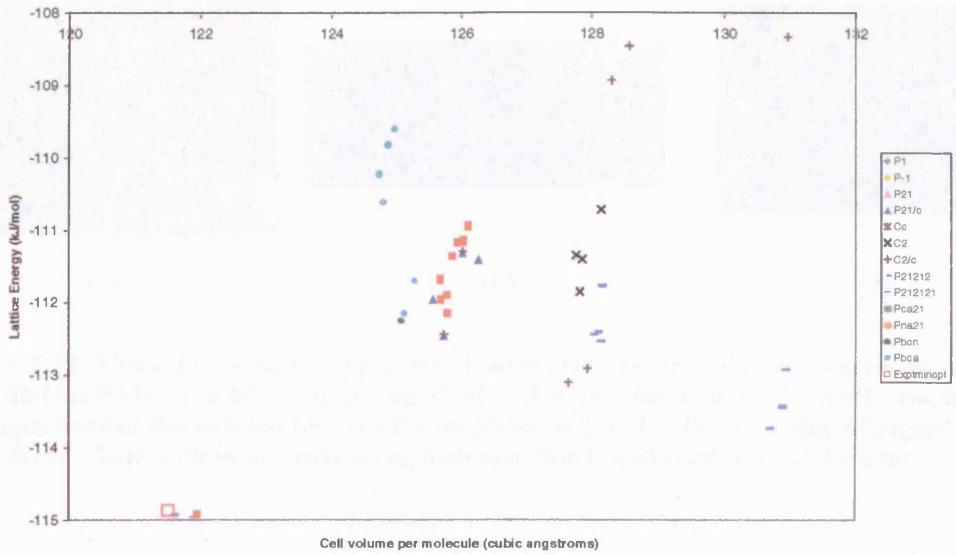


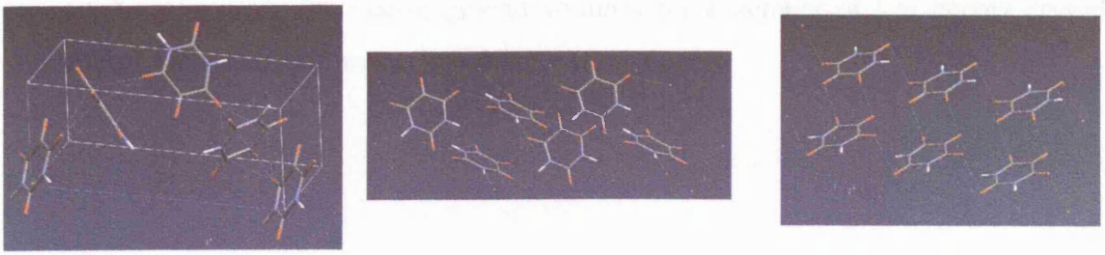
Figure 5.21 Graph showing the lattice energy vs cell volume per molecule for the minima found in the energy range of potential polymorphs (7 kJ mol^{-1}) for alloxan. ExptMinOpt is shown for comparison.

Table 5.11 The low energy hypothetical crystal structures of alloxan[a], within 7 kJ mol^{-1} of the global lattice energy minimum

Structure	Space	Lattice Energy	[b]Free energy	[c]Reduced Cell					[d]Hydrogen bond	[d]Hydrogen bond	[e]C involved in		[g]Elastic
	group	/kJ mol^{-1}	at 298 K/kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles[c]	acceptor < 2.1 Å	acceptor 2.1 - 2.5 Å	C O interactions	[f]O H structure	constant
298K Expt	$P4_22_12$			1.932	5.886	5.886	14.100			O2,O4,O6	C4,C5	3D	
123K neutron	$P4_22_12$			1.982	5.850	5.850	13.915			O2,O4,O6	C4,C5	3D	
42K neutron	$P4_22_12$			1.997	5.850	5.850	13.913			O2,O4,O6	C4,C5	3D	
ExptMinOpt	**$P4_22_12$**	**-114.875**	**-128.626**	**1.941**	**5.850**	**5.850**	**14.207**			**O2,O4,O6**	**C4,C5**	**3D**	**12.01**
BA16[sg]	$P2_12_12_1$	-114.938	-128.398	1.935	5.738	6.004	14.156			O2,O4,O6	C2,C4,C5	3D	11.43
AQ9	$P2_12_12_1$	-113.745	-127.465	1.805	6.194	7.909	10.673		O2,O6	O2,O6	C2,C4,C5	3D	5.09
DE39	C2/c	-113.107	-125.263	1.848	5.698	8.505	21.411	α 100.03	O4,O6		C2,C4,C5	Sheet	7.07
AQ38	$P2_12_12$	-112.548	-125.250	1.841	5.771	7.475	11.886		O4,O6	O2	C2,C4,C5	3D	5.59
AI45	$P2_1/c$	-112.464	-125.336	1.876	5.709	8.359	10.728	α 100.73	O4		C2,C4,C5	Sheet	8.31
CB41	Pbca	-112.146	-125.749	1.885	6.051	7.852	22.322			O2,O4,O6	C2,C4,C5	Jagged sheet	10.68
AK50	$P2_1/c$	-111.970	-125.064	1.878	5.725	8.352	10.720	α 101.45	O4,O6		C2,C4,C5	Sheet	4.05
BF25[sg]	$Pna2_1$	-111.650	-124.886	1.877	5.723	6.214	14.134		O2,O6		C4,C5	3D	9.50
CD44[sg]	Pbca	-111.485	-125.190	1.886	6.768	8.399	17.620			O2,O4,O6	C2,C4,C5	Jagged sheet	10.41
AF36	$P2_1$	-111.355	-124.572	1.846	5.808	6.773	6.807	α 107.35	O4,O6		C2,C4,C5	Sheet	9.79
BD12	$Pna2_1$	-110.934	-124.342	1.871	5.691	6.235	14.213		O2,O6		C4,C5	3D	9.43
DB45[sg]	$P2_1$	-110.692	-122.356	1.841	5.828	6.831	13.158	α 101.87	O4,O6		C2,C4,C5	Sheet	8.45
CC49	Pbca	-110.616	-124.916	1.891	5.772	8.298	20.843		O6	O2,O4	C2,C4,C5	Jagged sheet	6.14
DE43	C2/c	-108.485	-121.755	1.834	6.072	7.846	21.714	α 96.06	O4,O6		C2,C4,C5	Sheet	9.30
DE45	C2/c	-108.326	-121.103	1.802	5.350	12.857	15.721	α 104.30	O2,O4	O6	C2,C4,C5	3D	3.37

[a]All calculated structures are lattice energy minima calculated with the *ab initio* molecular model and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK co-ordination geometry and order of density, with 'sg' denoting a minimum that required a lowering of the original space group symmetry. [b]The Helmholtz Free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecular internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated by PLATON[158] are given for comparison. Only the reduced cells angles which are not 90 ° are tabulated. [d]O4 = O6. [e]Short (C=O)···(C=O) contacts, cutoff 3.5 Å. [f]Using all O···H interactions, < 2.5 Å. [g]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa

An analysis of the low energy structures shows that those consisting of molecules with conventional hydrogen bonds have O2 and/or O4 as the acceptors, and also have close (C=O)···(C=O) contacts, using the C5O5 carbonyl. The lack of low energy structures involving hydrogen bonds to C5O5 is almost certainly because of its geometrical position in the molecule, and its involvement in carbonyl-carbonyl interactions being more advantageous in the overall crystal packing. Figure 5.22 shows three low energy hypothetical structures found in the alloxan search, showing a mixture of strong hydrogen bonds and strong (C=O)···(C=O) interactions in the crystal.



(a) (b) (c)

Figure 5.22 Three low energy hypothetical structures in the alloxan search showing (a) ExptMinOpt/BA16, consisting of strong (C=O)···(C=O) interactions, (b) AI45, consisting of hydrogen bonded sheets tilted towards the *ab* plane, and (c) CC49, consisting of jagged ribbons along the *ac* plane, with both weak/strong hydrogen bonds and (C=O)···(C=O) contacts

5.7.4 Property calculations

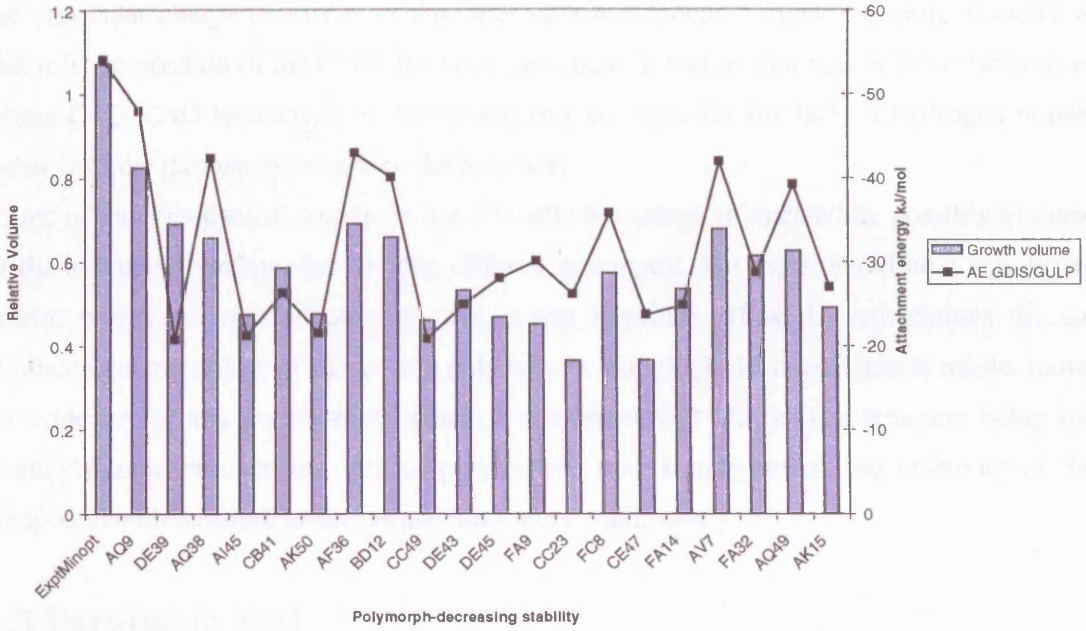


Figure 5.23 The relative growth volumes and minimum attachment energies calculated for the low energy crystal structures found in the alloxan search. Only the crystal structures that have not been symmetry reduced are shown.

The anhydrous crystal structure of alloxan was grown using sublimation[73,296], hence the attachment energy calculations which are based on a vapour grown crystal are immediately relevant. The morphology results, Figure 5.23, show that ExptMinOpt has the fastest growth volume and the fastest growth of the dominant face. The computational search and property results suggest that the known crystal structure is the most thermodynamically and kinetically favoured despite being 'exceptional' in its lack of conventional hydrogen bonds.

Nevertheless the predicted relative growth volumes for a number of low energy crystals containing hydrogen bonds are comparable.

5.7.5 Conclusions

The computational studies on alloxan show a variety of energetically feasible crystal structures, containing weak/strong hydrogen bonds and/or strong (C=O)···(C=O) interactions, using a variety of hydrogen bond donors and acceptors. It is clear that the acceptor C5O5 does not make a significant contribution to the intermolecular hydrogen bonding, despite the electrostatic potential around the molecule showing this acceptor being comparable with the other carbonyl groups, Figure 5.4. The potential around C5O5 includes through space effects of the electrostatic potential arising from the rest of the molecule, which may be compensating for any differences in the C5O5 charge distribution from its competition for the molecular charge density with the other carbonyl groups[74]. Hence it seems reasonable that it is the position of the C5O5 between two other carbonyls that makes its occurrence in strong C=O···C=O interactions so likely and that accounts for the lack of hydrogen bonds, rather than the intrinsic weakness of the acceptor.

If one of the hypothetical structures has a kinetic advantage in nucleation, possibly because of the hydrogen bonding, the packing differences suggest that there would be a significant kinetic barrier to transformation to the known structure. Thus the calculations do not eliminate the possibility of metastable polymorphs. Although the experimental results found no evidence for any polymorphs, which is consistent with the known structure being the thermodynamic product, the lack of polymorphs may simply reflect the reactivity of the compound with moisture under normal laboratory conditions.

5.8 Parabanic acid

Parabanic acid is a simple rigid molecule, with two distinct hydrogen bond acceptors, scheme 5.1. The known crystal structure[299;300] is hydrogen bonded through O4, there is no hydrogen bond to O2, forming a three-dimensional hydrogen bond network. The unused hydrogen bond acceptor phenomenon is also observed in molecular complexes of parabanic acid, including 9-ethyladenine-parabanic acid-oxaluric monohydrate, where no more than one of the six potential hydrogen bonding acceptor sites of the parabanic acid molecule participates in hydrogen bonding[301]; and in 9-ethyladenine-parabanic acid[302], where it was found that the parabanic acid molecule uses only one acceptor in an extremely long hydrogen bond (D-H···A distance being 3.313 Å). Parabanic acid is not a suitable drug

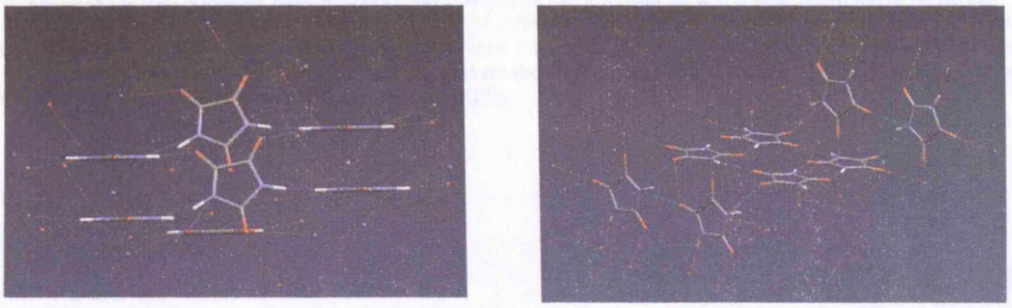
candidate as it is itself too easily hydrolysed to form oxaluric acid[303]. Nevertheless it is claimed to produce a soporific effect[304] and could be used to monitor oxygen radical activity in the human brain[305].

This study updates our previous computational and experimental polymorph search on parabanic acid[264] and is included for comparison.

5.8.1 Computational polymorph search

In the computational search using the *ab initio* molecular structure (of which up to 200 of the densest trial structures of each co-ordination type were used in the lattice energy minimisations, section 3.1.5), of the 3000 crystal structures lattice energy minimised, 31 unique crystal structures were found within 7 kJ mol^{-1} of the global lattice energy minimum. The results of the search are shown in Figure 5.25 and Table 5.12.

In the previous computational polymorph search[264] (using the W99 potential[99] with the carbon repulsion parameters decreased by 25 % and SCF DMA) ExptMinOpt was found at the global lattice energy minimum, with a plethora of low energy structures consisting of molecules that use O2 as a hydrogen bond acceptor with one of the non-unique acceptors unused. Repeating the calculations using the methodology in this thesis changed the relative ordering of the lattice energies, so that two structures with molecules that use the O2 acceptor are energetically more stable than the solid state crystal structure. Nevertheless when considering the room temperature energy estimates, the known structure is predicted to be the most stable by a small margin despite its relatively low density. Figure 5.24 shows the three-dimensional and sheet hydrogen bond networks present in two of the low energy crystal structures.



(a) (b)

Figure 5.24 The molecular packing in (a) ExptMinOpt, showing a three-dimensional hydrogen bonding network, and (b) FA40, showing hydrogen bonded jagged sheets

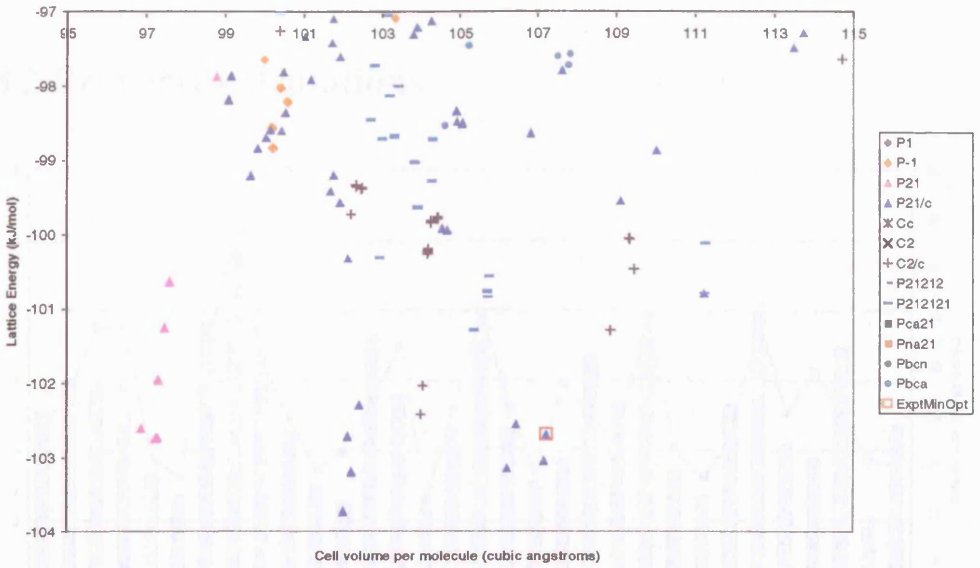


Figure 5.25 Graph showing the lattice energy vs cell volume per molecule for the minima found in the energy range of potential polymorphs (7 kJ mol^{-1}) for parabanic acid. ExptMinOpt is also shown for comparison

Table 5.12 The low energy hypothetical crystal structures of parabanic acid, within 5 kJ mol^{-1} from the global lattice energy minimium. The full structure list within 7 kJ mol^{-1} from the global lattice energy minimum is shown in Table 5.28 SI

Structure	Space group	Lattice Energy kJ mol^{-1}	[b]Free energy at 298 kJ mol^{-1}	Density/g cm^{-3}	[c]Reduced Cell a/Å	b/Å	c/Å	Angles/°	[d]Hydrogen Bond acceptors and motif		[e]Graph set Level 1	Level 2	Level 3	[f]Elastic constant
Expt 123K	$P2_1/n$			1.741	4.969	8.187	10.704	β 92.32	O4,O6	3D	C1,1(6)	C1,1(5)	C1,1(5)	
ExptMinOpt	**$P2_1/n$**	**-102.686**	**-115.647**	**1.766**	**5.247**	**7.765**	**10.550**	**β 93.440**	**O4,O6**	**3D**	**C1,1(6)**	**C1,1(6)**	**C1,1(6)**	**3.05**
FC18	$P2_1/c$	-103.739	-114.923	1.958	5.092	5.101	16.761	β 94.242	O2,O4	Sheets	C1,1(4)	R2,2(8)	C2,2(9)	1.75
AF116	$P2_1$	-102.728	-113.248	1.947	4.881	8.037	8.025	β 99.566	O2,O4	Sheets	C1,1(4)	C1,1(4)	C2,2(9)	4.72
AM84	**$P2_1/n$**	**-102.694**	**-115.661**	**1.766**	**5.247**	**7.765**	**10.550**	**β 93.452**	**O4,O6**	**3D**	**C1,1(6)**	**C1,1(6)**	**C1,1(6)**	**3.02**
DE36	C2/c	-102.417	-113.094	1.820	6.867	8.527	16.646	α 98.394	O2,O4	Chains	R2,2(8)	R2,2(8)	C2,2(9)	4.98
DE39	C2/c	-101.286	-112.038	1.740	5.286	10.225	16.148	α 93.484	O2,O4	3D	C1,1(6)	R2,2(8)	R2,2(8)	4.42
AZ74	$P2_12_12_1$	-100.779	-112.198	1.791	5.423	7.413	10.620		O2O4	3D	C1,1(4)	C1,1(4)	C2,2(9)	6.78
AQ49	$P2_12_12_1$	-100.316	-112.829	1.840	5.903	7.807	8.937		O4,O6	3D	C1,1(4)	C1,1(4)	C1,1(5)	4.48
FC56	$P2_1/c$	-100.316	-112.871	1.864	4.936	5.521	16.044	α 94.940	O2,O4,O6	3D	R2,2(8)	R2,2(8)	C1,1(4)	9.76
DE74	C2/c	-100.25	-111.512	1.818	5.421	10.908	14.959	α 109.519	O2,O4	Infinite ribbon of dimers	R2,2(8)	C1,1(5)	R4,4(16)	7.83
AK8	$P2_1/c$	-99.946	-111.976	1.809	5.124	8.120	10.079	α 93.087	O2,O4	Sheets	R2,2(8)	C1,1(4)	C2,2(9)	6.12
DE43	C2/c	-99.724	-110.734	1.852	5.093	10.610	16.701	α 103.303	O2,O4,O6	3D	C1,1(6)	R2,2(8)	R2,2(8)	4.74
AQ55	$P2_12_12_1$	-99.639	-113.136	1.822	5.022	5.243	16.792		O2,O4	3D	C1,1(4)	C1,1(4)	C2,2(9)	2.48
AM74	$P2_1/c$	-99.589	-111.053	1.858	4.918	8.942	9.286	α 93.036	O2,O4	Jagged sheets	R2,2(8)	R2,2(8)	C1,1(5)	3.84
AI88	$P2_1/c$	-98.863	-110.481	1.721	5.048	8.591	10.387	α 102.177	O4,O5	Sheets	C1,1(4)	R2,2(8)	C2,2(10)	6.89
CA34	P-1	-98.837	-109.594	1.890	5.038	5.067	8.296	α 94.849	O2,O4,O6	Sheets	C1,1(5)	R2,2(8)	R2,2(8)	7.88
								β 102.548						
								γ 101.932						
AQ117	$P2_12_12_1$	-98.722	-111.881	1.839	5.134	8.447	9.502		O2,O4	3D	C1,1(4)	C1,1(4)	C1,1(5)	5.63

[a]All calculated structures are lattice energy minima calculated with the *ab initio* molecular model and the same intermolecular potential. The hypothetical crystal structures are labelled according to the initial MOLPAK co-ordination geometry and order of density. [b]The Helmholtz free energy as estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52,54] [c]Niggli reduced cell parameters[157] (calculated during the MOLPAK/DMAREL search) are given for comparison. Only the reduced cell angles which are not 90 ° are tabulated. All structures have one molecule in the asymmetric unit. [d]O2 and O4 being the two unique hydrogen bond acceptors. [e]Only the first three levels shown, calculated using RPluto[168]. [f]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.

5.8.2 Property calculations

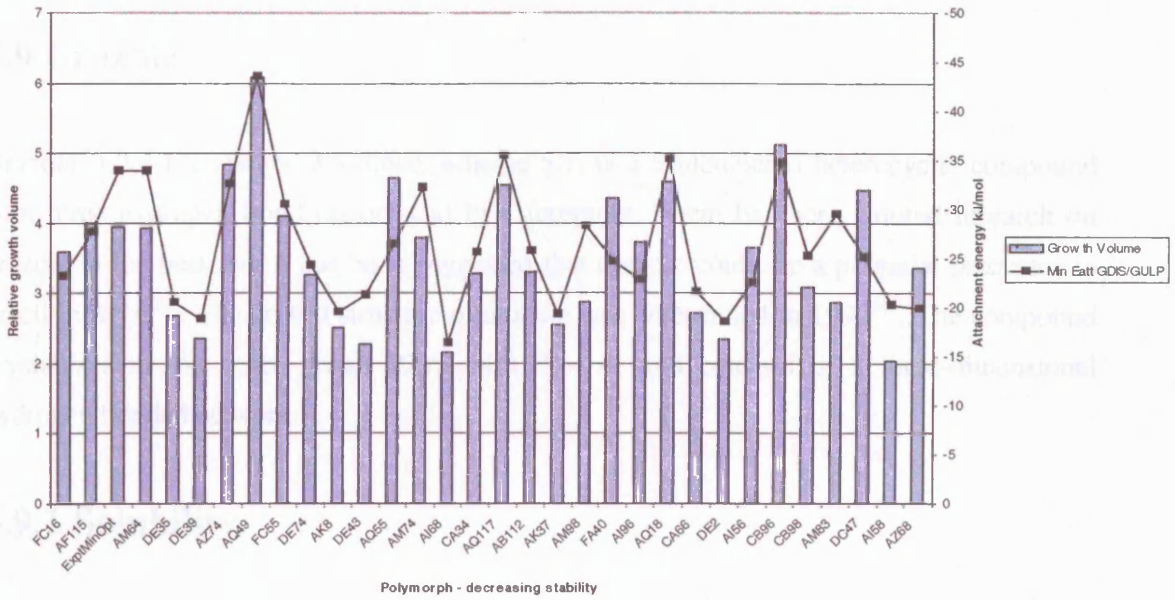


Figure 5.26 The relative growth volumes and minimum attachment energies calculated for the low energy crystal structures found in the parabanic acid computational polymorph search. Only the crystal structures that have not been symmetry reduced are shown, with AM64 the corresponding minimum to ExptMinOpt

The morphology calculations, Figure 5.26, show that the low energy crystal structures have similar minimum attachment energies, with ExptMinOpt predicted to have a relatively quick growth of the dominant face. The growth volumes show that AQ49 has a larger relative growth volume than the rest, suggesting that this has a kinetic advantage for growth from the vapour once nucleated. ExptMinOpt has an average growth rate.

5.8.3 Conclusions

The computational search on parabanic acid confirmed the results of the previous study[264], whereby the thermodynamic stability of the known form, relative to alternate structures which use O2 as a hydrogen bond acceptor, is not so great that the calculations exclude the possibility of such polymorphs being observed[264]. A parallel experimental screen failed to find any polymorphs of parabanic acid[264], and neither did a recent high pressure study[306]. However a new sesquihydrate form of parabanic acid was found at the increased pressure range[306]. Nevertheless the likelihood of a solvent promoting nucleation and growth of the significantly different structures with O2 and O4 as hydrogen bond acceptors, rather than just hydrogen bonding to O2, does seem remote. The calculations show that the known

structure is the most thermodynamically stable, nevertheless there are a few crystal structures, including AQ49, that could be kinetically favoured.

5.9 Urazole

Urazole, 1,2,4-triazolidine-3,5-dione, scheme 5.1, is a 5-membered heterocyclic compound with three hydrogen bond donors and two acceptors. There has been limited research on urazole in the past, but it has been suggested that urazole could be a potential precursor to uracil in RNA[307]. The crystal structure of urazole was determined in 1992[263]. The compound crystallises in the space group $P2_1/n$ with Z = 4, and consists of a three-dimensional hydrogen bonded network.

5.9.1 Solubility

A solubility screen was performed with the results shown in Table 5.13.

Table 5.13 The solubility of urazole in various solvents.

Observation	Solvent
Soluble	H_2O, methanol, ethanol, acetone, propan-1-ol, DMSO, DMF
Partially Soluble	Butan-2-ol, acetonitrile, aniline
Insoluble	Chloroform, dichloromethane, nitromethane, toluene, diethyl ether, ethyl acetate

The solvents H_2O, methanol, ethanol, acetone, acetonitrile, propan-1-ol, butan-2-ol, ethyl acetate, formaldehyde, DMSO and DMF were selected for the crystallisation studies, outlined in Table 5.29 SI.

5.9.2 Experimental results

The results of the experimental screen are shown in Table 5.29 SI. The majority of crystallisation experiments gave the known crystal structure[263], as a crystalline or a microcrystalline sample. The crystalline samples are of long needle morphology, typically around 1 – 1.5 mm in length, with the crystals growing on the side of the sample tube above the level of the solvent. The only evidence of a new solid state form, obtained by slow evaporation of a butan-2-ol solution of urazole, is a reaction product of urazole and butan-2-ol as shown by poor quality X-ray data.

5.9.3 Conformational *ab initio* analysis

The differences between the solid state[263] and *ab initio* molecular conformations (mainly in the puckering of the ring, section 5.2) suggests some intramolecular flexibility. To investigate this a conformational *ab initio* analysis was performed to see whether there are any local energy minima present on the potential energy surface that could correspond to other energetically plausible conformations. The conformational analysis involved a series of calculations keeping the dihedral angle N1N5C4N3 constant at 1 ° intervals between 0 ° and 17 °, and relaxing the rest of the molecule. The MP2 conformational energy profile is shown in Figure 5.27.

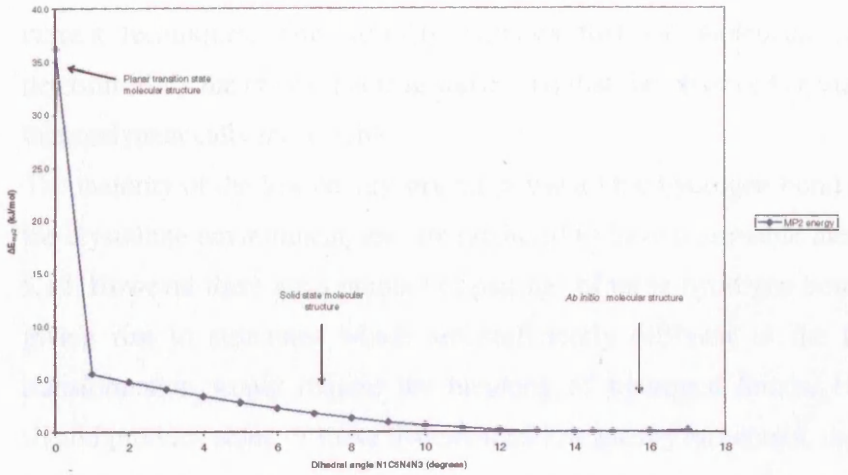


Figure 5.27 The MP2 energy torsion scan on urazole, with the ΔE_{MP2} compared to the *ab initio* molecular structure

It was found that the planar conformation of urazole exists at a transition state, around 36 kJ mol^{-1} above the global energy minimum. No other local minimum in the potential energy surface was found, with a 2.6 kJ mol^{-1} MP2 energy difference between the solid state[263] and *ab initio* molecular structures. Therefore to see whether these low energy conformational differences affect the results of the computational predictions, both the *ab initio* and solid state[263] molecular structures will be used.

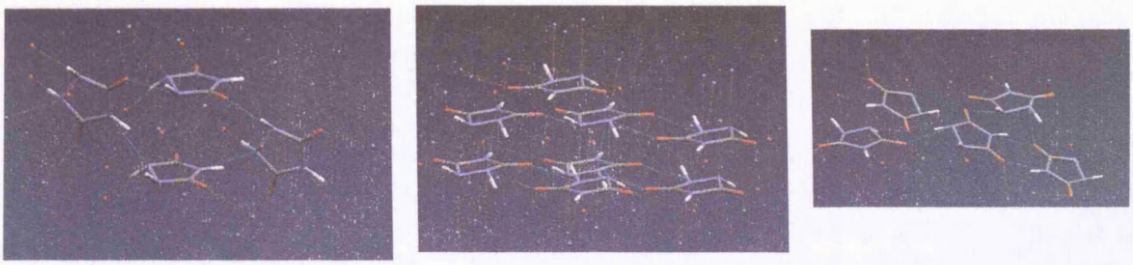
5.9.4 Computational polymorph searches

Using the *ab initio* molecular structure in the computational search, of the 1500 crystal structures that were lattice energy minimised, 39 unique crystal structures were found within 10 kJ mol^{-1} of the global lattice energy minimum. In the solid state molecular structure computational search (using only the space groups found within the low energy structures in

the *ab initio* molecular structure search), 27 unique structures were within 7 kJ mol^{-1} of the global lattice energy minimum. The results of the searches are shown in Figure 5.29 and Table 5.14.

When using the *ab initio* molecular structure in the computational search, ExptMinOpt is found around 9 kJ mol^{-1} above the global lattice energy minimum, reduced to 6 kJ mol^{-1} at room temperature estimates, with many more stable crystal structures with three-dimensional hydrogen bond networks. When using the solid state molecular structure, ExptMinExpt is found at the global lattice energy minimum. The difference in the lattice energies found with the two molecular conformers, U_{latt}(ExptMinExpt) - U_{latt}(ExptMinOpt), is almost 20 kJ mol^{-1}, much larger than the energy penalty involved in the distortion, estimated at less than 3 kJ mol^{-1} (section 5.2). This complimentary pair of results makes it difficult to estimate the relative stability of the known and hypothetical crystal structures within the limitations of current techniques. This strongly suggests that the molecular conformation is largely determined by the crystal packing forces and that the observed crystal structure could be the thermodynamically most stable.

The majority of the low energy structures use all the hydrogen bond donors and acceptors in the crystalline environment, and are predicted to have reasonable mechanical stability, Table 5.14. However there are a number of pairings of these hydrogen bond acceptors and donors, giving rise to structures which are sufficiently different to the known structure that a transformation would require the breaking of hydrogen bonds. Hence if kinetic factors should produce some of these hypothetical low energy structures, they could be observed as metastable polymorphs even if the known structure is the most thermodynamically stable. A selection of three-dimensional and sheet hydrogen bonding motifs present in the low energy crystal structures are shown in Figure 5.28.



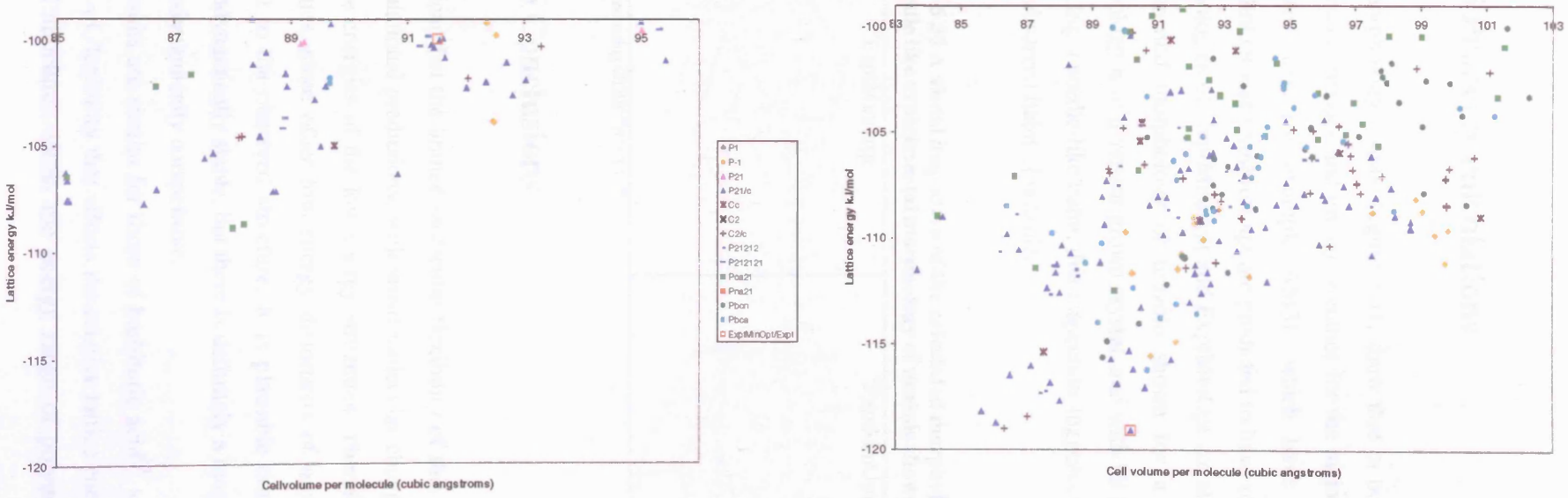
(a) (b) (c)

Figure 5.28 Three hypothetical crystal structures from the *ab initio* molecular structure search, showing (a) AY15 and (b) CA33 with a three-dimensional hydrogen bonding motif, and (c) AM3 showing hydrogen bonded sheets, Table 5.30 SI

Table 5.14 The low energy structures found in the computational polymorph searches on urazole[a] within 5 kJ mol^{-1} from the global lattice energy minimum, using both the *ab initio* and solid state molecular structures. All the structures use all of the hydrogen bond acceptors and donors in the crystal. All the structures consist of a three-dimensional hydrogen bond motif, except for AM12, AM33, CB10 and AM10 that contain sheet motifs. The full structure lists within 10 kJ mol^{-1} are shown in Tables 5.30 and 5.31 SI

Structure	Space	Lattice Energy	[b]Free energy		[c]Reduced Cell				[d]Graph set analysis			[e]Elastic
	group	/kJ mol^{-1}	at 298 K/kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles/°	Level 1	Level 2	Level 3	constant
Expt 105K	$P2_1/n$			1.861	3.462	9.513	10.995	β 95.06	C1,1(5)	R2,2(8)	C1,1(5)	
Gas phase molecular structure search												
ExptMinOpt	**$P2_1/n$**	**-99.939**	**-111.714**	**1.834**	**3.475**	**9.931**	**10.607**	**β 90.07**	**C1,1(5)**	**R2,2(8)**	**C1,1(5)**	**4.48**
AY15	$Pca2_1$	-108.714	-117.931	1.903	4.149	8.845	9.614		C1,1(5)	C1,1(4)	C1,1(5)	7.08
AK43	$P2_1/c$	-107.783	-116.845	1.940	4.464	7.861	10.467	α 109.62	C1,1(4)	C1,1(4)	C1,1(5)	13.20
AK23	$P2_1/c$	-107.607	-115.900	1.970	4.015	8.872	9.875	α 104.36	R2,2(8)	C1,1(4)	C1,1(4)	7.48
AK19	$P2_1/c$	-107.214	-117.433	1.891	4.172	8.823	9.725	α 97.38	C1,1(4)	C1,1(4)	C1,1(4)	5.16
Al21	$P2_1/c$	-106.638	-115.054	1.970	3.985	8.894	9.836	α 102.23	R2,2(8)	R2,2(8)	C1,1(4)	10.34
Al50	$P2_1/c$	-105.625	-115.116	1.917	4.070	8.878	10.111	α 106.59	C1,1(4)	R2,2(8)	C1,1(4)	5.88
Al46	$P2_1/c$	-105.075	-114.190	1.875	4.130	8.404	10.975	α 109.99	R2,2(8)	C1,1(5)	R2,2(8)	8.63
DA2	Cc	-105.045	-111.862	1.870	4.559	8.749	9.172	β 101.12	C1,1(5)	C1,1(4)	C1,1(4)	6.19
DE24	C2/c	-104.626	-113.374	1.904	4.797	8.492	17.619	α 100.69	C1,1(4)	C1,1(4)	R2,2(8)	6.34
AK4	$P2_1/c$	-104.620	-115.169	1.897	4.826	8.303	9.152	α 105.18	C1,1(4)	C1,1(4)	C1,1(4)	4.71
AQ34	$P2_12_12_1$	-104.205	-113.918	1.888	4.529	7.813	10.046		C1,1(4)	C1,1(5)	C1,1(4)	12.99
CA33	P-1	-103.918	-113.533	1.815	3.840	5.864	8.733	α 101.89	R2,2(8)	C1,1(5)	R2,2(8)	3.37
								β 95.70				
								γ 103.36				
Solid state molecular structure search												
Exptminexpt	**$P2_1/n$**	**-119.083**	**-129.918**	**1.862**	**3.461**	**9.681**	**10.765**	**β 90.88**	**C1,1(5)**	**R2,2(8)**	**C1,1(5)**	**6.60**
AM27	$P2_1/c$	-119.089	-129.923	1.862	3.461	9.681	10.765	β 90.88	C1,1(5)	R2,2(8)	C1,1(5)	6.59
DE31	C2/c	-119.004	-126.925	1.944	4.746	9.129	16.339	α 102.77	C1,1(4)	C1,1(4)	R2,2(8)	8.93
Al27	$P2_1/c$	-118.728	-127.921	1.95	4.529	7.838	10.304	α 109.77	C1,1(4)	C1,1(4)	C1,1(5)	11.10
DE42	C2/c	-118.437	-126.93	1.928	4.817	8.381	17.534	α 100.48	R2,2(8)	C1,1(4)	C1,1(4)	5.62
AM12	$P2_1/c$	-117.69	-127.284	1.907	4.102	8.588	10.014	γ 93.77	C1,1(5)	R2,2(8)	C1,1(5)	2.19
AM28	$P2_1/c$	-116.894	-126.637	1.87	3.678	9.62	10.262	γ 98.63	C1,1(5)	C1,1(5)	C1,1(5)	4.31
AM33	$P2_1/c$	-116.843	-126.684	1.913	3.997	8.748	10.05	γ 93.39	C1,1(5)	R2,2(8)	C1,1(5)	2.81
Al40	$P2_1/c$	-116.479	-126.351	1.941	4.859	8.54	8.643	α 105.32	C1,1(4)	C1,1(4)	C1,1(4)	8.10
CB10	Pbca	-116.228	-127.743	1.851	6.227	10.037	11.604		C1,1(5)	R2,2(8)	C1,1(5)	0.88
AM10	$P2_1/c$	-115.712	-125.776	1.874	5.395	6.774	10.025	γ 102.05	C1,1(5)	R2,2(8)	C1,1(5)	0.31
AB33	P-1	-115.566	-124.762	1.849	3.867	5.907	8.682	α 107.71	C1,1(5)	R2,2(8)	R2,2(8)	2.06
								β 92.22				
								γ 104.46				
DA45	Cc	-115.394	-120.133	1.918	4.58	8.712	8.966	β 101.9	C1,1(5)	C1,1(4)	C1,1(4)	5.54
FA27	$P2_1/c$	-115.321	-125.334	1.863	4.636	6.032	13.816	γ 111.14	C1,1(5)	C1,1(4)	C1,1(4)	6.17

All calculated structures are lattice energy minima calculated with the *ab initio* or solid state molecular model and same intermolecular potential. The hypothetical structures are labelling according to the initial MOLPAK coordination geometry and order of density. [b]The Helmholtz free energy as estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated using PLATON[158] are given for comparison. Only the reduced cell angles which are not 90 ° are tabulated. [d]Only the first three levels are shown, calculated using RPluto[168]. [e]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.



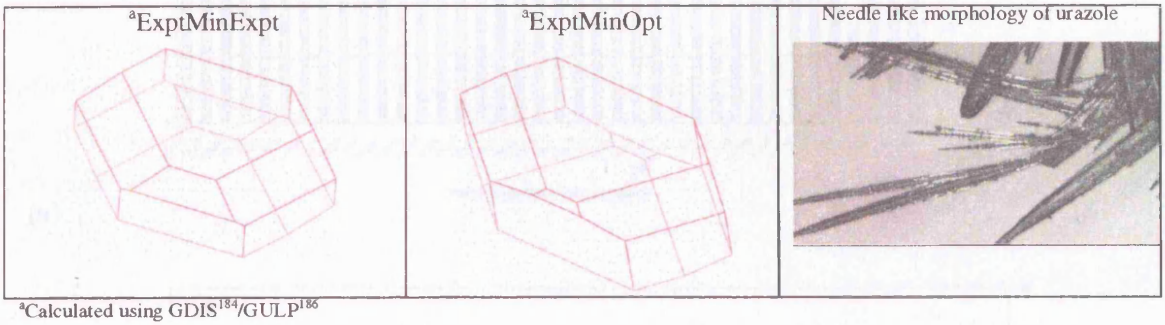
(a) (b)

Figure 5.29 Graph showing lattice energy vs cell volume per molecule for the lattice energy minima using the (a) *ab initio*, and (b) solid state molecular structure of urazole (selected space groups). ExptMinOpt and ExptMinExpt are also shown for comparison

5.9.5 Property calculations

The morphology results, Figure 5.31, show that in both computational searches the minimum attachment energies and growth volumes for the majority of structures are similar. There are a few structures, for example AM31, which have relatively low growth volumes. Both ExptMinOpt and ExptMinExpt are predicted to have relatively high growth rates. The predicted morphologies of ExptMinExpt and ExptMinOpt are shown in Figure 5.30, with the needle like experimental morphology of urazole shown for a visual inspection. Since the predicted morphology is of a vapour grown crystal, and with all the crystallisations of the anhydrous form producing a needle-like habit, this inspection suggests that solvent effects play a significant part in the observed habit of urazole.

Figure 5.30 A visual inspection of the calculated morphologies ExptMinExpt and ExptMinOpt, with the needle like experimental morphology of urazole shown for comparison.



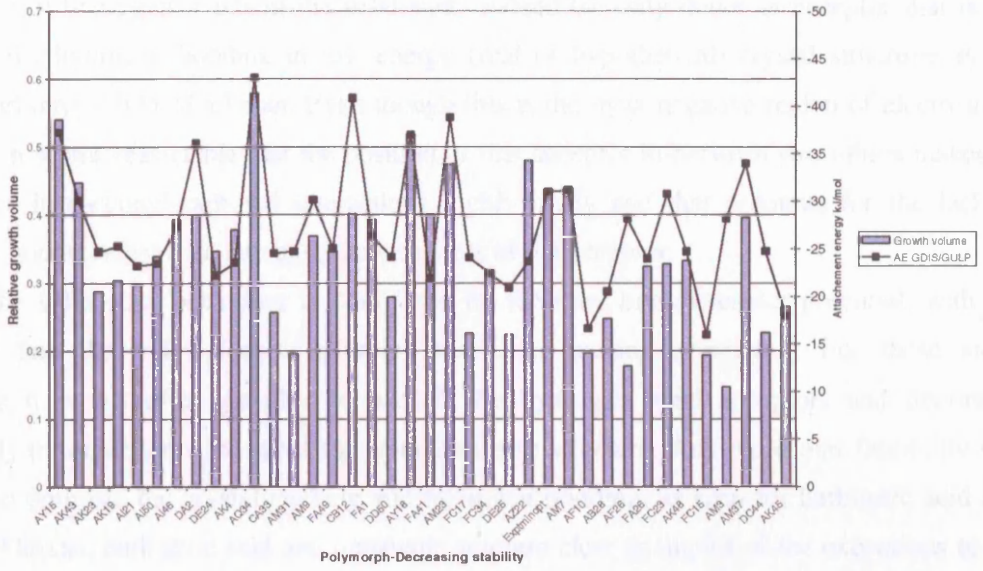
[a]Calculated using GDIS[184]/GULP[186]

5.9.6 Conclusions

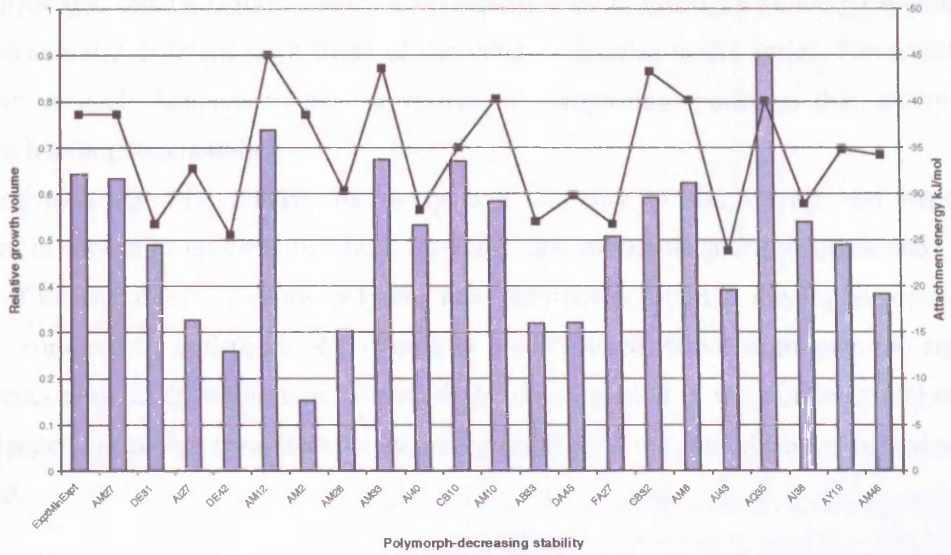
It is clear that the limited molecular flexibility of this compound has a significant effect on the computational predictions, with small molecular changes giving considerable differences in the relative energies of the low energy structures. Therefore it is difficult to estimate the relative stabilities, since other low energy distortions of molecule could improve this stability with respect to the observed structure. It is plausible that the known structure may be the most thermodynamically stable, but there is definitely a range of low energy crystal structures that are thermodynamically competitive.

The results are similar for those of barbituric acid[63], section 5.5, in that both molecules show a degree of flexibility that affects the relative lattice energy sufficiently that there is a plurality of crystal structures within the energy range of polymorphism. The calculations differ in that

whereas for urazole all the low energy structures consist of molecules that use all the hydrogen bond donors and acceptors, most of the low energy structures for barbituric acid consist of molecules that have one unused acceptor. The experimental screen only yielded the known crystal structure[263], and although the studies do not exclude the possibility of new polymorphs, this seems less likely than for barbituric acid.



(a)



(b)

Figure 5.31 The relative growth volumes and minimum attachment energies calculated for the low energy crystal structures found in the urazole computational search using the (a) *ab initio* and (b) solid state molecular structure[73].

5.10 General conclusions

The computational and experimental polymorph studies on these five similar, assumed rigid heterocyclic compounds show a range of patterns and relative energies of the low energy predicted structures. The variations in the electrostatic potential in the hydrogen bonding regions around these series of molecules, Figure 5.5, certainly do not show any correlation with the occurrence of hydrogen bonds in the solid state. Indeed the only donor or acceptor that is not found in the hydrogen bonding in low energy (real or hypothetical) crystal structures is the central carbonyl C5O5 of alloxan. Even though this is the most negative region of electrostatic potential, it seems reasonable that the position of this acceptor in-between two others makes its occurrence in carbonyl-carbonyl interactions highly likely and that accounts for the lack of hydrogen bonds, rather than any intrinsic weakness of the acceptor.

Overall the solid state behaviour is dominated by the total intermolecular potential, with the repulsion and dispersion forces enforcing the close-packing principle[33]. For these small molecules it is not often possible to use all the hydrogen bond acceptors and donors or completely optimize their geometries to obtain a dense structure. Any molecular flexibility will be used to optimize the crystal packing and hydrogen bonding, as seen for barbituric acid and urazole. Alloxan, barbituric acid and parabanic acid are clear examples of the exceptions to the normal hydrogen bonding rules[166] for crystal structures, with cyanuric acid forming solvates with unused hydrogen bond acceptors, which is consistent with its hydrogen bonding capabilities not being intrinsically different from those of the other molecules in the series. Nevertheless this particular molecule happens to have a variety of compromise packings that satisfy all the hydrogen bonding functionality.

For these molecules the relative thermodynamic stability of the known and hypothetical structures is subject to uncertainties from the inter- and intramolecular potentials used, and the neglect of thermal effects. Despite only one new polymorph found in the experimental screens on these compounds, a plethora of polymorphs are predicted, which highlights the significant role kinetics plays in crystallisation. Therefore the consideration of what other crystal structures are energetically feasible complements the understanding of the crystallisation behaviour of the molecule.

6. Uric acid as a further example of the sensitivity of crystal structure prediction to molecular conformation

6.1 Introduction

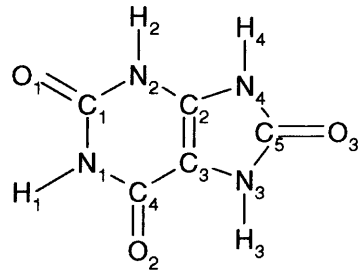
As seen in chapter 5, small molecular deviations can have a significant effect on the relative energies of the low energy crystal structures in the computational polymorph predictions. In this chapter a recent low temperature determination of uric acid is used to further explore the issue of molecular sensitivity to crystal structure prediction.

Uric acid (2,6-trioxypurine, scheme 6.1) was discovered by Scheele in 1776 as a major constituent of some mammalian concretions[308], and it has been found that it is by far the most abundant organic material found in urinary stones[309;310]. More recently the link between eating and drinking habits and urinary stone formation has been studied, including over-consumption of purine-rich foods[311] and beer, liquor and wine[312]. This high level of uric acid can also cause deposition of uric acid in joints, which results in pain and swelling causing gout[313].

Studies on uric acid crystals date back to 1899 when Brun[314] first examined the optical properties. However it was not until 1965 that limited crystallographic data of the anhydrous crystal structure was determined[315], along with a full X-ray structure determination a year later[316]. It was noted that the molecule is slightly bent with the three oxygen atoms and N3 below the ring plane, with a significant deviation of N2-H2 from the planarity of the ring to which it is attached. The solubility of pure uric acid in aqueous solution depends on pH[317] and ionic strength. It is virtually insoluble in the majority of organic solvents and only slightly soluble in water. The only other crystal structure of uric acid is the dihydrate form[308;315;318].

Since the original room temperature X-ray determination[316] has a 17 ° deviation from ring planarity of one of the N-H bonds, in which the study of this phenomenon formed the basis of work by Miss. Pinky Pridhanai Jethani as part of her 4th year degree project. This involved both computational and experimental polymorph searches, in the hope that a better, low temperature X-ray determination could be obtained. The experimental polymorph screen found the dihydrate

but no anhydrous crystal structure, mainly due to the poor solubility in many organic solvents outlined in Table 6.1 SI.



Scheme 6.1 Molecular structure of uric acid

After the experimental polymorph screen was completed, further attempts were made to crystallise the anhydrous crystal structure in collaboration with Dr. Robert Lancaster. A suitable small single crystal was finally grown from slow evaporation of a water solution over a number of months. A 120 K data collection was performed at the University of Southampton, using the EPSRC national crystallography service, with the crystallographic data shown in Table 6.2 SI. In this new determination the molecular structure is more planar with the deviation in the N2-H2 bond from the ring plane around 7 °, compared to around 17 ° found previously[316]. The metric parameters are not significantly different between the two determinations, shown in Table 6.3 SI. With these two determinations of the solid state molecular structure of uric acid, computational analysis was performed to compare these two crystal and molecular structures and the effect the molecular differences have on the computational polymorph predictions.

The new computational searches used the extended version of MOLPAK[148], outlined in chapter 3. All the searches used the FIT potential[100-102] with a 35 % reduction in the carbon repulsion parameters (for the reasons given in section 6.4), and MP2 DMA, unless otherwise stated.

6.2 *Ab initio* molecular studies

To obtain the *ab initio* molecular structure, both determinations of the solid state molecular structure were optimised using a MP2/6-31G** wave function using Gaussian98[116], resulting in the same *ab initio* minimum which is planar. The ΔE_{MP2} between the *ab initio* isolated and the room and low temperature solid state molecular structures is 25 and 39 kJ mol^{-1} respectively. A comparison of the metric parameters between the different conformations is shown in Table 6.3 SI.

To further investigate the effect of the deviations of the N2-H2 bond on the intramolecular energies, two conformations of uric acid were derived from the room temperature solid state

molecular structure[316], as defined in Table 6.1. ConoptNH is when the N2-H2 bond is fixed at the approximate room temperature solid state deviation from planarity[316] (17 °) with the rest of the molecule allowed to relax during optimisation. Conoptring is when the N2-H2 bond is relaxed with the rest of the molecule fixed. The MP2/6-31G** wavefunction was used in the *ab initio* optimisations. All the conformations are shown in Figure 6.1.

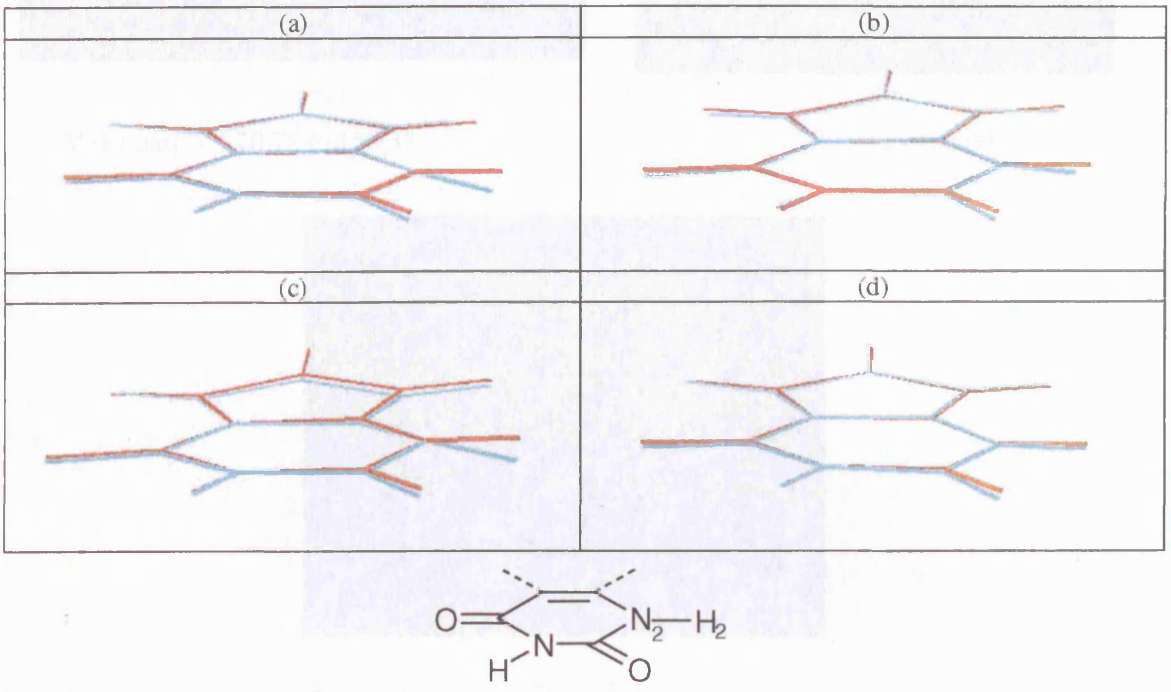
Table 6.1 Information on the conoptNH$_2$ and conoptring molecular structures of uric acid

Conformation	Optimisation	Resulting molecular structure	ΔE_{MP2} from *ab initio* isolated molecular structure/ kJ mol^{-1}
ConoptNH	N1C1N2H2 and C1N2H2 angles fixed at 164 ° and 116 ° respectively, giving the N2-H2 deviation from planarity fixed approximately at the room temperature solid state value[316] of 17 °, whilst relaxing rest of molecule	Ring structure becomes planar, except for slight puckering of the ring for which the N2-H2 is attached	1.4
Conoptring	N1C1N2H2 and C1N2H2 angles and N2-H2 bond allow to relax, rest of the molecule fixed	N2-H2 becomes close to planar with the ring	23

The difference in the intramolecular energy between ConoptNH and the *ab initio* isolated molecular structure, Table 6.1 is relatively small. This shows that the deviation in the N2-H2 bond does not have a significant effect on the intramolecular energy of the system, and therefore suggests that it could be plausible that the N2-H2 bond could exhibit this large deviation from ring planarity in the solid state. The large ΔE_{MP2} associated with Conoptring is due to the molecular deviations associated with the five-membered ring and carbonyl groups which were not allowed to relax during the optimisation, having a large effect on the intramolecular energy. The optimisations involved using the correlated MP2/6-31G** level of theory, which gives a marked pyramidilisation of the amino groups and very slight non-planarity to the structures of guanine and adenine[212], in contrast to less realistic SCF wavefunctions. Thus the *ab initio* method used is certainly likely to be correct in that the large deviation of the N2-H2 from the plane would not be found in the true gas phase structure, nevertheless we cannot be confident the level of theory used has 'converged' to give an accurate representation of this molecular structure. Therefore it is difficult to give a reliable estimate of the energy penalty for any conformational distortion[319]. In addition this estimate is also hampered by the precision in the

single crystal X-ray refinement of the molecular structure used in the calculations. The original room temperature determination[316] was refined to an R value of 6.6 %. This is relatively high and suggests that the molecular refinement could be more accurate, which in turn is reflected in the errors in the intramolecular energies.

Figure 6.1 Comparison of the *ab initio* isolated (red) molecular structure of uric acid with the (a) room temperature solid state[316], (b) low temperature solid state, (c) ConoptNH, and (d) Conoptring molecular structures, shown in blue



6.3 Electrostatic potentials

To determine whether the slight conformational differences have a significant effect on the intrinsic electrostatic contribution to the hydrogen bonding energy, the electrostatic potentials (calculated on the water accessible surface from a DMA derived from a MP2/6-31G** wave function, Figure 6.2) for the *ab initio* and both determinations of the solid state molecular structure were compared. The V_{max} for all conformations is associated close to the electrostatic potential around C2, in-between N2-H2 and N4-H4. The comparison between V_{min} and V_{max} between the different conformations shows that there is some variation of around 10 kJ mol^{-1} difference from the molecular structure going to planar. This could indicate a sufficient change to the electrostatic contribution to the lattice energy, which in turn could affect the relative energies of the low energy structures in the computational predictions.

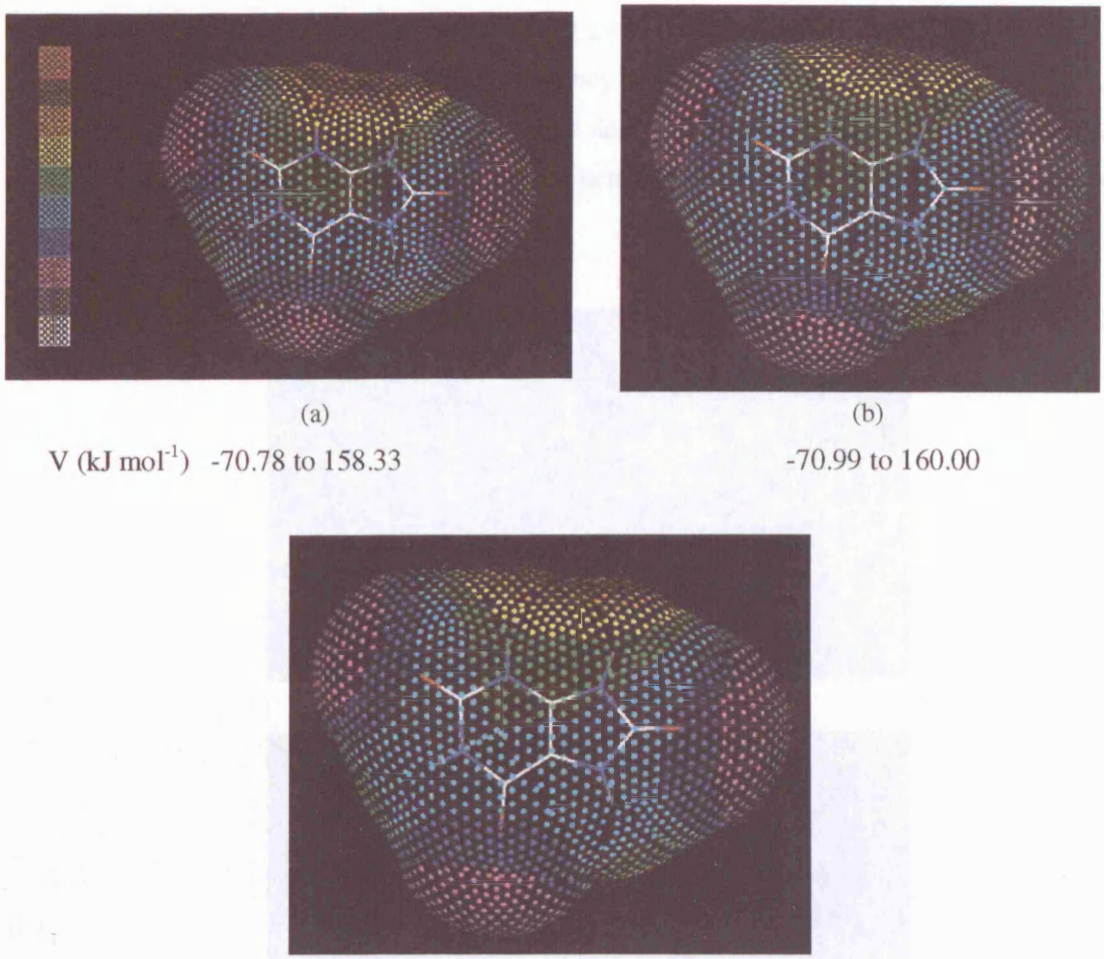


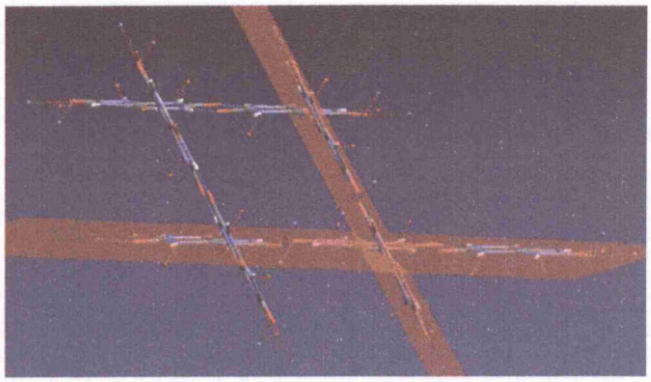
Figure 6.2 The electrostatic potential V (kJ mol^{-1}) on the water accessible surface of uric acid, using (a) the room temperature solid state, (b) the low temperature solid state, and (c) the *ab initio* molecular structure, as calculated from DMA derived from the MP2/6-31G wave function, colour coded: white < -75 < grey < -50 < magenta < -25 < blue < 0 < cyan < 25 < green < 50 < yellow < 75 < orange < 100 < brown < 125 < red**

6.4 Reproduction of the known crystal structure

It was found in the previous study[319] that using the FIT potential gave an unacceptable reproduction of the room temperature solid state crystal structure after lattice energy minimisation using the *ab initio* molecular structure, shown in Table 6.2. The *b* axis elongates by 29 % due to the relative twist of the molecules, with the angle between the two planes (Figure

6.3, associated with the chains of molecules), being 63 ° and 80 ° in the solid state and lattice energy minimised structures respectively. Whilst these are qualilatively different structures, this molecular motion corresponds to a low frequency mode (approximately 80 cm^{-1} for the low temperature ExptMinOpt structure) implying that despite some change in the hydrogen bonding, the potential energy surface is relatively flat and hence is very sensitive to changes in the model potential.

(a)



(b)

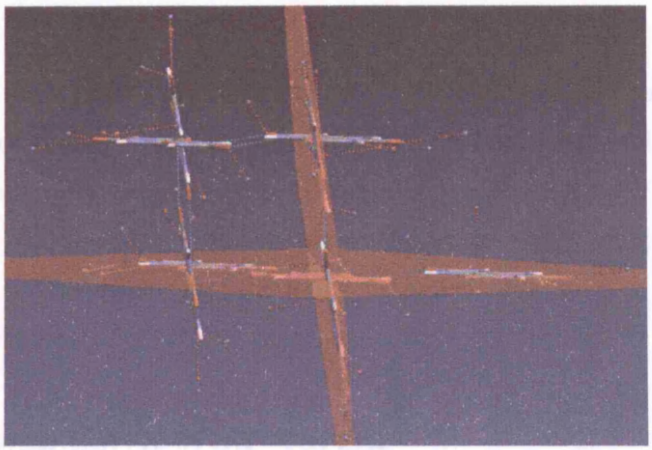


Figure 6.3 The relative twist of the hydrogen bonded chains of uric acid molecules, represented by the two red planes, in (a) the low temperature solid state and (b) the lattice energy minimised crystal structure (using the FIT potential without any carbon repulsion manipulation and the *ab initio* molecular structure)

When the low temperature solid state crystal structure is used in the lattice energy minimisations, Table 6.2, the results are very similar with a 35 % reduction in the carbon repulsion parameters needed for a satisfactory reproduction of the crystal structure (the full results given in Table 6.4 SI). This is similar to chapter 5, in which a 25 % decrease in the

carbon repulsion parameters was needed for a satisfactory reproduction of several similar heterocyclic compounds after lattice energy minimisation.

Table 6.2 Comparison of the lattice energy minimisations using the room[316;319], low temperature, conoptNH and conoptring molecular structures of uric acid. All minimistions used MP2 DMA, with the FIT potential with or without the carbon repulsion parameters decreased by 35 %. The cell setting of the two determinations of the solid state crystal structure are different. The values in brackets are the % errors compared to the associated solid state determination

	[i]Solid state								
Molecular structure		Solid state	Solid state	*Ab initio*	*Ab initio*	ConoptNH	ConoptNH	Conoptring	Conoptring
Model potential		FIT	FIT -35%	FIT	FIT -35%	FIT	FIT -35%	FIT	FIT -35%
a/Å	14.464	14.77 *(2.09)*	14.77 *(2.09)*	14.74 *(1.88)*	14.86 *(2.71)*	14.75 *(1.95)*	14.84 *(2.62)*	14.72 *(1.80)*	14.76 *(2.07)*
b/Å	7.403	7.75 *(4.73)*	7.15 *(-3.40)*	9.51 *(28.58)*	7.39 *(-0.17)*	8.99 *(21.48)*	7.17 *(-3.20)*	8.87 *(19.87)*	7.19 *(-2.90)*
c/Å	6.208	6.03 *(-2.80)*	6.19 *(-0.34)*	5.59 *(9.89)*	6.30 *(1.42)*	5.82 *(-6.26)*	6.74 *(2.70)*	5.65 *(-9.04)*	6.15 *(-0.99)*
β/°	65.099	62.21 *(-4.44)*	63.76 *(-2.06)*	52.22 *(-19.78)*	61.29 *(-5.85)*	53.57 *(-17.70)*	61.90 *(-4.91)*	56.18 *(-13.70)*	64.57 *(-0.82)*
Cell volume/Å^3	602.926	611.14 (1.36)	586.02 *(-2.81)*	620.18 *(2.56)*	606.23 *(0.55)*	620.95 *(2.99)*	598.14 *(-0.79)*	613.00 *(1.67)*	589.09 *(-2.29)*
F value		55	27	1344	44	820	57	698	21
Lattice energy/ kJ mol^{-1}		-171.80	-181.60	-158.05	-164.90	-157.29	-166.25	-171.07	-180.21
	[ii]Solid state								
Molecular structure		Solid state	Solid state	*Ab initio*	*Ab initio*				
Model potential		FIT	FIT -35%	FIT	FIT -35%				
a/Å	6.230	5.99 *(-3.90)*	6.14 *(-1.39)*	5.60 *(-10.06)*	6.30 *(1.14)*				
b/Å	7.237	7.79 *(7.64)*	7.16 *(-1.04)*	9.50 *(31.22)*	7.74 *(1.84)*				
c/Å	13.110	13.15 *(0.34)*	13.35 *(1.80)*	12.15 *(-7.35)*	13.07 *(-0.31)*				
β/°	90.849	86.59 *(-4.69)*	89.30 *(-1.70)*	73.68 *(-18.91)*	86.46 *(-4.84)*				
Cell volume/Å^3	591.007	612.44 *(3.63)*	587.16 *(-0.65)*	620.25 *(4.95)*	605.80 *(2.51)*				
F value		115	14	1800	41				
Lattice energy/ kJ mol^{-1}		-172.42	-182.15	-158.06	-166.72				

[i]Room temperature solid state crystal structure[316]
[ii]Low temperature solid state crystal structure

Using the ConoptNH and Conoptring conformations in the lattice energy minimisations, Table 6.2, there is not much difference in the reproduction of the solid state crystal structure. Again it was necessary to decrease the carbon repulsion parameters in the potential by 35 % to gain a better reproduction of the solid state crystal structure after lattice energy minimisation.
These computational results suggest that the lattice energy minimisation is sensitive to all the molecular deviations present, not just those associated with the N2-H2 bond.

6.5 Computational polymorph searches

To see whether the small molecular deviations affect the computational polymorph predictions, the computational searches using the *ab initio*, room temperature solid state, low temperature solid state, conoptNH and conoptring conformations were compared, shown in Figure 6.4. The searches using the *ab initio* and room temperature solid state molecular structures were taken from the previous study[319], and the results are shown in Tables 6.5 SI and 6.6 SI. The results of the low temperature solid state molecular structure computational search are shown in Table 6.3, with a selection of low energy structures found in this search shown in Figure 6.5. The results of both Conopt molecular structure searches are shown in Table 6.8 SI.

Table 6.3 The low energy crystal structures[a] found within 4 kJ mol^{-1} of the global lattice energy minimum for uric acid, using the low temperature solid state molecular structure. All the structures use all the hydrogen bond acceptors in the crystal, except for AY14 that only uses O1 and O3. The full list of structures within 7 kJ mol^{-1} of the global lattice energy minimum is shown in Table 6.7 SI. ExptMinExpt is also shown for comparison.

Structure	Space	Lattice Energy	[b]Free energy		[c]Reduced Cell				Hydrogen bond	[d]Graph set			[e]Elastic
	group	/kJ mol^{-1}	at 298 kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles/°	motif	Level 1	Level 2	Level 3	constant
ExptMinExpt	**$P2_1/n$**	**-182.153**	**-193.443**	**1.902**	**6.144**	**7.162**	**13.346**	**β 90.7**	**3D**	**C1,1(6)**	**R2,2(8)**	**R2,2(8)**	**2.65**
am74	$P2_1/c$	-182.156	-193.405	1.902	6.143	7.161	13.346	β 90.7	3D	C1,1(6)	R2,2(8)	R2,2(8)	2.65
de87[sg]	P-1	-181.829	-194.591	1.965	3.74	13.121	23.309	α 94.41	3D	D1,1(2)	D1,1(2)	R2,2(8)	2.79
								β 93.4					
								γ 92.6					
ca62	P-1	-181.707	-194.004	1.959	3.782	6.74	11.568	α 84.24	3D	R2,2(8)	C1,1(6)	R2,2(8)	3.44
								β 86.2					
								γ 76.41					
da62	Cc	-180.525	-189.249	1.957	3.678	11.758	13.216	γ 93.1	3D	C1,1(6)	C1,1(6)	C1,1(7)	2.24
ai28	$P2_1/c$	-179.668	-193.346	1.959	3.759	11.856	11.912	α 97.26	3D	C1,1(4)	R2,2(8)	R2,2(8)	3.87
am29	$P2_1/c$	-179.245	-193.556	1.98	3.596	11.992	13.144	β 95.7	3D	C1,1(6)	C1,1(6)	R2,2(8)	2.26
am19	$P2_1/c$	-178.937	-190.268	1.892	5.32	9.438	11.893	β 98.69	3D	C1,1(6)	R2,2(8)	R2,2(8)	9.72
ay14	$Pca2_1$	-178.936	-194.075	1.998	3.534	12.129	13.039		3D	C1,1(6)	C1,1(6)	C1,1(7)	1.46

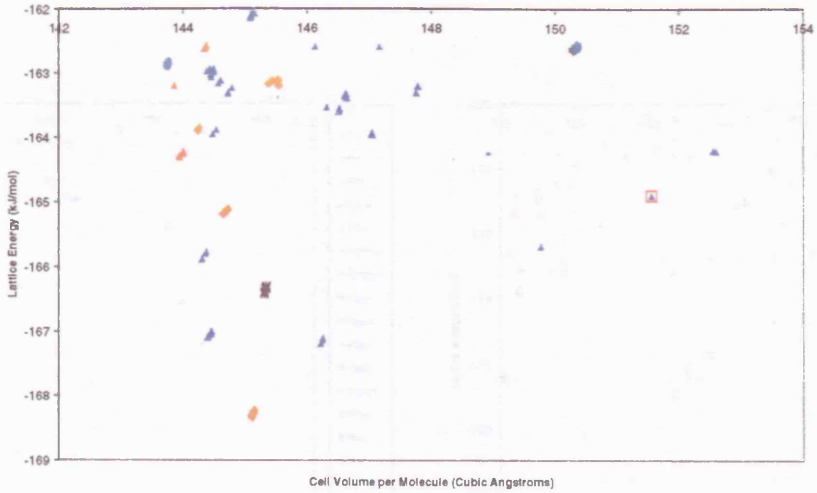
[a]All calculated structures are lattice energy minima calculated with the low temperature solid state molecular model and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK coordination geometry and order of density, with 'sg' denoting a minimum that required a lowering of the original space group symmetry. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated by PLATON[158] are given for comparison. Only the reduced cell angles which are not 90° are tabulated. [d]Only the first three levels shown, calculated using RPluto[168]. [e]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.

The small molecular changes have a significant effect on the relative stabilities of the low energy structures. In the computational search using the *ab initio* isolated molecular structure, ExptMinOpt is found 4 kJ mol^{-1} above the global lattice energy minimum (5 kJ mol^{-1} for room temperature free energy estimates) with seven structures which are energetically more favourable. Many of the low energy structures have three-dimensional hydrogen bond networks, with many of the denser more energetically favourable structures using different combinations of hydrogen bond acceptors and donors. There are some low energy structures that consist of a double hydrogen bond motif using N4-H4···O3 instead of N3-H3···O3 found in the solid state, which implies that should a route be found to form such crystal structures, there is likely to be a significant barrier for a solid state transformation to the known form.

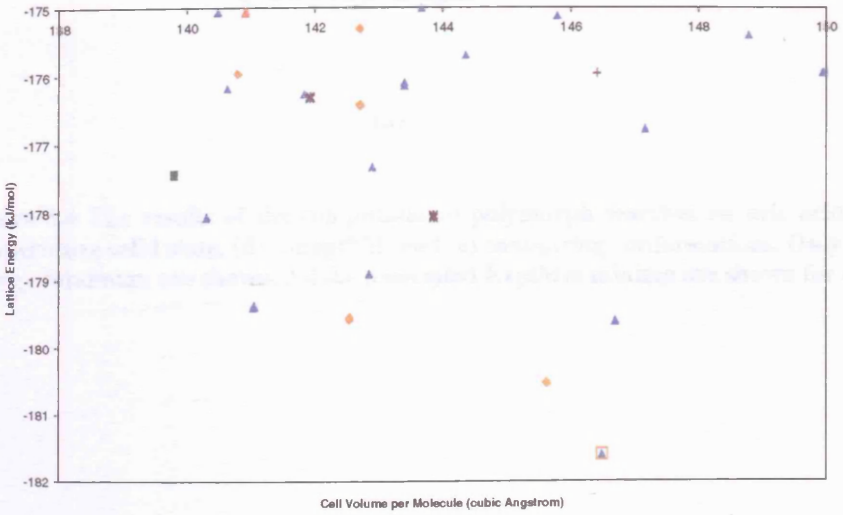
The lattice energy difference between ExptMinOpt and both ExptMinExpts (Figure 6.4(a), (b), (c)) is around 17 kJ mol^{-1}, much larger than the intramolecular energy penalty of around 1.4 kJ mol^{-1} estimated by the *ab initio* methods (section 6.2). Both ExptMinExpts are found at the global lattice energy minimum. The ConoptNH computational search (Figure 6.4(d)) has relative energies between the low energy structures that are comparable to the *ab initio* isolated computational search, with a similar lattice energy scale. This suggests that the lattice energies are not that sensitive to the deviations in the N2-H2 bond. Nevertheless there is a change in the relative stabilisation of the solid state structure compared to the other energetically feasible alternatives, with ExptMinConoptNH found 0.9 kJ mol^{-1} above the global lattice energy

minimum. The Conoptring computational search has relative energies between the low energy structures that are comparable to the ConoptNH search, with ExptMinConoptring found 1.6 kJ mol^{-1} above the global lattice energy minimum. However the lattice energies of the low energy structures in the Conoptring search are comparable to the solid state molecular structure searches. This is a crucial observation which shows that the deviations in the five membered ring and the carbonyl groups have a significant effect on the lattice energies. Nevertheless there are errors associated with the calculation of the intramolecular energies, section 6.2, and therefore it would be advantageous to optimise the geometry of uric acid using a higher level of theory to gain more accurate comparative lattice energies of the low energy structures.

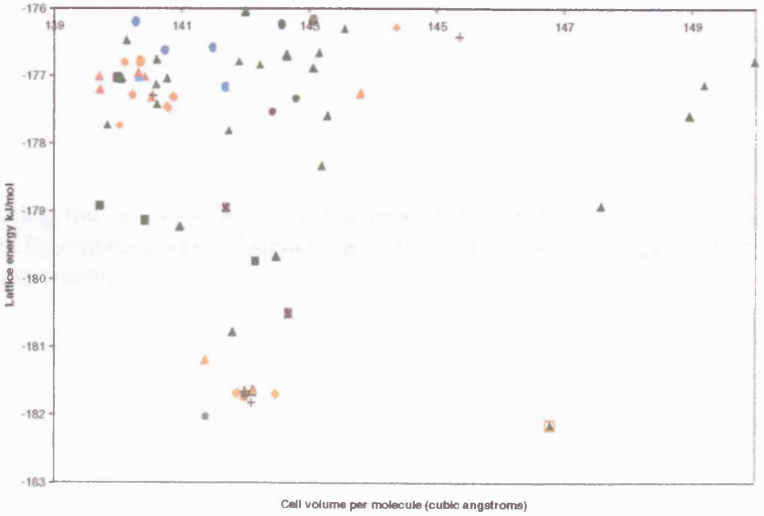
These computational searches show that the N2-H2 deviation from planarity does have an effect on the qualitative energy landscape, with the solid state crystal structure stablilised more compared to the other low energy structures with a greater deviation of this bond from planarity. Nevertheless all the other molecular deviations (including the five-membered ring and carbonyl groups) have a significant quantitative effect on the lattice energies of the low energy structures.


Lattice Energy (kJ/mol)
Cell Volume per Molecule (Cubic Angstroms)

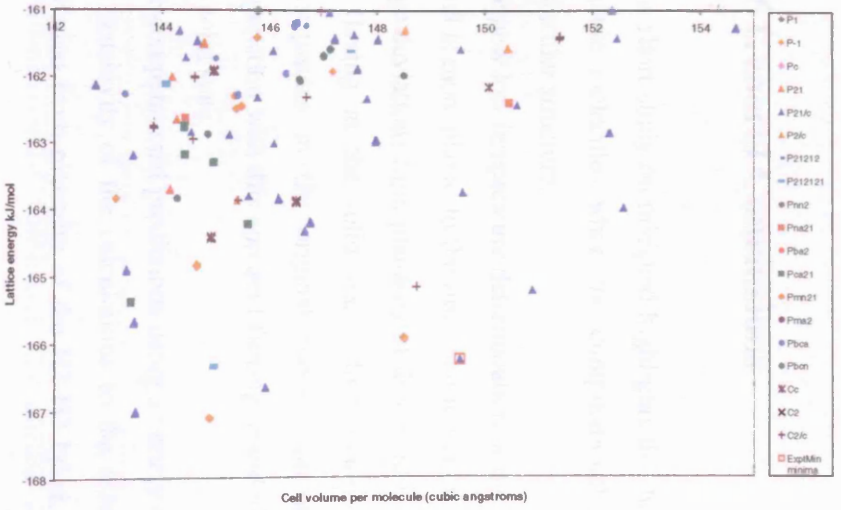
(a)


Lattice Energy (kJ/mol)
Cell Volume per Molecule (cubic Angstrom)

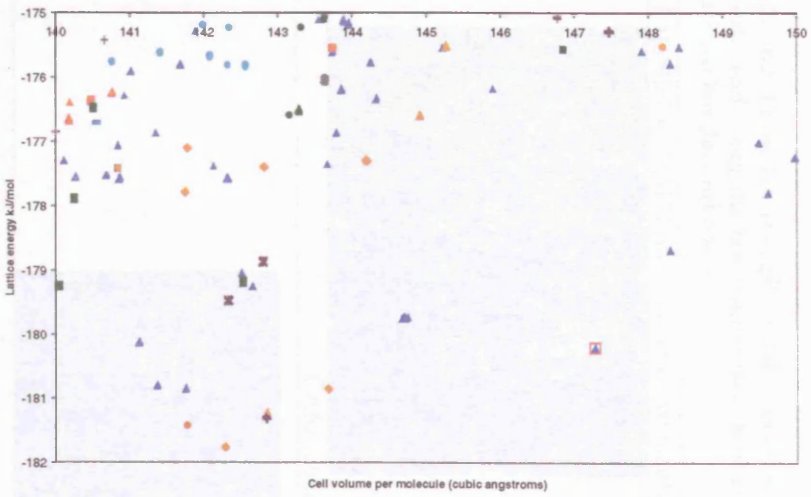
(b)


Lattice energy kJ/mol
Cell volume per molecule (cubic angstroms)

(c)



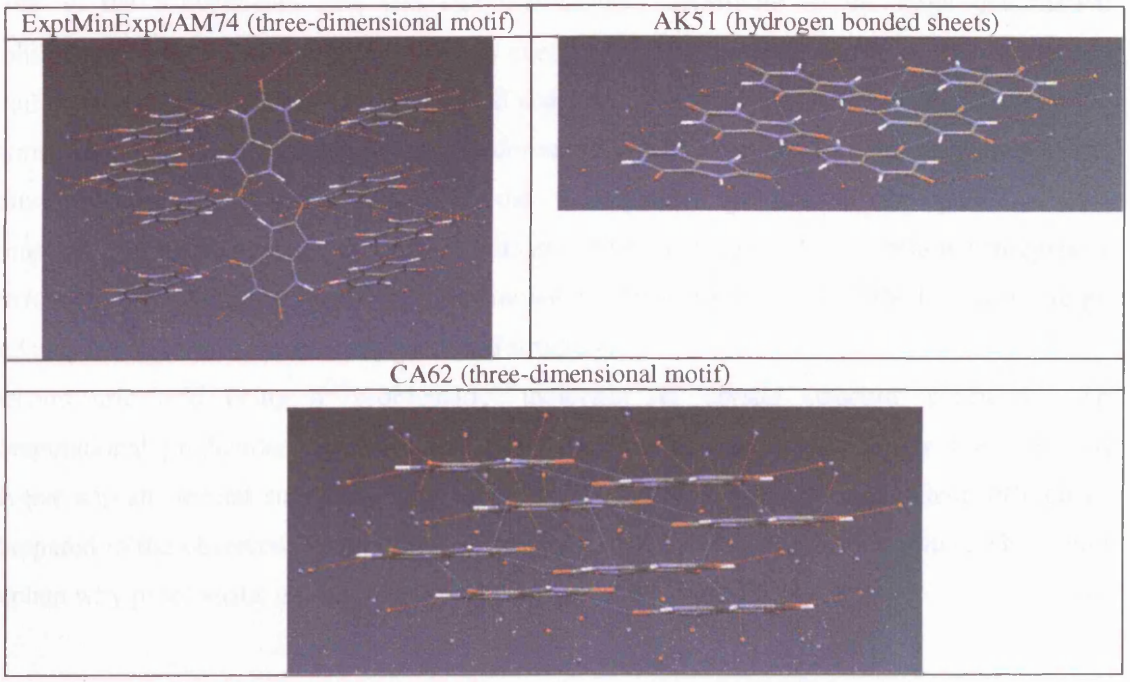
(d)



(e)

Figure 6.4 The results of the computational polymorph searches on uric acid, using the (a) *ab initio*, (b) room temperature solid state[316;319], (c) low temperature solid state, (d) conoptNH, and (e) conoptring conformations. Only the hypothetical crystal structures within 7 kJ mol^{-1} of the global lattice energy minimum are shown. All the associated ExptMin minima are shown for comparison.

Figure 6.5 Three low energy crystal structures (shown in Table 6.7 SI) in the computational search on uric acid using the low temperature molecular structure, showing three-dimensional and sheet hydrogen bonding networks



6.6 General Conclusions

This short study on uric acid highlights the difficulties in crystal structure prediction for small, organic molecules when the computational predictions are very sensitive to the assumed molecular structure.

The new low temperature determination of the anhydrous crystal structure shows that the N2-H2 bond is more planar to the ring than found previously[316]. The *ab initio* calculations suggest that large deviations from planarity of this bond is energetically plausible, therefore this bond could be flexing in the solid state which could well be a low frequency mode. Nevertheless inadequacies in the original room temperature single crystal X-ray determination[316], in conjunction with this apparent flexing, could also be a factor in observing this large deviation in the solid state.

The computational predictions using a variety of molecular conformations of uric acid show that the sensitivity of the calculations to the assumed molecular structure is not just due to the deviation from planarity of the N2-H2 bond, but due to all the molecular deviations present.

Specifically, the deviations in the N2-H2 bond from planar stabilises the observed structure relative to the energetically feasible alternatives, whilst all the molecular deviations (including those in the 5-membered ring and carbonyl groups) contribute to the large quantitative stabilisation of the lattice energies of the low energy structures between the *ab initio* isolated and solid state molecular structure computational searches. However it is difficult to give a relative estimate of the energy penalty for the conformational distortion, and as the reduction in the carbon repulsion parameters highlights the inadequacies present in the potential, this computational model seriously limits the confidence that can be placed in the relative energies of the low energy structures. Nevertheless the known crystal structure is one of the lowest in energy and may be the thermodynamically preferred structure.

Despite uric acid being a 'problematic' molecule for crystal structure prediction, the computational predictions do show that polymorphism is thermodynamically feasible with diverse supramolecular structures. Many of these crystal structures show sufficient differences compared to the observed crystal structure, giving a plaurity of low energy minima. This could explain why problematic growth of anhydrous uric acid has been observed[319].

7. Crystal structure prediction blind test 2003

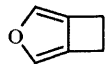
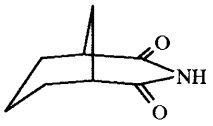
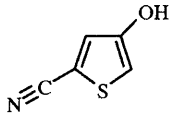
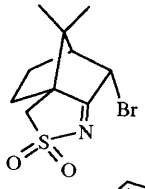
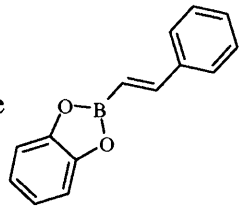
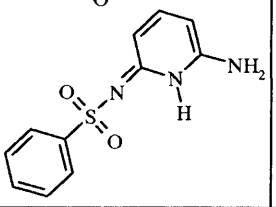
7.1 Introduction

This chapter explores the computational predictions on molecule VIII, hydantoin, performed as part of the third blind test on crystal structure prediction[28]. This will give an insight into which particular areas of theory need to be improved to take this area of scientific research forward, with the goal of being able to predict the observed crystal structures of organic compounds just by knowing the molecular structure. The methodology used in this chapter is the same as throughout the thesis, unless stated otherwise.

7.2 Previous blind tests

The first collaborative workshop on crystal structure prediction was held in 1999[30], and was followed by a second workshop in 2001[29]. There were 11 participants in the first blind test and 17 in the second, with the molecules chosen for the crystal structure prediction shown in Table 7.1. The crystal structures selected were restricted to Z' = 1, with the results shown in Table 7.1. For both these blind tests the occurrence of some predictions with an accuracy of a few percent in the cell dimensions represented significant success. If the arbitrary rule of submitting the best three structures from each method had been extended to six structures the success quota would have been much higher[29]. However the methods used for crystal structure prediction struggled to predict the experimentally observed structures for more flexible molecules (III and VI), with only one success with molecule III and none for VI. It was therefore concluded that it cannot yet be claimed that any of the methods used is consistently reliable.

Table 7.1 The molecules selected for the two previous CCDC blind tests of crystal structure prediction[29;30], with the number of successful predictions also shown

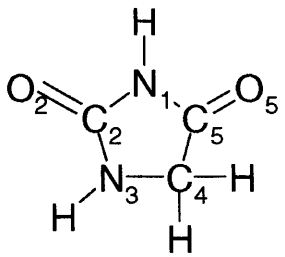
I. Rigid		4[a]	IV. Rigid		2
II. Rigid		1	V. Rigid		4[b]
III. Flexible		1	VI. Flexible		0
1999[30]			2001[29]		

[a]Compound I is polymorphic, all 4 predictions were of the metastable form
[b]van Eijck's result has the correct packing, but a large RMSD owing to differences in molecular model conformation.

7.3 Third blind test of crystal structure prediction

In early October 2003 the criteria and molecules were set out to all the applicants for the third crystal structure prediction blind test. The criteria for the third blind test were: (1) a small rigid molecule containing only elements C, H, O and N allowed, (2) small rigid molecule, elements C, H, O, N allowed, plus some less common atoms (eg. halogen) as a challenge for the model potential, and (3) small flexible molecule with a maximum of 2-3 torsion angles and common atom types. The molecule in category (1) is hydantoin, shown in Table 7.2. In this blind test scenario the molecules chosen could crystallise in any space group, with $Z' \leq 2$.

Table 7.2 The information given to the applicants regarding the crystal structure prediction blind test on molecule VIII, hydantoin

Molecule (hydantoin)	Crystallisation conditions
	The compound was dissolved in methanol with heat, filtered and evaporated slowly at room temperature

Hydantoin (2,4-imidazolidinedione), was discovered by Baeyer in 1861 as a hydrogenation product of allantoin[320,321]. Hydantoin is of interest as it is the parent compound of the anti-epileptic drug diphenylhydantoin[322]. This and related compounds with aliphatic side groups are commonly used as sedatives, whereas phenyl substitution is used to obtain a drug effective against certain types of epilepsy[322]. The hydantoin ring system rarely occurs in nature, however a significant number of synthetic derivatives have been prepared[323], including 5,5-dimethylhydantoin[324].

7.4 *Ab initio* study

To determine which molecular conformation(s) will be used in the computational polymorph predictions, an *ab initio* study was performed. A slightly puckered conformation was built using MOLDEN[117], and then optimized (using a MP2/6-31G** wave function using Gaussian98) to the *ab initio* minimum puckered conformation, Figure 7.1. It was also found that a near planar conformation optimizes, using the same wave function, with the ring planar and the CH_2 out of the plane. This conformation (deemed the planar molecular structure for the rest of this structure) is only 0.3 kJ mol^{-1} higher in energy than the *ab initio* conformation. Therefore a variety of conformations are energetically plausible in the solid state. It was decided to use both the *ab initio* and planar conformers in the computational polymorph predictions. The molecular parameters for both these conformations are shown in Table 7.1 SI.

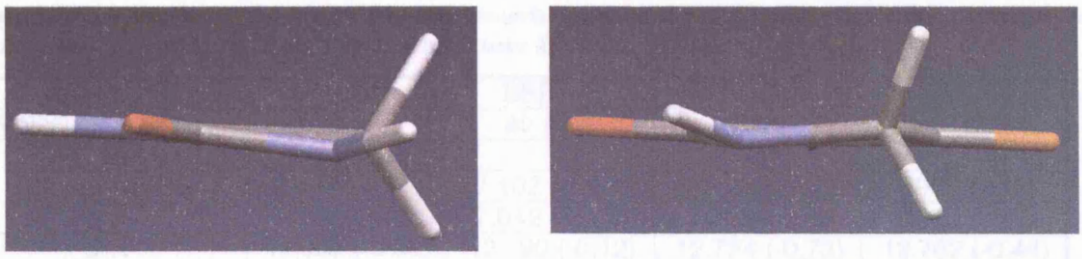
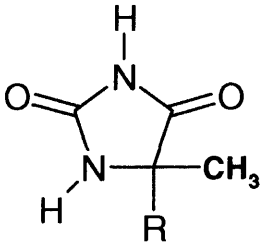


Figure 7.1 The puckered *ab initio* conformation of hydantoin

7.5 Related crystal structures

To give an indication of how well the method of modelling the intermolecular forces might be expected to reproduce the crystal structure of hydantoin, crystal structures of structurally similar

molecules were chosen from the Cambridge Structural Database[114] for testing using lattice energy minimisation. These are 5,5-dimethylhydantoin[324] (refcode BEPNIT) and 5-ethyl-5-methylhydantoin[325] (refcode ADUQOF) , shown in Scheme 7.1. Both crystal structures belong in the space group $P2_12_12_1$, Z' = 1.



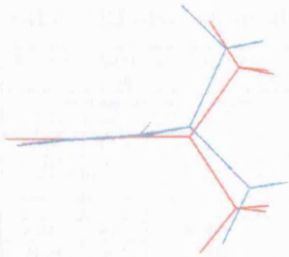
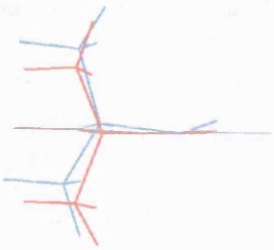
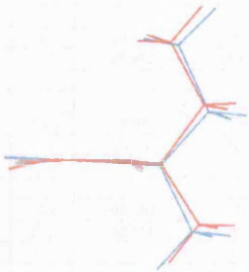
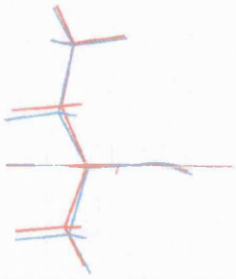
Scheme 7.1 The structurally similar molecules to hydantoin showing BEPNIT (R = methyl), and ADUQOF (R = ethyl)

Both the solid state and *ab initio* (calculated from a MP2/6-31G** wave function using Gaussian98[116] from the solid state molecular structure) molecular structures were used in the lattice energy minimisations, Table 7.3. The results show satisfactory reproductions of both crystal structures, despite the 4.7 % change in the *a* axis when using the solid state molecular structure in BEPNIT. There is some molecular flexibility present in both molecules, as shown with the slight differences in the orientations of the ethyl and methyl groups between the *ab initio* and solid state molecular structures, Table 7.4. These results suggest that the model potential and DMA would be adequate for use in the computational studies on hydantoin.

Table 7.3 The results of the lattice energy minimisations on BEPNIT and ADUQOF, with the % errors compared to the solid state crystal structure shown in brackets

Crystal structure	BEPNIT	BEPNIT	ADUQOF	ADUQOF
Molecular structure	Expt	*ab initio*	Expt	*ab initio*
a/ Å	6.877 (-4.70)	7.102 (-1.58)	7.819 (-2.02)	7.989 (0.12)
b/ Å	7.074 (-1.80)	7.042 (-2.23)	7.296 (1.06)	7.240 (0.29)
c/ Å	12.965 (-0.31)	12.990 (-0.12)	12.724 (-0.73)	12.762 (-0.44)
(all cell angles 90°)				
Cell Volume / Å^3	630.664 (-6.70)	649.713 (-3.88)	725.819 (-1.71)	738.172 (-0.03)
F value	**33**	**25**	**10**	**7**
Lattice Energy kJ/mol	-100.302	-92.601	-105.847	-97.788

Table 7.4 The superimposed solid state (red) and the *ab initio* (blue) molecular structures of BEPNIT and ADUQOF.

		BEPNIT[324]
		ADUQOF[325]

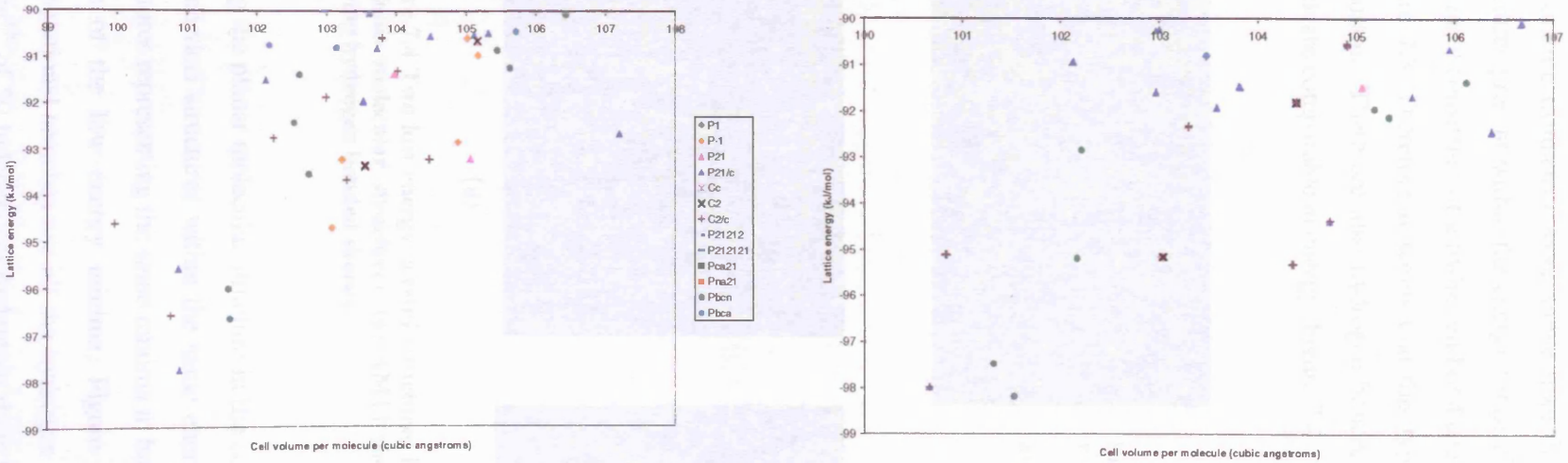
7.6 Computational polymorph searches

In the computational search using the *ab initio* molecular structure, of the 1500 structures that were found, there are 26 unique crystal structures within 7 kJ mol^{-1} of the global lattice energy minimum. Using the planar molecular structure, there were 21 unique crystal structures within this same energy range. The results of the computational searches are shown in Figure 7.2 and Table 7.5.

Table 7.5 The low energy crystal structures of hydantoin[a], using both the *ab initio* (gas phase) and planar molecular structures, within 5 kJ mol^{-1} of the global lattice energy minimum. All the crystal structures use all the hydrogen bonding acceptors and donors in the crystal. The full structure lists within 7 kJ mol^{-1} of the global lattice energy minimum are shown in Tables 7.2 and 7.3 SI

Structure	Space group	Lattice Energy /kJ mol^{-1}	[b]Free energy at 298 K/kJ mol^{-1}	Density/g cm^{-3}	[c]Reduced Cell a/Å	b/Å	c/Å	Angles/°	Hydrogen Bond motif	[d]Elastic constant
Gas phase molecular structure search										
AM14	$P2_1/n$	-97.766	-109.936	1.647	4.028	8.325	12.248	γ 100.67	Chains	2.53
CD34[sg]	$P2_1/n$	-96.686	-108.217	1.635	7.778	8.767	12.185	γ 101.93	Chains	1.47
DD19[sg]	P-1	-96.609	-106.893	1.649	8.212	8.221	12.174	α 89.06	Chains	4.61
								β 88.14		
								γ 78.97		
AM33	$P2_1/c$	-95.601	-107.262	1.647	4.258	7.994	12.106	γ 101.68	Chains	3.16
AB34	P-1	-94.732	-106.531	1.612	4.063	6.375	8.343	α 87.65	Chains	3.35
								β 79.21		
								γ 76.18		
DE21	C2/c	-94.626	-104.97	0.831	4.864	10.809	15.766	α 105.23	Chains	7.45
DD27[sg]	P-1	-93.687	-105.334	1.609	8.271	8.31	12.56	α 77.81	Dimers and Chains	2.80
								β 89.58		
								γ 78.51		
CD14[sg]	P2/n	-93.564	-106.079	1.617	6.47	10.56	12.03	γ 90.30	Sheets	0.02
AB22	P-1	-93.251	-104.727	1.61	4.32	6.178	8.075	α 91.54	Chains	4.13
								β 102.35		
								γ 100.63		
DE2	C2/c	-93.237	-105.167	0.795	4.151	12.008	16.811	α 94.14	Chains	2.94
AB48	P-1	-92.855	-103.899	1.585	5.921	6.243	6.752	α 75.30	Chains	1.06
								β 74.95		
								γ 61.72		
DE23	C2/c	-92.777	-103.369	0.813	4.945	11.019	15.462	α 103.86	Chains	6.57
AM9	$P2_1/n$	-92.66	-106.395	1.55	3.834	9.571	11.712	β 93.85	Chains	4.02
Planar molecular structure search										
CD36[sg]	P2/n	-98.188	-109.716	1.637	7.62	8.908	12.229	γ 101.96	Chains	1.58
AM10	$P2_1/c$	-97.99	-109.78	1.651	4.087	8.212	12.21	γ 100.75	Chains	2.73
CD22[sg]	$P2_1/m$	-97.474	-109.328	1.64	6.349	10.472	12.196	α 91.64	Chains	0.45
DE47	C2/c	-95.319	-107.171	0.796	4.061	12.089	17.028	α 92.81	Chains	3.26
CD6[sg]	Pcmn	-95.196	-107.482	1.627	6.439	10.529	12.055	γ 90.01	Sheets	0.18
AB1	P-1	-95.18	-106.75	1.613	4.142	6.267	8.277	α 90.24	Chains	3.72
								β 101.40		
								γ 101.67		
DE5	C2/c	-95.126	-105.348	0.824	4.901	10.895	15.628	α 104.87	Chains	7.34
AB42	P-1	-94.379	-105.46	1.587	6.077	6.081	6.755	α 75.09	Chains	1.90
								β 75.09		
								γ 61.57		
DD17	C2/c	-94.376	-105.514	0.793	6.222	6.756	10.446	α 107.43	Chains	0.53

[a]All calculated structures are lattice energy minima calculated with the *ab initio* or planar molecular model and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK co-ordination geometry and order of density, with 'sg' denoting a minimum that required a lowering of the original space group symmetry. All symmetry reduced structures are Z' = 2. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] are calculated using PLATON[158] are given for comparison. Only the reduced cell angles which are not 90 ° are tabulated. [d]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants (C_{ij}, ij = 4,5,6).



(a) (b)

Figure 7.2 Graph showing the lattice energy vs cell volume per molecule for the minima found in the energy range of polymorphism (7 kJ mol^{-1}) for hydantoin using the (a) *ab initio*, and (b) planar molecular structures.

In the *ab initio* molecular structure computational search, the crystal structure found at the global lattice energy minimum, AM14, is around 1 kJ mol^{-1} (1.7 kJ mol^{-1} at room temperature estimates) more stable than the nearest rival. There are a variety of crystal structures present within the energy range of polymorphism, with just less than half of the structures consisting of a hydrogen bond chain motif with the same centrosymmetric dimer, Figure 7.3. Therefore it seems that this type of hydrogen bonding motif is energetically favourable. There are also hydrogen bonded sheets and infinite ribbons or dimers present which are comparable in energy, Figure 7.4.

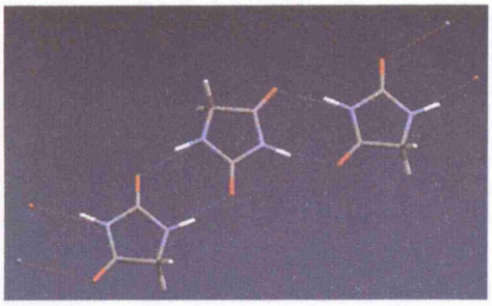
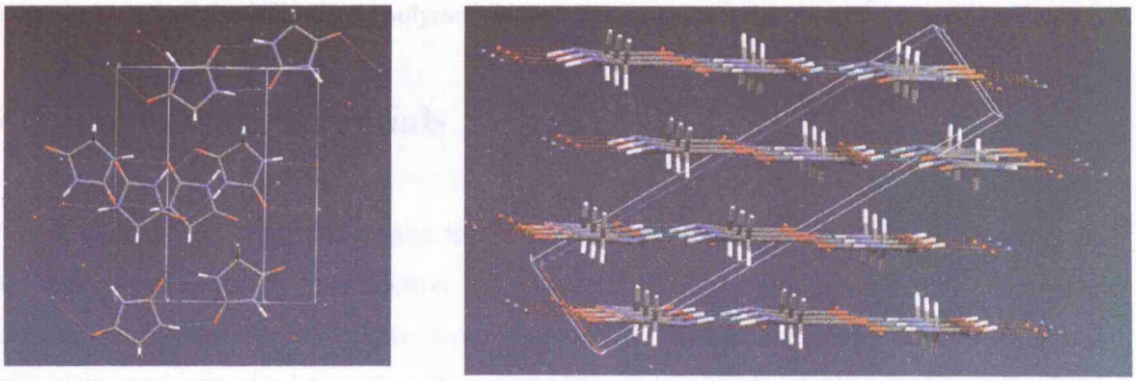


Figure 7.3 The hydrogen bonding motif present in many of the lowest energy hypothetical structures for hydantoin (using the *ab initio* molecular structure), consisting of centrosymmetric chains of molecules.



(a) (b)

Figure 7.4 Two low energy crystal structures found in the computational search on hydantoin (*ab initio* molecular structure), (a) AM14 showing hydrogen bonded chains, and (b) FC24 showing hydrogen bonded sheets.

Using the planar molecular structure in the computational search, there are significantly less hypothetical structures within the same energy range of polymorphism, with a number of structures representing the same minima in both searches. There is a significant reordering of some of the low energy minima, Figure 7.5. All the low energy structures in both computational searches use all the hydrogen bond acceptors, which is in stark contrast to a survey[326] of 50 independent hydantoin rings in the Cambridge Structural Database[114]. In the reported experimental structures more than half have unused hydrogen bond acceptors. In the

planar molecular structure computational search the hydrogen bonded chain motif is still predicted to be the most energetically favourable, with the sheet motif more energetically competitive when using this molecular structure.

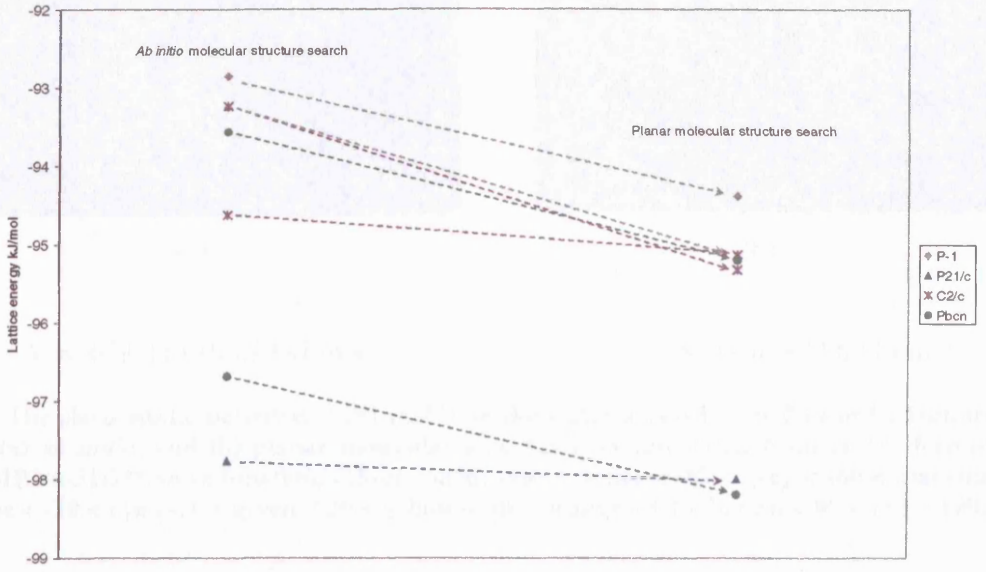
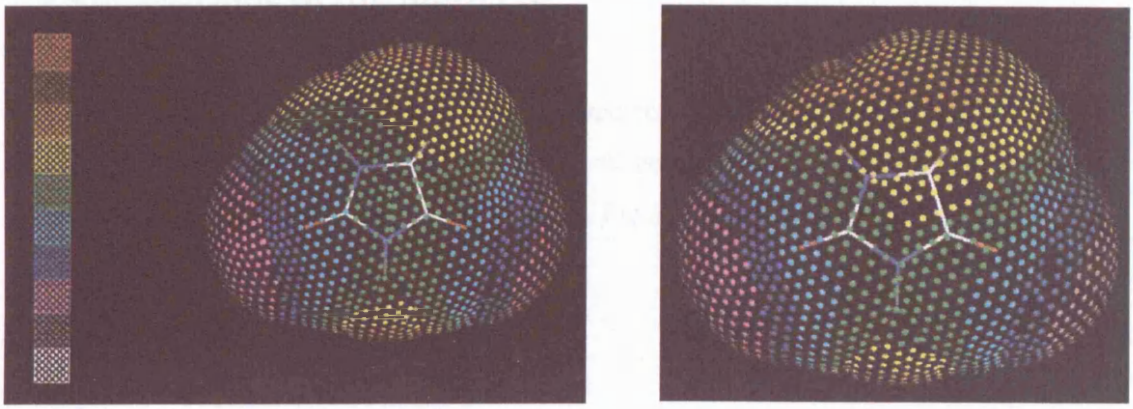


Figure 7.5 A comparison of the relative energies within a select number of equivalent low energy structures in both computational polymorph searches on hydantoin.

7.7 Electrostatic potentials

To see whether the small differences in the two molecular conformations have an affect on the intrinsic electrostatic contribution to the intermolecular hydrogen bonding, the water accessible electrostatic potentials were calculated, Figure 7.6. The two electrostatic potentials are very similar with only small differences in the relative strengths of the potential around the molecules, which is also highlighted in the similarities between V_{min} and V_{max}. This suggests that the differences in the molecular structure do not have much of an affect on the electrostatic contribution to the lattice energy in the crystalline environment.



(a) (b)

V = -63.08 to +93.99 kJ mol^{-1} -64.14 to +94.60 kJ mol^{-1}

Figure 7.6 The electrostatic potential V (kJ mol^{-1}) on the water accessible surface of hydantoin, using the (a) *ab initio*, and (b) planar molecular structures as calculated from DMA derived from the MP2/6-31G wave function, colour coded: -100 < white < -80 < grey < -60 < magenta < -40 < blue < -20 < cyan < 0 < green < 20 < yellow < 40 < orange < 60 < brown < 80 < red < 100.**

7.8 Conference abstract

In January 2004 around three months before the deadline for submission to the blind test, limited data regarding the crystal structure of hydantoin was found in a conference abstract[28]. This information was made freely available and it was decided that the submissions to the blind test should still go ahead, despite the predictions now not being technically 'blind'. The unit cell from the abstract was *a = 9.339(7), b = 12.187(2), c = 7.304(4), β = 104.91(2)*. The hydrogen bonding observed experimentally is shown in Figure 7.7. Although these results were known to my supervisor, they were not disclosed to me until after the submissions took place.

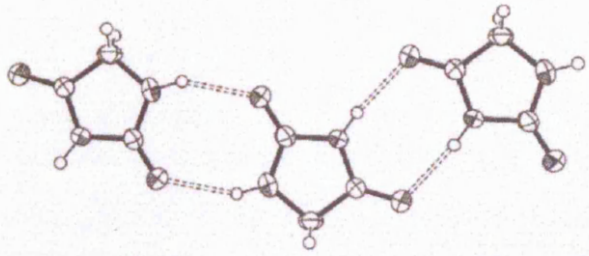


Figure 7.7 The information contained in the conference abstract on the packing in the crystal structure of hydantoin, consisting of two centrosymmetric N-H···O dimers per molecule

7.9 Submissions to the blind test

Three hypothetical crystal structures were selected for submission to the blind test. It was decided to choose one crystal structure from both computational polymorph searches and one from each based on the property calculations, Figure 7.8 and Table 7.6.

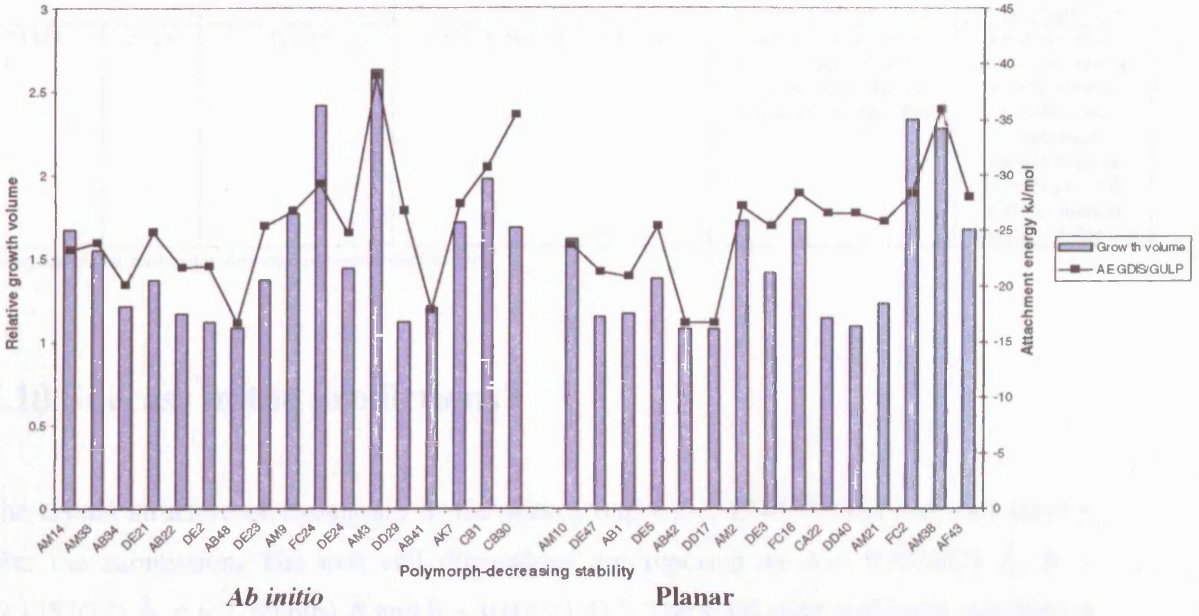


Figure 7.8 Comparison of the growth volume calculations and the minimum attachment energies for the low energy structures found in the *ab initio* and planar molecular structure polymorph searches. Only the structures that were not symmetry reduced are shown

Based on the thermodynamic and kinetic information available, the structures selected for the blind test submissions are shown in Table 7.6. In addition extended lists of the low energy structures present in both computational searches were also submitted.

Table 7.6 The three hypothetical crystal structures of hydantoin selected for submission to the blind test 2003

Structure	Space group	Lattice energy kJ mol^{-1}	Reduced cell parameters/ Å,°	Hydrogen bonding	Why chosen
AM14	$P2_1/c$	-97.766	4.028 8.325 12.248 γ100.67	Centrosymmetric dimer chains, diagonal to the *bc* plane	Based on thermodynamic grounds from both searches, Table 7.5
AM3	$P2_1/c$	-91.512	3.887 9.950 10.598 γ94.60	Hydrogen bonded sheet structure, diagonal to the *bc* plane	Highest growth volume and fastest growth of the dominant morphological face, Figure 7.8
[a]**AM2**	$P2_1/n$	-92.448	3.869 9.545 11.616 β97.29	Centrosymmetric dimer chains, diagonal to the *bc* plane with adjacent chains 45 ° to this plane	Balance between properties: average growth volume, growth of the dominant morphological face, Figure 7.8, and mechanical stability, Table 7.5

[a]From the planar molecular structure computational search

7.10 Success in the predictions?

The crystal structure of hydantoin, in the spacegroup C2/c, Z'=1[326], was received shortly after the submission. The unit cell dimensions are reported as a = 9.3538(7) Å, b = 12.1757(11) Å, c = 7.2286(6) Å and β = 104.593(4) °. The solid state molecular structure is essentially planar, except that the N1-H1 bond deviates around 7 ° from the plane of the ring, shown in Figure 7.9.

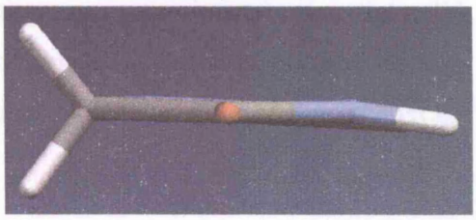
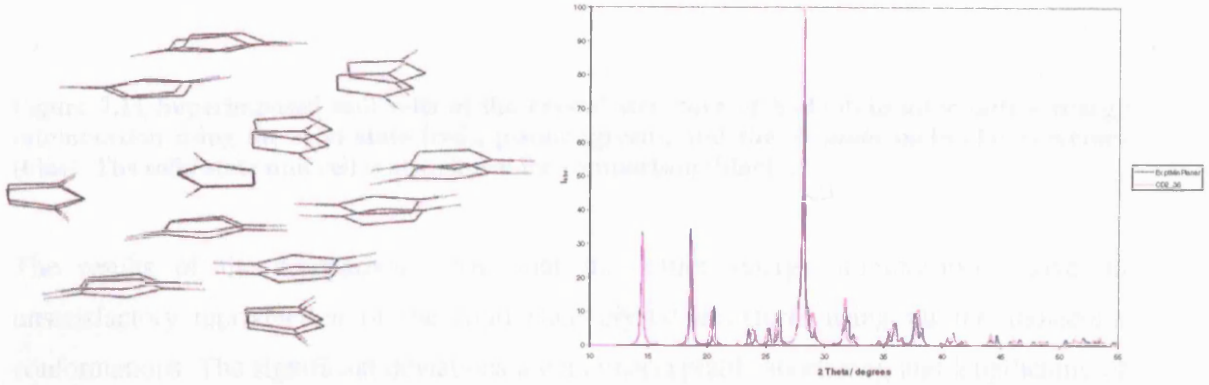


Figure 7.9 The solid state molecular structure of hydantoin, showing the deviation from planarity of the N2-H2 bond (far right of the molecule).

After careful comparisons of the low energy structures, it was found that CD34 and CD36 represent the same minima (found at the global lattice energy minimum and as the 2nd ranked structure in the *ab initio* and planar molecular structure searches respectively) and are visually close to the solid state crystal structure. Both these structures have been symmetry reduced from a Pbca to a $P2_1/n$ Z' = 2 structure. Further analysis showed that comparison of the coordination sphere about a single molecule shows a good agreement between CD36 and the solid state crystal structure (the RMS error in the atomic positions is 0.42 for the 12 molecule co-ordination sphere) Figure 7.10 (a), with a slightly inferior agreement for CD34

(the RMS error being 0.92). In addition the simulated powder patterns derived from ExptMinPlanar (section 7.11) and CD36 are in very good agreement, Figure 7.10 (b). The computational searches did not find this solid state crystal structure in the native space group C2/c. After useful discussion with Herman Ammon it was found that MOLPAK was not programmed with the particular co-ordination type present in this particular crystal structure. Therefore the trial structures generated were too far away from the correct lattice energy minimum.



(a) (b)

Figure 7.10 (a) The comparison for the co-ordination spheres for the ExptMinPlanar and CD36 structures, produced using COMPACK[161;162], and (b) comparison of the simulated powder diffractograms (calculated using Cerius2[164]) of ExptMinPlanar and CD36.

7.11 Modelling of the solid state structure

The experimental solid state crystal structure was lattice energy minimised using the solid state molecular structures. In addition, two other minimisations were performed on this crystal structure which contained the *ab initio* and planar molecular structures respectively. The results are shown in Table 7.7 and Figure 7.11.

Table 7.7 The results of the lattice energy minimisations on the solid state crystal structure of hydantoin using the solid state, planar and *ab initio* molecular structures. The values in brackets are the % errors compared to the experimental crystal structure.

Molecular Structure	Solid State	Planar (section 7.4), ExptMinPlanar	Ab initio (section 7.4), ExptMinOpt
Lattice energy/ kJmol^{-1}	-103.794	-97.292	-97.063
a/ Å	8.813(-5.78)	8.777(-6.17)	8.798(-5.94)
b/ Å	12.172(-0.03)	12.240(0.53)	12.162(-0.11)
c/ Å	7.593(5.05)	7.708(6.61)	7.713(6.69)
β/ °	101.639(-2.82)	101.497(-2.96)	101.827(-2.64)
Cell Volume/ Å^3	797.819(0.14)	811.476(1.86)	807.779(1.39)
F value	**77**	**107**	**108**

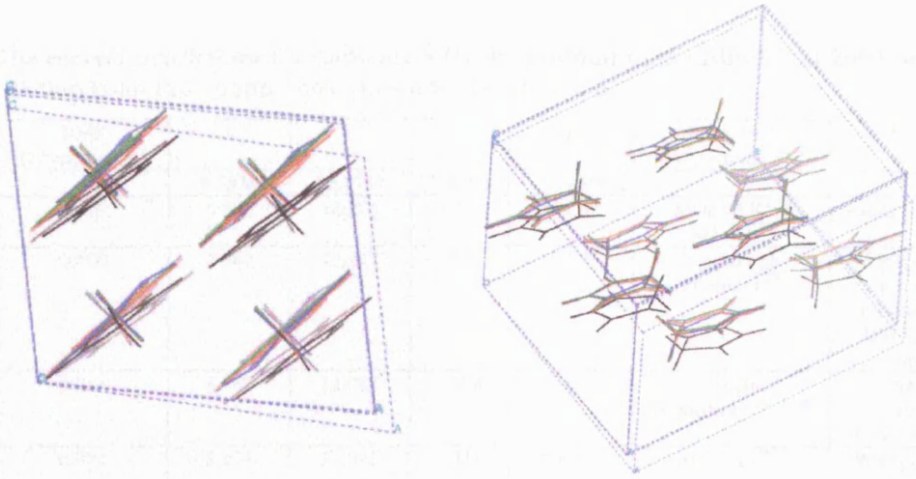


Figure 7.11 Superimposed unit cells of the crystal structure of hydantoin after lattice energy minimisation using the solid state (red), planar (green), and the *ab initio* molecular structure (blue). The solid state unit cell is also shown for comparison (black).

The results of the comparison show that the lattice energy minimisations give an unsatisfactory reproduction of the solid state crystal structure, using all the molecular conformations. The significant deviations are an unacceptable shortening and lengthening of the *a* and *c* axes respectively, thus causing the molecules to tilt and distorting the unit cell, Figure 7.11. The hydrogen bonded chains are essentially stacking along the *c* axis and are further apart suggesting an overestimation of the carbon repulsion interactions between the molecules. U_{latt}ExptMinPlanar – U_{latt}ExptMinOpt is only 0.2 kJ mol^{-1}, suggesting that the slight conformational changes in the molecular structure do not have a significant effect on the relative energies of the low energy structures.

7.12 Other participants results

Seven participants correctly predicted the crystal structure of hydantoin. These are summarised in Table 7.8.

Table 7.8 The correct predictions for molecule VIII (hydantoin) in the Blind Test 2003, with the similar prediction from this chapter also shown for comparison

	RMS (12 molecules)	a	B	c	β	Prediction program/method	Force Field
Expt		**9.3538**	**12.1757**	**7.2286**	**104.59**		
Ammon	0.479	9.008	12.283	7.758	102.62	MOLPAK[148]/ WMIN[327]	Atom centred charges
Day	0.500	8.962	12.287	7.857	102.96	Polymorph Predictor[328]	W99 pots[99], atomic point charges, reminimised using DMAREL[115] with multipoles
Facelli	0.444	8.769	12.087	7.598	101.27	Genetic Algorithm[125;126;128]	AMBER[329], restrained electrostatic potential
Pantelides	0.392	8.974	12.091	7.751	102.55	Ab Initio[39;71]	W99 pots[99], optimal site charges
Schweizer	0.552	8.665	11.836	7.481	101.81	ZIP-PROMET[139]	PIXEL[330]
van Eijck	0.377	8.745	12.216	7.722	103.29	UPACK[42]	OPLS[331;332], atomic charges
[a]Leusen	1.101	8.212	12.074	8.559	103.77	Materials Studio Polymorph[328]	CVFF, atomic charges
[b]*This research*	*0.42*	*8.908*	*12.229*	*7.620*	*101.96*	*MOLPAK[148]/ DMAREL[115]*	*Modified FIT potential[100-102], with multipoles*

[a]only matched are increasing the normal tolerances on contact distances and angles
[b]Corresponding to the CD36 crystal structure, containing the planar molecular conformation. The cell setting has been altered so that the unit cell is comparable

The majority of the correct predictions were ranked first in the relevant submissions, with the exception of Facelli (ranked 2nd) and Leusen (ranked 3rd). It should be noted that Herman Ammon used the information in the conference abstract, section 7.8, inputting the solid state hydrogen bonded dimer unit into the predictions.

A number of interesting points arose from the study. Scheranga tested the model potential used in his predictions on a variety of structures from the Cambridge Structural Database[114], including parabanic acid[72] and alloxan[73], which gave fairly acceptable reproductions of the solid state crystal structures after lattice energy minimisation. However as reported in this chapter (section 7.11), when using the crystal structure of hydantoin the results are unsatisfactory. Several other groups made similar observations hence it seems a variety of methods of modelling the intermolecular forces struggle to accurately reproduce the crystal structure of this molecule. The analysis of the submitted extended lists of the hypothetical crystal structures of hydantoin established that many of the participants found the same low energy structures, but no list appeared to be complete[333]. Therefore any consideration of thermal effects in the predictions is hindered because a fairly complete set of hypothetical structures with reliable energies is needed for more refined work[333].

The other significant point to note is that the predictions show that using more elaborate multipole techniques, rather than point charge models, did result in more low energy crystal structures containing the solid state hydrogen bonding chain motif. Approximately 45 % of the low energy structures in the computational predictions in this chapter consist of this motif, including all the structures within 3 kJ mol^{-1} of the global lattice energy minimum in

both searches. This does show that this method of modelling the electrostatic interactions, which gives a more accurate reproduction of the orientation dependence of hydrogen bonds, can give more success in predicting the correct hydrogen bonding motif in the crystalline environment. However this does depend to some extent on the prediction methods used in conjunction with these electrostatic models, as some of the successful predictions, Table 7.8, used point charges in the calculations.

Despite the inadequacies in the methods of modelling the intermolecular forces, these results do show that hydantoin is predictable by a range of crystal structure prediction techniques.

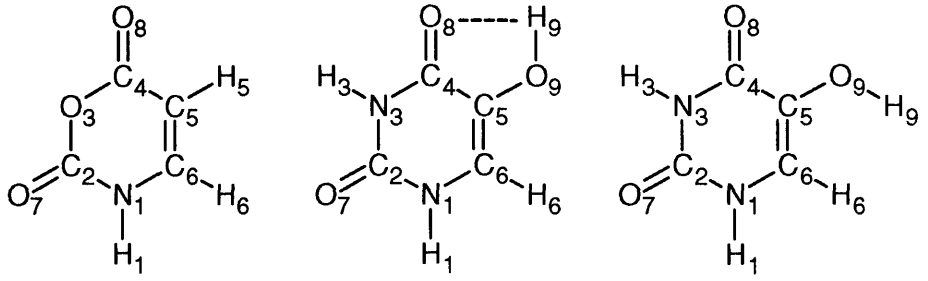
7.13 General Conclusions

The predictions on molecule VIII, hydantoin, in the third international blind test on crystal structure prediction would have been successful if the criterion to choose the three hypothetical crystal structures for submission was based solely on the lattice energy of the system. The computational predictions did not find the solid state crystal structure in the native space group C2/c. Nevertheless a very similar crystal structure was found as the same minimum in both computational searches via symmetry reduction. These correspond to the lowest and second lowest energy crystal structures for the planar and *ab initio* molecular structure computational searches respectively. It is encouraging that more elaborate electrostatic models give more accurate modelling and hence predictions of energetically favourable hydrogen bond motifs. However selecting one crystal structure through the modelling of kinetic effects, as emphasised in section 7.9, is difficult to achieve. This does highlight the current limitations in modelling the kinetic effects in crystallisation, so that crystal structure prediction can be made without relying solely on the lattice energy of the system.

8. 3-oxauracil and 5-hydroxyuracil – an informal blind test

8.1 Introduction

This chapter explores computational and experimental polymorph studies on two small, rigid molecules as part of an informal blind test scenario. The computational studies were performed at University College London, whilst Dr. Royston Copley and Ms. Lucie Deprez at GlaxoSmithKline performed a manual experimental screen to provide both crystal structures and an indication as to whether polymorphs can be readily obtained. This allowed the results of the approach to crystal structure prediction used in this thesis to be interpreted in relation to the polymorphism observed in the laboratory. The molecules that will be studied in this chapter are 3-oxauracil and 5-hydroxyuracil, shown in scheme 8.1.



(a) (b)

A **B**

Scheme 8.1 The molecular structures of (a) 3-oxauracil, and (b) 5-hydroxyuracil, showing the two conformations A and B considered in the computational work

In this chapter the computational polymorph searches used the extended version of MOLPAK[148] (section 3.1.5) with the other methodologies as before, unless stated otherwise. The predicted crystal structures found to be at a transition state in the computational searches were discarded. The experimental procedures carried out at GlaxoSmithKline are described elsewhere[49].

8.2 3-oxauracil

3-oxauracil consists of a six-membered ring, containing both a C = C bond and an anhydride group, scheme 8.1. There are a variety of hydrogen bond donors and acceptors, giving a range of hydrogen bond functionality within the crystalline environment. There was a flurry of research on 3-oxauracil in the 1970's and early 1980's, as it was found to inhibit *E Coli* growth[334], and be be effective against the herpes simplex virus type 2[335] and a variety of leukemias[336;337]. While 3-oxauracil is present in the World Drug Index[226], there were no anhydrous crystal structures known prior to these investigations.

8.2.1 *Ab initio* molecular structure

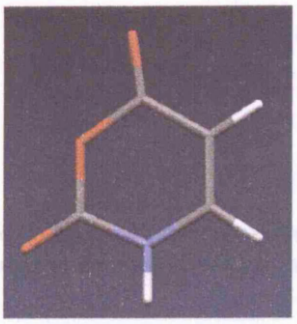


Figure 8.1 The *ab initio* molecular structure of 3-oxauracil.

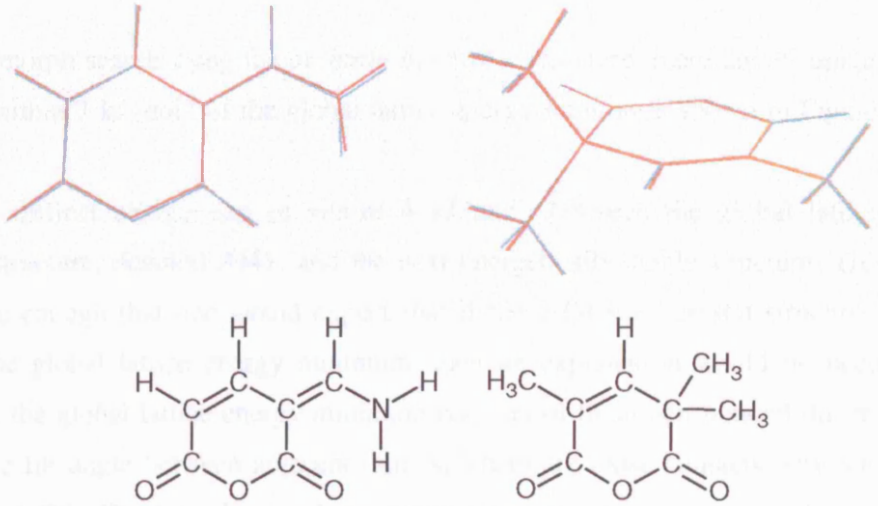
The *ab initio* molecular structure, Figure 8.1, for use in the computational polymorph screen was calculated using a MP2/6-31G** wave function (by optimisation starting from an approximate molecular structure built using MOLDEN[117]) using Gaussian98[116]. The resulting molecular structure is planar, with no obvious flexibility. The C6-H6 bond is positioned slightly towards the polar hydrogen bond donor N1-H1.

8.2.2 Related crystal structures

To investigate how well the method of modelling the intermolecular forces might be expected to reproduce the crystal structure of 3-oxauracil two crystal structures of structurally similar molecules were found in the Cambridge Structure Database[114] and tested by lattice energy minimisation. These were α-aminomethylene-glutaconic anhydride[338] (refcode AMYGLA, $P2_1/c$ Z' = 1) and 3,3,5-trimethyl-3H-pyran-2,6-dione[339;340] (refcode FIWVEM, $P2_1/m$ Z' = 0.5), both containing the characteristic anhydride group, shown in Figure 8.2.

The *ab initio* molecular structures (calculated with a MP2/6-31G** wave function using Gaussian 98[116] from the solid state molecular structure) of AMYGLA and FIWVEM are

shown in Figure 8.2, with the comparison of the solid state and *ab initio* molecular parameters shown in Table 8.1 SI.



(a) (b)

Figure 8.2 The qualitative overlays of the solid state molecular structure (blue) compared with the *ab initio* molecular structure (red) of (a) AMYGLA, and (b) FIWVEM.

For both molecules the lattice energy minimisation satisfactory reproduces the solid state crystal structures despite the molecular flexibility present, Figure 8.2 and Table 8.1. The results suggest that the modelling of the intermolecular forces is adequate for these structurally similar 3-oxauracil molecules.

Table 8.1 The results of the lattice energy minimisations on the AMYGLA and FIWVEM crystal structures. The % errors associated with the solid state crystal structure are shown in brackets.

CSD refcode	AMYGLA	AMYGLA	FIWVEM	FIWVEM
Molecular structure	Solid state	*Ab initio*	Solid state	*Ab initio*
Lattice energy/ kJ mol^{-1}	-128.343	-114.821	-90.275	-87.828
a/Å	3.663 *(-1.76)*	3.698 *(-0.80)*	6.439 *(0.13)*	6.452 *(0.33)*
b/Å	14.127 *(0.06)*	14.052 *(-0.47)*	6.594 *(-0.83)*	6.575 *(-1.12)*
c/Å	10.972 *(-0.84)*	11.038 *(-0.24)*	9.511 *(-1.25)*	9.665 *(0.34)*
β/°	89.752 *(-1.85)*	88.529 *(-3.18)*	88.582 *(-2.38)*	88.684 *(-2.26)*
Volume/Å^3	567.708 *(-2.50)*	573.439 *(-1.51)*	403.723 *(-1.97)*	409.887 *(-0.47)*
F	**9**	**13**	**9**	**8**

8.2.3 Computational search results using the *ab* initio molecular structure

In the polymorph search using the *ab initio* molecular structure, there are 30 unique crystal structures within 7 kJ mol^{-1} of the global lattice energy minimum, shown in Figure 8.3 and Table 8.2.

There is a distinct energy gap of almost 4 kJ mol^{-1} between the global lattice energy minimum structure, denoted AI41, and the next energetically stable structure. This energy gap is large enough that one would expect that if the solid state crystal structure was not found at the global lattice energy minimum, then an explanation would be needed. The structure at the global lattice energy minimum consists of hydrogen bonded dimers, with a considerable tilt angle between adjacent dimers. There are close contacts between O3 and both O7 and O8. The hypothetical low energy structures are comprised of dimers held together by two N1-H1···O7 hydrogen bonds. In the region around 5 kJ mol^{-1} above the global lattice energy minimum there are contrasting structures that form chains of molecules through N1-H1···O8 hydrogen bonds. Both the dimer and chain motifs combine to form jagged sheets or infinite ribbons in other structures. In some structures there are hydrogen bonds to the ethereal O3 (nine out of the thirty hypothetical structures). There are even hypothetical crystal structures just involving hydrogen bonds to this acceptor, forming simple chains of molecules, Figure 8.4 (c). These structures are relatively high in lattice energy, suggesting that this type of hydrogen bonding arrangement is not that energetically favourable. A selection of the different types of hydrogen bond motifs present within the low energy crystal structures are shown in Figure 8.4.

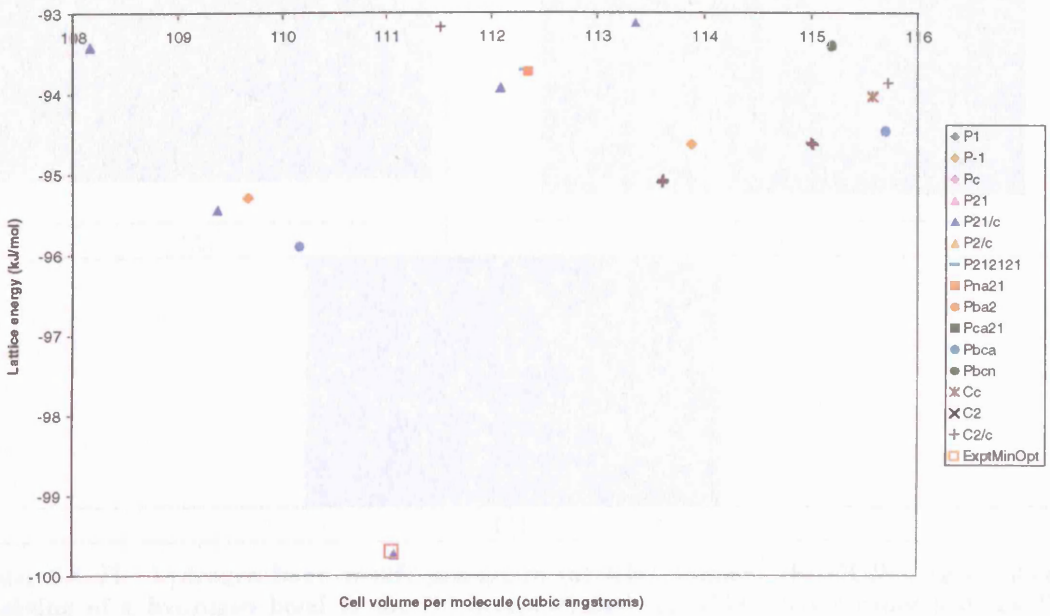


Figure 8.3 The results of the computational polymorph search on 3-oxauracil, using the *ab initio* molecular structure. Only the hypothetical crystal structures within 7 kJ mol^{-1} of the global lattice energy minimum are shown. The ExptMinOpt structure is also shown for comparison, section 8.2.4

Table 8.2 The low energy crystal structures[a] found within 5 kJ mol^{-1} of the global lattice energy minimum for 3-oxauracil, using the *ab initio* molecular structure. The full structure list within 7 kJ mol^{-1} is shown in Table 8.2 SI. The ExptMinOpt minimum is also shown for comparison, section 8.2.4

Structure	Space	Lattice Energy	[b]Free energy		[c]Reduced Cell				Hydrogen Bonding		[d]Graph set			[e]Elastic
	group	/kJ mol^{-1}	at 298 kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles/°	and motif		Level 1	Level 2	Level 3	constant
AI41	$P2_1/c$	-99.736	-113.155	1.69	5.9906	7.0975	10.5621	α 98.343	N1-H1···O7	Dimer	R2,2(8)			2.82
ExptMinopt	**$P2_1/c$**	**-99.693**	**-113.117**	**1.691**	**5.991**	**7.096**	**10.561**	**α 98.34**	**N1-H1···O7**	**Dimer**	**R2,2(8)**			**2.82**
CC28	Pbca	-95.901	-110.112	1.704	6.5229	10.5147	12.849		N1-H1···O3	Jagged sheet	C1,1(4)	R2,2(8)	C2,2(8)	1.30
									N1-H1···O7					
AK7	$P2_1/c$	-95.444	-108.753	1.717	6.6109	7.7002	8.7658	α 101.348	N1-H1···O3	Jagged sheet	C1,1(4)	R2,2(8)	C2,2(8)	1.96
									N1-H1···O7					
AB99	P-1	-95.296	-109.029	1.712	6.0797	6.6072	6.7485	α 70.941	N1-H1···O8	Infinite ribbon of dimers	C1,1(6)	R2,2(8)	R4,4(20)	1.58
								β 67.674	N1-H1···O7					
								γ 62.908						
DD64	C2/c	-95.112	-108.591	1.652	5.8322	11.0235	14.4117	α 101.189	N1-H1···O7	Dimer	R2,2(8)			2.76
AB24	P-1	-94.641	-108.469	1.649	4.1022	6.0574	9.7024	α 90.667	N1-H1···O7	Dimer	R2,2(8)			2.16
								β 95.781						
								γ 108.099						
DE57	C2/c	-94.636	-108.271	1.632	4.1434	11.1736	19.8757	α 90.413	N1-H1···O7	Dimer	R2,2(8)			6.35
CC73	Pbca	-94.465	-109.608	1.623	6.7515	7.0385	19.4769		N1-H1···O7	Ribbon	C1,1(4)			3.00
DA50	Cc	-94.039	-105.526	1.624	6.2062	6.8664	10.8491	α 90.272	N1-H1···O8	Chain	C1,1(6)			1.29

[a]All calculated structures are lattice energy minima calculated with the *ab initio* molecular structure and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK coordination geometry and order of density. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the k=0 second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated during the MOLPAK[148]/DMAREL[115] process are given for comparison. Only the reduced cell angles which are not 90° are tabulated. All structures have one molecule in the asymmetric unit. [d]Only the first three levels shown, calculated using RPluto[168]. [e]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.

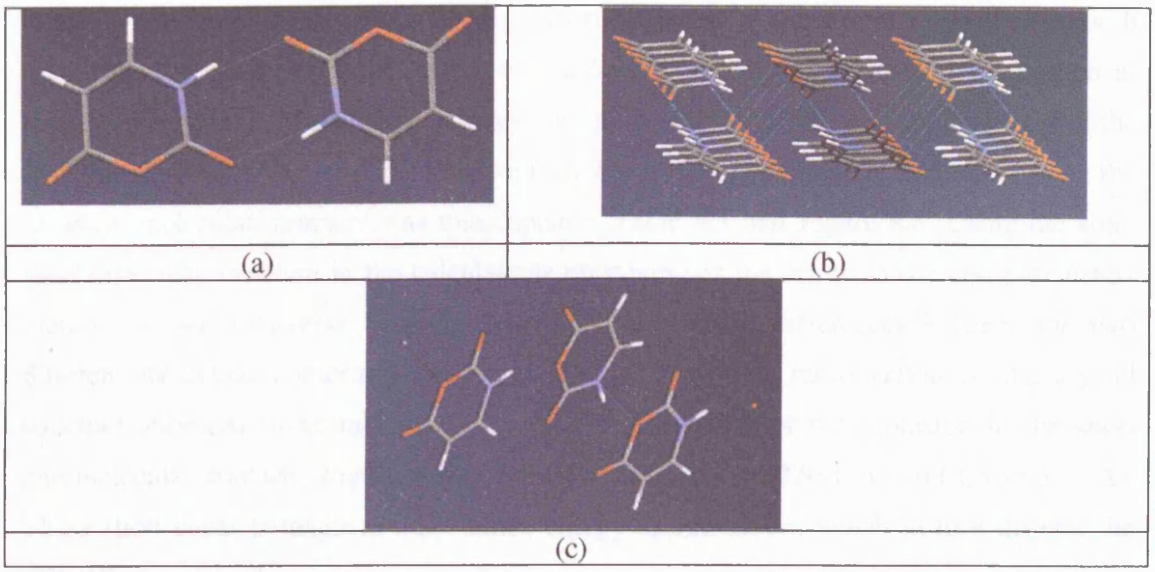


Figure 8.4 The hydrogen bond motifs present in (a) AI41 (dimer), (b) CC28 (jagged sheet, consisting of a hydrogen bond to the O3 acceptor) and (c) AI49 (chains using just the O3 acceptor)

8.2.4 Post Analysis – crystallisation of 3-oxauracil

The experimental polymorph screen at GlaxoSmithKline found both an anhydrous and a monohydrate crystal structure of 3-oxauracil[49]. The anhydrous crystal structure forms characteristic hydrogen bonded dimer units, whilst the monohydrate forms a hydrogen bonded sheet structure, Figure 8.5.

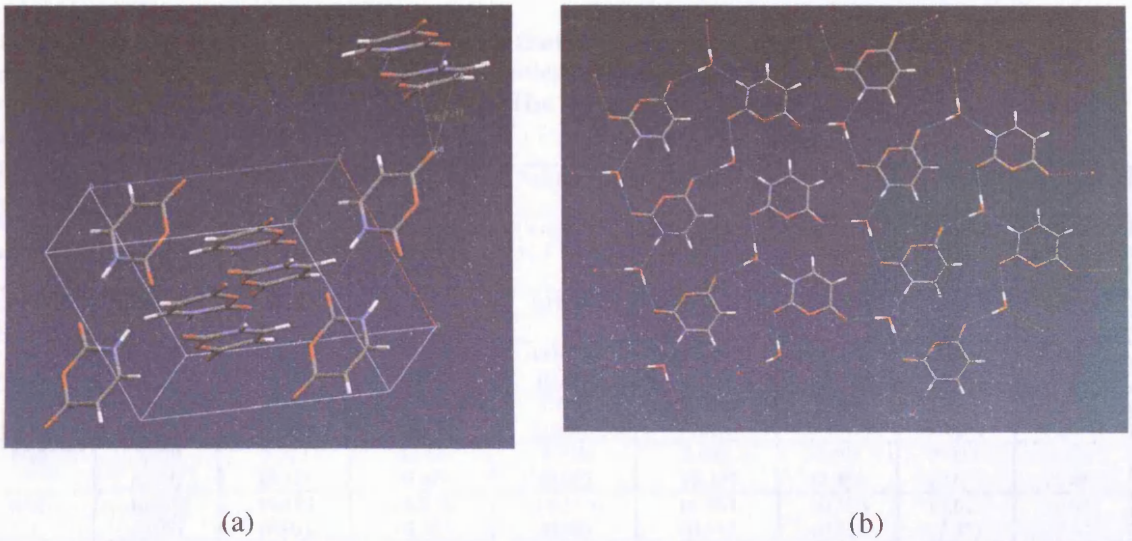


Figure 8.5 The hydrogen bonding present in (a) the anhydrous crystal of 3-oxauracil showing the hydrogen bonded dimers and the short intermolecular contacts between O8 and both C4 and C5 (shown in green) and (b) the monohydrate crystal structure showing the sheet structure.

After careful analysis of the low energy structures found in the computational polymorph search it was found that the anhydrous crystal structure corresponded to the hypothetical structure found at the global lattice energy minimum, AI41, with the packing diagrams of the two structures qualitatively comparing well. However the quantitative reproduction using the *ab initio* molecular structure was unacceptable, Table 8.3 and Figure 8.6. Using the solid state molecular structure in the calculations only reduced the errors in the cell parameters marginally, and so cannot be attributed to the very small differences between the two different molecular conformations, Table 8.3 SI. The poor reproduction of the crystal structure appeared to be mainly due to the overestimation of the repulsion in the short intermolecular contacts, Figure 8.5(a), between O8 and C4 (2.981 Å) and C5 (3.217 Å). These short contacts lengthen after lattice energy minimisation, which in turn distorts the unit cell.

8.2.5 Altering the potential

With these inadequacies in the reproduction of the solid state crystal structure (section 8.2.4), lattice energy minimisations were performed with the carbon repulsion potential parameters modified, and using both a SCF and MP2 DMA. This is to see whether this has an affect on the reproduction after lattice energy minimisation. The results are shown in Table 8.3 and Figure 8.6.

Table 8.3 Results of the lattice energy minimisations using both the solid state and *ab initio* molecular structures of 3-oxauracil. The FIT potential with and without a 25 % reduction in the carbon repulsion parameters has been used. The % errors compared to the solid state crystal structure are shown in brackets.

Molecular structure	*Solid state*	*Solid state*	*Solid state*	*Solid state*	*Ab initio*	*Ab initio*	*Ab initio*	*Ab initio*
Potential and DMA	*FITdec25, MP2 DMA*	*FITdec25, SCF DMA*	*Normal FIT, MP2 DMA*	*Normal FIT, SCF DMA*	*FITdec25, MP2 DMA*	*FITdec25 , SCF DMA*	*Normal FIT, MP2 DMA*	*Normal FIT, SCF DMA*
Lattice E kJ mol^{-1}	-106.218	-127.370	-101.353	-121.882	-99.693	-119.559	-95.005	-114.267
Cell volume/Å^3	436.766 *(-0.12)*	428.617 *(-1.99)*	449.940 *(2.89)*	441.058 *(0.86)*	444.233 *(1.59)*	436.360 *(-0.21)*	458.091 *(4.75)*	449.384 *(2.76)*
a/ Å	7.093 *(-8.37)*	7.258 *(-6.24)*	7.313 *(-5.52)*	7.439 *(-3.90)*	7.096 *(-8.32)*	7.321 *(-5.43)*	7.333 *(-5.27)*	7.510 *(-2.98)*
b/ Å	5.956 *(7.53)*	5.721 *(3.25)*	5.911 *(6.69)*	5.708 *(3.02)*	5.991 *(8.13)*	5.695 *(2.80)*	5.931 *(7.05)*	5.677 *(2.46)*
c/ Å	10.502 *(0.09)*	10.494 *(0.01)*	10.570 *(0.74)*	10.556 *(0.60)*	10.561 *(0.65)*	10.578 *(0.81)*	10.645 *(1.45)*	10.654 *(1.54)*
β/°	100.207 *(-3.31)*	100.324 *(-3.20)*	100.059 *(-3.46)*	100.230 *(-3.29)*	98.343 *(-5.11)*	98.343 *(-5.11)*	98.309 *(-5.15)*	98.358 *(-5.10)*
F value	**159**	**70**	**104**	**45**	**194**	**84**	**134**	**65**

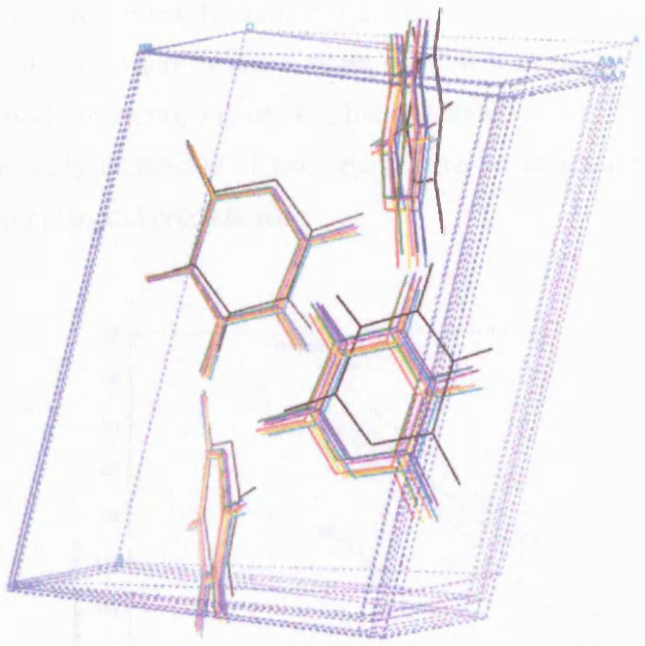


Figure 8.6 Overlay of the solid state crystal structure of 3-oxauracil (black) with the lattice energy minima using the solid-state molecular structure and the FITdec25/MP2 (green), FITdec25/SCF (blue), FIT/MP2 (pink), and FIT/SCF (purple). The lattice energy minima using the *ab initio* molecular structure are also shown, FITdec25/MP2 (red), FITdec25/SCF (brown), FIT/MP2 (yellow), and FIT/SCF (turquoise).

The results show that when using both experimental and *ab initio* molecular structures the potential without the reduction in the carbon repulsion and a SCF DMA gives the best reproduction of the solid state crystal structure. This improvement can be partly rationalised by the fact that 3-oxauracil has a characteristic anhydride group which was not present in the molecular structures against which the potential was tested, outlined in chapter 5. This demonstrates that there is still considerable potential to improve the modelling of the intermolecular forces for this molecule.

8.2.6 Computational polymorph search using the original FIT potential

To see whether using a different model for the intermolecular forces has a significant effect on the relative energies of the low energy crystal structures an additional polymorph search was performed. This used the FIT potential without the reduction in the carbon repulsion parameters and a SCF DMA (section 8.2.5), along with the *ab initio* molecular structure. Only a select number of space groups were explored, giving the results shown in Figure 8.7. The solid state crystal structure (deemed ExptMinOpt2), is still found at the global lattice energy minimum and is 3.9 kJ mol^{-1} more stable than the next ranked structure. It is now the densest structure. The search has significantly fewer hypothetical crystal structures within energy range of polymorphism than the previous computational screen. Nevertheless with the hydrogen bonded dimer motif still predicted to be the most energetically stable. The

hydrogen bonded chain motif is around 6 kJ mol^{-1} (4 kJ mol^{-1} at room temperature estimates) less stable. There is also a slight reordering of the relative energies for the other predicted crystal structures, Figure 8.7, but the stability of ExptMinOpt2 compared to the others is relatively unaffected. These results give an increased confidence in the reliability of the computational predictions.

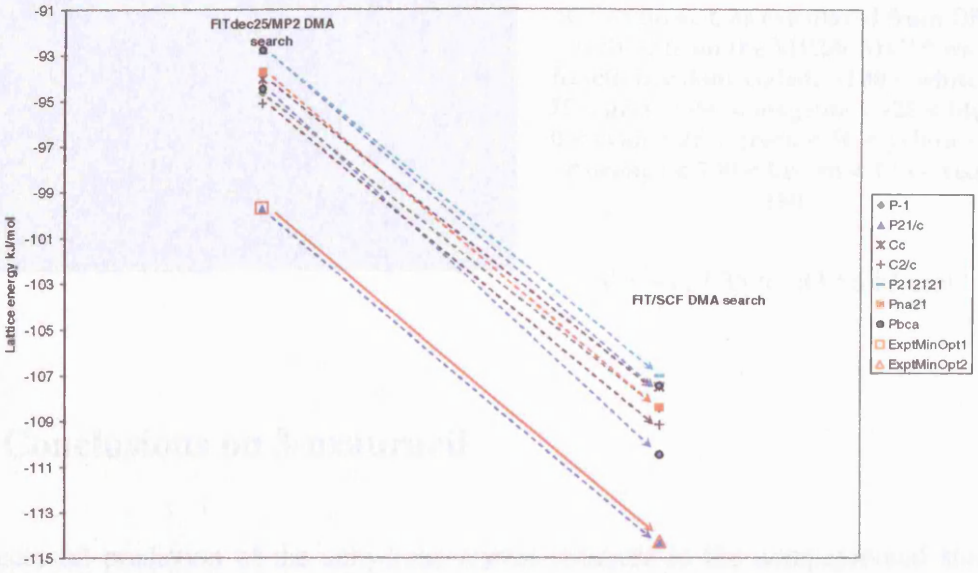


Figure 8.7 A comparison of the relative energies within the equivalent low energy structures in both computational polymorph searches on 3-oxauracil, within 7 kJ mol^{-1} of the global lattice energy minimum. The ExptMinOpt structures (denoted 1 and 2 respectively) are also shown for comparison.

8.2.7 Electrostatic potential

The electrostatic potential on the water accessible surface was calculated (Figure 8.8) to give an indication as to why the characteristic dimer unit is predicted to be the most stable motif in the computational predictions. The electrostatic potential shows a large negative region around the anhydride group in the molecule. This is similar to the situation observed for alloxan[64], chapter 5, which contains three adjacent carbonyl groups. The computational predictions show some hydrogen bonds to the middle ethereal oxygen, despite these crystal structures being less energetically competitive than the hydrogen bonded dimers. The ether oxygen has competing interactions between two adjacent carbonyl groups, and for similar ester type compounds it was found that the oxygen atoms adjacent to carbonyl groups are not intrinsically worse as a hydrogen bond acceptor[341]. However if hydrogen bonds to the

ethereal oxygen atom are weak for steric reasons then one might imagine that the compound crystallises in the hydrogen bonded dimer motif as unfavourable interactions are minimised.

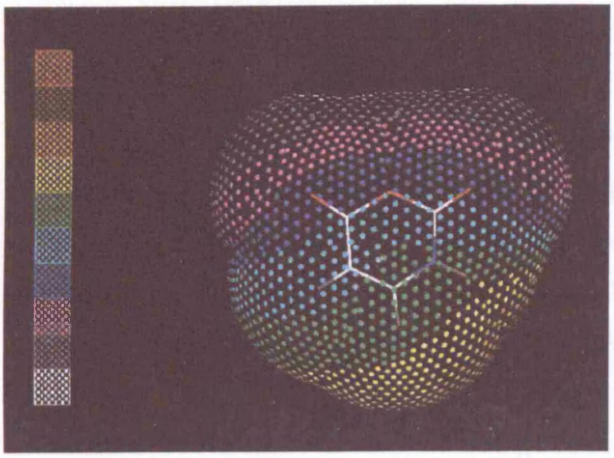


Figure 8.8 The electrostatic potential V (kJ mol^{-1}) on the water accessible surface of 3-oxauracil, as calculated from DMA derived from the MP2/6-31G wave function, colour coded: -100 < white < -75 < grey < -50 < magenta < -25 < blue < 0 < cyan < 25 < green < 50 < yellow < 75 < orange < 100 < brown < 125 < red < 150**

V = +127.35 to -83.56 kJ mol^{-1}

8.2.8 Conclusions on 3-oxauracil

The successful prediction of the anhydrous crystal structure in the computational studies with an energy difference of 4 kJ mol^{-1} between this structure at the global lattice energy minimum and the next most stable structure is unusually large[342] for small rigid molecules. However the energy gap is not so large to completely rule out the possibility of polymorphism on purely thermodynamic grounds, especially since the gap is reduced to 3 kJ mol^{-1} at room temperature. The prediction that this should be the observed crystal structure could be made with some confidence, as the energy gap is significant compared with the likely errors in the relative lattice energies. It is notable that this energy gap is fairly insensitive to the model potential used in the calculations. Alternative hypothetical structures with a hydrogen bonding motif of a N1-H1···O8 chain are nearly 5 kJ mol^{-1} less stable in lattice energy. Therefore the computational search and experimental screen gives reasonable confidence that further long lived polymorphs are unlikely.

8.3 5-hydroxyuracil

5-hydroxyuracil, Scheme 8.1, consists of a six-membered ring, containing a C = C double bond. There are three hydrogen bond donors and acceptors, giving a range of possibilities for hydrogen bonding. 5-hydroxyuracil has been found to be a weak mutagene inducing

intergenic mutations in *Vicia faba* seeds[343]. There were no anhydrous or solvate crystal structures known for this compound prior to these investigations.

8.3.1 *Ab initio* conformational analysis

The hydroxyl group allows some molecular flexibility and hence two conformations were considered for the computational work. In conformation A, scheme 8.1, there is an intramolecular hydrogen bond to the adjacent carbonyl group, while in conformation B this hydrogen bond is absent. The question arises as to whether conformation B can form more favourable intermolecular hydrogen bonds to give increased stability within the crystal lattice.

An *ab initio* conformational analysis was performed, using Gaussian98[116] at SCF/6-31G** level, to investigate the potential energy surface associated with this molecular flexibility. This analysis involved altering the torsion angle C4C5O9H9 (scheme 8.1) from 0 ° to 360 ° in 10 ° steps, and relaxing the rest of the molecule, shown in Figure 8.9. The global energy minimum is located when the torsion angle C4C5O9H9 is 180 °, corresponding to conformation A. There is also a shallow local minimum present at 360 °, corresponding to conformation B. At the MP2 level of calculation, there is also a 31 kJ mol^{-1} difference in the intramolecular energy between the two conformations (ΔE_{intra}), of the order of the energy of a single hydrogen bond (i.e. $\Delta E_{intra}{}^{A} = 0$ kJ mol^{-1} and $\Delta E_{intra}{}^{B} = 31$ kJ mol^{-1}).

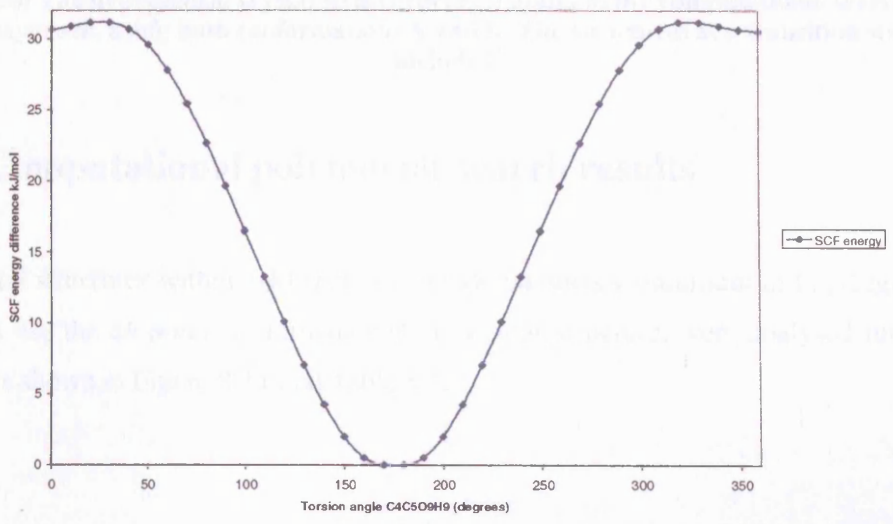


Figure 8.9 Conformation energy scan of the 5-hydroxyuracil molecule, altering the torsion angle C4C5O9H9, showing the SCF energy difference compared to conformation A.

8.3.2 Relative energies of the low energy structures

Both conformations A and B, scheme 1, were used in the computational polymorph search to see whether the differences in ΔE_{intra} could be compensated for by differences in the relative lattice energies of the low energy structures. The results of the two computational polymorph searches were compared by considering the total energy of the crystals where $E_{tot} = E_{latt} + \Delta E_{intra}$. The results are shown in Figure 8.10. These results show that conformation A produces better crystal lattice energies, over 10 kJ mol^{-1} more than stable than the lowest found for conformation B.

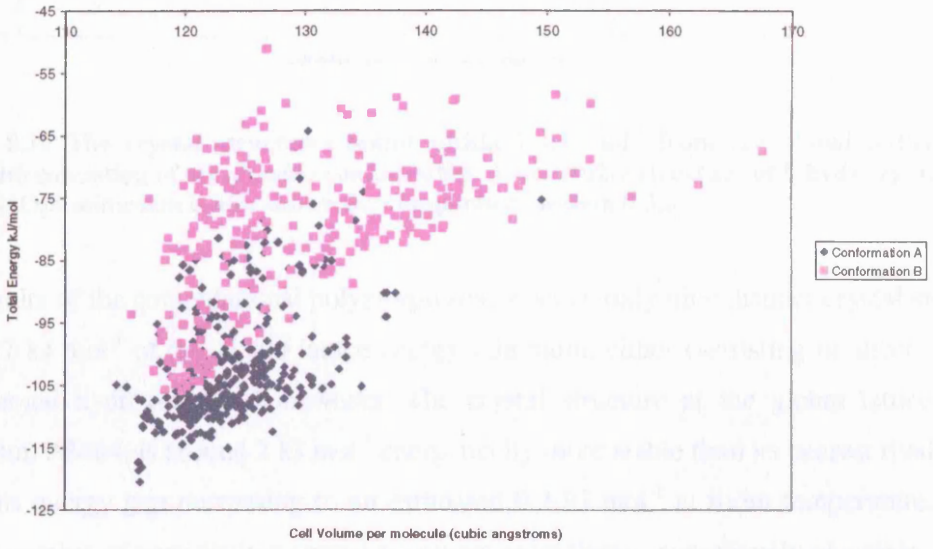


Figure 8.10 The hypothetical crystal structures (E_{tot}) found in the computational searches on 5-hydroxyuracil, using both conformations A and B. The structures at a transition state are included

8.3.3 Computational polymorph search results

The crystal structures within 7 kJ mol^{-1} of the global energy minimum in E_{tot} (Figure 8.10), which all use the *ab initio* conformation A molecular structure, were analysed further with the results shown in Figure 8.11 and Table 8.4.

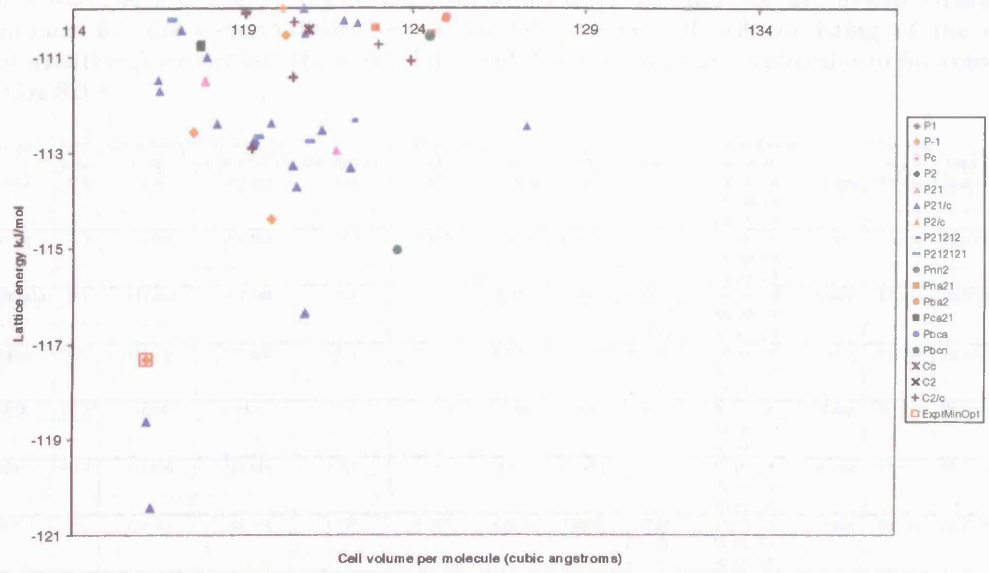


Figure 8.11 The crystal structures found within 7 kJ mol^{-1} from the global lattice energy minimum consisting of the *ab initio* conformation A molecular structure of 5-hydroxyuracil. The ExptMinOpt minimum is also shown for comparison, section 8.3.5

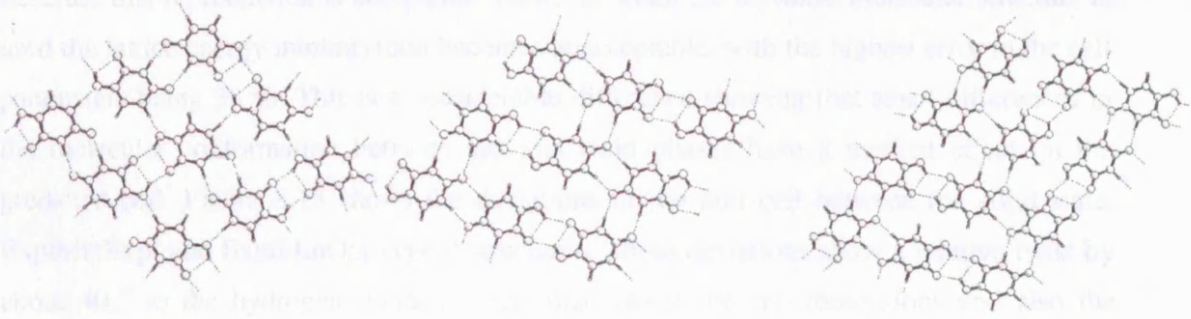
The results of the computational polymorph search show only nine distinct crystal structures within 7 kJ mol^{-1} of the global lattice energy minimum, either consisting of sheet or three-dimensional hydrogen bond networks. The crystal structure at the global lattice energy minimum, AM64, is around 2 kJ mol^{-1} energetically more stable than its nearest rival AM10, with this energy gap decreasing to an estimated 0.2 kJ mol^{-1} at room temperature. Thus a limited number of very distinct crystal structures seem thermodynamically plausible.

The intermolecular hydrogen bond acceptors and donors used differ between the low energy crystal structures, with some consisting of molecules that use all the hydrogen bond acceptors in the crystal, whilst the hydroxyl O9 acceptor is not used in the others. The majority of the low energy structures consist of hydrogen bonded sheets, showing subtle variations that give rise to three different motifs, shown in Figure 8.12.

Table 8.4 The low energy crystal structures within 7 kJ mol^{-1} of the global lattice energy minimum for the computational search on 5-hydroxyuracil, all consisting of the *ab initio* conformation A molecular structure[a]. The ExptMinOpt minimum is also shown for comparison, section 8.3.4

Structure	Space	Lattice Energy	[b]Free energy		[c]Reduced Cell				Hydrogen Bonding		[d]Graph set			[e]Elastic
	group	/kJ mol^{-1}	at 298 K/ kJ mol^{-1}	Density/g cm^{-3}	a/Å	b/Å	c/Å	Angles/°	and motif		Level 1	Level 2	Level 3	constant
AM64	$P2_1/c$	-120.427	-133.223	1.829	3.4572	10.5288	12.881	β 97.18	N1-H1···O9	Sheets	S1,1(5)	R2,2(10)	C1,1(5)	2.07
									O9-H9···O8					
									N3-H3···O7					
AM10	$P2_1/c$	-118.632	-133.026	1.831	3.3806	10.0209	13.9051	γ 99.41	N1-H1···O9	3D	S1,1(5)	R2,2(10)	R2,2(10)	2.40
									O9-H9···O8					
									N3-H3···O7					
ExptMinOpt	**P-1**	**-117.323**	**-127.364**	**1.831**	**6.151**	**6.371**	**7.232**	**α 66.78**	**O9-H9···O8**	**Sheets**	**S1,1(5)**	**R2,2(10)**	**R2,2(8)**	**1.92**
								β 80.05	**N3-H3···O7**					
								γ 63.12	**N1-H1···O7**					
AB90	P-1	-117.339	-127.092	1.831	6.1514	6.3701	7.2316	α 66.80	O9-H9···O8	Sheets	S1,1(5)	R2,2(10)	R2,2(8)	3.15
								β 80.07	N3-H3···O7					
								γ 63.13	N1-H1···O7					
AM39	$P2_1/c$	-116.382	-128.11	1.76	5.7353	7.8482	10.8837	γ 99.42	N1-H1···O9	Sheets	S1,1(5)	R2,2(10)	C1,1(5)	0.14
									O9-H9···O8					
									N3-H3···O7					
CD97	Pbca	-115.058	-129.455	1.721	6.057	10.8504	15.0413		N1-H1···O9	Sheets	S1,1(5)	R2,2(10)	C1,1(5)	0.12
									O9-H9···O8					
									N3-H3···O7					
CA76	P-1	-114.415	-125.184	1.775	5.1982	6.0124	7.8866	α 82.26	O9-H9···O8	Sheets	S1,1(5)	R2,2(10)	R2,2(8)	7.71
								β 89.59	N3-H3···O7					
								γ 78.98	N1-H1···O7					
FC6	$P2_1/c$	-113.734	-124.631	1.764	5.4293	7.2586	12.4002	β 99.20	O9-H9···O8	Sheets	S1,1(5)	R2,2(10)	C1,1(4)	0.51
									N3-H3···O7					
									N1-H1···O7					
AM98	$P2_1/c$	-113.344	-125.397	1.741	5.1852	7.8563	12.0026	β 91.774	N3-H3···O7	3D	S1,1(5)	C1,1(7)	C1,1(6)	2.16
									N1-H1···O8					
									O9-H9···O7					
AK1	$P2_1/c$	-113.301	-125.875	1.765	5.3006	7.5281	12.3354	α 107.72	O9-H9···O8	3D	S1,1(5)	C1,1(5)	C1,1(4)	5.44
									N3-H3···O7					
									N1-H1···O7					

[a]All calculated structures are lattice energy minima calculated with the *ab initio* molecular model of conformation A and the same intermolecular potential. The hypothetical structures are labelled according to the initial MOLPAK coordination geometry and order of density. [b]The Helmholtz free energy is estimated from the lattice energy, zero point intermolecular energy and temperature dependence of the rigid molecule internal energy and entropy, as derived from the $k = 0$ second derivative properties[52]. [c]The Niggli reduced cell parameters[157] as calculated during the MOLPAK/DMAREL procedure are given for comparison. Only the reduced cell angles which are not 90° are tabulated. All structures have one molecule in the asymmetric unit. [d]Only the first three levels shown, calculated using RPluto[168]. [e]The smallest eigenvalue of the lower right sub-matrix of the elastic stiffness constants, GPa.



(a) (b) (c)

Figure 8.12 The variations in the hydrogen bonding sheets present in (a) AM64, (b) AB90, and (c) FC6

The different hydrogen bond acceptors and donors used in these sheet structures suggests that they are unlikely to transform to one another as these rearrangements would involve breaking hydrogen bonds. The computational search also predicts three-dimensional hydrogen bonded structures, only around 2 kJ mol^{-1} less stable than the global lattice energy minimum, indicating another type of hydrogen bonding motif might be energetically competitive.

8.3.4 Post analysis – crystallisation of 5-hydroxyuracil

The experimental polymorph screen at GlaxoSmithKline yielded an anhydrous form, and a DMSO solvate structure. This solvate structure could not be refined adequately due to twinning and disorder[49].

8.3.5 Modelling of the solid state structure

It was found that the solid state crystal structure is visually close to that of AB90, Table 8.4, which was the third most stable in the computational search. There are two more energetically stable structures, AM10 and AM64, which have N1-H1···O9 hydrogen bonds in contrast to the N1-H1···O7 present in the solid state.

The solid state molecular conformation is very similar to the one used for the low energy structures in the computational predictions. However the ring is non-planar, with the torsion angle C6C5O9H9 being around 4 °, and the C4 = O8 bond around 4 ° out of planarity compared to the planar *ab initio* conformation. A comparison of the molecular parameters for this *ab initio* and solid state molecular structures are shown in Table 8.4 SI.

The new solid state crystal structure was used in the lattice energy minimisations to see how well this crystal structure is reproduced, Table 8.5. When using the solid state molecular structure this reproduction is acceptable. However when the *ab initio* molecular structure is used the lattice energy minimisation becomes unacceptable, with the highest error in the cell parameters being 33 %. This is a considerable difference showing that small differences in the molecular conformation between gas and solid phases have a marked effect on the predicted cell. Figure 8.13 shows the deviations in the unit cell between the solid state, ExptMinExpt and ExptMinOpt crystal structures. These deviations show a relative twist by about 40 ° to the hydrogen bonded sheets that makes the cell dimensions and also the simulated powder patterns disparate. The hydrogen bonded sheets show some degree of expansion, Figure 8.13, however there is not a significant change in the relative sheet stacking.

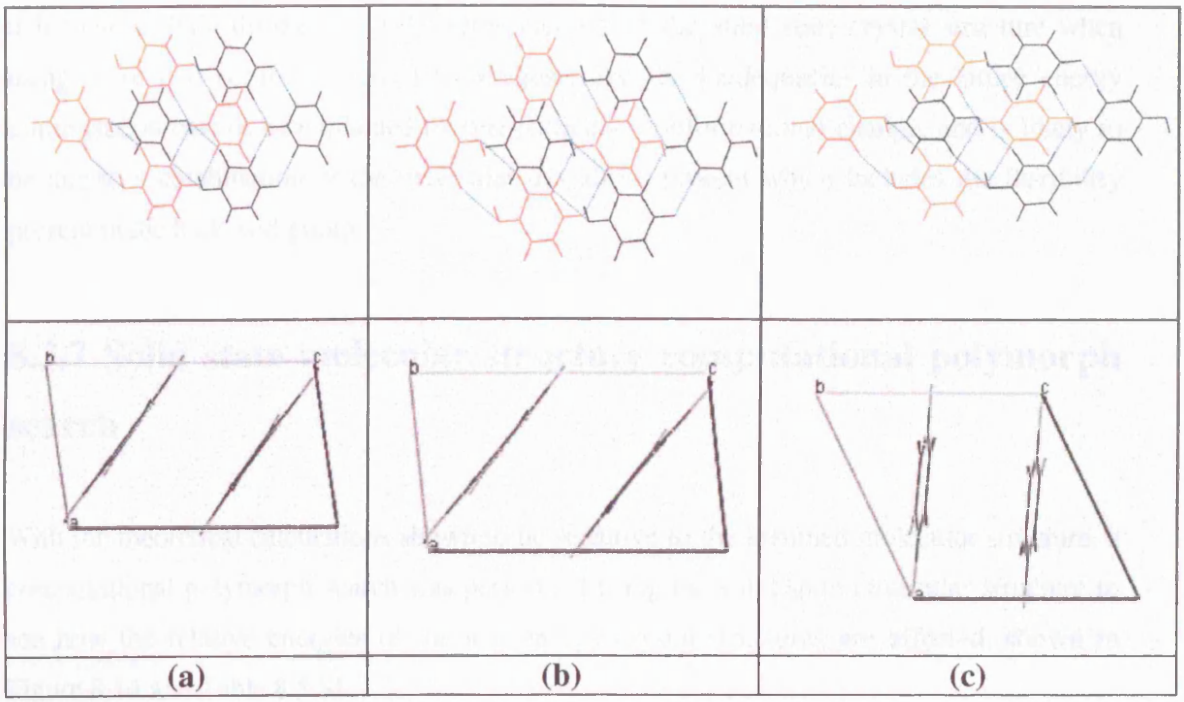


Figure 8.13 The overlay of two hydrogen bonded sheets and the view along the *b* cell axis (not to scale) for the (a) solid state, (b) ExptMinExpt, (c) and ExptMinOpt crystal structures.

8.3.6 Sensitivity to the molecular deviations

Table 8.5 The results of the lattice energy minimisations on the solid state crystal structure using both the *ab initio*, modified *ab initio* and solid state molecular structures.

	Expt	ExptMinExpt[a]	ExptMinOpt[a]	Modified ExptMinOpt[a]	Modified ExptMinOpt[a]
Molecular structure		Solid state	*Ab initio*	*Ab initio* ∠ C6C5C4O8 fixed[b]	Ab initio ∠ N3C4C5O9 fixed[c]
Lattice energy kJ mol^{-1}		-123.633	-117.232	-117.445	-117.387
a/Å	4.620	4.510 *(-2.37)*	6.151 *(33.16)*	6.161 *(33.36)*	6.166 *(33.48)*
b/ Å	7.049	7.173 *(1.75)*	7.232 *(2.59)*	7.241 *(2.71)*	7.235 *(2.64)*
c/ Å	7.345	7.289 *(-0.77)*	6.371 *(-13.26)*	6.367 *(-13.32)*	6.361 *(-13.39)*
α/°	88.168	89.751 *(1.80)*	113.221 *(28.42)*	113.283 *(28.49)*	113.264 *(28.46)*
β/°	81.503	83.986 *(3.05)*	63.123 *(-22.55)*	62.904 *(-22.82)*	62.876 *(-22.85)*
γ/°	86.245	88.279 *(2.39)*	99.949 *(15.89)*	99.944 *(15.88)*	99.984 *(15.93)*
Volume/ Å^3	235.997	234.396 *(-0.68)*	232.314 *(-1.56)*	232.226 *(-1.60)*	232.055 *(-1.67)*
F value		**29**	**2995**	**3013**	**3033**

[a]After lattice energy minimisation. [b]Keeping the C6C5C4O8 torsions angle fixed to 4 ° during optimisation. [c]Keeping the N3C4C5O9 torsion angle fixed to 177 ° during optimisation.

To determine whether the lattice energy minimisations are sensitive to the small molecular deviations, two different *ab initio* conformations were used, outlined in Table 8.5. Conopt1 has the C6C5C4O8 torsion angle at the solid state value of 4 °, with the rest of the molecule relaxed during the *ab initio* optimisation. Conopt2 has the torsion angle N3C4C5O9 fixed at the solid state value of 177 °, with the rest of the molecule allowed to relax. This will highlight whether the small deviations in C4 = O8 and C5-O9 from ring planarity (as seen in the solid state) have a significant effect on the lattice energy minimisation. The results show

that there is little difference in the reproductions of the solid state crystal structure when using these two conformations. This suggests that the inadequacies in the lattice energy minimisation cannot be explained by one particular conformational change, and is likely to be due to a combination of the molecular deviations present which includes the flexibility present in the hydroxyl group.

8.3.7 Solid state molecular structure computational polymorph search

With the theoretical calculations shown to be sensitive to the assumed molecular structure, a computational polymorph search was performed using the solid state molecular structure to see how the relative energies of the low energy crystal structures are affected, shown in Figure 8.14 and Table 8.5 SI.

The non-planar molecular structure generated significantly more low energy minima, with ExptMinExpt found as the forth most stable crystal structure, 2 kJ mol^{-1} above the global lattice energy minimum and increasing to 3 kJ mol^{-1} at room temperature estimates. This is similar to the relative energies found in the *ab initio* molecular structure polymorph search. The relative stability of the low energy crystal structures to the known structure is maintained in the majority of cases, shown in Figure 8.14, even when the small molecular distortions found in the solid state structure are included, and these hypothetical structures might be further stabilized by further distortions of the molecule from planarity. Again, the lower energy structures include sheet and three-dimensional networks, with the majority of hypothetical crystal structures consisting of molecules that use only O7 and O8 as hydrogen bond acceptors.

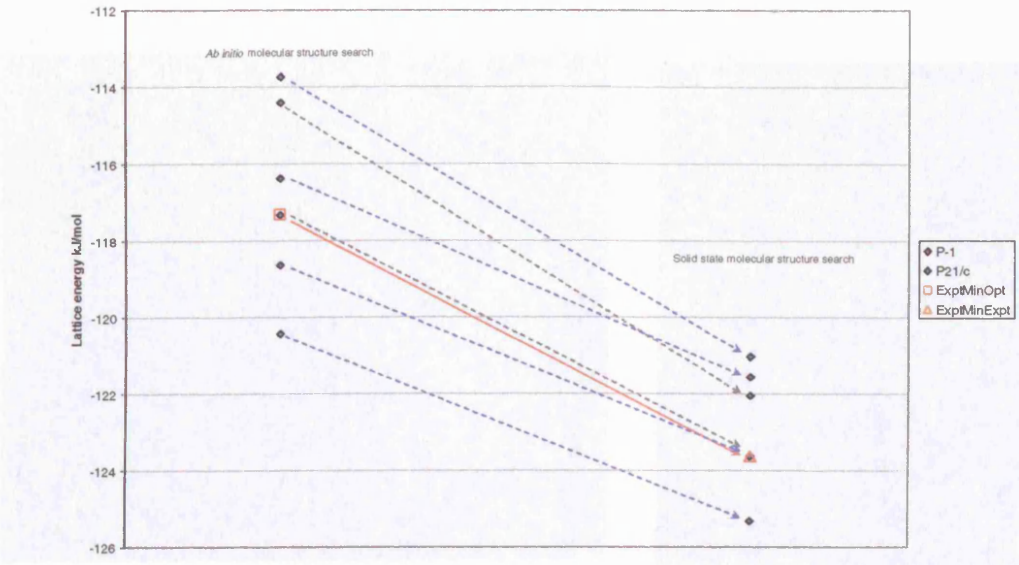
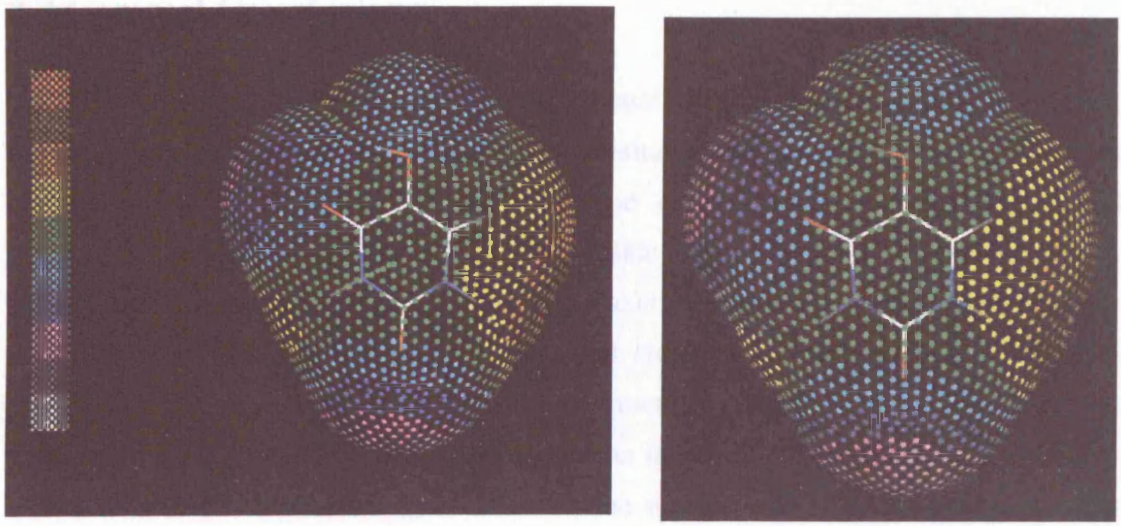


Figure 8.14 A comparison of the relative energies within the equivalent low energy structures in both the computational searches on 5-hydroxyuracil, within 7 kJ mol^{-1} of the global lattice energy minimum. The ExptMinOpt and ExptMinExpt are also shown for comparison.

8.3.8 Electrostatic potential

To determine whether the limited molecular flexibility present has a significant effect on the intrinsic electrostatic contribution to the intermolecular hydrogen bonding, the electrostatic potential of the water accessible surface was calculated for both the *ab initio* and solid state conformations, Figure 8.15. These results show that these small molecular changes do have an effect on the relative strengths of the potential. One example is shown in the *ab initio* conformation where there is a slightly more negative electrostatic potential around the hydroxyl group than in the solid state (when looking at this particular area on the potential). This is also reflected in the V_{min} and V_{max}, with differences of around 6 kJ mol^{-1} between the two conformations. This could suggest significant differences in the electrostatic contribution to the lattice energy, and hence could have an affect on the relative energies of the low energy structures.



(a) V = +106.77 to -57.46 kJ mol^{-1}

(b) +112.01 to -64.28 kJ mol-1

Figure 8.15 The electrostatic potential V (kJ mol^{-1}) on the water accessible surface around 5-hydroxyluracil, for the (a) *ab initio* and (b) solid state conformations, as calculated from DMA derived from the MP2/6-31G wave function, colour coded: white < -80 < grey < -60 < magenta < -40 < blue < -20 < cyan < 0 < green < 20 < yellow < 40 < orange < 60 < brown < 80 < red.**

8.3.9 Conclusions on 5-hydroxyuracil

In the computational predictions the anhydrous crystal structure was predicted qualitatively correctly as the third most stable structure in the search. The deficiencies in the computational model, mainly due to the sensitivity of the calculations to slight molecular deviations, could be due to the limited molecular flexibility associated with the hydroxyl group. This has a significant effect on the electrostatic potential around the molecule.

The alternative energetically feasible structures differ from the observed structure in such a way that it is plausible that such structures could be formed. The kinetics of transformation between the polymorphs may be sufficiently slow to allow the observation of a metastable form. However although the solvent crystallisation scheme did not find any polymorphs this does not exclude their formation by alternative crystallisation methods.

8.4 General Conclusions

The computational predictions on 3-oxauracil and 5-hydroxyuracil would have been successful in a blind test scenario despite the quantitative inaccuracies in the reproductions of the solid state crystal structures. This type of collaborative computational and experimental study greatly aids the predictive side of this area of scientific research. Nevertheless this study does highlight how sensitive crystal structure prediction can be to the model potential used and the assumed molecular structure, giving significantly different patterns and relative energies of the hypothetical structures. Being able to predict the crystal structure of a given molecule is very dependent on the energy distribution of the possible structures, and this distribution is so reliant on the specific molecule that the calculations need to be performed before the assessment of whether the crystal structures can be predicted with any confidence can be made. Two anhydrous crystal structures were obtained relatively easily in the experimental polymorph screens, despite simple non-solvated crystal structures proving elusive for other small molecules. Nevertheless these two molecules add to the examples of successful predictions based on lattice energy minimisation without prior knowledge of the crystal structures[28-30].

9. Conclusions and recommendations for future work

This thesis has used both computational and experimental polymorph studies to investigate a variety of small organic molecules, with the results summarised in Table 9.2. These chemically related molecules, with mainly N-H hydrogen bond donors and C=O acceptors, show a range of distributions of the relative energies of the predicted hypothetical crystal structures.

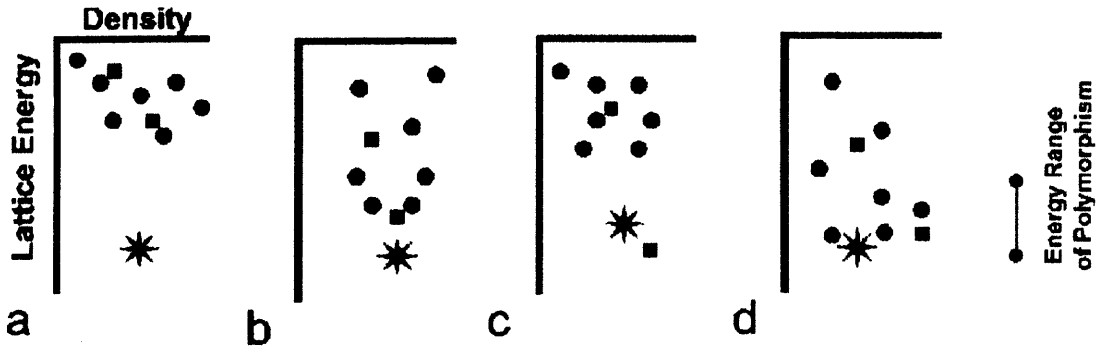


Figure 9.1 Four possible scenarios for the crystal energy landscape[65], compared with the energy range of polymorphism. The star represents the experimental solid state crystal structure, with the other symbols representing the other low energy structures generated in the computational search

Figure 9.1 shows four possible scenarios for the crystal energy landscape of the low energy structures[65], ignoring the errors in the lattice energy calculations and contrasting the relative energies with the energy range of polymorphism.

In Figure 9.1, scenario *a* is when one form is much lower in energy than the other hypothetical structures, which suggests that finding polymorphs is very unlikely. This would be the ideal scenario to find during polymorph studies on a particular compound in the pharmaceutical industry, as this would be some conformation that the known structure is the only likely polymorph. The computational predictions on 3-oxauracil in chapter 8 is the closest example to this scenario. Scenario *b* is when the experimental crystal structure is found at the global lattice energy minimum, with a few systems close in energy. This suggests that metastable polymorphs could be observed if there was a barrier of transformation to the known form. In scenario *c*, the experimental crystal structure is not found at the global lattice energy minimum warning that a more stable polymorph might be

found given the right kinetic route. This would be the nightmare scenario for the pharmaceutical industry, nevertheless the structure predicted more thermodynamically stable would be known through these predictions and hence the crystallisation experiments could be tailored in an attempt to find this polymorph[190;191]. Scenario *d* is when a number of predicted structures have virtually identical energies, in which the kinetic factors will determine the polymorphs found. Table 9.1 outlines the molecules studied for crystal structure prediction in this thesis and which of the scenarios they correspond to.

Table 9.1 The molecules used for crystal structure prediction in this thesis, corresponding to one of the four possible scenarios for the crystal energy landscape. The *ab initio* molecular structure is used unless otherwise stated

Scenario A	Closest example being 3-oxauracil
Scenario B	3-oxauracil, cytosine (solid state)
Scenario C	[a]Barbituric acid, cyanuric acid, parabanic acid, urazole, hydantoin, 5-hydroxyuracil, uric acid, 6-methyluracil, thymine, cytosine
Scenario D	Alloxan, urazole (solid state), hydantoin (planar), uric acid (solid state)

[a]ExptMinOpt/ExptMinExpt not found in the computational polymorph search, however would have corresponded to this scenario if a more complete search had been perfomed. Adenine and guanine are not included as they have no known solid state crystal structures

The energy range of polymorphism for use in the computational polymorph searches is difficult to gauge as the approximate energy range for real polymorphs[5] is often 0 – 10 kJ mol^{-1}. For all the scenarios shown in Figure 9.1 increasing the energy range by 1 kJ mol^{-1} can give a substantial increase in the number of unique crystal structures found. This trend is usually observed for other lattice energy polymorph searches for small, organic molecules[50], and hence a cutoff of 7 kJ mol^{-1} above the global lattice energy minimum was used as a compromise for the computational searches in this thesis.

With the survey of molecules in this thesis (Table 9.2) it has been shown that the inadequacies in the model potential is still a major factor in how reliable crystal structures can be predicted, highlighted by the fact that the carbon repulsion parameters needed to be reduced to satisfactorilly reproduce a range of similar heterocyclic crystal structures after lattice energy minimisation, shown in chapters 5 and 6. Even though it is worthwhile to increase the accuracy of modelling these intermolecular forces[344], it was found that in just under half of the molecules studied in this thesis (taking into account all the molecular conformations considered) the solid state crystal structures were found at the global lattice energy minimum. This success using these DMA based potentials is consistant with other work[50;345] performed during the duration of this PhD. Nevertheless improving the model potential[50] and the method of modelling the electrostatic forces (from point charges to using a DMA[113]) usually increases the number of possible hydrogen bond donor-acceptor combinations leading to less reliable predictions. It is also encouraging that the use of atomic

multipoles gives a fairly accurate representation of the electrostatic interactions[113], as highlighted in the blind test predictions on hydantoin[28] in chapter 7, which is a clear example of the DMA stabilising the correct solid state hydrogen bond motif in many of the low energy structures. Nevertheless there is still room for improvement in these model potentials, which could include modelling of the polarization effects in the calculations, chapter 2.

It is clear that any molecular flexibility could have a significant effect on the computational predictions, highlighted by cytosine in chapter 4 where slight conformational changes in the amine group has a considerable effect on the stabilisation of the observed crystal structure compared to the other energetically feasible alternatives. However the molecules that are assumed rigid and are later found to have some slight intramolecular flexibility can prove more of a challenge, as shown for barbituric acid and urazole in chapter 5, where gauging the relative stabilities is problematic. This is a fundamental problem as if the energies cannot be relied upon for accuracy, then it is difficult to make predictions regarding polymorphism. This is further emphasised for uric acid, chapter 6, when even slight molecular deviations of this fairly rigid molecule (including the positions of the hydrogen atoms) have a detrimental effect on the confidence in the relative stabilities of the low energy structures. Despite this there has been some success in predicting crystal structures of larger conformationally flexible molecules[46;47;346], including salts[347], mainly due to the development of computational resources which allows more energetically plausible conformers to be studied.

For these computational predictions the crystal energy landscape needs to be known before any confidence in the predictions can be made. If the relative energy gaps between the low energy structures are sufficiently large then there is a greater confidence in predicting structures which are thermodynamically the most stable. However if these energy gaps are small, it is difficult to judge this stability. In addition the existence of experimental polymorphism suggests a greater importance of kinetic effects in the predictions. This is emphasised more for small organic molecules which have a tendency to generate more low energy structures in a small energy range[50], highlighted by thymine in chapter 4, which has over sixty structures within 7 kJ mol^{-1} of the global lattice energy minimum. In some computational polymorph searches many of the low energy crystal structures are sufficiently different (i.e. differences in the hydrogen bond motifs and acceptor/donors used) than the known form(s), as shown for 5-hydroxyuracil in chapter 8, which could suggest the possibility of finding trapped metastable structures which will not readily transform to the observed crystal structure. The free energies usually show some reordering of the crystal stabilities at room temperature, however the majority of other simple kinetic effects considered are similar between structures and therefore the information gained is fairly limited. It is crucial that there is more experimental data for some of these property calculations[54;55] to improve the understanding of how they relate to experimental behaviour.

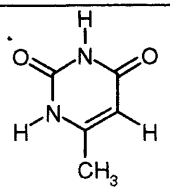
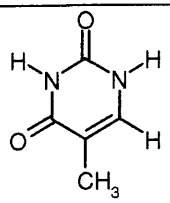
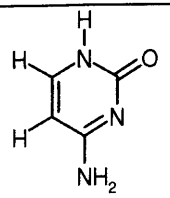
One major conclusion from this thesis is that even though polymorphs can be computationally predicted, the limitations in varying the crystallisation conditions is a major factor in finding these crystal structures experimentally. The experimental polymorph screens can be limited by the solubility of the compound, as shown for guanine and adenine in chapter 4, and the sensitivity to moisture, shown by cytosine in chapter 4 and alloxan in chapter 5. In some cases alternative crystallisation methods are needed to find other crystal forms, as highlighted for 6-methyluracil in chapter 4 in which sublimation of the compound produced a new polymorph, and by using templates during crystallisation, the method in which Form *ii* of progesterone has recently been crystallised[348]. Therefore the confidence in finding any experimental polymorphs through manual crystallisation screens is very much molecule specific.

This thesis has shown a greater scope of scientific research using crystal structure prediction, highlighted by the study of unused hydrogen bond acceptors in the crystalline environment, chapter 5. These crystal structure prediction methods can be an aid to thinking about the crystal system, which is shown for adenine (chapter 4) where one of the low energy structures has a very similar simulated X-ray powder pattern to the experimental, in which powder X-ray refinement seems promising to determine the elusive anhydrous crystal structure[254]. It is crucial that these computational polymorph predictions can be evaluated using results obtained from experimental polymorph screens, and promotes interest from pharmaceutical companies regarding the evaluation and understanding of crystal structure prediction methods, as shown in chapter 8.

Dame Kathleen Lonsdale stated that crystals are like a class of children in which each individual child is a little fidgety[1], chapter 1. The research in this thesis, with each molecule studied a different child, has shown that in some cases the children can find a more comfortable environment and are therefore less fidgety, having a well defined crystal structure. Others cannot find a comfortable arrangement and are therefore more fidgety, and can adopt a greater range of structures i.e. are polymorphic.

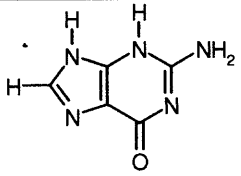
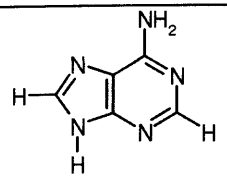
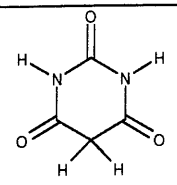
It is clear that there is still a vast improvement needed in the theoretical methods to confidently predict polymorphs of a variety of compounds just by knowing the molecular structure. Five of the compounds studied in this thesis would have been successfully predicted in a blind test scenario, based solely on lattice energy, in which three predictions are allowed. Therefore the success of predicting some of the crystal structures of the small organic molecules in this thesis is particularly encouraging. The range of experimental and computational polymorphic behaviour for this series of chemical related compounds highlights the challenges faced by crystal structure prediction, with this thesis helping to develop this exciting field of scientific research.

Table 9.2 Summary of molecules whose crystal structures were studied in this thesis

Molecule	Molecular structure	Crystal structure(s) found in the experimental screen[a]	Computational search results (energies in kJ mol^{-1})	Comments
6-methyluracil		**Form *i*[209]** Form *ii*[210] **New polymorph Form *iii***	ExptMinOpt1 (Form *i*) ΔE_{global} = 1.7 Rank = 4 ExptMinOpt2 (Form *ii*) ΔE_{global} = 2.5 Rank = 7	Form *i* and Form *ii* found in the computational search, Form *ii* around 1 kJ mol^{-1} less stable than Form *i*. New polymorph Form *iii* not found in search as Z' = 2 with two different conformations
Thymine		**Anhydrous[207]** **Monohydrate[231]**	ExptMinOpt ΔE_{global} = 0.2 Rank = 4	Many structures within 2 kJ mol^{-1} of the global lattice minimum. Known form probably the most thermodynamically stable structure, although other stackings of hydrogen bonded chains possible
Cytosine		**Anhydrous[208]** **Monohydrate[233]** **Hydrochloride[235]** **Oxalic acid dihydrate[349]** Ammonium hydrogen carbonate[241]	ExptMinOpt ΔE_{global} = 8.2 Rank = 6 ExptMinExpt Rank = 1	Search sensitive to the amine group conformation. Known structure probably the most thermodynamically stable. Limited range of experimental crystallisations possible due to sensitivity to moisture

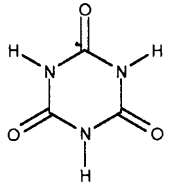
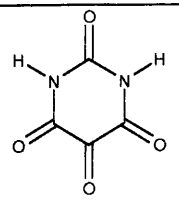
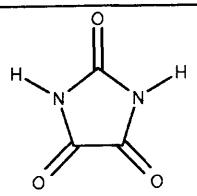
[a]The structures in bold correspond to those solved and refined by single crystal X-ray diffraction in this thesis

Table 9.2 continued. Summary of molecules whose crystal structures were studied in this thesis

Molecule	Molecular structure	Crystal structure(s) found in the experimental screen[a]	Computational search results (energies in kJ mol^{-1})	Comments
Guanine		**Guaninium chloride monohydrate**[248]; **Guaninium chloride dihydrate**[350] **Butryamide**[61] Triethylenediaminium dichloride[351] **Triethylenediaminium dichloride dihydrate**[62] 4-dimethylaminopyridine hydrochloride dihydrate[352]	N/A	No anhydrous crystal structure known, predictions show a range of three dimensional hydrogen bonded structures. Search sensitive to the assumed molecular structure
Adenine		**Adeninium dichloride**[60] Adeninium chloride hemihydrate[250]	Possible match for CC17 (planar molecular structure) using powder X-ray data[254] $\Delta E_{global} = 1.1$ Rank = 4	No anhydrous crystal structure known, predictions show a range of hydrogen bonded sheets and stackings of these sheets. Possible match between CC17 and solid state crystal structure using powder X-ray data[254]
Barbituric acid		**Form *i***[63] **New polymorph Form *ii***[63] Dihydrate[279] **5-isopropylidene-barbituric acid**[63]	ExptMinOpt (Form *i*) $\Delta E_{global} = 6.2$ ExptMinExpt (Form *i*) $\Delta E_{global} = 2.8$	ExptMinOpt (Form *i*) and ExptMinExpt (Form *i*) not found in the computational searches because of extreme sensitivity to conformation. Form *ii* has two molecular conformations in the asymmetric unit

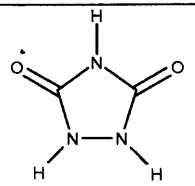
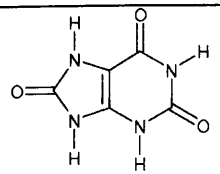
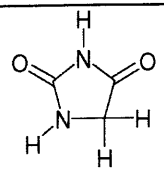
[a]The structures in bold correspond to those solved and refined by single crystal X-ray diffraction in this thesis

Table 9.2 continued. Summary of molecules whose crystal structures were studied in this thesis

Molecule	Molecular structure	Crystal structure(s) found in the experimental screen[a]	Computational search results (energies in kJ mol^{-1})	Comments
Cyanuric acid		Anhydrous[286] **Dihydrate**[64] **DMF solvate**[64]	ExptMinOpt $\Delta E_{global} = 5$ Rank = 7	Three-dimensional hydrogen bonded networks more energetically stable than solid state sheet motif. Possible kinetic factor favouring sheets or errors in the potential
Alloxan		**5,5-dihydroxybarbituric acid**[353] **5,5-dihydroxybarbituric acid monohydrate**[57] [b]**5,5-dihydroxybarbituric acid trihydrate**[58]	ExptMinOpt Rank = 1	Hypothetical structures consist of a mixture of weak/strong hydrogen bonding and strong carbonyl-carbonyl contacts. Limited range of crystallisations, alloxan sensitive to moisture[296]
Parabanic acid		Anhydrous[72] Oxo-ureido-acetic acid methyl ester[264] (Structures found in the experimental screen reported previously[264])	ExptMinOpt $\Delta E_{global} = 1.1$ Rank = 3	Results similar to that found previously[264], with the relative energies very sensitive to the model potential. The known form is the most thermodynamically stable

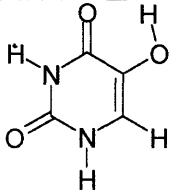
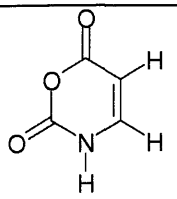
[a]The structures in bold correspond to those solved and refined by single crystal X-ray diffraction in this thesis
[b]Obtained from Aldrich and redetermined at low temperature to compliment the results of the experimental screen

Table 9.2 continued. Summary of molecules whose crystal structures were studied in this thesis

Molecule	Molecular structure	Structures found in experimental screen[a]	Computational search results (energies in kJ mol^{-1})	Comments
Urazole		Anhydrous[263] **3,5-Dioxo-[1,2,4]triazolidine-1-carboxylic acid**[64]	ExptMinOpt $\Delta E_{global} = 9$ Rank = 26 ExptMinExpt Rank = 1	Search sensitive to assumed molecular structure of urazole. The known structure probably the most thermodynamically stable
Uric acid		[c]**Anhydrous**[316;319] Dihydrate[318;319;354]	Previous ExptMinOpt[316;319] $\Delta E_{global} = 3.4$ Rank = 8 Previous ExptMinExpt[316;319] Rank = 1 New ExptMinExpt Rank = 1	Small changes in the molecular structure have a significant effect on the relative energies. Known structure probably the most thermodynamically stable
Hydantoin		N/A	ExptMinOpt $\Delta E_{global} = 1.1$ Rank = 2 ExptMinPlanar Rank = 1	Part of blind test 2003, experimental crystal structure found via symmetry reduction in the computational predictions

[a]The structures in bold correspond to those solved and refined by single crystal X-ray diffraction in this thesis
[c]Determined in collaboration with Dr Robert Lancaster at UCL

Table 9.2 continued. Summary of molecules whose crystal structures were studied in this thesis

Molecule	Molecular structure	Structures found in experimental screen[a]	Computational search results (energies in kJ mol^{-1})	Comments
5-hydroxyuracil		Anhydrous[49] DMF solvate[49] (Both crystal structures found in an experimental screen at GlaxoSmithKline[49])	ExptMinOpt ΔE_{global} = 3.1 Rank = 3 ExptMinExpt ΔE_{global} = 1.7 Rank = 4	Computational predictions sensitive to the intramolecular flexibility present, possible polymorphism if right kinetic route found
3-oxauracil		Anhydrous[49] Monohydrate structure[49] (Both crystal structures found in an experimental screen at GlaxoSmithKline[49])	ExptMinOpt Rank = 1	Experimental crystal structure found 3.9 kJ mol^{-1} more energetically stable than nearest rival. Polymorphism unlikely

[a]The structures in bold correspond to those solved and refined by single crystal X-ray diffraction in this thesis

Appendix

All this supplementary information is on the CD attached inside the back cover. This information is included in the supplementaryinfo.doc file in the root folder, with clickable links.

Chapter 3

Chapter3.doc

Table 3.1	Dipole moments and dielectric constants

Chapter 4

Chapter4summary.doc

Table 4.1	Comparison of the solid state/*ab initio* molecular structures

6-methyluracil.doc

Table 4.2	Experimental polymorph screen
Table 4.5	Metric parameters of 6-methyluracil Form *iii*
Table 4.6	Crystallographic data for 6-methyluracil Form *iii*
Table 4.7	Results of the computational polymorph search
Section 4.3	Crystallisation of 6-methyluracil Form *i*
Section 4.4	6-methyluracil Form *ii*

Thymine.doc

Table 4.8	Experimental polymorph screen
Table 4.11	Results of the computational polymorph search
Section 4.9	Crystallisation of anhydrous thymine
Section 4.10	Crystallisation of thymine monohydrate

Cytosine.doc

Table 4.12	Experimental polymorph screen
Section 4.13	Crystallisation of anhydrous cytosine
Section 4.14	Crystallisation of cytosine hydrochloride
Section 4.15	Crystallisation of oxalic acid dihydrate

Guanine.doc

Table 4.16	Experimental polymorph screen
Section 4.17	Crystallisation of guaninium chloride monohydrate
Section 4.18	Crystallisation of guaninium chloride dihydrate
Section 4.19	Crystallisation of butryamide

Adenine.doc

Section 4.20 Crystallisation of adeninium dichloride

Chapter 5

Chapter5summary.doc

Table 5.1 Metric parameters of the solid state/*ab initio* molecular structures of barbituric acid Form *i*, cyanuric acid, urazole, alloxan and parabanic acid

Table 5.2 Lattice energy minimisations on barbituric acid

Table 5.3 Lattice energy minimisations on cyanuric acid

Table 5.4 Lattice energy minimisations on alloxan

Table 5.5 Lattice energy minimisations on parabanic acid

Table 5.6 Lattice energy minimisations on urazole

Barbituricacid.doc

Table 5.7 The crystal structures of selected barbiturates

Table 5.9 Experimental polymorph screen

Table 5.10 Unit cell parameters of barbituric acid Form *i*

Table 5.11 Metric parameters of barbituric acid Form *i*

Table 5.12 Unit cell and metric/hydrogen bond parameters of barbituric acid Form *ii*

Table 5.14 The crystal structures found in the experimental polymorph screen

Table 5.15 Crystallographic data for barbituric acid Form *i*

Table 5.16 Crystallographic data for barbituric acid Form *ii*

Table 5.18 Results of the computational polymorph search using the *ab initio* molecular structure

Table 5.19 Results of the computational polymorph search using the solid state molecular structure

Figure 5.8 Hydrogen bond motifs in the barbiturates

Section 5.13 Crystallisation of 5-isopropylidene-barbituric acid

Cyanuricacid.doc

Table 5.20 Cyanuric acid solvates and adducts

Table 5.21 Experimental polymorph screens

Section 5.22 Crystallisation of cyanuric acid dihydrate

Section 5.23 Crystallisation of cyanuric acid DMF solvate

Alloxan.doc

Table 5.24 Experimental polymorph screen

Section 5.25 Crystallisation of 5,5-dihydroxybarbituric acid (alloxan monohydrate)

Section 5.26 Crystallisation of 5,5-dihydroxybarbituric acid monohydrate (alloxan dihydrate)

Section 5.27 Redetermination of 5,5-dihydroxybarbituric acid trihydrate (alloxan tetrahydrate)

Parabanicacid.doc

Table 5.28 Results of the computational polymorph search

Urazole.doc

Table 5.29 Experimental polymorph screen

Table 5.30 Results of the computational polymorph search using the *ab initio* molecular structure

Table 5.31 Results of the computational polymorph search using the solid state molecular structure

Chapter 6

Chapter6.doc

Table 6.1 Previous experimental polymorph screen

Table 6.2 Crystallographic data of the low temperature determination of anhydrous uric acid

Table 6.3 Metric parameters of the *ab initio* and both determinations of the solid state molecular structure of uric acid

Table 6.4 Lattice energy minimisations on the low temperature determination of anhydrous uric acid

Table 6.5 Results of the previous computational polymorph search using the *ab initio* molecular structure

Table 6.6 Results of the previous computational polymorph search using the room temperature solid state molecular structure

Table 6.7 Results of the computational polymorph search using the low temperature solid state molecular structure

Table 6.8 Results of the computational polymorph searches using the ConoptNH and Conoptring molecular structures

Chapter 7

Chapter7.doc

Table 7.1 Selected metric parameters in the *ab initio*/planar molecular structures

Table 7.2 Results of the computational polymorph search using the *ab initio* molecular structure

Table 7.3 Results of the computational polymorph search using the planar molecular structure

Chapter 8

Chapter8summary.doc

Table 8.1 Metric parameters of the solid state/*ab initio* molecular structures for AMGLA and FIWVEM

3-oxauracil.doc

Table 8.2 Results of the computational polymorph search

Table 8.3 Metric parameters of the *ab initio*/solid state molecular structures of 3-oxauracil

5-hydroxyuracil.doc

Table 8.4 Metric parameters of the *ab initio* conformation A/solid state molecular structures of 5-hydroxyuracil

Table 8.5 Results of the computational polymorph search using the solid state molecular structure

Bibliography

1. Lonsdale, K. *Crystals and X-Rays,* van Nostrand: 1949.

2. Rowe, R. C. *DDT* **2001,** *6*, 395-97.

3. Lonsdale, K. *Fifty Years of X-ray diffraction*, Ewald, P. P., Ed.; International Union of Crystallography: Utrecht, The Netherlands, 1962; Chapter 14.

4. Caira, M. R. *Top.Curr.Chem* **1998,** *198*, 164-208.

5. Bernstein, J. *Polymorphism in Molecular Crystals,* Oxford Science Publications: 2002.

6. Vippagunta, S. R; Brittain, H. G; Grant, D. J. W *Adv.Drug.Deliver.Rev* **2001,** *48*, 3-26.

7. ICH Harmonised Tripartite Guideline - Specifications: Test procedures and acceptance criteria for new drug substances and new drug products: Chemical substances. 1999. ICH Steering Committee.

8. Morissette, S. L.; Almarsson, O.; Peterson, M. L.; Remenar, J. F.; Read, M. J; Lemmo, A. V.; Ellis, S; Cima, M. J; Gardner, C. R *Adv.Drug.Deliver.Rev* **2004,** *56*, 275-300.

9. Chemburkar, S. R.; Bauer, J.; Deming, K.; Spiwek, H.; Patel, K.; Morris, J.; Henry, R.; Spanton, S.; Dziki, W.; Porter, W.; Quick, J.; Bauer, P.; Donaubauer, J.; Narayanan, B. A.; Soldani, M.; Riley, D.; McFarland, K. *Org.Process.Res.Dev* **2000,** *4*, 413-17.

10. Morissette, S. L.; Soukasene, S; Levinson, D; Cima, M. J; Almarsson, O. *Proc.Natl.Acad.Sci.U.S.A* **2003,** *100*, 2180-84.

11. Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. *Pharmaceut.Res* **2001,** *18*, 859-66.

12. Snider, D. A; Addicks, W; Owens, W. *Adv.Drug.Deliver.Rev* **2004,** *56*, 391-95.

13. Almarsson, O.; Hickey, M. B; Peterson, M. L.; Morissette, S. L.; Soukasene, S; McNulty, C; Tawa, M; MacPhee, J. M; Remenar, J. F. *Cryst.Growth.Des* **2003,** *3*, 927-33.

14. Gardner, C. R; Almarsson, O.; Chen, H; Morissette, S; Peterson, M; Zhang, Z; Wang, S; Lemmo, A; Gonzalez-Zugasti, J.; Monagle, J.; Marchionna, J.; Ellis, S; McNulty, C; Johnson, A.; Levinson, D; Cima, M. *Comput.Chem.Eng* **2004**.

15. Kohlbeck, J. A *Microscope* **1982,** *30*, 249.

16. Teetsov, A; McCrone, W. C. *Microscope* **1965,** *15*, 13.

17. Threfall, T. L. *Analyst* **1995,** *120*, 2435-60.

18. Dunitz, J. D *ChemComm* **2003,** 545-48.

19. Gavezzotti, A. *Acc.Chem.Res.* **1994,** *27*, 309-14.

20. Blagden, N.; Davey, R. J. *Cryst.Growth.Des* **2003,** *3*, 873-85.

21. Ostwald, W. *Z.Phys.Chem* **1897,** *22*, 289.

22. Hussain, K; Thorsen, G; Malthe-Sorenssen, D *Chem.Eng.Sci* **2001,** *56*, 2295-304.

23. Ischenko, V; Englert, U; Jansen, M *Chem.Eng.J* **2005,** *11*, 1375-83.

24. Weissbuch, I; Torbeev, V. Y; Leiserowitz, L; Lahav, M *Angew.Chem.Int.Ed.Engl* **2005,** *44*, 3226-29.

25. Threfall, T. *Org.Process.Res.Dev* **2003,** *7*, 1017-27.

26. Desiraju, G. R. *Nat.Mater* **2002,** *1*, 77-79.

27. Maddox, J. *Nature* **1988,** *335*, 201.

28. Day, G. M; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, X. M; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A; Dunitz, J. D; Schweizer, B; Van Eijck, B. P.; Erk, P; Facelli, J. C; Bazterra, V. E; Ferraro, M. B; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C; Pantelides, C. C; Karamertzanis, P. G; Price, S. L.; Lewis, T. C.; Nowell, H; Torrisi, A; Scheraga, H. A; Arnautova, Y. A; Schmidt, M. U; Verwer, P *Acta.Cryst* **2005,** *B61*, 511-27.

29. Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D; Dzyabchenko, A; Erk, P; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Lommerse, J. P. M; Mooij, W. T. M.; Price, S. L.; Scheraga, H; Schweizer, B; Schmidt, M. U; Van Eijck, B. P.; Verwer, P.; Williams, D. E. *Acta.Cryst* **2002,** *B58*, 647-61.

30. Lommerse, J. P. M; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooij, W. T. M.; Price, S. L.; Schweizer, B; Schmidt, M. U; Van Eijck, B. P.; Verwer, P.; Williams, D. E. *Acta.Cryst* **2000,** *B56*, 697-714.

31. Cody, V; DeJarnette, F; Duax, W; Norton, D. A *Acta.Cryst* **1971,** *B27*, 2458-68.

32. Allen, F. H.; Harris, S. E; Taylor, R *J.Comput.Aid.Mol.Des* **1996,** *10*, 247-54.

33. Kitaigorodsky, A. I. *Organic Chemical Crystallography,* Consultants Bureau: New York, 1961.

34. Rodriguez-Spong, B; Price, C. P; Jayasankar, A; Matzger, A. J; Rodriguez-Hornedo, N *Adv.Drug.Deliver.Rev* **2004,** *56*, 241-74.

35. Gavezzotti, A. *CrystEngComm* **2003,** *5*, 429-38.

36. Gavezzotti, A. *CrystEngComm* **2003,** *5*, 439-46.

37. Pillardy, J; Wawak, R. J; Arnutova, Y. A; Czaplewski, C; Scheraga, H. A *J.Am.Chem.Soc* **2000,** *122*, 907-21.

38. Gavezzotti, A. *J.Phys.Chem.B* **2003,** *107*, 2344-53.

39. Karamertzanis, P. G; Pantelides, C. C *Mol Simul.* **2004,** *30*, 413-36.

40. Gavezzotti, A. *Curr.Opin.Solid.St.M* **1996,** *1*, 501-05.

41. Brock, C. P; Dunitz, J. D *Chem.Mater* **1994,** *6*, 1118-27.

42. Van Eijck, B. P.; Kroon, J. *Acta.Cryst* **2000,** *B56*, 535-42.

43. Van Eijck, B. P. *J.Comput.Chem* **2002,** *23*, 456-62.

44. Steiner, T *Acta.Cryst* **2000,** *B56*, 673-76.

45. Mooij, W. T. M.; Van Eijck, B. P.; Kroon, J. *J.Am.Chem.Soc* **2000,** *122*, 3500-05.

46. Nowell, H; Price, S. L. *Acta.Cryst* **2005,** *B61*, 558-68.

47. Ouvrard, C; Price, S. L. *Cryst.Growth.Des* **2004,** *4*, 1119-27.

48. Coombes, D. S.; Nagi, G. K.; Price, S. L. *Chem.Phys.Lett* **1997,** *265*, 532-37.

49. Copley, R. C. B; Deprez, L. S; Lewis, T. C.; Price, S. L. *CrystEngComm* **2005,** *7*, 421-28.

50. Day, G. M.; Chisholm, J; Shan, N; Motherwell, W. D. S.; Jones, W *Cryst.Growth.Des* **2004,** *4*, 1327-40.

51. Hulme, A. T; Price, S. L.; Tocher, D. A. *J.Am.Chem.Soc* **2005,** *127*, 1116-17.

52. Day, G. M.; Price, S. L.; Leslie, M. *J.Phys.Chem.B* **2003,** *107*, 10919-33.

53. Gavezzotti, A.; Filippini, G. *J.Am.Chem.Soc* **1995,** *117*, 12299-305.

54. Day, G. M.; Price, S. L.; Leslie, M. *Cryst.Growth.Des* **2001,** *1*, 13-27.

55. Day, G. M.; Price, S. L. *Handbook of Elastic Properties of Solids, Liquids and Gases. Volume III: Elastic Properties of Solids: Biological and Organic Materials, Earth and Marine Sciences.,* Academic Press: New York, 2001.

56. Coombes, D. S.; Catlow, C. R. A.; Gale, J. D; Rohl, A. L.; Price, S. L. *Cryst.Growth.Des* **2005,** *5*, 879-85.

57. Lewis, T. C.; Tocher, D. A. *Acta.Cryst* **2004,** *E60*, o1689-o1690.

58. Lewis, T. C.; Tocher, D. A. *Acta.Cryst* **2004,** *E60*, o1748-o1750.

59. Lewis, T. C.; Tocher, D. A. *Acta.Cryst* **2005,** *E61*, o1023-o1025.

60. Lewis, T. C.; Tocher, D. A. *Acta.Cryst* **2005,** *E61*, o1052-o1054.

61. Lewis, T. C.; Tocher, D. A. *Acta.Cryst* **2005,** *E61*, o1985-o1986.

62. Lewis, T. C.; Tocher, D. A. *Acta.Cryst* **2005,** *E61*, o2202-o2204.

63. Lewis, T. C.; Tocher, D. A.; Price, S. L. *Cryst.Growth.Des* **2004,** *4*, 979-87.

64. Lewis, T. C.; Tocher, D. A.; Price, S. L. *Cryst.Growth.Des* **2005,** *5*, 983-93.

65. Price, S. L. *Adv.Drug.Deliver.Rev* **2004,** *56*, 301-19.

66. Price, S. L. *J.Chem.Soc., Faraday Trans* **1996,** *92*, 2997-3008.

67. Besler, B. H; Merz, K. M; Kollman, P. A *J.Comput.Chem* **1990,** *11*, 431-39.

68. Spackman, M. A *J.Comput.Chem* **1998,** *17*, 1-18.

69. Wiberg, K. B.; Rablen, P. R *J.Comput.Chem* **1993,** *14*, 1504.

70. Coombes, D. S.; Catlow, C. R. A.; Gale, J. D; Hardy, M. J; Saunders, M. R. *J.Pharm.Sci* **2002,** *91*, 1652-58.

71. Karamertzanis, P. G; Pantelides, C. C *J.Comput.Chem* **2005,** *26*, 304-24.

72. He, X. M.; Swaminathan, S.; Craven, B. M. *Acta.Cryst* **1988,** *B44*, 271-81.

73. Swaminathan, S.; Craven, B. M. *Acta.Cryst* **1985,** *B41*, 113-22.

74. Price, S. L. *CrystEngComm* **2004,** *6*, 344-53.

75. Stone, A. J. *Chem.Phys.Lett* **1981,** *83*, 233-39.

76. Stone, A. J.; Alderton, M. *Mol.Phys* **1985,** *56*, 1047-64.

77. Stone, A. J. *The Theory of Intermolecular Forces,* Clarendon Press: Oxford, 1996.

78. Moller, C; Plesset, M. S *Phys.Rev* **1934,** *46*, 618-22.

79. Winn, P. J; Ferenczy, G. G; Reynolds, C. A *J.Phys.Chem.A* **1997,** *101*, 5437-45.

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

81. Price, S. L.; Richards, N. G *J.Comput.Aid.Mol.Des* **1991,** *5*, 41-54.

82. Price, S. L., Vinter, A, and Stone, A. J. Estgen: A program to display electrostatic potentials and fields around a molecule. 1993. http://www-stone.ch.cam.ac.uk/programs.html#Estgen.

83. Chipot.C; Luque, F. J *Chem.Phys.Lett* **2000,** *332*, 190-98.

84. Buckingham, R. A. *Proceedings of the Royal Society of London* **1938,** *A168*, 264.

85. Stone, A. J. *Chem.Phys.Lett* **1993,** *211*, 101-09.

86. Kestner, N. R.; Combariza, J. E. *Reviews in Computational Chemistry*, Lipkowitz, K. B.; Boyd, D. B., Eds.; Wiley-VCH: New York, 1999.

87. Price, S. L.; Stone, A. J. *Chem.Phys.Lett* **1979,** *65*, 127-31.

88. Umeyama, H; Morojuma, K. *J.Am.Chem.Soc* **1976,** *99*, 1316.

89. Kollman, P. A; Allen, L. C. *J.Am.Chem.Soc* **1970,** *92*, 6101.

90. Kollman, P. A; Allen, L. C. *Chem.Rev* **1972,** *72*, 283.

91. Pauling, L. *The nature of the chemical bond,* Cornell University Press: Ithaca, New York, 1939.

92. Desiraju, G. R.; Steiner, Th. *The Weak Hydrogen Bond.,* Oxford University Press: Oxford, 1999.

93. Gilli, P; Bertolasi, V; Ferretti, V; Gastone, G. *J.Am.Chem.Soc* **1994,** *116*, 909-15.

94. Gilli, G.; Gilli, P *J.Mol.Struct* **2000,** *552*, 1-15.

95. Buckingham, A. D; Fowler, P. W *J.Chem.Phys* **1983,** *79*, 6426-28.

96. Desiraju, G. R. *Encyclopedia of Supramolecular Chemistry*, Marcel Dekker, Inc: 2004.

97. Van Eijck, B. P. *Phys.Chem.Chem.Phys* **2002,** *4*, 4789-94.

98. Stone, A. J.; Tong, C. S. *J.Comput.Chem* **1994,** *15*, 1377-92.

99. Williams, D. E. *J.Comput.Chem* **2001,** *22*, 1154-66.

100. Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J.Phys.Chem* **1996,** *100*, 7352-60.

101. Cox, S. R.; Hsu, L.; Williams, D. E. *Acta.Cryst* **1981,** *A37*, 293-301.

102. Williams, D. E.; Cox, S. R. *Acta.Cryst* **1984,** *B40*, 404-17.

103. Mitchell, J. B. O.; Price, S. L. *J.Comput.Chem* **1990,** *11*, 1217-33.

104. Averkiev, B. B; Antipin, M. Y; Sheremetev, A. B; Timofeeva, T. V *Cryst.Growth.Des* **2005,** *5*, 631-41.

105. Williams, D. E. *J.Mol.Struct* **1999,** *485-486*, 324-47.

106. Williams, D. E. *J.Comput.Chem* **2000,** *22*, 1-20.

107. Allen, F. H.; Kennard, O; Watson, D. G. *J.Chem.Soc.Perk.T.2* **1987,** *12*, S1-S18.

108. Pillardy, J; Arnautova, Y. A; Czaplewski, C; Gibson, K. D; Scheraga, H *P.Natl.Acad.Sci.USA* **2001,** *98*, 12351-56.

109. Mayo, S. L.; Olafson, B. D.; Goddard III, W. A. *J.Phys.Chem* **1990,** *94*, 8897-909.

110. Hwang, M. J; Stockfisch, T. P; Hagler, A. T *J.Am.Chem.Soc* **1994,** *116*, 2515-25.

111. Peng, Z; Ewig, C. S; Hwang, M; Waldman, M; Hagler, A. T *J.Phys.Chem.A* **1997,** *101*, 7243-52.

112. Sun, H. *J.Phys.Chem.B* **1998,** *102*, 7338-64.

113. Day, G. M.; Motherwell, W. D. S.; Jones, W *Cryst.Growth.Des* **2005,** *5*, 1023-33.

114. Allen, F. H. *Acta.Cryst* **2002,** *B58*, 380-88.

115. Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J.Comput.Chem* **1995,** *16*, 628-47.

116. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery Jr., J.A, Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Baboul, A. G., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S., and Pople, J. A. Gaussian98, Revision A.9. 1998. Pittsburgh PA, Gaussian, Inc.

117. Schaftenaar, G; Noordik, J. H. *J.Comput.Aid.Mol.Des* **2000,** *14*, 123-34.

118. Filippini, G.; Gavezzotti, A. *Acta.Cryst* **1993,** *B49*, 868-80.

119. Beyer, T.; Price, S. L. *CrystEngComm* **2000,** *34*, 1-8.

120. Ouvrard, C; Mitchell, J. B. O. *Acta.Cryst* **2003,** *B59*, 676-85.

121. Clydesdale, G.; Roberts, K. J.; Walker, E. M. *Theoretical Aspects and Computer Modeling of the Molecular Solid State*, Gavezzotti, A., Ed.; John Wiley and Sons: Chichester, 1997; Chapter 7.

122. Williams, D. E *J.Chem.Phys* **1966,** *45*, 3770-78.

123. Kitaigorodsky, A. I. *Molecular crystals and molecules,* Academic Press: New York, 1973.

124. Chaka, A. M; Zaniewski, R; Youngs, W; Tessier, C; Klopman, G. *Acta.Cryst* **1996,** *B52*, 165-83.

125. Bazterra, V. E; Ferraro, M. B; Facelli, J. C *J.Chem.Phys* **2002,** *116*, 5984-91.

126. Bazterra, V. E; Ferraro, M. B; Facelli, J. C *J.Chem.Phys* **2002,** *116*, 5992-95.

127. Bazterra, V. E; Ona, O; Caputo, M. C; Ferraro, M. B; Fuentealba, P; Facelli, J. C *Phys.Rev.A* **2004,** *69*, 053202.

128. Bazterra, V. E; Ferraro, M. B; Facelli, J. C *Int.J.Quantum.Chem* **2004,** *96*, 312-20.

129. Williams, D. E. *Acta.Cryst* **1996,** *A52*, 326-28.

130. Williams, D. E. mpa/mpg, Molecular Packing Analysis/Molecular Packing Graphics. 1996. Louisville, KY, Department of Chemistry, University of Louisville.

131. Shoda, T; Yamahara, K; Okazaki, K; Williams, D. E *J.Mol.Struc-Theochem* **1995,** *333*, 267-74.

132. Shoda, T; Williams, D. E *J.Mol.Struc-Theochem* **1995,** *357*, 1-8.

133. Payne, R. S.; Roberts, R. J.; Rowe, R. C.; Docherty, R. *Int.J.Pharm* **1999,** *177*, 231-45.

134. Payne, R. S.; Rowe, R. C.; Roberts, R. J.; Charlton, M. H.; Docherty, R. *J.Comput.Chem* **1999,** *20*, 262-73.

135. Leusen, F. J. J. *J.Cryst.Growth* **2003,** *3*, 189-92.

136. Perlstein, J *J.Am.Chem.Soc* **1992,** *114*, 1955-63.

137. Perlstein, J *J.Am.Chem.Soc* **1994,** *116*, 455-70.

138. Perlstein, J *J.Am.Chem.Soc* **1994,** *116*, 11420-32.

139. Gavezzotti, A. *J.Am.Chem.Soc* **1991,** *113*, 4622-29.

140. Gavezzotti, A.; Filippini, G. *J.Am.Chem.Soc* **1996,** *118*, 7153-57.

141. Gavezzotti, A. *Acta.Cryst* **1996,** *B52*, 201-08.

142. Dapporto, P.; Mealli, P.; Paoli, P.; Virtuani, M. *Struct.Chem* **1998,** *9*, 129.

143. Filippini, G,; Gavezzotti, A.; Novoa, J. J. *Acta.Cryst* **1999,** *B55*, 543-53.

144. Van Eijck, B. P.; Kroon, J. *J.Comput.Chem* **1999,** *20*, 799-812.

145. Mooij, W. T. M.; Van Eijck, B. P.; Price, S. L.; Verwer, P.; Kroon, J. *J.Comput.Chem* **1998,** *19*, 459-74.

146. Rovira, C; Novoa, J. J *J.Phys.Chem.B* **2001,** *105*, 1710-19.

147. Mooij, W. T. M.; Van Eijck, B. P.; Kroon, J. *J.Phys.Chem.A* **1999,** *103*, 9883-90.

148. Holden, J. R.; Du, Z.; Ammon, H. L. *J.Comput.Chem* **1993,** *4*, 422-37.

149. Beyer, T.; Day, G. M.; Price, S. L. *J.Am.Chem.Soc* **2001,** *123*, 5086-94.

150. Price, S. L.; Wibley, K. S. *J.Phys.Chem.A* **1997,** *101*, 2198-206.

151. Potter, B. S.; Palmer, R. A.; Withnall, R.; Chowdhry, B. Z.; Price, S. L. *J.Mol.Struct* **1999,** *485-486*, 349-61.

152. Mitchell, J. B. O.; Price, S. L.; Leslie, M.; Buttar, D.; Roberts, R. J. *J.Phys.Chem.A* **2001,** *105*, 9961-71.

153. Toukmaji, A. Y; Board Jr, J. A *Comput.Phys.Comm* **1996,** *95*, 73-92.

154. Ewald, P *Ann.Phys* **1921,** *64*, 253.

155. Niggli, P. *Handbuch der Experimentalphysik*, Leipzig: Akademische Verlagsgesellschaft: 1928; Chapter 1.

156. Santoro, A; Mighell, A. D *Acta.Cryst* **1970,** *A26*, 124-27.

157. Krivy, I. *Acta.Cryst* **1976,** *A32*, 297-98.

158. Spek, A. L. PLATON, A Multipurpose Crystallographic Tool. 2003. Utrecht, The Netherlands, Utrecht University.

159. van de Streek, J. Personal Communication. 2005.

160. van de Streek, J; Motherwell, S *Acta.Cryst* **2005,** *B61*, 504-10.

161. Chisholm, J. A; Motherwell, S *J.Appl.Crystallogr* **2004,** *37*, 331-34.

162. Chisholm, J. A; Motherwell, S *J.Appl.Crystallogr* **2004,** *38*, 228-31.

163. Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. *Acta.Cryst* **2002,** *B58*, 389-97.

164. Cerius2. 1997. San Diego, CA., Molecular Simulations Inc.

165. Kuleshova, L. N.; Zorky, P. M. *Acta.Cryst* **1980,** *B36*, 2113-15.

166. Etter, M. C. *Acc.Chem.Res.* **1990,** *23*, 120-26.

167. Etter, M. C.; MacDonald, J. C. *Acta.Cryst* **1990,** *B46*, 256-62.

168. This program is available free of charge for non-commercial use and may be downloaded from the CCDC website at http://www.ccdc.cam.ac.uk/free_services/rpluto/. 2003.

169. Payne, R. S.; Roberts, R. J; Rowe, R. C.; McPartlin, M; Bashal, A. *Int.J.Pharm* **1996,** *145*, 165-73.

170. Roberts, R. J; Payne, R. S.; Rowe, R. C. *Eur.J.Pharm.Sci* **2000,** *9*, 277-83.

171. Yoshihara, A; Bernstein, E. R *J.Chem.Phys* **1982,** *77*, 5319-26.

172. Zunger, A; Huler, E. *J.Chem.Phys* **1975,** *62*, 3010.

173. Capelli, S. C; Fortsch, M; Burgi, H. B *Acta.Cryst* **2000,** *A56*, 413-24.

174. Grimbergen, R. F. P; Reedijk, M. F; Meekes, H; Bennema, P. *J.Phys.Chem.B* **1998,** *102*, 2646-53.

175. Liu, X. Y; Boek, E. S; Briels, W. J; Bennema, P. *J.Chem.Phys* **1995,** *103*, 3747.

176. Bravais, A. Etudes Crystallographiques . 1913. Academie des Sciences, Paris.

177. Freidel, G *Bull.Soc.Fr.Miner* **1907,** *30*, 326.

178. Yanson, I. K; Verkin, B. I; Shklyare, O. I; Teplitsky, A. B *Studia.Biophys* **1974,** *46*, 29-44.

179. Hartman, P; Perdok, W. G *Acta.Cryst* **1955,** *8*, 49-52.

180. Berkovitch-Yellin, Z. *J.Am.Chem.Soc* **1985,** *107*, 8239-53.

181. Clydesdale, G; Roberts, K. J; Lewtas, K; Docherty, R. *J.Cryst.Growth* **1994,** *141*, 443-50.

182. Engkvist, O; Price, S. L.; Stone, A. J. *Phys.Chem.Chem.Phys* **2000,** *2*, 3017-27.

183. Rohl, A. L. *Curr.Opin.Solid.St.M* **2003,** *7*, 21-26.

184. Fleming, S. GDIS. [0.80]. 2003. Curtin University of Technology, Perth.

185. Donnay, J. D. H; Harker, D *Am.Mineral* **1937,** *22*, 446.

186. Gale, J. D; Rohl, A. L. *Mol Simul.* **2003,** *29*, 291-341.

187. Breneman, C. M; Wiberg, K. B. *J.Comput.Chem* **1990,** *11*, 361.

188. Brunsteiner, M.; Price, S. L. *Cryst.Growth.Des* **2001,** *1*, 447-53.

189. Day, G. M. Calcvol. 2002. University College London.

190. Blagden, N.; Cross, W. I.; Davey, R. J.; Broderick, M; Pritchard, R. G; Roberts, R. J; Rowe, R. C. *Phys.Chem.Chem.Phys* **2001,** *3*, 3819-25.

191. Cross, W. I.; Blagden, N.; Davey, R. J. *Cryst.Growth.Des* **2003,** *3*, 151-58.

192. De Rosa, C; Ruiz de Ballesteros, O; Gennaro, M. D; Auriemma, F *Polymer* **2003,** *44*, 1861-70.

193. Sakamoto, M; Kobaru, S; Kasashima, Y; Mino, T; Fujita, T *Tetrahedron.Lett* **2005,** *46*, 4439-42.

194. Lide, D. R. *CRC Handbook of Chemisty and Physics,* 83rd ed.; 2003.

195. Clegg, W.; Blake, A. J.; Gould, R. O.; Main, P. *Crystal Structure Analysis, Principles and Practice,* Oxford Science Publications: Oxford, 2001.

196. Gu, C. Young; V; Grant, D. J. W. *J.Pharm.Sci* **2001,** *90*, 1878-90.

197. Trask, A. V; Shan, N; Motherwell, W. D. S.; Jones, W; Feng, S; Tan, R. B. H; Carpenter, K. J *ChemComm* **2005,** 880-82.

198. Gilmore, C. J; Barr, G; Paisley, J *J.Appl.Crystallogr* **2004,** *37*, 242.

199. Bergese, P; Bontempi, E; Colombo, I; Depero, L. E *J.Appl.Crystallogr* **2001,** *34*, 663-65.

200. Roberts, S. N. C; Williams, A. C; Grimsey, I. M; Booth, S. W *J.Pharmaceut.Biomed* **2002,** *28*, 1149-59.

201. Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. *Acta.Cryst* **2002,** *B58*, 389-97.

202. SAINT: Program for Integration of Area Detector Data. [4]. 1994. Madison WI, USA, Bruker AXS.

203. Sheldrick, G. M. SADABS Program for Bruker Area Detector Adsorption Correction. 1996. Germany, University of Grottingen.

204. Sheldrick, G. M. SHELXTL. [6.12]. 1997. Germany, University of Grottingen.

205. Watson, J. D; Crick, F. H. C *Nature* **1953,** *171*, 737-38.

206. Voet, D; Rich, A *Prog.Nucleic.Acid.Res.Mol.Biol* **1970,** *10*, 183-265.

207. Portalone, G; Bencivenni, L; Colapietro, M; Pieretti, A.; Ramondo, F. *Acta.Chem.Scand* **1999,** *53*, 57-68.

208. McClure, R. J; Craven, B. M *Acta.Cryst* **1973,** *B29*, 1234-38.

209. Leonidov, N. B; Zorkii, P. M; Masunov, A. E; Gladkikh, O. P; Bel'skii, V. K; Dzyabchenko, A. V; Ivanov, S. A *Russ.J.Phys.Chem* **1993,** *67*, 2220-23.

210. Reck, G; Kretschmer, R.-G; Kutschabsky, L *Acta.Cryst* **1988,** *A44*, 417-21.

211. Stewart, E. L; Foley, C. K; Allinger, N. L; Bowen, J. P *J.Am.Chem.Soc* **1994,** *116*, 7282-86.

212. Sponer, J; Hobza, P *J.Phys.Chem* **1994,** *98*, 3161-64.

213. Reck, G. Personal communication. 2004.

214. Scharfenberg-Pfeiffer, D; Kretschmer, R. G; Hoffmann, S *Cryst.Res.Technol.* **1988,** *23*, 881-87.

215. Rasmussen, H; Sletten, E *Acta.Chem.Scand* **1973,** *27*, 2757-68.

216. Spek, A. L. and Smeets.W.J.J. 9-methylhypoxanthine. Cambridge Structure Database. 2000.

217. Schmalle, H. W; Hanggi, G; Dubler, E *Acta.Cryst* **1988,** *C44*, 732-36.

218. Ferro, D; Bencivenni, L; Teghil, R; Pelino, M *J.Indian.Chem.Soc* **1980,** *57*, 629-32.

219. Burkinshaw, P. M; Mortimer, C. T *J.Chem.Soc., Dalton.Trans* **1984,** *1*, 75-77.

220. Ferro, D; Bencivenni, L; Teghil, R; Mastromarino, R *Thermochim.Acta* **1980,** *42*, 75-83.

221. Nabavian, M; Sabbah, R; Chastel, R; Laffitte, M *J.Chem.Phys.Phys-Chim.Biol* **1977,** *74*, 115-26.

222. Teplitsky, A. B; Yanson, I. K; Glukhova O.T; Zielenkiewicz, A; Zielenkiewicz, W; Wierzchowski, K. L *Biophys.Chem* **1980,** *11*, 17-21.

223. Sabbah, R *Thermochim.Acta* **1980,** *35*, 73-77.

224. Clark, LB; Peschel, G. G; Tinoco, I *J.Phys.Chem* **1965,** *69*, 3615.

225. Stephenson, R. M.; Malanowski, S. *Handbook of Thermodynamics of Organic Compounds,* Elsevier: New York, 1987.

226. Derwent World Drug Index. 2005. http://thomsonderwent.com/products/lr/wdi/.

227. Gero, A. M; O'Sullivan, W. J; Brown, D. J *Biochem.Med* **1985,** *34*, 60-69.

228. Srikrishnan, T; Parthasarathy, R *ACA, Ser.2* **1984,** *12*, 55.

229. Colarusso, P; Zhang, K; Guo, B; Bernath, P. F *Chem.Phys.Lett* **1997,** *269*, 39-48.

230. Rastogi, V. K; Singh, C; Jain, V; Palafox, M. A *J.Raman.Spectrosc* **2000,** *31*, 1005-12.

231. Gerdil, R *Acta.Cryst* **1961,** *14*, 333-44.

232. Barker, D. L; Marsh, R. E. *Acta.Cryst* **1964,** *17*, 1581-87.

233. Weber, H. P; Craven, B. M *Acta.Cryst* **1990,** *B46*, 532-38.

234. Jeffrey, G. A.; Kinoshita, Y *Acta.Cryst* **1963,** *16*, 20-28.

235. Mandel, N. S *Acta.Cryst* **1977,** *B33*, 1079-82.

236. Estrin, D. A; Paglieri, L; Coronglu, G *J.Phys.Chem* **1994,** *98*, 5653-60.

237. Aleman, C *Chem.Phys.Lett* **1999,** *302*, 461-70.

238. Aleman, C *Chem.Phys* **1999,** *244*, 151-62.

239. van Mourik, T; Benoit, D. M; Price, S. L.; Clary, D. C *Phys.Chem.Chem.Phys* **2000,** *2*, 1281-90.

240. Voet, D; Rich, A *J.Am.Chem.Soc* **1969,** *91*, 3069-75.

241. Brooks, R; Alcock, T. C *Nature* **1950,** *166*, 435-36.

242. Pertlik.F *Tscher.Miner.Petrog* **1981,** *29*, 67-74.

243. Lyakhov, A. S; Gaponik, P. N; Voitekhovich, S. V *Acta.Cryst* **2001,** *C57*, 185-86.

244. Fukuyo, M; Hirotsu, K; Higuchi, T *Acta.Cryst* **1982,** *B38*, 640-43.

245. Cox, P. J. *Acta.Cryst* **2001,** *E57*, o1203-o1205.

246. Thewalt, U; Bugg, C. E; Marsh, R. E. *Acta.Cryst* **1971,** *B27*, 2358-63.

247. Matkovic-Calogovic, D; Sankovic, K *Acta.Cryst* **1999,** *C55*, 467-69.

248. Maixner, J; Zachova, J *Acta.Cryst* **1991,** *C47*, 2474-76.

249. Tret'yak, S.M; Mitkevich, V. V; Sukhodub, L. F *Sov.Phys.Crystallogr* **1987,** *32*, 748-50.

250. Cunane, L. M; Taylor, M. R *Acta.Cryst* **1993,** *B49*, 524-30.

251. Langer, V; Huml, K *Acta.Cryst* **1978,** *B34*, 1881-84.

252. Broomhead, J. M *Acta.Cryst* **1948,** *1*, 324-29.

253. Cochran, W *Acta.Cryst* **1951,** *4*, 81-92.

254. Vella-Zarb, L and Tremayne, M. Personal Communication. 2005.

255. Henry, M. *ChemPhysChem* **2002,** *3*, 607-16.

256. Leonard, G. A *Acta.Cryst* **1995,** *D51*, 136-39.

257. Spencer, M *Acta.Cryst* **1959,** *12*, 66-71.

258. Desiraju, G. R. *CrystEngComm* **2002,** *4*, 499.

259. Etter, M. C. *J.Am.Chem.Soc* **1982,** *104*, 1095-96.

260. Allen, F. H.; Motherwell, W. D. S.; Raithby, P. R; Shields, G. P.; Taylor, R *New.J.Chem* **1999,** *23*, 25-34.

261. Price, S. L. *J.Chem.Soc., Faraday Trans* **1992,** *88*, 1755-63.

262. Verschoor, G. C.; Keulen, E. *Acta.Cryst* **1971,** *B27*, 134-45.

263. Belaj, F. *Acta.Cryst* **1992,** *C48*, 1088-90.

264. Lewis, T. C.; Tocher, D. A.; Day, G. M.; Price, S. L. *CrystEngComm* **2003,** *5*, 3-9.

265. Price, S. L.; Andrews, J. S; Murray, C. W; Amos, R. D *J.Am.Chem.Soc* **1992,** *114*, 8268-76.

266. von Baeyer, A. *Annalen* **1864,** *130*, 129.

267. von Baeyer, A. *Annalen* **1863,** *127*, 199.

268. Conrad, M.; Guthzeit, M. *berichte* **1882,** *15*, 2844.

269. Geissler, W. *M?nch.Med.Wochenschr* **1912,** *59*, 922.

270. Hauptmann, A. *M?nch.Med.Wochenschr* **1908,** *55*, 400-02.

271. Mesley, R. J; Clements, R. L *J.Pharm.Pharmacol* **1968,** *20*, 341-47.

272. Craven, B. M.; Vizzini, E. A.; Rodrigues, M. M. *Acta.Cryst* **1969,** *B25*, 1978-93.

273. Caillet, J.; Claverie, P. *Acta.Cryst* **1980,** *B36*.

274. Williams, P. P. *Anal.Chem* **1959,** *31*, 140.

275. Otsuka, M.; Onoe, M.; Matsuda, Y. *Drug.Dev.Ind.Pharm* **1994,** *20*, 1453-70.

276. Williams, P. P. *Acta.Cryst* **1973,** *B29*, 1572-79.

277. Mesley, R. J; Clements, R. L; Flaherty, B; Goodhead, K. *J.Pharm.Pharmacol* **1968,** *20*, 329-40.

278. Bolton, W. *Acta.Cryst* **1963,** *16*, 166-73.

279. Al-karaghouli, A. R.; Abdul-Wahab, B.; Ajaj, E.; Al-Asaff, S. *Acta.Cryst* **1977,** *B33*, 1655-60.

280. Jeffrey, G. A.; Ghose, S.; Warwicker, J. O. *Acta.Cryst* **1961,** *14*, 881-87.

281. Al-Saqqar, S; Falvello, L. R; Soler, T. *J.Chem.Crystallogr* **2004,** *34*, 61-65.

282. Nichol, G. S; Clegg, W *Acta.Cryst* **2005,** *B61*, 464-72.

283. *Synthesis of cyanuric acid,* http://42.1911encyclopedia.org/C/CY/CYANIC_ACID_AND_CYANATES.htm **2003**.

284. Seifer, G. B. *Russ.J.Coord.Chem* **2001,** *28*, 301-24.

285. Wiebenga, E. H.; Moerman, N. F. *Z.Krist* **1938,** *99*, 217.

286. Coppens, P. *Acta.Cryst* **1971,** *B27*, 146-58.

287. Kutoglu, A.; Hellner, E. *Acta.Cryst* **1978,** *B34*, 1617-23.

288. Kutoglu, A.; Scheringer, C. *Acta.Cryst* **1979,** *A35*, 458-62.

289. Newman, R.; Badger, R. M. *J.Am.Chem.Soc* **1952,** *74*, 3545-48.

290. Wiebenga, E. H. *J.Am.Chem.Soc* **1952,** *74*, 6156-57.

291. DeWit, H. G. M; Van Miltenburg, J. C; DeKruif, C. G. *J.Chem.Thermodyn* **1983,** *15*, 651.

292. Kozyro, A. A.; Kabo, G. Y.; Soldatova, T. V.; Simirskii, V. V. *Russ.J.Phys.Chem* **1992,** *66*, 2583-90.

293. Pedireddi, V. R.; Belhekar, D. *Tetrahedron* **2002,** *58*, 2937-41.

294. Liebig, J. *Justus.Liebigs.Ann.Chem* **1862,** *121*, 81.

295. Dunn, S. J.; Sheehan, H. L.; Mcletchie, N. G. B. *Lancet* **1943,** *244*, 484-87.

296. Bolton, W. *Acta.Cryst* **1964,** *17*, 147-52.

297. Allen, F. H.; Baalham, C. A; Lommerse, J. P. M; Raithby, P. R *Acta.Cryst* **1998,** *B54*, 320-29.

298. Rubin, M. B. *Chem.Rev* **1975,** *75*, 177.

299. Craven, B. M.; McMullan, R. K *Acta.Cryst* **1979,** *B35*, 934-45.

300. Davies, D. R.; Blum, J. J. *Acta.Cryst* **1955,** *8*, 129-36.

301. Shieh, H.; Voet, D. *Acta.Cryst* **1975,** *B31*, 2192-201.

302. Shieh, H.; Voet, D. *Acta.Cryst* **1976,** *B32*, 2361-67.

303. Andrews, J. C.; Sell, I. T. *Arch.Biochem.Biophys* **1955,** *56*, 405-11.

304. Makarina-Kibak, L. Y.; Korablev, M. V.; Vredenskii, V. M. *Zdravookhr.Beloruss* **1972,** *18*, 32-34.

305. Hillered, L.; Persson, L. *Neuroreport* **1995,** *6*, 1816-20.

306. Fabbiani, F. P. A; Allan, D. R; Marshall, W. G; Parsons, S; Pulham, C. R.; Smith, R. I *J.Cryst.Growth* **2004,** *275*, 185-92.

307. Kolb, V. M.; Dworkin J.P.; iller, S. L. *J.Mol.Evol* **1994,** *38*, 549-57.

308. Shirley, R *Science* **1966,** *152*, 1512-13.

309. Frincu, M. C; Fogarty, C. E; Swift, J. A *Langmuir* **2004,** *20*, 6524-29.

310. Lonsdale, K.; Sutor, D; Wooley, S *Brit.J.Urol* **1968,** *40*, 33-36.

311. Coe, F. L; Moran, E; Kavalich, A. G *J.Chron.Dis* **1976,** *29*, 793-800.

312. Choi, H. K; Curhan, G *Arthritis.Rheum* **2004,** *51*, 1023-29.

313. Kalkura, S. N; Vaidyan, V. K; Kanakavel, M; Ramasamy, P *J.Cryst.Growth* **1993,** *132*, 617-20.

314. Brun, A *Arch.Sci.Phys.Nat* **1899,** *7*, 284.

315. Ringertz, H *Acta.Cryst* **1965,** *19*, 286-87.

316. Ringertz, H *Acta.Cryst* **1966,** *20*, 397-403.

317. Sours, R. E; Fink, D. A; Swift, J. A *J.Am.Chem.Soc* **2002,** *124*, 8630-36.

318. Parkin, S; Hope, H *Acta.Cryst* **1998,** *B54*, 339-44.

319. Price, S. L.; Patel, B; Pridhanani-Jethani, P; Torrisi, A *ACA, Ser.2* **2005,** *In Press.*

320. Blagoeva, I. B; Pojarlieff, I. G; Dimitrov, V. S *J.Chem.Soc.Perk.T.2* **1978,** 887-92.

321. Bateman, J. H. *Kirk-Othmer Encycl. Chem. Technol*, Grayson, M.; Eckroth, D., Eds.; 3 ed.; Wiley: New York, 1980.

322. Camerman, A; Camerman, N *Acta.Cryst* **1971,** *B27*, 2205.

323. Avendano, J.; Menendez, J. C. *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th ed.; Wiley: New York, 1999.

324. Cassady, R. E; Hawkinson, S. W. *Acta.Cryst* **1982,** *B38*, 1646-47.

325. Beilles, S; Cardinael, P; Ndzie, E; Petit, S; Coquerel, G. *Chem.Eng.Sci* **2001,** *56*, 2281-94.

326. Yu, F; Schwalbe, C. H; Watkin, D. J *Acta.Cryst* **2004,** *C60*, o714-o717.

327. Busing, W. R. WMIN A Computer Program to Model Molecules and Crystals in Terms of Potential Energy Functions. 1981. Tennessee, USA, Oak Ridge National laboratory.

328. Verwer, P.; Leusen, F. J. J. *Reviews in Computational Chemistry*, Lipkowitz, K. B.; Boyd, D. B., Eds.; Wiley-VCH: New York, 1998.

329. Weiner, S. J; Kollman, P. A; Case, D. A; Singh, U. C; Ghio, C; Alagona, G; Profeta Jr, S; Weiner, P *J.Am.Chem.Soc* **1984,** *106*, 765-84.

330. Dunitz, J. D; Gavezzotti, A.; Schweizer, W. B *Helv.Chim.Acta* **2003,** *86*, 4073-92.

331. Damm, W.; Fontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J.Comput.Chem* **1997,** *18*, 1955-70.

332. Jorgensen, W. L.; Madura, J. D; Swenson, C. J *J.Am.Chem.Soc* **1984,** *106*, 6638-46.

333. Van Eijck, B. P. *Acta.Cryst* **2005,** *B61*, 528-35.

334. Skoda, J; Flegelova, Z; Farkas, J *Biochem.Biophys.Res.Comm* **1973,** *50*, 80-83.

335. Jamison, R. M; Seal, L. A *Clin.Res* **1977,** *25*, A29.

336. Reiner, P; Mardiak, J; Reinerova, M; Ujhazy, V *Neoplasma* **1981,** *28*, 533-39.

337. Ujhazy, V; Reiner, P; Farkas, J; Skoda, J *Neoplasma* **1977,** *24*, 259-60.

338. Tsai, L; Silverton, J. V; Lingh, H. T *J.Org.Chem* **1978,** *43*, 4415.

339. Martin, H. D; Kummer, M; Martin, G; Bartsch, J; Bruck, D; Heinrichs, A; Mayer, B; Rover, S; Steigel, A; Mootz, D; Middelhauve, B; Scheutzow, D *Chem.Br* **1987,** *120*, 1133-49.

340. Martin, H. D. Personal communication of the hydrogen positions. 2004.

341. Lommerse, J. P. M; Price, S. L.; Taylor, R *J.Comput.Chem* **1997,** *18*, 757-74.

342. Beyer, T.; Lewis, T.; Price, S. L. *CrystEngComm* **2001,** *3*, 178-213.

343. Pesina, K; Pleticha-Lansky, R; Sosna, M *Biol.Plantarum* **1977,** *19*, 453-56.

344. Mooij, W. T. M.; Leusen, F. J. J. *Phys.Chem.Chem.Phys* **2001,** *3*, 5063-66.

345. Rice, B. M; Sorescu, D. C *J.Phys.Chem.B* **2004,** *108*, 17730-39.

346. Van Eijck, B. P.; Mooij, W. T. M.; Kroon, J. *J.Comput.Chem* **2000,** *22*, 805-15.

347. Karamertzanis, P. G; Price, S. L. *J.Phys.Chem.B* **2005,** *In press.*

348. Lancaster, R. W. Personal Communication. 2005.

349. Zobel, D; Luger, P; Dreissig, W *Acta.Cryst* **1992,** *B48*, 837-48.

350. Iball, J; Wilson, H. R *Proc.R.Soc.Lond.Ser-A* **1965,** *288*, 418-39.

351. Kennedy, S. W; Schultz, P. K; Slade, P. G; Tiekink, E. R *Z.Krist* **1987,** *180*, 211.

352. Nakai, H; Saito, T; Yamakawa, M *Acta.Cryst* **1988,** *C44*, 533-35.

353. Harrowfield, J. M.; Skelton, B. W.; Soudi, A. A.; White, A. H. *Aust.J.Chem* **1989,** *42*, 1795-98.

354. Artioli, G; Masciocchi, N; Galli, E *Acta.Cryst* **1997,** *B53*, 498-503.